Partir d'un bon pied

Voir corrigés en fin de manuel

A Interpréter un intervalle de fluctuation à 95 % d'une fréquence

Voir AP page 250

Une étude, réalisée en France en 2010, affirme que 30 % des malades atteints de la rougeole ont été hospitalisés.
Dans une région A, on compte 39 malades hospitalisés parmi les 100 personnes atteintes de la maladie.
Dans la région B, 89 des 150 malades ont été hospitalisés.
On rappelle que l'intervalle de fluctuation à 95 % d'une fréquence est $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$.

Vrai ou faux ? Justifier la réponse.

a. Pour la région A, la taille de l'échantillon est 39.
b. $p=0{,}3$.
c. D'après les résultats obtenus dans la région A, on rejette l'affirmation de l'étude avec un risque d'erreur de 5 % .
d. D'après les résultats obtenus dans les deux régions, on rejette l'affirmation de l'étude avec un risque d'erreur de 5 % .

B Déterminer un intervalle de fluctuation à 95 % à l'aide de la loi binomiale

Voir AP page 250

Vrai ou faux ? Justifier la réponse.

1 Soit X une variable aléatoire suivant la loi binomiale $\mathscr{B}(150\,;0{,}6)$.
L'intervalle $\left[\frac{a}{n}\,;\frac{b}{n}\right]$ est l'intervalle de fluctuation à 95 % de la fréquence correspondant à X.

a. $a=86$ **b.** $b=104$
c. $P(X\geqslant b+1)>0{,}25$ **d.** $P(a\leqslant X\leqslant b)\geqslant 0{,}95$

2 Un journal régional affirme qu'en 2011, la probabilité de réussite au Bac en Terminale ES dans cette région est 0,6 .
On interroge au hasard 150 candidats de ES dans cette région.
95 d'entre eux ont obtenu le Bac.
a. La différence entre la fréquence observée et celle annoncée par le journal est significative : on rejette l'affirmation du journal avec un risque d'erreur de 5 % .
b. Étant donnés les résultats observés sur cet échantillon, on ne peut pas rejeter l'hypothèse que, dans cette région en 2011, la probabilité d'obtenir le Bac ES est 0,6 .

k	$P(X\leqslant k)$	k	$P(X\leqslant k)$
77	0,019	91	0,596
78	0,028	92	0,659
79	0,04	93	0,718
80	0,057	94	0,772
81	0,079	95	0,82
82	0,106	96	0,86
83	0,139	97	0,895
84	0,179	98	0,922
85	0,225	99	0,944
86	0,278	100	0,961
87	0,337	101	0,961
88	0,399	102	0,972
89	0,464	103	0,976
90	0,531	104	0,98

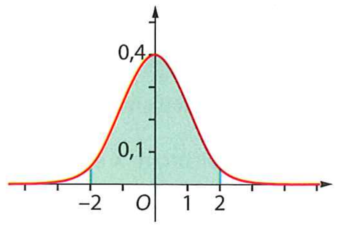
C Reconnaître les probabilités usuelles pour la loi normale $\mathcal{N}(0\,;1)$

Voir AP 37 page 249

La variable aléatoire Z suit la loi normale $\mathcal{N}(0\,;1)$.

Vrai ou faux ? Justifier la réponse.

a. L'aire coloriée représente $P(-2\leqslant Z\leqslant 2)$.
b. $P(-1\leqslant Z\leqslant 1)\approx 0{,}68$.
c. Si $P(-a\leqslant Z\leqslant a)\approx 0{,}95$, alors $a\approx 1{,}96$.
d. $P(Z\geqslant 3)=0{,}01$.

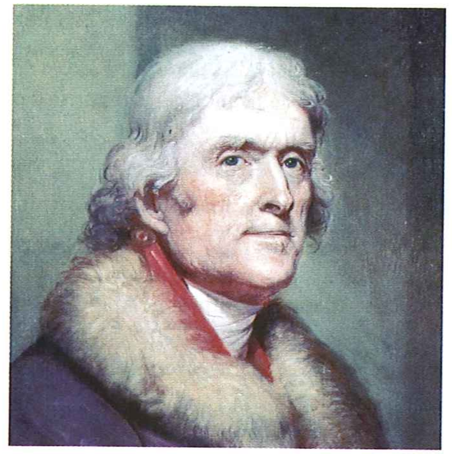


CHAPITRE 8 Échantillonnage

D'hier à aujourd'hui

Recensement : faut-il compter ou estimer ? L'exemple des États-Unis

Après un séjour de cinq ans en France, durant lequel il côtoie les élites des Lumières, Thomas Jefferson retourne aux États-Unis et participe à la rédaction de la Constitution américaine en 1787. Le pouvoir et l'autorité du gouvernement y sont fondés sur le peuple, dont la représentativité est organisée par le *Census Act*, qui stipule que les représentants à la Chambre sont répartis proportionnellement au nombre d'habitants dénombrés dans chaque état, à l'issue d'un recensement décennal. Jefferson réalisa le premier d'entre eux en 1790. Parallèlement, en France, dès 1785, Pierre-Simon de Laplace a estimé la population française en s'appuyant seulement sur les registres de baptêmes de quelques paroisses.
Le développement des théories statistiques, notamment après la Seconde Guerre Mondiale, permit de mettre au point des outils pour estimer la fiabilité d'un recensement. Ainsi, lors de cette guerre, le nombre de soldats noirs s'engageant dans l'armée américaine surprit les autorités car le nombre d'Américains noirs avait été systématiquement sous-évalué dans les recensements. Des outils sophistiqués (et coûteux !) ont, depuis, permis aux statisticiens de quantifier l'erreur commise lors d'un recensement et de prouver que cette erreur touche plus particulièrement certaines catégories de la population (pauvres, communautés spécifiques, etc.), celles-ci étant plus difficiles d'accès.
Le développement, depuis 50 ans, des enquêtes statistiques sur échantillons a parallèlement permis de mettre en place des méthodes bien plus efficaces d'estimation de la taille d'une population. Jugeant que la discrimination les pénalisait, certaines villes américaines ont ainsi demandé, en 1996, de remplacer le recensement par une enquête statistique de type sondage. Une controverse politique en a découlé.



