

Échantillonnage

Connaissances du collège nécessaires à ce chapitre

- ▶ Calculer une fréquence
- ▶ Interpréter une fréquence
- ▶ Calculer une probabilité
- ▶ Interpréter une probabilité



Auto-évaluation

Des ressources numériques pour préparer le chapitre sur manuel.sesamath.net



1 Aux congrès des héros, on trouve des Jedis (J), des chevaliers de la Table Ronde (T) et des elfes des Terres du milieu (E).

Héros	J	T	E
Effectifs à Londres	37	77	23
Effectifs à Baltimore	53	48	40

- 1) Dans quelle ville de congrès, les Jedis sont-ils les plus présents ?
- 2) Même question pour les chevaliers.
- 3) Pour ces deux relevés, calculer les fréquences de présence de chaque type de héros à 0,01 près.

2 Un jeu consiste à miser sur le doigt des deux mains que va faire apparaître un animateur (qui ne connaît pas les paris).

- 1) Quelle est la probabilité de gagner ? Répondre sans justifier.
- 2) Si je joue 10 fois, suis-je sûr de gagner ?
- 3) Si je joue 100 fois, suis-je sûr de gagner 10 fois ?
- 4) 10 personnes jouent. Elles ne connaissent pas le vote des autres. Est-on sûr qu'au moins l'une d'entre elles va gagner ?
- 5) 10 personnes jouent. Chacune a un vote différent des autres. Est-on sûr qu'au moins l'une d'entre elles va gagner ?



Voir solutions p. 49



ACTIVITÉ 1 Lancer de dé cubique

INFO

On utilise un dé bien équilibré. On le lance pour noter le numéro de la face supérieure.

1) a) On effectue 20 lancers. Proposer une répartition possible de ces lancers.

Faces	1	2	3	4	5	6
Effectifs						

b) Plusieurs solutions sont-elles possibles ?

c) Celle que vous avez choisie vous paraît-elle :

- probable ?
- improbable ?
- difficile à dire ?

2) Pour reproduire cette expérience, on utilise la feuille de calcul d'un tableur.

La fonction ALEA donne un nombre aléatoire dans l'intervalle $[0; 1[$.

La fonction ENT nous donne la partie entière d'un nombre.

a) Expliquer pourquoi $ENT(1+6*ALEA())$ simule le lancer d'un dé équilibré.

b) Entrer cette fonction dans la case A1. La copier (Ctrl C)

Sélectionner la plage A2:A100 (en écrivant A2:A100 dans le sélectionneur de plage en haut à gauche) puis coller la formule (Ctrl V).

c) La formule $NB.SI(A1:A100;1)$ permet de savoir combien il y a de 1 sur la plage A1:A100. Donner l'effectif de chaque face.

d) Construire le diagramme en barres.

e) Appuyer sur F9. Qu'observe-t-on ?

Expliquer le terme de fluctuation d'échantillonnage.

ACTIVITÉ 2 Lancé de dé décagonal

INFO

On lance un dé équilibré à 10 faces et on note le numéro de la face supérieure.

1) a) En utilisant un tableur, faire une colonne de 100 lancers d'un dé à 10 faces.

b) Afficher en cellule A102 la fréquence des lancers supérieurs ou égaux à 4.

On utilisera la fonction $NB.SI(A1:A100;">=4")$.

c) Recalculer plusieurs fois et noter les résultats.

d) Quelle semble être la probabilité d'obtenir une face supérieure ou égale à 4 ?

2) On va maintenant faire 1 000 simulations de 100 lancers.

a) Sélectionner la plage A1:A102, puis la copier et la coller dans la plage B1:ALL102.

b) Sélectionner la plage des fréquences et faire un graphique de type ligne (points seuls).

c) Dans quel intervalle se situe la plupart des fréquences ?

3) a) Déterminer la proportion des fréquences comprises entre les 2 valeurs trouvées en 2c.

On utilisera deux fois la formule $NB.SI$ pour l'encadrement.

b) « Au moins 95 % des fréquences de nos 1 000 simulations sont dans cet intervalle ». Est-ce plausible ?

Expliquer le terme de fiabilité des simulations.



1. Échantillon, simulation et fluctuation

■ DÉFINITION : Expérience aléatoire

Une **expérience aléatoire** est une expérience renouvelable dont les résultats possibles sont connus sans qu'on puisse déterminer lequel sera réalisé.

REMARQUE : exemples d'expériences aléatoires :

- le lancer de dé ;
- un sondage d'opinion avant une élection ;
- le tirage de jetons dans une urne ou de cartes dans un jeu.

■ DÉFINITION : Échantillon

Un **échantillon** de taille n est constitué des résultats de n répétitions indépendantes de la même expérience.

REMARQUE : exemples d'échantillons.

- on lance une pièce 50 fois et on regarde si on obtient pile ;
- on tire 20 fois une carte d'un jeu de 32 cartes en la remettant et on regarde si c'est un cœur ;
- on interroge 1 000 personnes et on leur demande si elles voteront.

■ DÉFINITION : Fluctuation d'échantillonnage

Deux échantillons de même taille issus de la même expérience aléatoire ne sont généralement pas identiques.

On appelle **fluctuation d'échantillonnage** les variations des fréquences des valeurs relevées.

NOTATION :

- n est le nombre d'éléments de l'échantillon. C'est l'**effectif** ou la **taille de l'échantillon**.
On dit que l'échantillon est de taille n .
- f_o est la **fréquence** du caractère observé dans l'échantillon.
- p est la **proportion effective** du caractère observé dans la population.

REMARQUE :

Plus la taille de l'échantillon augmente, plus les fréquences observées se rapprochent de p .

2. Prise de décision : intervalle de fluctuation (p est connu)

Protocole Soit une population pour laquelle on étudie la proportion d'un caractère.

On émet une hypothèse sur la proportion p du caractère étudié dans la population. On considère donc p comme connu car il a une valeur conjecturée.

Un échantillon de taille n de cette population est prélevé et on observe une fréquence f_o du caractère étudié.

La question Peut-on, à partir de l'observation de f_o , valider la conjecture faite sur p ?

La fréquence observée, f_o , est-elle proche ou éloignée de la probabilité ou proportion théorique, p ?



■ DÉFINITION : Intervalle de fluctuation

L'intervalle de fluctuation au seuil de 95 %, relatif aux échantillons de taille n , est l'intervalle centré autour de p qui contient la fréquence observée f_o dans un échantillon de taille n avec une probabilité égale à 0,95.

REMARQUES :

- Il n'existe pas d'intervalle dans lequel on trouverait f_o avec certitude (à moins de prendre l'intervalle $[0; 1]$...) à cause de la fluctuation d'échantillonnage.
- Cet intervalle peut être obtenu de façon approchée à l'aide de simulations.

■ PROPRIÉTÉ

Soit p la proportion effective d'un caractère d'une population comprise entre 0,2 et 0,8 et f_o la fréquence du caractère dans un échantillon de taille n supérieure ou égale à 25.

f_o appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ avec une probabilité d'environ 0,95.

REMARQUE : La taille de l'intervalle de fluctuation $\left(\frac{2}{\sqrt{n}} \right)$ diminue si n augmente.

MÉTHODE 1 Prendre une décision

► Ex. 18 p. 35

Dans les conditions de la définition et de la propriété :

- On émet une hypothèse sur la proportion du caractère de la population p .
- On détermine l'intervalle de fluctuation au seuil de 95 % de la proportion p dans des échantillons de taille n .
 - Si f_o n'appartient pas à cet intervalle, on rejette l'hypothèse faite sur p avec un risque d'erreur de 5%.
 - Si f_o appartient à cet intervalle, on ne rejette pas l'hypothèse faites sur p .

Exercice d'application

Dans la réserve indienne d'Aamjiwnaag, située au Canada, à proximité d'industries chimiques, il est né entre 1999 et 2003, 132 enfants dont 46 garçons. Est ce normal ?