Thomas Jefferson (1743-1826), rédacteur d'une partie de la Déclaration d'Indépendance et 3e président des États-Unis de 1801 à 1809.

Aujourd'hui, après bien des polémiques, la Cour suprême a tranché en 1999 en défaveur des enquêtes statistiques malgré leur plus grande efficacité sur le plan scientifique. En revanche, beaucoup d'autres pays ont passé le cap : le Canada, où le recensement a un poids politique important, utilise depuis les années 80 des enquêtes statistiques pour améliorer ses recensements.
Et depuis 2004, en France, l'Insee a remplacé son recensement exhaustif de la population française tous les neuf ans par une enquête statistique annuelle touchant environ 14 % de la population.

Les statistiques sont vraies quant à la maladie et fausses quant au malade.
Elles sont vraies quant aux populations et fausses quant à l'individu.

Léon Schwartzenberg (1923-2003)

Activité 1 Découvrir une nouvelle variable aléatoire

TICE

Dans une pâtisserie industrielle, un robot fabrique des petites meringues.
Un contrôleur affirme que 20 % des meringues sont cassées et que le robot doit être remplacé. Pour vérifier son affirmation, on prélève au hasard plusieurs échantillons de 10 meringues.

1 a. Soit X la variable aléatoire égale au nombre de meringues cassées dans un échantillon. Quelle est la loi suivie par la variable aléatoire X ?
Quelles sont les valeurs prises par la variable X ? Calculer l'espérance de X.

b. À l'aide d'un tableur, on simule le tirage de 100 échantillons de 10 meringues, notées de B1 à B10, en entrant en **A2** :
=ENT(ALEA()+0,20)
que l'on recopie sur la plage **A2:J101** .
Expliciter la signification de la réponse 0 ou 1 par cette formule.

	A	B	C	D	E	F	G	H	I	J	K
1	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	f_i
2	0	1	1	0	1	1	1	0	0	1	0,6
3	0	0	0	1	0	0	0	0	0	0	0,1

2 a. On calcule la fréquence d'apparition d'une meringue cassée en colonne K par saisie en **K2** de la formule =SOMME(A2:J2)/10 et on recopie vers le bas jusqu'en ligne 101.

On définit alors la variable aléatoire « fréquence » $F = \frac{1}{10}X$.

Les valeurs prises par F sont les **fréquences** d'apparition d'une meringue cassée au cours des 10 tirages.

b. Estimer l'espérance de F en calculant en cellule **M1** =MOYENNE(K2:K101) .

c. On obtient une autre simulation en tapant $\boxed{\text{F9}}$. Observer les fluctuations des fréquences et de l'espérance.
Recommencer la simulation avec un prélèvement de 50 meringues et observer les fluctuations des fréquences.

3 Soit X une variable aléatoire suivant une loi binomiale $\mathscr{B}(n\,;p)$. Rappeler l'espérance et l'écart type de X.

Soit $F = \frac{1}{n}X$. L'espérance E(F) dépend-elle de la taille de l'échantillon ?

Activité 2 Découvrir un nouvel intervalle de fluctuation à 95 %

Philémon se présente aux élections des délégués au conseil de la vie lycéenne de son lycée. Il est élu avec 60 % des voix. Donc la proportion d'élèves ayant voté pour Philémon est $p = 0{,}6$. On interroge n élèves du lycée.
Soit X_n la variable aléatoire égale au nombre d'élèves ayant voté pour Philémon, parmi les n élèves interrogés.

1 La variable aléatoire X_n suit la loi binomiale $\mathscr{B}(n\,;0{,}6)$.

a. Calculer l'espérance et l'écart type de X_n .

b. Soit Z_n la variable aléatoire **centrée réduite** de X_n , définie par $Z_n = \frac{X_n - 0{,}6n}{\sqrt{0{,}24n}}$.

Montrer que : $a \leqslant Z_n \leqslant b \iff 0{,}6n + a\sqrt{0{,}24n} \leqslant X_n \leqslant 0{,}6n + b\sqrt{0{,}24n}$.

c. On considère la variable aléatoire « fréquence » $F_n = \frac{1}{n}X_n$ qui, à un échantillon de taille n, associe la fréquence du caractère dans l'échantillon. Montrer que $a \leqslant Z_n \leqslant b \iff 0{,}6 + a \times \frac{\sqrt{0{,}24}}{\sqrt{n}} \leqslant F_n \leqslant 0{,}6 + b \times \frac{\sqrt{0{,}24}}{\sqrt{n}}$.

d. On prend pour valeurs de a et b : $a = -1{,}96$ et $b = 1{,}96$. Donner l'encadrement de F_n .

2 Le **théorème de Moivre-Laplace** énonce que, lorsque n prend de très grandes valeurs, la variable aléatoire Z_n suit approximativement la loi normale $\mathscr{N}(0\,;1)$. Voir chapitre 7 page 210

a. Déterminer la valeur arrondie au centième près du nombre a tel que $P(Z \in [-a\,;a]) \approx 0{,}95$.

b. En utilisant **1 c.**, donner, lorsque n prend de très grandes valeurs, un encadrement probable de F_n .

Soit I_n l'intervalle $\left[0{,}6 - 1{,}96 \times \frac{\sqrt{0{,}24}}{\sqrt{n}}\,;\,0{,}6 + 1{,}96\frac{\sqrt{0{,}24}}{\sqrt{n}}\right]$. Alors, $P(F_n \in I_n) \approx 0{,}95$.

Cet intervalle est appelé **intervalle de fluctuation asymptotique à 95 % de la variable aléatoire F_n** .

3 On interroge 10, 200, 400, puis 1 000 et 1 200 élèves du lycée. Déterminer, dans chacun des cas, l'intervalle de fluctuation asymptotique à 95 % de F_n . *On arrondira les bornes au millième près.*
Comparer les bornes et les longueurs des intervalles obtenus quand n prend de grandes valeurs.

Activité 3 Comparer deux intervalles de fluctuation

TICE

On se propose de comparer, lorsque n prend de très grandes valeurs, les deux intervalles de fluctuation à 95 % de la fréquence, notés I_n et J_n . L'intervalle J_n est défini par $\left[p - \frac{1}{\sqrt{n}}\,;p + \frac{1}{\sqrt{n}}\right]$ et l'intervalle I_n de fluctuation asymptotique est défini par $\left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

1 On reprend **l'Activité 2**. La proportion d'élèves ayant voté pour Philémon est $p = 0{,}6$.
X_n est la variable aléatoire égale au nombre d'élèves ayant voté pour Philémon, parmi les n élèves interrogés. On pose $F_n = \frac{1}{n}X_n$.

	A	B	C	D	E
1	n	100		probabilité que F_n soit dans les Intervalles de fluctuation	
2	p	0,6	calcul des probabilités	I_n	J_n
3	k		$P(X_n=k)=P(F_n=k/n)$	$P(a_n \leq F_n \leq b_n)$	$P(c_n \leq F_n \leq d_n)$
4	0				

On étudie les lois de probabilité des variables aléatoires X_n et F_n sur tableur, pour $n = 10$, 200 , 400 et 1 000.

a. De **A4** à **A1004**, entrer les valeurs de k de 0 à 1 000.

b. Entrer en cellule **B4** = A4/$B$1 ,
puis en **C4** = LOI.BINOMIALE (A4 ; $B$1 ; $B$2 ; FAUX) .

c. En **F2**, **G2** et en **F5**, **G5**, calculer les bornes des intervalles de fluctuations, puis en **H2** et **H5**, les amplitudes de ces intervalles.

d. Entrer en **D4** = SI (((B4>=$F$2)*(B4<=$G$2)) ; C4 ; 0)
et en **E4** = SI (((B4>=$F$5)*(B4<=$G$5)) ; C4 ; 0) .
Expliquer le résultat obtenu dans ces cellules.

F	G	H
borne a_n	borne b_n	$b_n - a_n$
$P(F_n \in I_n) =$		
borne c_n	borne d_n	$d_n - c_n$
$P(F_n \in J_n) =$		

e. Recopier la ligne 4 jusqu'à la ligne 1 004.
Calculer alors en cellules **G3** et **G6**, la probabilité que F_n appartienne à chacun des intervalles de fluctuation.

2 a. En changeant les valeurs de n, en **B1**, compléter le tableau ci-contre.

b. Vers quelle valeur semblent se rapprocher les probabilités $P(F_n \in I_n)$ et $P(F_n \in J_n)$ lorsque n prend de grandes valeurs ?
Les comparer.

n	a_n	b_n	$P(F_n \in I_n)$	$b_n - a_n$	c_n	d_n	$P(F_n \in J_n)$	$d_n - c_n$
10								
200								
400								
1 000								

Activité 4 Simplifier l'intervalle de fluctuation asymptotique

On se propose de montrer que, à certaines conditions sur la proportion p d'un caractère dans une population, on peut « simplifier » l'intervalle asymptotique I_n et prendre pour intervalle J_n , intervalles vus à l'**Activité 3**.

1 On considère la fonction du second degré g , définie sur l'intervalle $[0\,;1]$ par $g(x) = x(1-x)$.

a. Montrer que, pour tout réel x de $[0\,;1]$, on a $g(x) \leqslant \frac{1}{4}$.
En remarquant que $1{,}96 \leqslant 2$, en déduire que, pour tout réel p de $[0\,;1]$, $1{,}96 \times \sqrt{p(1-p)} \leqslant 1$.

b. Montrer alors l'inclusion $I_n \subset J_n$ et comparer les probabilités $P(F_n \in I_n)$ et $P(F_n \in J_n)$.