Correction On fait ici l'hypothèse P suivante : « le sexe d'un enfant qui naît dans cette réserve est un garçon avec une probabilité de 0,5 ».

La taille de l'échantillon est $n = 132$ ($n \geq 25$) et

la fréquence observée est $f_o = \frac{46}{132} \approx 0,34$

avec $0,2 \leq f_o \leq 0,8$.

L'intervalle de fluctuation au seuil de 95 % est :

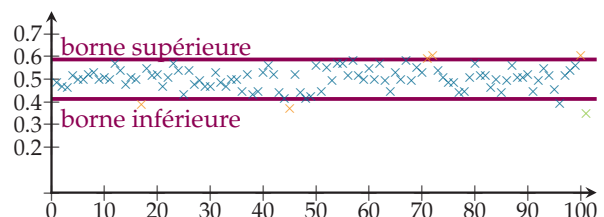
$$IF = \left[0,5 - \frac{1}{\sqrt{132}}; 0,5 + \frac{1}{\sqrt{132}} \right] \approx [0,41; 0,58].$$

$f_o \notin IF$ et on rejette l'hypothèse P .

La probabilité qu'un garçon naisse dans cette réserve n'est pas de 0,5.

Les 95% sont illustrés avec le graphique qui suit :

On simule 100 fois le comptage de garçons sur 132 naissances. Dans 94 simulations, la proportion des garçons nés se trouve dans l'intervalle de fluctuation.





3. Estimation : Intervalle de confiance (p est inconnu)

L'intervalle de fluctuation permet d'avoir un intervalle où se situe la proportion inconnue p avec une probabilité de 0,95%.

PROPRIÉTÉ

On considère un échantillon de taille n ($n \geq 25$) tel que $f_o \in [0, 2; 0, 8]$.

Alors p appartient à l'intervalle $\left[f_o - \frac{1}{\sqrt{n}}; f_o + \frac{1}{\sqrt{n}} \right]$ avec une probabilité de 0,95.

DÉFINITION : Intervalle de confiance

Un **intervalle de confiance au seuil de 95%**, relatif aux échantillons de taille n , est un intervalle centré autour de f_o où se situe la proportion p du caractère dans la population avec une probabilité égale à 95 %.

L'intervalle $\left[f_o - \frac{1}{\sqrt{n}}; f_o + \frac{1}{\sqrt{n}} \right]$ est donc appelé intervalle de confiance au seuil de 95 %.

PREUVE Cette symétrie dans les définitions d'intervalles de confiance et de fluctuation provient des inégalités suivantes :

$$\begin{aligned}
 f_o \in \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] &\Leftrightarrow p - \frac{1}{\sqrt{n}} \leq f_o \text{ et } f_o \leq p + \frac{1}{\sqrt{n}} \\
 &\Leftrightarrow p \leq f_o + \frac{1}{\sqrt{n}} \text{ et } f_o - \frac{1}{\sqrt{n}} \leq p \\
 &\Leftrightarrow p \in \left[f_o - \frac{1}{\sqrt{n}}; f_o + \frac{1}{\sqrt{n}} \right]
 \end{aligned}$$

MÉTHODE 2 Estimer la proportion d'un caractère

► Ex. 40 p. 38

- On réalise un échantillon de taille n et on y obtient une fréquence observée f_o .
- On construit l'intervalle de confiance à partir de n et f_o .

La proportion réelle dans la population se situe dans cet intervalle avec une probabilité d'environ 0,95.

Exercice d'application

Le 4 mai 2007 soit deux jours avant le second tour des élections présidentielles, on publie le sondage suivant réalisé auprès de 992 personnes :

S. Royal : 45%
N. Sarkozy : 55%

Interpréter ce sondage.

Correction On calcule l'intervalle de confiance pour N. Sarkozy.

$$\begin{aligned}
 I &= \left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right] \\
 I &= \left[0,55 - \frac{1}{\sqrt{992}}; 0,55 + \frac{1}{\sqrt{992}} \right]
 \end{aligned}$$

soit $I \approx [0,518; 0,582]$.

La proportion des votants en faveur de N. Sarkozy se trouvant dans $[0,518; 0,582]$ avec 95% de chance, on peut en déduire qu'il avait de grande chance d'être élu.

REMARQUE : Les sondages sont souvent réalisés auprès d'environ 1000 personnes car cela permet de connaître la proportion d'un candidat à 3% près.



Activités mentales

On donnera les résultats sous forme fractionnaire.

1 Une entreprise pharmaceutique souhaite savoir si une de ses machines dose correctement des gélules de paracétamol de 1g.

Pour cela, le responsable qualité prélève un lot de 513 pastilles et constate que 437 sont conformes.

La machine est considérée comme fonctionnelle si, sur 100 gélules, au moins 97 sont conformes.

Dans cette étude, quelle est :

- 1) la taille n de l'échantillon ?
- 2) la proportion théorique p ?
- 3) la fréquence observée f_o ?

2 On estime qu'il y a 119 garçons pour 100 filles nées en Chine. Une chinoise est enceinte. Quelle est la probabilité que son enfant soit un garçon ?

3 Dans une usine automobile, une machine fabrique 230 pommeaux de levier de vitesse par heure.

Pour tester si les dimensions du pommeau sont bonnes, on mesure durant 30 min les produits fabriqués et on constate que 73 pommeaux ont des dimensions conformes.

Donner la proportion des pommeaux conformes parmi les pommeaux prélevés.

4 Dans une classe de seconde, on constate qu'il y a 12 garçons pour 24 filles. La répartition garçon/fille de cette classe est-elle conforme à la population française ?

5 Lors d'un sondage auprès de 1 000 personnes aux États-Unis avant les élections de 2012, on a recueilli une intention de vote de 52,2% pour M. Obama contre 47,8% pour M. Romney.

L'équipe de campagne de M. Obama pouvait-elle être sereine ?

6 Pour l'élection des délégués, Hermione fait un sondage sur 100 élèves.

Quelle est la précision de ses résultats ?

7 H.Potier fait effectuer un sondage et obtient que f_o est dans l'intervalle $[32, 3; 32, 5]$.

- 1) Quelle valeur de f_o a-t-il trouvé ?
- 2) Quelle est la taille de l'échantillon ?

Échantillon, simulation, fluctuation

8 Dédé possède un dé équilibré à quatre faces numérotées de 1 à 4.

On s'intéresse à la sortie du nombre 4.

1) Imaginer un échantillon de taille 30 puis compléter le tableau suivant qui résume votre échantillon.

Issue				
Effectif				

2) Proposer un tableau qui pourrait résumer un échantillon de taille 237 de l'expérience du lancer de dé, qui selon vous, est raisonnable.

3) Proposer un autre tableau pour un autre échantillon de taille 237, qui selon vous, est très improbable.

Expliquer pourquoi.

9 Une urne contient 60 jetons, 20 blancs et 40 noirs.

1) On tire successivement, sans regarder et sans remise 6 jetons de cette urne.

On s'intéresse au nombre de jetons blancs.

A-t-on obtenu un échantillon ? Justifier.

2) On tire successivement sans regarder et avec remise 6 jetons de cette urne.

On s'intéresse au nombre de jetons blancs.

A-t-on obtenu un échantillon ? Justifier.

10 Programmer l'aléatoire

ALGO

En informatique, tout langage a une commande qui simule un nombre aléatoire entre 0 et 1 exclu.

Par exemple :

- AlgoBox : random()
- Scilab : rand()
- Python : random()
- LibreOfficeCalc : alea()

En version papier, on utilisera la notation aleatoire().

Que fait l'algorithme suivant ?

1. *Algorithme* : algo_mystere
2. *Liste des variables utilisées*
3. alea : nombre
4. resultat : nombre
5. *Traitements*
6. Donner à alea la valeur de aleatoire()
7. Donner à resultat la valeur de 10*alea
8. Afficher la valeur resultat
9. *Fin de l'algorithme*



11 Pile ou face

ALGO

Écrire un algorithme qui simule le lancer d'une pièce de monnaie équilibrée.