2 On cherche à quantifier l'erreur commise en remplaçant I_n par J_n et à la limiter.

a. Montrer que la différence des amplitudes des intervalles I_n et J_n est égale à $\frac{2\left(1 - 1{,}96 \times \sqrt{g(p)}\right)}{\sqrt{n}}$.

b. À l'aide de la calculatrice, déterminer les valeurs de p pour avoir $2\left(1 - 1{,}96 \times \sqrt{g(p)}\right) \leqslant 0{,}1$.

c. En déduire l'intervalle où se situe la proportion p pour avoir $\frac{2\left(1 - 1{,}96 \times \sqrt{g(p)}\right)}{\sqrt{n}} \leqslant \frac{0{,}1}{\sqrt{n}}$.

3 Justifier que, à condition que p appartienne à l'intervalle trouvé en **2**, lorsque n prend de grandes valeurs, on peut prendre l'intervalle « simplifié » J_n au lieu de l'intervalle de fluctuation I_n .

Activité 5 Estimer p à partir d'un échantillon TICE

Un ami a prêté à Quentin son lecteur mp3 contenant 1 000 chansons.
La proportion p de chansons françaises contenues dans le lecteur est inconnue.
Quentin se propose d'écouter vingt fois 100 chansons et de calculer la fréquence d'apparition d'une chanson française.
On désire savoir s'il peut estimer ainsi la proportion de chansons françaises contenues dans le lecteur mp3.

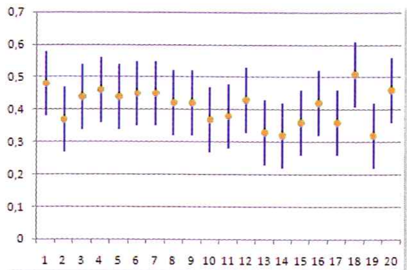
1 Montrer l'équivalence : $p-\frac{1}{\sqrt{n}} \leqslant f \leqslant p+\frac{1}{\sqrt{n}} \iff f-\frac{1}{\sqrt{n}} \leqslant p \leqslant f+\frac{1}{\sqrt{n}}$.

2 On a simulé sur un tableur, le tirage de 20 échantillons de taille 100 et, pour chacun des 20 échantillons, on a calculé en ligne 3 les fréquences f d'apparition d'une chanson française, puis en lignes 1 et 2, on a calculé les bornes de l'intervalle $K_n=\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]$.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	*f*-1/racine(*n*)	0,38	0,27	0,34	0,36	0,34	0,35	0,35	0,32	0,32	0,27	0,28	0,33	0,23	0,22	0,26	0,32	0,26	0,41	0,22	0,36
2	*f*+1/racine(*n*)	0,58	0,47	0,54	0,56	0,54	0,55	0,55	0,52	0,52	0,47	0,48	0,53	0,43	0,42	0,46	0,52	0,46	0,61	0,42	0,56
3	**Fréquence *f***	**0,48**	**0,37**	**0,44**	**0,46**	**0,44**	**0,45**	**0,45**	**0,42**	**0,42**	**0,37**	**0,38**	**0,43**	**0,33**	**0,32**	**0,36**	**0,42**	**0,36**	**0,51**	**0,32**	**0,46**

À l'aide d'un tableur, on a représenté les « fourchettes » correspondant à chacun des échantillons, en sélectionnant toute la plage de B1 à L3 et en insérant un graphique Boursier.
En appuyant sur F9, on obtient d'autres échantillons.
À la lecture du graphique ci-contre, proposer une valeur pour p.

Boursier



3 On admet que la proportion de chansons françaises est $p=0{,}4$.
Sur ces 20 échantillons représentés ci-contre, quelle est la proportion d'intervalles ne contenant pas la proportion $p=0{,}4$?

Activité 6 Déterminer un intervalle de confiance à 95% de p

Un laboratoire pharmaceutique fabrique des comprimés. On considère un stock important de comprimés dont une proportion p, inconnue, est conforme aux normes imposées. On prélève au hasard et avec remise 100 comprimés.

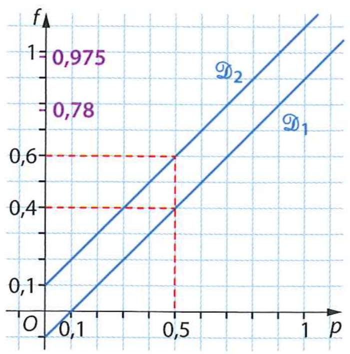
X est la variable aléatoire égale au nombre de comprimés conformes et $F=\frac{1}{100}X$ est la variable aléatoire fréquence.

On considère $J_{100}=\left[p-\frac{1}{10}\,;\,p+\frac{1}{10}\right]$, où la variable aléatoire F prend ses valeurs avec une **probabilité de 0,95** (vu dans les Activités 3 et 4).

1 On a représenté, sur l'intervalle $[0\,;1]$, la droite $\mathscr{D}_1$ d'équation $y=p-\frac{1}{10}$ et la droite $\mathscr{D}_2$ d'équation $y=p+\frac{1}{10}$. On suppose que $p=0{,}5$.
Lire, sur l'axe des ordonnées, l'intervalle J_{100} de fluctuation de la fréquence F.

2 a. On suppose que la **proportion p** de médicaments du stock est **inconnue**.
On réalise le tirage d'un échantillon de 100 comprimés prélevés au hasard : 78 % sont conformes. Ainsi, $f=0{,}78$ est un résultat observé pour la variable fréquence F. On parle de *réalisation* de F.
Déterminer graphiquement les bornes de l'intervalle $\left[f-\frac{1}{10}\,;\,f+\frac{1}{10}\right]$.

b. Si $f=0{,}975$, peut-on déterminer graphiquement $\left[f-\frac{1}{10}\,;\,f+\frac{1}{10}\right]$?



En faisant « glisser » les droites $\mathscr{D}_1$ et $\mathscr{D}_2$ vers le haut, déterminer graphiquement une valeur de n, autre que 100, pour que l'intervalle puisse être déterminé.

3 D'après l'Activité 5, en généralisant à toute fréquence observée f, on peut écrire :

$$F\in\left[p-\frac{1}{10}\,;\,p+\frac{1}{10}\right] \iff p\in\left[F-\frac{1}{10}\,;\,F+\frac{1}{10}\right].$$

Que peut-on en déduire pour la probabilité de l'intervalle aléatoire $\left[F-\frac{1}{10}\,;\,F+\frac{1}{10}\right]$ de contenir p ?

L'intervalle $\left[f-\frac{1}{10}\,;\,f+\frac{1}{10}\right]$ est une *réalisation* de l'intervalle aléatoire $\left[F-\frac{1}{10}\,;\,F+\frac{1}{10}\right]$.

L'intervalle $\left[f-\frac{1}{10}\,;\,f+\frac{1}{10}\right]$, qui n'est pas aléatoire, s'appelle l'**intervalle de confiance de la proportion p** au niveau de confiance 0,95. On parle aussi de « fourchette ».

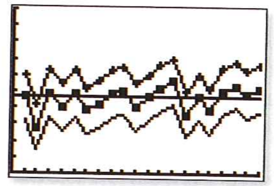
Activité 7 Simuler un sondage politique

Lors d'une élection, un candidat A est assuré qu'il remportera entre 40 % et 60 % des voix.
20 sondages sont effectués auprès de n personnes afin de connaître leurs intentions de vote pour le candidat A.
Pour chaque sondage, on observe une fréquence f de votants pour le candidat A.
À l'aide d'un programme sur calculatrice (ci-contre), on simule ces sondages et on représente les « fourchettes » $\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]$ au niveau de confiance 0,95.

```
PROGRAM:CONF
:EffListe L1,L2,
L3,L4
:suite(X,X,1,20,
1)→L1
:Prompt N
:0.4+0.2*NbrAléa
t→P
:For(J,1,20)
:0→E
:For(I,1,N)
:partEnt(NbrAléa
t+P)+E→E
:End
:E/N→L2(J)
:E/N-1/√(N)→L3(J
)
:E/N+1/√(N)→L4(J
)
:End
```

1 a. Dans ce programme, que calcule la formule =0,4+0,2*NbrAléat ?
b. PartEnt(NbrAléat+p) simule l'intention de vote d'une personne tirée au hasard : 1 s'il vote A et sinon 0. Que représente la variable E ?
c. Que contiennent les listes L2, L3 et L4 à la fin de l'exécution du programme ?

2 On a représenté ci-contre les trois listes (L1, L2), (L1, L3) et (L1, L4), pour $n=50$, ainsi que la droite d'équation $Y1=p$.
La proportion des intentions de vote est-elle toujours comprise dans la fourchette ? Estimer la proportion de « fourchettes » ne contenant pas p.
Quel lien a-t-elle avec le niveau de confiance ?



1 Fluctuation d'échantillonnage et prise de décision

a Intervalle de fluctuation au seuil de 95 %

On s'intéresse à un caractère de **proportion p** dans une population donnée. On considère la variable aléatoire X_n qui, à un **échantillon aléatoire de taille n** de cette population, associe le nombre d'individus de cet échantillon ayant ce caractère et la variable aléatoire F_n qui, à ce même échantillon, associe la **fréquence** du caractère étudié dans cet échantillon.

Remarque Si la variable aléatoire X_n prend la valeur k, alors la variable aléatoire F_n prend les valeurs $\frac{k}{n}$.

Définition On appelle **intervalle de fluctuation de F_n au seuil de 95 %**, tout intervalle $[\alpha ; \beta]$ tel que $P(F_n \in [\alpha ; \beta]) \geqslant 0{,}95$.

k	$P(X_{100} = k)$
60	0,020988576
61	0,033978998
62	0,053045586
63	0,079880042
64	0,116078606
65	0,162858288
66	0,220742239
67	0,289281444
68	0,366892014
69	0,450876399
70	0,537660264
71	0,623221821
72	0,703633839
73	0,77560076
74	0,836869896
75	0,886429818
76	0,924469233
77	0,952134261
78	0,971168747
79	0,983537147
80	0,991112792

EXEMPLE

Dans un supermarché, 70 % des produits mis en vente respectent les normes d'étiquetage ; ainsi, $p = 0{,}7$. Dans le chariot de plusieurs clients, on prélève au hasard cent produits ; ainsi $n = 100$. On note X_n le nombre d'articles bien étiquetés lors de ce prélèvement et F_n la proportion correspondante pour ces 100 produits. X_n suit une loi binomiale $\mathcal{B}(n ; p)$.
D'après le cours de Première et la table ci-contre, on peut déterminer l'**intervalle de fluctuation à 95 %** de F_n. Il s'agit de l'intervalle $\left[\frac{61}{100} ; \frac{79}{100}\right]$.
Donc, dans **au moins 95 % des cas**, le prélèvement dans les chariots fournira entre **61 % et 79 %** des articles correctement étiquetés.