On pourra utiliser la commande « `aleatoire()` » qui renvoie un nombre aléatoire dans l'intervalle $[0; 1[$.

12 Piocher une carte

ALGO

Écrire un algorithme qui simule la pioche d'une carte dans un paquet de 32 cartes et renvoie la valeur de la carte.

13 Pile ou face enchaîné

ALGO

En utilisant l'algorithme de l'exercice 11 : écrire un algorithme qui simule 127 lancers d'une pièce de monnaie équilibrée.

14 Essaie encore

ALGO

En utilisant l'algorithme de l'exercice 12 : écrire un algorithme papier qui simule des lancers de dé à quatre faces tant que l'utilisateur répond « oui » à la question « Voulez vous continuer ? ».

15 Simulation

ALGO INFO

Ouvrir le logiciel de votre choix.

- 1) Programmer la simulation de 100 lancers de dés et compter le nombre de 6 obtenus.
- 2) Effectuer 10 fois cette simulation et noter les résultats obtenus.

16 Simulation jusqu'à plus soif

ALGO INFO

Reprendre l'exercice 15 avec 1 000 lancers et en comptant le nombre de 2 obtenus.

17 On a lancé plusieurs simulations d'un lancer de dé et on a obtenu les résultats suivants.

Nombre de lancers	1 000	2 000	5 000	10 000
Nombre de 1	194	371	839	1 663

- 1) Calculer les fréquences d'apparition du 1 pour chaque échantillon.
- 2) Quelle est la proportion théorique du résultat ?
- 3) Calculer l'écart de la fréquence obtenue et la fréquence théorique dans chaque cas.
- 4) Quel phénomène a-t-on mis en évidence ?

Intervalle de fluctuation

18 ► MÉTHODE 1 p. 32

La répartition des groupes sanguins dans le monde est donnée dans le tableau ci-dessous.

Groupes	O	A	B	AB
Fréquences en %	45	40	11	4

Elle est cependant variable selon les ethnies.

- 1) On a testé le sang de 480 esquimaux et on a trouvé que 211 d'entre eux sont du groupe A.
 - a) Déterminer n, p, f_0 .
 - b) Les conditions de calcul de l'intervalle de fluctuation sont-elles réunies ?
 - c) Si oui, déterminer l'intervalle de fluctuation.
 - d) La proportion du groupe A chez les esquimaux est-elle conforme à la population mondiale ?
- 2) On a trouvé 62 esquimaux du groupe B. Que peut-on dire ?

19 D'après l'INSEE (www.insee.fr), la proportion de 0-17 ans en France est de 21,95 % en 2010.

La ville de Nice compte 64 242 jeunes âgé de 17 ans ou moins parmi 343 304 habitants.

- 1) Calculer la fréquence f_0 des 0-17 ans à Nice.
- 2) Les conditions de détermination de l'intervalle de fluctuation sont-elles réunies ?
- 3) Que peut-on en conclure ?

20 Au 20 h d'une grande chaîne de télévision française publique, la journaliste conclut un reportage sur le divorce par cette phrase :

« En 2012, le divorce des femmes a augmenté de 20 % ». Commenter cette conclusion.

21 Un article de « PC impact » de janvier 2013 est intitulé « Une sociologue pointe l'inefficacité des avertissements de la Hadopi » .

Il se base sur le travail d'une sociologue utilisant « une enquête qualitative menée entre mars et octobre 2012 auprès de 8 individus (5 hommes et 3 femmes, âgés de 27 à 55 ans, de catégories socioprofessionnelles et de formations variées) ».

3 des 8 personnes ont reçu un avertissement de la part d'Hadopi.

Que peut-on dire de cet article ?



22 Fonctions ?

1) Tracer sur un même graphique les courbes représentatives des deux fonctions sur l'intervalle $]0; 10\ 000]$:

a) $f(x) = 0,4 + \frac{1}{\sqrt{x}}$ b) $g(x) = 0,4 - \frac{1}{\sqrt{x}}$

2) Quel phénomène ce graphique illustre-t-il ?

23 Le maire d'une ville vient d'installer un feu rouge sur l'artère principale et demande que le feu soit réglé de la manière suivante :

Couleur	Rouge	Orange	Vert
Durée	20 s	5 s	35 s

Il observe ensuite pendant une journée la couleur du feu lorsqu'une voiture arrive.

Le tableau répertorie ses résultats :

Couleur	Rouge	Orange	Vert
Nombres	263	64	429

- Calculer la proportion de temps p de la couleur verte sur un cycle.
- Calculer la fréquence f des voitures passées au vert.
- On s'intéresse ici au réglage du feu vert et à l'hypothèse « le feu est bien réglé ».
 - Déterminer les bornes de l'intervalle de fluctuation au seuil d'environ 95 %.
 - Commenter le réglage du feu.

24 On lance 35 fois une pièce de monnaie et on obtient 12 fois face.

- Après avoir énoncé l'hypothèse faite sur la pièce, donner l'intervalle de fluctuation au seuil de 95 %.
- La pièce de monnaie est-elle équilibrée ?

25 Truquée ?

On lance 3 500 fois une pièce de monnaie et on obtient 1 810 fois pile. La pièce de monnaie est-elle équilibrée ?

26 Poids

Le poids médian d'un homme en France est 77,3 kg.

- Un institut de sondage interroge 649 agriculteurs. 357 répondent qu'ils pèsent moins de 77,3 kg. Que peut-on conclure d'un tel sondage ?
- L'institut interroge, ensuite, 526 cadres. 235 répondent de même. Conclure.
- Comment expliquer ces différences ?

27 Au Casino

Au « Lion Vert », le jeu de la roulette française consiste à lancer une bille dans un récipient circulaire tournant et muni d'encoches numérotées et colorées sur la périphérie. Les cases, numérotées de 0 à 36 sont alternativement rouge et noire (sauf le 0 qui est vert).

Le croupier souhaite vérifier si sa roulette est équilibrée pour la mise sur chance simple Pair-Impair (Le joueur mise Pair ou Impair et gagne une fois sa mise si le numéro qui sort correspond à son choix. 0 est considéré comme ni pair, ni impair.)



Pendant une soirée, le croupier consigne les lancers :

Sortie	Pair	Impair ou 0
Nombre	242	295

Que peut-on conclure ?

28 Naissance dans le monde

Un village chinois comptabilise dans un tableau le nombre de naissances par sexe en 2012 :

Sexe	Garçon	Fille
Nombre	75	49

- Calculer la fréquence de garçons nés en 2012.
- La situation dans le village vous paraît-elle raisonnable ? Justifier.
- Sachant que le rapport de masculinité en Chine est de 112/100 (il naît 112 garçons pour 100 filles), la situation vous paraît-elle raisonnable (Justifier) ?
- Le rapport de masculinité de la France est de 105/100.

Comment peut-on expliquer la différence par rapport à la chine ?

29 Un fabricant de dés à 4 faces a lancé le même dé un grand nombre de fois.

Il a regardé si la face numéroté 3 était équilibrée.

Sa fréquence d'apparition a été de 0,267.

Le fabricant a conclu (à 95 %) que son dé n'était pas équilibré car la fréquence du nombre 3 n'était pas compatible avec son intervalle de fluctuation.

Quelle est la taille minimale de l'échantillon utilisé ?



30 Répression des fraudes

La répression des fraudes a enquêté au casino « Le Lion Vert » (voir exercice 27).

Elle a estimé que la roulette française n'était pas équilibrée et que le casino avait fraudé avec une probabilité d'environ 95 %.

Ils ont, pour cela, effectué une analyse informatique des tirages de l'année précédente et remarqué que la couleur rouge était sortie avec une fréquence de 49 % sur un certain nombre de lancers au jeu de la mise sur chance simple Noir-Rouge.