Voir AP 37 page 249

REMARQUE

Pour déterminer un tel intervalle de fluctuation, il est nécessaire de connaître la table de la loi binomiale correspondante (ou de disposer d'un ordinateur ou d'une calculatrice).

b Intervalle de fluctuation asymptotique au seuil de 95 %

Définition On appelle **intervalle de fluctuation asymptotique au seuil de 95 %** de la variable aléatoire F_n, l'intervalle :

$$I_n = \left[p - 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right].$$

En pratique La probabilité $P(F_n \in I_n)$ peut être raisonnablement approchée par 0,95 quand **les critères** $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ **sont respectés.**

C'est un intervalle **déterministe**, c'est-à-dire non aléatoire, construit à partir de la proportion p et de la taille de l'échantillon n, contenant la variable aléatoire F_n avec une probabilité d'autant plus proche de 0,95 que la taille de l'échantillon n est grande.

EXEMPLE

Avec $p = 0{,}7$ et $n = 100$, alors $np = 70 \geqslant 5$ et $n(1-p) = 30 \geqslant 5$.

Dans ce cas, les bornes de l'intervalle sont :

$$p - 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}610 \quad \text{et} \quad p + 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}799.$$

Donc, avec une probabilité **proche de 0,95**, le prélèvement dans les chariots fournira entre 61 % et 80 % d'articles correctement étiquetés.

Note Cet intervalle est simple à déterminer, mais $P(F_n \in I_n)$, bien que proche de 0,95, peut lui être inférieure ou supérieure. Il faut donc choisir avec soin les valeurs approchées.

Déterminer et interpréter un intervalle de fluctuation asymptotique

Exercice corrigé

Énoncé

Une enquête a révélé que 60 % des ventes de cahiers de format 24×32 sont des cahiers de la marque C. Une petite papeterie prévoit de vendre 30 cahiers 24×32. Elle désire connaître, avec une probabilité proche de 0,95, le nombre de cahiers 24×32 de la marque C qu'elle vendra. Un gérant d'un supermarché qui prévoit de vendre 1 000 cahiers 24×32 se pose la même question.

1 Pour chacun des points de vente, déterminer l'intervalle de fluctuation asymptotique, au seuil de 95 %, de la fréquence de cahiers 24×32 de la marque C vendus : I_{30} et $I_{1\,000}$.
Comparer les amplitudes des deux intervalles.

2 Interpréter les résultats obtenus.

Points méthode

1 Avant de déterminer ces intervalles, vérifier les **critères d'approximation** qui permettent d'utiliser l'intervalle de fluctuation asymptotique à 0,95.
$n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.
On peut calculer les bornes de l'intervalle :

Sur TI™ :

```
PROGRAM:FLUCTU
:Prompt N,P
:P-1.96*√(P(1-P)
)/√(N)→A
:P+1.96*√(P(1-P)
)/√(N)→B
:Disp A,B
```

Sur Casio :

```
======FLUCTU ======
"N":?→N:"P":?→P↵
P-1.96√(P(1-P))÷√N→A↵
P+1.96√(P(1-P))÷√N→B↵
A◢
B◢
```

2 Pour interpréter, revenir au concret en multipliant les bornes obtenues par la taille de l'échantillon.

Solution

1 La proportion de cahiers de la marque C dans les ventes est $p = 0{,}6$.
Soit I_n l'intervalle asymptotique à 95 % de la proportion de cahiers de la marque C.
Pour la papeterie : $n = 30$ et $p = 0{,}6$, $np > 5$ et $n(1-p) > 5$, donc n et p vérifient les critères.

$$1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}175. \quad \text{Ainsi, } I_{30} \approx [0{,}425 ; 0{,}775].$$

L'amplitude de l'intervalle I_{30} est $0{,}775 - 0{,}425 = 0{,}35$.
Pour l'hypermarché : $n = 1\,000$ et $p = 0{,}6$, donc n et p vérifient largement les critères.

$$1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}03. \quad \text{Ainsi, } I_{1\,000} \approx [0{,}57 ; 0{,}63].$$

L'amplitude de l'intervalle $I_{1\,000}$ est $0{,}63 - 0{,}57 = 0{,}06$.
L'amplitude est bien plus faible quand l'échantillon est de grande taille.

2 Dans une papeterie ayant en stock 30 cahiers de ce format, le représentant de la marque C peut vendre entre 42,5 % et 77,5 % du stock de cahiers, soit entre **13 et 23** cahiers de la marque C, avec une **probabilité proche de 0,95**.
Dans un hypermarché ayant un stock de 1 000 cahiers de ce format, le représentant de la marque C peut vendre entre 57 % et 63 % du stock de 1 000 cahiers, soit entre **570 et 630** cahiers de la marque C, avec une **probabilité proche de 0,95**.

Exercices d'application

Voir exercices 17 à 21

1 Les acariens sont responsables de 70 % des asthmes allergiques de l'enfant. On choisit au hasard un groupe de 300 enfants asthmatiques.
a. Déterminer l'intervalle de fluctuation asymptotique à 95 % de la proportion des asthmes dus aux acariens.
b. Déterminer le nombre d'enfants présentant un asthme allergique dû aux acariens que l'on peut obtenir avec une probabilité proche de 0,95.

2 Un fabricant affirme qu'au moins 80 % des pièces qui sortent de son usine ne contiennent aucun défaut. On prélève, au hasard, un échantillon de 300 pièces.
a. Déterminer l'intervalle de fluctuation asymptotique à 95 % de la proportion de pièces sans défaut.
b. Déterminer le nombre de pièces sans défaut que l'on peut obtenir avec une probabilité proche de 0,95.

c Décision à partir de la fréquence d'un échantillon

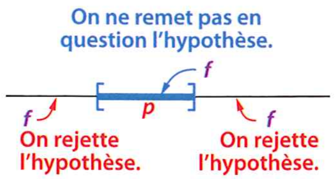
On fait l'hypothèse que la **proportion** d'un caractère étudié dans une population est p. On note F_n la variable aléatoire « fréquence », qui, à un échantillon aléatoire quelconque de taille n, associe la fréquence du caractère étudié dans cet échantillon.

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

$$I_n = \left[p - 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \ ; \ p + 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right].$$

On prélève dans la population **un échantillon de taille n** sur lequel on calcule la **fréquence observée f du caractère** : f est une *réalisation* de F_n.
On vérifie que les critères d'approximation permettant d'utiliser I_n sont respectés.
D'après la partie b page 236, dans ces conditions, la probabilité que F_n soit dans l'intervalle I_n est proche de 0,95.
Cet intervalle permet alors de fixer des seuils de décision : **la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, est proche de 0,05**.

Propriété ▸ Si la **fréquence observée f n'appartient pas** à l'intervalle I_n, alors la fréquence est trop éloignée de la proportion p : la différence est jugée significative.
On **rejette** donc l'hypothèse selon laquelle la proportion est p dans l'ensemble de la population avec un **risque d'erreur d'environ 5 %**.
▸ Si la **fréquence observée f appartient** à l'intervalle I_n, l'hypothèse selon laquelle la proportion est p dans la population **n'est pas remise en question**.



EXEMPLE

Fin 2010, le taux de chômage en France s'élevait, selon l'INSEE, à 9,7 % de la population active. Lors d'un reportage, un journaliste a interrogé 100 personnes au hasard sur un marché de Bergerac. 15 sont au chômage. Il affirme alors que Bergerac est plus touchée par le chômage.
Son raisonnement est-il satisfaisant ?
On fait l'hypothèse que le taux de chômage à Bergerac est identique au taux français, c'est-à-dire que la proportion de chômeurs est $p = 0{,}097$.
100 personnes sont interrogées au hasard, donc la taille de l'échantillon est $n = 100$.
Or $n \geqslant 30$, $np = 9{,}7 \geqslant 5$, $n(1-p) = 90{,}3 \geqslant 5$. Les critères sont respectés.

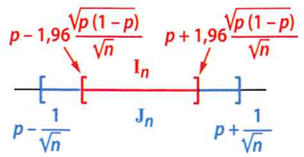
$$p - 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{100}} \approx 0{,}039 \quad \text{et} \quad p + 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{100}} \approx 0{,}155 ;$$

donc l'intervalle de fluctuation asymptotique est : $I_n = [0{,}039 ; 0{,}155]$.
La fréquence de chômeurs observée parmi les 100 personnes interrogées est $f = \frac{15}{100} = 0{,}15$. Or $f \in I_n$, donc **l'hypothèse** selon laquelle la proportion est $p = 0{,}097$ dans la population **n'est pas remise en question**.
La différence n'est **pas suffisamment significative** pour pouvoir affirmer que Bergerac est plus touchée : la fréquence observée peut être due à l'échantillonnage.

Remarque
Comparaison avec l'intervalle :

$$J_n = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right].$$

On a montré, dans l'Activité 4 page 233, que $I_n \subset J_n$, donc : $P(F_n \in J_n) \geqslant P(F_n \in I_n)$.
L'intervalle de fluctuation vu en Seconde est moins précis que l'intervalle de fluctuation asymptotique au seuil de 95 %, mais il peut être utilisé de la même manière et dans les mêmes conditions.



Décider à partir d'un intervalle de fluctuation asymptotique

Exercice corrigé

Énoncé

En 2011, le taux de réussite d'un concours national d'enseignement dans une matière littéraire est 53,7 %.
On étudie les résultats de deux centres de formations.
Dans le centre A :
48 candidats sur les 100 inscrits sont admis.
Dans le centre B :
252 candidats sur 515 inscrits sont admis.

1 Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % pour chacun des centres.
Prendre des valeurs approchées des bornes de ces intervalles, au millième près.

2 Ces centres sont-ils représentatif des résultats nationaux ?

Points méthode

❶ Pour déterminer un intervalle de fluctuation asymptotique au seuil de 95 % :
• on vérifie que n et p satisfont aux **critères d'approximation** ;
• pour calculer les **bornes de l'intervalle de fluctuation**, on utilise la formule du cours ou le programme vu à la page 237 ;
• on prend les valeurs approchées pour contenir l'intervalle exact.

❷ Pour décider du rejet ou non de l'hypothèse que l'échantillon A est représentatif :
• on calcule la fréquence du caractère dans l'échantillon ;
• on observe si cette fréquence appartient ou non à l'intervalle trouvé en **1**.

On conclut suivant la règle de décision donnée dans le cours.