- 1) Quelle est la taille minimale de l'échantillon pour que la répression des fraudes puisse arriver à cet conclusion ?
- 2) La veille, un client notant le déséquilibre en a profité pour l'utiliser.
Quelle stratégie le client a-t-il pu mettre en place ?

31 Taille de l'intervalle de fluctuation

On lance 50 fois une pièce de monnaie équilibrée.

- 1) Déterminer la taille de l'intervalle de fluctuation au seuil de 95 % de la fréquence d'apparition de « pile ».
- 2) Comment diviser par 2 la taille de l'intervalle de fluctuation.

32 Algorithme intervalle fluctuation

ALGO

Écrire un algorithme qui, à partir de la taille de l'échantillon et la proportion théorique, donne l'intervalle de fluctuation au seuil de 95 %.

33 Pratique

ALGO

Programmer l'algorithme de l'exercice 32

- sur la calculatrice ;
- sur un ordinateur.

34 Perfectionnement

ALGO

Modifier l'algorithme de l'exercice 32 pour qu'il vérifie si les conditions sur n et p sont respectées avant de faire les calculs et qu'il affiche si besoin un message d'erreur à l'utilisateur.

35 In or out ?

ALGO

Écrire un algorithme qui connaissant la taille de l'échantillon, la proportion supposée et la fréquence observée, indique si le test se révèle concluant ou non.

36 Inspiré par les comics Marvel

Double Face, célèbre ennemi de Batman, utilise une pièce fétiche pour choisir s'il doit gracier ou tuer ses ennemis. Spiderman, fan de Batman, a regardé tous les comics et a constaté que la pièce était tombée 117 fois sur la face vie pour 256 lancers.

Clark Kent pense qu'il y a autant de chances à ce tirage de tomber sur la face mort que sur la face vie.

Catwoman, et son esprit vif comme un félin, pense, elle, que Double Face a truqué la pièce pour laisser seulement à ses proies une chance de survie de 40 %.

Qui de Clark ou de Catwoman peut avoir raison ?

- 37 On a lancé 1 000 fois un dé électronique à trois faces et on a obtenu les résultats suivants :

	1	2	3
Nombres	359	342	299

- 1) La face 1 est-elle équilibrée ?
La face 2 est-elle équilibré ? Et la face 3 ?
- 2) En quoi les résultats sont-ils surprenants ?
- 3) Que peut-on donc conclure ?

38 Réussite

Dans le lycée Sophie Germain, les élèves viennent de passer le bac.

- 1) 64 % des 34 élèves de la TL1 ont eu leur bac alors que la moyenne nationale est de 78 %.
 - a) Peut-on déterminer l'intervalle de fluctuation ?
 - b) Ces résultats sont-ils anormalement bas ?
- 2) Une autre classe, les TL2, de 28 élèves a obtenu 98 % de réussite cette même année.
 - a) Peut-on calculer l'intervalle de fluctuation ?
 - b) Ces résultats sont-ils anormalement élevés ?

39 Loterie

Lors d'une kermesse d'école, des billets de loterie sont vendus avec l'annonce : « 1 billet sur 4 est gagnant ». Le papa de Cunégonde est très joueur et en achète 28.

- 1) Va-t-il gagner 7 lots ?
- 2) En fait, il en a obtenu 4. Déterminer l'intervalle de fluctuation associé à cet échantillon.
- 3) Ce papa peut-il crier au scandale ?
- 4) Pour quels résultats pourrait-il crier au scandale ?



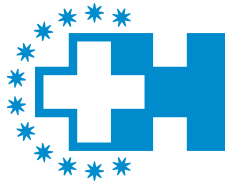
Estimation : intervalle de confiance

40 ► MÉTHODE 2 p. 33

En septembre 2013, un sondage réalisé auprès de 1 297 joueurs américains, révèle que 26 % des sondés achèteront la console SP2 à sa sortie.

- 1) Les conditions de validité de l'intervalle de confiance sont-elles réunies ?
- 2) Si oui, dans quelle fourchette peut-on estimer, avec une probabilité de 95 %, le pourcentage de joueurs américains qui feront l'acquisition d'une SP2 ?

41 En février 2013, un sondage a été réalisé par téléphone sur un échantillon national représentatif de 1 000 personnes résidant en France âgées de 18



ans et plus. Cet échantillon a été constitué d'après la méthode des quotas (sexe, âge, catégorie socioprofessionnelle du chef de ménage), par région et taille d'agglomération. Le sondeur donne 257 personnes n'ayant pas ou peu confiance dans les hôpitaux publics.

- 1) Les conditions de validité de l'intervalle de confiance sont-elles réunies ?
- 2) Si oui, dans quelle fourchette peut-on estimer le pourcentage de Français qui n'ont pas ou peu confiance dans les hôpitaux publics avec une probabilité de 95 % ?

42 En décembre 2012, un sondage été réalisé auprès de 1 003 personnes résidant en France, âgées de 18 ans et plus. L'échantillon a été constitué d'après la méthode des quotas (sexe, âge, catégorie socioprofessionnelle du répondant) par région et taille d'agglomération.

772 personnes interrogées ont déclaré avoir déjà été confrontées à une arnaque ou une tentative d'arnaque sur Internet.

Dans le même temps, 211 personnes interrogées déclarent avoir déjà été piégées sur Internet par un mail ou un site Internet leur demandant leurs coordonnées personnelles.

- 1) Estimer le pourcentage de personnes en France confrontées à une arnaque sur Internet.
- 2) Estimer le pourcentage de personnes en France ayant déjà été piégées sur internet.

43 Algorithme, intervalle confiance ALGO

Écrire un algorithme qui, connaissant la taille de l'échantillon et la fréquence observée, donne l'intervalle de confiance au seuil 0,95.

44 Pratique ALGO

Programmer l'algorithme de l'exercice 43, selon votre choix, soit sur la calculatrice, soit sur un ordinateur.

45 Perfectionnement ALGO

Modifier l'algorithme de l'exercice 43 pour qu'il vérifie avant les calculs si les conditions sur n et f_0 sont respectées et qu'il affiche un message d'erreur si besoin.

46 Avant les municipales de Sésalondes, le maire sortant, M. Aissekro, commande un sondage auprès de 500 personnes. D'après ce sondage, son adversaire obtiendrait un score de 48 %.

- 1) Déterminer l'intervalle de confiance de cet échantillon.
- 2) M. Aissekro peut-il fêter prématurément sa victoire ?
- 3) Si non, quelle taille d'échantillon minimale aurait-il dû prendre pour être rassuré ?

47 On demande aux 230 spectateurs de la rétrospective en 3D de *Blanche-Neige et les 7 nains* le nom du nain le plus amusant. 34 % des spectateurs votent pour Simplet et 28 % pour Grognon.

- 1) Peut-on déterminer des intervalles de confiance ?
- 2) Si oui, déterminer les deux intervalles de confiance.
- 3) Peut-on affirmer avec un risque d'erreur inférieur à 0,05 que Simplet est plus drôle que Grognon ?

48 Un institut de sondage interroge un groupe de filles sur leur acteur préféré.

- 1) Sur un premier échantillon de 800 filles, 38 % ont répondu : Léonard Dupacre.
Détermine l'intervalle de confiance de ce premier échantillon.
- 2) Sur un deuxième échantillon de 650 filles, 42 % ont, elles, répondu Brad Flip.
Déterminer l'intervalle de confiance de ce deuxième échantillon.
- 3) Ces deux intervalles sont-ils disjoints ?
- 4) Peut-on en déduire que chez les filles Brad Flip a plus de succès que Léonard Dupacre ?

49 Le pari du Chevalier de Méré

INFO

Le problème dit du « pari du chevalier de Méré » opposa ce dernier à Fermat et à Pascal :

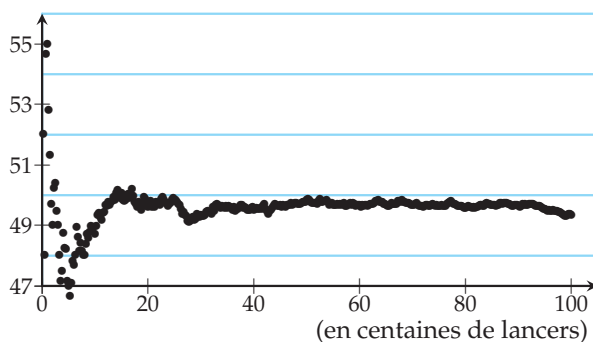
- *Pari 1* : si l'on jette 4 fois un dé à 6 faces, il y a plus de chances qu'on obtienne un six plutôt qu'on n'en obtienne pas.
- *Pari 2* : si l'on jette 24 fois deux dés à 6 faces, il y a aussi plus de chances qu'on obtienne un double six qu'on n'en obtienne pas.

Que penser de ces deux paris ?

- 1) À l'aide d'un tableur, on simule 500 lancers de 4 dés à 6 faces. Voici une capture d'écran.

	A	B	C	D	E	F
1	dé 1	dé 2	dé 3	dé 4		
2	3	5	6	3		0,498
3	5	1	1	4	0	

- a) Proposer une formule pour simuler le lancer d'un dé, à saisir en A2, B2, C2 et D2.
- b) Que calcule l'instruction `NB.SI(A2:D2;6) ?`
Qu'affiche la cellule E2 si on entre `=SI(NB.SI(A2:D2;6) >1;1;0) ?`
- c) Que calcule la formule `=NB.SI(E2:E501;1)/500`, à saisir en cellule F2 ?
- d) Compléter la feuille de calcul.
Donner la fréquence d'apparition d'un six et l'intervalle de confiance associé.
Que pouvez-vous conclure du *Pari 1* ?
- 2) Simuler 1 000 puis 10 000 expériences.
Que peut-on conclure du *Pari 1* ?
- 3) On simule 1 000 lancers de 24 fois deux dés à 6 faces. Le graphique ci-dessous donne l'évolution des fréquences d'apparition en % d'un double six. Que penser du *Pari 2* ?



50 Intervalle de fluctuation

ALGO

On considère une urne contenant :

- 3 boules blanches ;
- 5 boules noires.

On tire au hasard une boule de l'urne.

On note sa couleur avant de la remettre dans l'urne .

- 1) Quelle est la probabilité p d'obtenir une boule blanche ?

On simule 1 000 expériences de 100 tirages.

- 2) Avec un risque de 5%, dans quel intervalle vont se trouver les fréquences ainsi obtenues ?

On propose l'algorithme ci-dessous.

```

1. Liste des variables utilisées
2. nb100, nb1000 : nombre
3. p100, p1000 : nombre
4. Traitements
5. Donner à nb1000 la valeur de 0
6. Pour i variant de 1 à 1000 faire
7. Donner à nb100 la valeur de 0
8. Pour k variant de 1 à 100 faire
9. Si Alea_entier(1;8) ≤ 3 Alors
10. Donner à nb100 la valeur de nb100+1
11. Fin Si
12. Fin Pour
13. Donner à p100 la valeur de nb100/100
14. Si 0,275 ≤ p100 ≤ 0,475 Alors
15. Donner à nb1000 la valeur de nb1000+1
16. Fin Si
17. Fin Pour
18. Donner à p1000 la valeur de nb1000/1000
19. Affichage
20. Afficher p1000
21. Fin de l'algorithme
    
```

- 3) À partir de l'algorithme ci-dessus :

- a) Que représente la valeur de sortie ?
- b) Programmer l'algorithme.

Le résultat affiché est-il conforme à la théorie ?

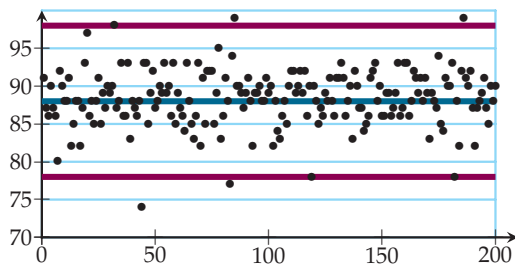
- c) À partir de l'algorithme ci-dessus, en écrire un qui demande en entrée le nombre n de tirages successifs et qui affiche en sortie le pourcentage d'expériences donnant une fréquence située dans l'intervalle de fluctuation associé.



51 Campagne publicitaire

ALGO

Un grossiste affiche sur ses plaquettes : « 86 % de notre thé est garanti sans trace de pesticides ». L'inspectrice de la DGCCRF souhaite étudier la validité de l'affirmation. Elle prélève 100 boîtes au hasard du stock du grossiste et en trouve 26 avec des traces de pesticides. On suppose que la proportion de boîtes sans trace de pesticides est bien égale à 0,86 du stock du grossiste. À l'aide d'un logiciel, on simule 200 prélèvements de 100 boîtes. Le graphique ci-dessous présente la fréquence en % des boîtes de thé ne contenant pas de pesticides pour chaque échantillon.



- Combien de fréquences n'appartiennent pas à l'intervalle $[0, 76; 0, 96]$? Interpréter.
- L'inspectrice peut-elle décider, au seuil de 95%, que la publicité est mensongère ?
- Proposer un algorithme simulant cette situation.

52 Grippe

En France, les consultations pour syndromes grippaux représentent 34 % des consultations chez les médecins généralistes.

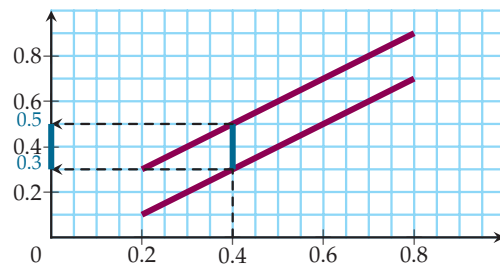
- La première semaine d'octobre, un cabinet médical du réseau Sentinelles, reçoit 967 patients dont 368 présentent les symptômes de la grippe.
 - Quelle part de ses consultations concerne des patients présentant des symptômes grippaux ?
 - Quel est l'intervalle de fluctuation associé à cet échantillon ?
 - Le médecin doit-il prévenir les autorités sanitaires d'un risque d'épidémie ? Argumenter.
- La même semaine, dans une autre région, 36 % des consultations d'un « médecin Sentinelle » concerne des syndromes grippaux. Il n'alerte pas les autorités sanitaires. Quel est le nombre maximum de patients reçus par ce praticien cette semaine-là ?

53 Abaque

On considère une urne où la proportion de boules noires est p (inconnue) et on s'intéresse à des échantillons de taille 100.

- Vérifier que l'intervalle de fluctuation est donné par $[p - 0,1; p + 0,1]$.

Sur le graphique sont tracées les droites d'équations $y = x - 0,1$ et $y = x + 0,1$. On utilise ce graphique pour lire l'intervalle de fluctuation pour une proportion p donnée. Exemple : pour $p = 0,4$, $IF = [0,3; 0,5]$.



- On suppose que $p = 0,6$. Donner l'intervalle de fluctuation IF au seuil de 95%. Utiliser le graphique ci-dessus pour retrouver l'intervalle IF .
- Dans cette question, on ne connaît pas p . La fréquence f_0 observée dans un échantillon de taille 100 est de 0,7. Proposer une méthode graphique pour trouver l'intervalle de confiance et estimer une valeur de p . Retrouver ce résultat par le calcul.

54 Auto focus

Un photographe vend des appareils photographiques. Il veut estimer par un intervalle de confiance le pourcentage p d'acheteurs d'appareils autofocus avec zoom dans sa clientèle

- Dans un échantillon de 100 clients, 60 achètent un tel appareil. Donner une estimation de p par un intervalle de confiance avec un seuil de 95 %.
- On considère l'affirmation suivante : « la fréquence p est obligatoirement dans l'intervalle de confiance obtenu à la question précédente ». Est-elle vraie ?
- Déterminer la taille n , n étant supérieur à 30, d'un échantillon de clients pour qu'un intervalle de confiance de p , au seuil de 95 % soit $[0,557; 0,643]$.



À la fin de ce chapitre, je dois être capable de :

Connaître le sens de :

- ▶ population, caractère, proportion
- ▶ échantillon, taille d'un échantillon

Savoir la définition

- ▶ de fluctuation d'une proportion lors d'un échantillonnage
- ▶ d'un intervalle de confiance à 95 %
- ▶ d'un intervalle de fluctuation à 95 %

Savoir calculer et utiliser

- ▶ la proportion d'un caractère
- ▶ un intervalle de confiance à 95 %
- ▶ un intervalle de fluctuation à 95 %



QCM d'auto-évaluation

Des ressources numériques
pour préparer le chapitre sur
manuel.sesamath.net



Pour chaque question, plusieurs réponses sont proposées. Déterminer celles qui sont correctes.

55 La proportion de gauchers dans le monde est 12 %. Dans une classe de 25 élèves :

- a on doit avoir 3 gauchers
- b on peut avoir 10 gauchers
- c il est probable que l'on ait 3 gauchers
- d on ne peut pas avoir 10 gauchers

56 Lors du tirage d'un échantillon, on parle de fluctuation d'échantillonnage pour indiquer :

- a qu'on choisit l'échantillon au hasard
- b qu'une proportion dans un échantillon n'est pas forcément la même que dans la population totale
- c que la proportion dans 2 échantillons de même taille peut être différente

57 Dans une école où un tiers des élèves croient encore au père Noël, on interroge 25 élèves au hasard (un élève pouvant être interrogé plusieurs fois). L'intervalle de fluctuation à 95 % de la proportion d'élèves qui croient encore au père Noël, dans cet échantillon, est :

- a $[0, 133; 0, 533]$
- b $\left[\frac{22}{75}; \frac{28}{75}\right]$
- c $[0, 29; 0, 38]$
- d $\left[\frac{2}{15}; \frac{8}{15}\right]$

58 Dans une tombola, la proportion de billets gagnants (c'est à dire permettant de gagner un lot, petit ou gros) est de 20 %. Hervé achète 64 billets de cette tombola. Sachant qu'il y a un très grand nombre de billets mis en vente, la proportion de billets gagnants parmi ceux qu'a achetés Hervé se situera dans l'intervalle de fluctuation à 95 %

- a $[0, 075; 0, 325]$
- b $[0, 08; 0, 32]$
- c $[0, 08; 0, 32]$
- d $[0, 07; 0, 33]$.

59 Karim qui est philatéliste (collectionneur de timbres) commande une série de 49 timbres choisis dans un catalogue publicitaire qui annonce une proportion de 42 % de timbres sur le thème des célébrités dans ces séries. À la réception de sa commande, Karim compte 15 timbres montrant des célébrités dans la série qu'il a reçu. Peut-il affirmer que la publicité était mensongère ?

- a Oui, car la proportion de timbres "célébrités" dans sa série n'est pas de 42 %
- b Non, car le calcul de 42 % de 49 ne donne pas un entier
- c Non, car on peut considérer qu'il y a fluctuation d'échantillonnage
- d Oui, car la proportion de timbres "célébrités" dans sa série n'est pas dans l'intervalle de fluctuation

60 Dans un groupe de 15 élèves d'une classe de seconde, le professeur principal a demandé à chaque élève s'il avait accès à Internet sans restriction chez lui. Trois élèves ont répondu Non, les autres ont répondu Oui. Ce professeur peut-il estimer la proportion d'élèves ayant accès à Internet sans restriction chez eux ?

- a** Oui, en donnant un intervalle de confiance à 95 %
- b** Non, car il y a fluctuation
- c** Oui, c'est tout simplement $\frac{12}{15} = \frac{4}{5}$, soit 80 %
- d** Non, car la proportion trouvée dans le groupe est trop forte
- e** Non, car la taille de l'échantillon est trop faible

61 Un maraîcher en gros vend de jeunes plants de poireaux à d'autres maraîchers. Le ramassage et le conditionnement des plants de poireaux est mécanisé. Il se doit d'indiquer à ses clients la proportion de plants de petite taille. Pour pouvoir donner cette indication, il procède au comptage des plants de petite taille dans un lot de 10 000 poireaux et en trouve 21 %. À l'aide ce comptage, il peut ensuite :

- a** fournir un intervalle de confiance à 95 % pour la proportion de poireaux de petite taille
- b** fournir un intervalle de fluctuation à 95 % pour la proportion de poireaux de petite taille
- c** indiquer qu'il y a 21 % de poireaux de petite taille dans les plants qu'il vend.
- d** autre réponse

62 Un candidat à une élection pense qu'il bénéficie d'environ 52 % d'intentions de vote. Il commande un sondage. De quelle taille doit être l'échantillon pour qu'il puisse être rassuré ?

- a** 500
- b** 1 000
- c** 3 000
- d** 10 000

63 On a procédé à une enquête de satisfaction auprès de 200 élèves demi-pensionnaires. 72 % sont contents de la cantine. Le lycée compte 1 300 demi-pensionnaires.

Le nombre de demi-pensionnaires satisfaits se situe, avec une probabilité de 95 %, dans l'intervalle :

- a** [900; 972]
- b** [1 143; 1 300]
- c** [1 199; 1 271]
- d** [845; 1 027]

64 54 % des français déclarent pratiquer un sport. À Sésaville, haut lieu de la Drôme, 2 324 habitants, ils sont 1 236 à pratiquer. L'intervalle de fluctuation à 95 % de cet échantillonnage, arrondi au millième, est :

- a** [0, 512; 0, 568]
- b** [0, 511; 0, 553]
- c** [0, 519; 0, 561]

65 Environ 78 % des français ont un ordinateur. À Bourg-de-Sésaville, plus bas dans la vallée, 1 286 habitants, ils sont 823 à en posséder un. L'amplitude de l'intervalle de fluctuation au seuil de 95 % est environ :

- a** 0,17
- b** 0,07
- c** 0,056

TP 1 Affaire Partida

INFO

En Novembre 1976 dans le comté de Hidalgo, Rodrigo Partida était condamné à huit ans de prison pour cambriolage et tentative de viol.

Il attaqua ce jugement affirmant que la désignation des jurés de ce comté était discriminatoire pour les américains d'origine mexicaine : 79,1 % de la population du comté était d'origine mexicaine mais, sur les 870 personnes convoquées pour être jurés les 11 années précédentes, seules 339 d'entre elles étaient d'origine mexicaine.

1 Simulation de la désignation d'un juré

On étudie une fonction du tableur qui choisit un juré en tenant compte de ses origines.

- 1) Quel nombre de jurés d'origine mexicaine peut-on espérer en choisissant au hasard 870 personnes dans la population de ce comté ?
- 2) Avec un tableur, la fonction ALEA() génère un nombre aléatoire dans $[0; 1[$. Que renvoie $SI(Alea() < p; 1; 0)$?
- 3) En prenant $p = 0,791$ expliquer comment cette formule permet de simuler la désignation d'un juré de ce comté en respectant les fréquences. On pourra s'aider du schéma ci-dessous :



2 Simulation

On procède à une simulation de 100 séries de désignation de jury.

- 1) Compléter la feuille de calcul à l'aide des instructions.

- Saisir en cellule A1 la formule $SI(ALEA() < 0,791; 1; 0)$
- Copier sur la plage A2:A870
- Saisir en A871 la formule $SOMME(A1:A870)/870$
- Sélectionner la plage A1:A871
- Copier sur la plage B1:CV871

- 2) Que représentent les nombres de la plage de cellules A1:A870 ?
- 3) Que représente le nombre affiché dans la cellule A871 ?
- 4) Que représentent les valeurs extrêmes obtenues dans la plage de cellules A871:CV871 ?
- 5) Représenter avec un nuage de points la série de données de la plage A871:CV871.
- 6) A-t-on obligatoirement 688 jurés d'origine mexicaine ?
Calculer le nombre maximal de jurés d'origine mexicaine dans un jury, obtenu lors de la simulation.

3 Intervalle de fluctuation

Il s'agit d'interpréter les résultats de cette simulation.

- 1) Donner l'intervalle de fluctuation correspondant à la simulation antérieure.
Celui-ci confirme-t-il les observations précédentes ?
- 2) Dans les simulations faites sur tableur, obtient-on un nombre de jurés mexicains égal à celui de l'affaire Partida ?
- 3) Comment expliquer cette situation ?



TP 2 Élections présidentielles de 2012

Voici un récapitulatif des sondages parus en France peu de jours avant le second tour des élections présidentielles de 2012. (source : sondages-en-france.fr)

Numéro	1	2	3	4	5	6	7	8	9
Date en mai	1	2	2	3	3	3	3	3	4
Hollande	53,5	53	52,5	53,5	52,5	53	53	52,5	52
Sarkozy	46,5	47	47,5	46,5	47,5	47	47	47,5	48
Nombre de personnes interrogées	1387	1077	2009	1000	1018	1072	1123	2161	1255

À l'issue du second tour des élections présidentielles 2012, M. Hollande a été élu président de la République Française avec 51,64 % des voix contre 48,36 % pour M. Sarkozy.

1 Un modèle

On considère une urne contenant 10 000 boules, 5 164 roses et 4 836 bleues.

On effectue 1000 tirages successifs et sans remise dans l'urne.

On note f_r la fréquence de boules roses obtenues.

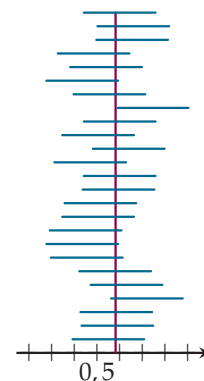
- 1) Que modélise cette expérience ? Quelle est la proportion p de boules roses dans l'urne ?
- 2) Quel est l'intervalle de fluctuation au seuil de 95 % ?

a) Démontrer que si $f_r \in \left[p - \frac{1}{\sqrt{1000}}; p + \frac{1}{\sqrt{1000}} \right]$ alors $p \in \left[f_r - \frac{1}{\sqrt{1000}}; f_r + \frac{1}{\sqrt{1000}} \right]$.

b) Comment appelle-t-on l'intervalle $\left[f_r - \frac{1}{\sqrt{1000}}; f_r + \frac{1}{\sqrt{1000}} \right]$?

- 3) On simule 25 échantillons de taille 1 000 à l'aide d'un logiciel de calcul numérique. Le graphique obtenu donne les 25 intervalles de confiance correspondants.

- a) Quelle est la proportion d'intervalles de confiance ne contenant pas p ?
- b) Modifier le programme disponible sur le site pour simuler 25 échantillons de taille n (nombre à saisir par l'utilisateur).
Exécuter votre programme avec $n = 2\,000$ et $n = 10\,000$.
Commenter.



2 Les sondages

On applique maintenant le modèle étudié à l'élection présidentielle de 2012.

- 1) Calculer, pour chaque sondage, les intervalles de confiance correspondant aux intentions de vote pour les deux candidats.
- 2) Tous ces sondages assuraient-ils une victoire de M. Hollande au second tour ?
Argumenter votre réponse.
- 3) a) Comment sont constitués les échantillons utilisés dans ces sondages ?
Faire une recherche documentaire.
b) Que dire alors de notre modèle ?

TP 3 Des listes pour le chocolat

ALGO

La marque de chocolat Nesrhonna, sponsor de l'équipe de France de Football, vend des tablettes contenant chacune une image d'un des onze principaux joueurs de l'équipe.

On suppose que les images sont réparties au hasard.

Thaïs, collectionneuse dans l'âme, se met en tête d'obtenir les onze images. Elle demande à sa grande sœur Mélia si elle sait combien de tablettes ses parents devront acheter pour avoir toute la collection.

1 Simulation à deux, avec une calculatrice

Mélia a trouvé comment générer un nombre entier aléatoire entre 1 et 11 à l'aide de sa calculatrice, elle demande à Thaïs de compléter un tableau d'effectifs et de lui dire à partir de quand chaque numéro est apparu au moins une fois.

À l'aide du tableau complété, comment peut-on répondre au problème posé ?

Tester cette solution. Quels sont les inconvénients de cette méthode ?

2 Simulation à l'aide d'un algorithme

Mélia, qui vient de terminer le chapitre sur les statistiques, a bien noté une remarque de son professeur : « *au lycée, on ne pourra pas exiger de vous l'utilisation de listes en algorithmique, c'est bien triste car c'est très pratique !* ».

Elle fait des recherches et s'aperçoit que ces listes lui permettent de stocker des résultats. Elle va pouvoir s'en servir pour simuler le problème. L'algorithme qu'elle a conçu est donné ci-contre.

- 1) quoi sert la première boucle pour ?
- 2) Expliquer l'utilité de la variable test.
- 3) Mélia programme son algorithme et le fait tourner. Celui-ci affiche en sortie « *Tu possèdes tous les joueurs* » et s'aperçoit qu'il ne répond pas au problème posé. Le compléter en conséquence.
- 4) Compléter l'algorithme pour répéter 1 000 fois la simulation. Calculer le nombre moyen de tablettes à acheter.

```

1. Liste des variables utilisées
2. T : liste
3. test : nombre
4. Traitements
5. Donner à test la valeur de 0
6. Pour i variant de 1 à 11 faire
7.   Donner à T[i] la valeur de 0
8. Fin Pour
9. Tant que (test=0) faire
10.   Donner à j la valeur de Alea(1,11)
11.   Donner à T[j] la valeur de T[j]+1
12.   Donner à test la valeur de T[1]
      × T[2] × T[3] × T[4] × T[5] ×
      T[6] × T[7] × T[8] × T[9] × T[10]
      × T[11]
13. Fin Tant que
14. Affichage
15. Afficher "Tu possèdes tous les
      joueurs"
16. Fin de l'algorithme
    
```

3 Concrètement...

Une coupe du monde de Football dure environ un mois, la société Nesrhonna propose son offre deux mois avant le début, jusque deux mois après la fin.

Sachant que Thaïs mange une tablette de chocolat par semaine, que penser de ses chances d'obtenir la collection complète ?



TP 4 Surbooking : une réalité statistique

ALGO

Lors d'un vol Madrid-Barcelone pouvant accueillir 250 passagers, la compagnie aérienne s'aperçoit qu'environ 20 % des passagers ayant acheté leur billet ne se présentent pas à l'aéroport. Elle vend alors plus de billets que de places dans le vol.

L'objectif de ce TP est de mettre en évidence un critère statistique utilisé par la compagnie aérienne pour déterminer le nombre de billets supplémentaires qu'elle émettra.

1 Simulation du remplissage d'un avion

On étudie le remplissage de l'avion suivant le nombre de billets vendus.

1) Dans cette question, on suppose que la compagnie n'a pas eu recours à la sur-réservation et qu'elle a vendu exactement 250 billets.

- a) Dans l'algorithme ci-contre qui simule le remplissage d'un vol, que désigne la variable P ?
- b) Le programmer, l'exécuter 10 fois et noter à chaque fois les résultats. Commenter.

```
1. Liste des variables utilisées
2. k, P : entiers
3. Traitements
4. Donner à P la valeur de 0
5. Pour k variant de 1 à 250 faire
6. Donner à P la valeur de
P+ent(alea()+0,8)
7. Fin Pour
8. Affichage
9. Afficher P
10. Fin de l'algorithme
```

2) Dans cette question, la compagnie aérienne vend un nombre de billets supérieur à 250.

- a) Modifier l'algorithme pour simuler le remplissage de ce vol. (On prendra en entrée le nombre de billets vendus.) Le programmer et l'exécuter.
- b) Quel nombre de billets supplémentaires la compagnie peut-elle vendre ? Justifier votre choix et détailler votre démarche.

2 Étude d'un critère pour décider du nombre de billet à vendre

La proportion de passagers se présentant à l'embarquement ayant acheté leur billet est de 80 %.

1) Dans cette question, on suppose que la compagnie émet 255 billets.

- a) On s'intéresse à la fréquence du nombre de passagers se présentant au comptoir. Au seuil de 95%, dans quel intervalle va-t-elle se trouver ? Comment appelle-t-on cet intervalle ?
- b) Quel pourcentage maximal de passagers se présentant au comptoir de la compagnie, au seuil de 95 %, peut-on prévoir ?
- c) En déduire le nombre maximum, avec un risque de 5 %, de passagers se présentant à la porte de l'embarquement.

2) Dans cette question, on suppose que la compagnie émet N billets.

- a) Donner l'intervalle de fluctuation associé à cet échantillon.
- b) Vérifier que le nombre maximum de passagers, au seuil de 95 % est donné par : $0,8N + \sqrt{N}$.
- c) Le critère choisi par la compagnie aérienne pour émettre les billets sur ce vol est que le nombre maximum de passagers, au seuil de 95%, ne doit pas dépasser 250. Combien de billets supplémentaires la compagnie proposera-t-elle alors à la vente ?



TP 5

Piège à fourmis sur une idée de C. Schwartz et B. Cailhol, <http://www.statistix.fr>

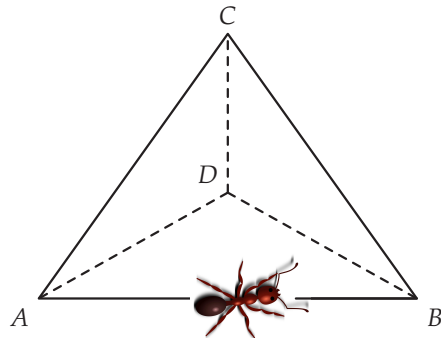
INFO

1 Le principe

Un piège à fourmi se présente sous la forme d'un tétraèdre $ABCD$. L'entrée se fait en A .

À chaque étape, la fourmi se dirige vers un autre sommet. Si elle arrive en A elle sort. Si elle n'est pas sorti au bout de 4 déplacements, elle meurt.

- 1) Calculer le nombre de parcours possibles et le nombre de parcours où la fourmi meurt.
- 2) Conjecturer la proportion de fourmis qui meurent.



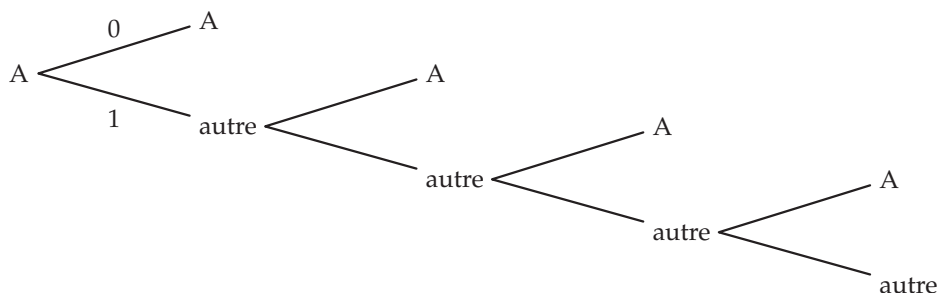
2 Simulation à l'aide d'un tableur

Ouvrir un tableur.

- 1) On pose A pour 0, B pour 1, C pour 2 et D pour 3.
Le premier sommet est A donc dans A1, on met 0.
Les autres sont choisis aléatoirement mais différent du précédent : $\text{MOD}(A1 + \text{ENT}(\text{ALEA}() * 3 + 1); 4)$.
Compléter de B1 à E1.
- 2) Dans G1 : déterminer le minimum de B1 : E1.
- 3) Recopier A1 : G1 jusqu'à A1000 : G1000.
- 4) Calculer la proportion de 0 (sortie de la fourmi) sur G1 : G1000.
- 5) Rafraîchir plusieurs fois. La conjecture du 2 de la partie 1 est-elle plausible ?

3 Analyse

Pour calculer la probabilité que la fourmi meure, on utilise un arbre pondéré.



- 1) Compléter les probabilités sur chaque branche.
- 2) En déduire la probabilité que la fourmi meure puis la probabilité pour qu'elle survive.
- 3) Votre résultat est-il cohérent avec la simulation : est-il dans l'intervalle de confiance ?



Paradoxe de Simpson

Dans la ville de Mathyville, il y a deux lycées, le lycée Sophia Kovalski et le lycée Ada Byron (toutes deux mathématiciennes célèbres). On donne les résultats du baccalauréat dans ces deux lycées en fonction du sexe :

Lycée Sophia Kovalski		
	Effectif	Bacheliers
Garçons	203	161
Filles	842	592

Lycée Ada Byron		
	Effectif	Bacheliers
Garçons	423	258
Filles	131	77

- 1) Calculer la proportion de bacheliers chez les garçons et les filles dans les deux lycées. Peut-on en déduire que les garçons sont meilleurs que les filles ?
- 2) Choquée par cette assertion, Mme Pita Gaure, maire de la ville, demande qu'on refasse les calculs sur l'ensemble de la population.
Faire ces calculs.

Que peut-on en conclure ?

Le paradoxe de Condorcet

Les 30 membres du Comité de Vie Lycéenne doivent choisir une salle pour héberger le foyer : la salle F qu'ils occupent actuellement, ou la salle H ou la salle B. Chaque membre classe ces 3 salles par ordre de préférence. On obtient :

- 12 voix pour FHB ;
- 8 voix pour HBF ;
- 1 voix pour HFB ;
- 5 voix pour BFH ;
- 4 voix pour BHF.

Un élève demande le nombre de voix de ceux qui préfèrent B à F. Le CVL se déplace donc en B.

Un autre élève demande le nombre de voix de ceux qui préfèrent H à B. Le CVL se déplace donc en H.

un 3ème demande alors combien préfèrent F à H ...

Que se passe-t-il ?

Le sens critique

Le 1er mars 2013, le maire de Mathyville se félicitait devant la presse : « ce mois-ci, le nombre d'actes criminels a baissé de près de 10 pour cent par rapport au mois précédent. »

Y a-t-il vraiment de quoi pavoiser ?

Un échantillon biaisé

En 2012, l'indice de fécondité de la France, c'est-à-dire le nombre d'enfants par femme, était d'environ 2,01. Une enquête sur les mères de votre lycée donne un indice supérieur.

Comment expliquer cela ?

Surreprésentativité

En 2012, l'alcoolémie positive d'un conducteur est présente dans des accidents causant 30,9 pour cent des tués sur la route. Elle est donc absente dans 69,1 pour cent des cas.

Qu'est-ce qui est le plus dangereux ?



SOLUTIONS

Chapitre SP2

Échantillonnage

Auto-évaluation

- 1) 1) Baltimore
2) Londres
3)

héros	J	T	E
fréquence à Londres	0,27	0,56	0,17
fréquence à Baltimore	0,38	0,34	0,28

- 2) 1) 1/10 4) non
2) non 5) oui
3) non

S'entraîner

- 1) 1) 513 3) 437/513
2) 97/100

2) 119/219

3) 73/115

4) oui

5) non

6) 0,1

7) 1) 32,4 2) 100

18) 1)

a) $n = 480, p = 0,4, f_0 = 0,44$

b) oui

c) $[0,35; 0,45]$

d) oui

2) conforme

40

1) oui

2) $p \in [23\%; 29\%]$

Auto-évaluation QCM

55) (b) (c)

57) (d)

59) (c)

61) (a)

63) (d)

65) (c)

56) (c)

58) (a)

60) (e)

62) (c) (d)

64) (c)