Solution

1 • On fait l'hypothèse que le taux de réussite est $p = 0{,}537$ dans les deux centres A et B.

• **Dans le centre A**, la taille de l'échantillon est $n = 100$.
$np = 53{,}7 > 5$ et $n(1-p) \approx 46{,}3 > 5$; donc les critères sont respectés.

Comme $1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} = 1{,}96 \times \frac{\sqrt{0{,}537 \times 0{,}463}}{10} \approx 0{,}09773$,

l'intervalle de fluctuation asymptotique au seuil de 95 % est :
$I_{100} \approx [0{,}537 - 0{,}09773 ; 0{,}537 + 0{,}09773] = [0{,}43927 ; 0{,}63473]$.
Donc $\mathbf{I_{100} \approx [0{,}439 ; 0{,}635]}$.

• **Dans le centre B**, la taille de l'échantillon est $n = 515$.
$np \approx 276$ et $n(1-p) \approx 238$; donc les critères sont respectés.

Comme $1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} = 1{,}96 \times \frac{\sqrt{0{,}537 \times 0{,}463}}{\sqrt{515}} \approx 0{,}04306$,

alors $\mathbf{I_{515} \approx [0{,}493 ; 0{,}581]}$.

2 • **Pour le centre A**, le taux de réussite au concours est $f_A = 0{,}48$.
Or $0{,}48 \in [0{,}439 ; 0{,}635]$. La fréquence appartient à l'intervalle.
L'hypothèse $p = 0{,}537$ **n'est pas remise en question** pour cet échantillon.
La différence n'est pas suffisamment significative pour affirmer que le centre A n'est pas représentatif des résultats nationaux.

• **Pour le centre B**, le taux de réussite est $f_B = \frac{252}{515} \approx 0{,}4893$.
Or $0{,}4893 \notin [0{,}493 ; 0{,}581]$. La fréquence n'est pas dans l'intervalle.
On **rejette l'hypothèse** $p = 0{,}537$ **avec un risque d'erreur de 5 %** pour cet échantillon.
Le centre B n'est pas représentatif des résultats nationaux.

Exercices d'application

Voir exercices 22 à 24

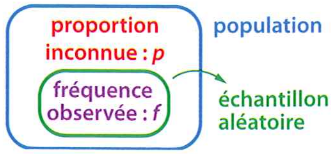
3 Au premier trimestre 2011, le taux d'équipement en France en *smartphone* est de 31,4 %.
On interroge au hasard 180 personnes à l'entrée d'une salle de cinéma :
72 possèdent un *smartphone*.
Peut-on en conclure que cet échantillon n'est pas représentatif ?

4 Le mildiou est une maladie qui touche environ 65 % des plants de tomates en période d'humidité.
Une jardinerie propose deux produits pour traiter cette maladie. Après application du produit A, 87 plants sur les 150 observés sont encore atteints par la maladie.
Après application du produit B, 223 des 378 plants sont encore malades.
Peut-on considérer que le produit A est efficace ? et le produit B ?

2 Estimation et intervalle de confiance

a Estimation ponctuelle

On considère une population donnée, dans laquelle la probabilité p de réalisation d'un caractère est **inconnue**. On définit la variable aléatoire F_n qui, à **un échantillon aléatoire quelconque de taille n**, associe la **fréquence f**, dans cet échantillon, du caractère étudié.



Définition Une **estimation ponctuelle** d'une proportion p **inconnue** est la réalisation f de la variable aléatoire F_n sur un échantillon aléatoire, c'est-à-dire la fréquence observée dans cet échantillon.

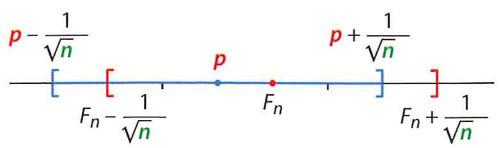
EXEMPLE

Sur 980 conducteurs français interrogés : 850 sont prêts à utiliser des biocarburants. Une **estimation ponctuelle** de la proportion p des conducteurs français prêts à utiliser des biocarburants est $f=\frac{850}{980}\approx 0{,}87$, soit 87 %.

b Intervalle de confiance au niveau de confiance 0,95

Lorsque les conditions d'approximation sont vérifiées, un intervalle de fluctuation à 95 % de F_n est $J_n=\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$.

Or $F_n\in J_n \Leftrightarrow p\in\left[F_n-\frac{1}{\sqrt{n}}\,;F_n+\frac{1}{\sqrt{n}}\right]$.



Ce dernier intervalle est aléatoire et, si n est grand :

$$P\left(p\in\left[F_n-\frac{1}{\sqrt{n}}\,;F_n+\frac{1}{\sqrt{n}}\right]\right)=P\left(F_n\in\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]\right)\geqslant 0{,}95.$$

Si f est une réalisation de la variable aléatoire F_n sur un échantillon, $\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$ est une réalisation de $\left[F_n-\frac{1}{\sqrt{n}}\,;F_n+\frac{1}{\sqrt{n}}\right]$.

Définition Sous les conditions d'approximation $n\geqslant 30$, $nf\geqslant 5$ et $n(1-f)\geqslant 5$, où f est la **fréquence observée** du caractère dans un **échantillon de taille n** :

l'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$, centré en la fréquence observée, est un **intervalle de confiance de p au niveau de confiance 0,95**.

Note **Précision de l'estimation et détermination d'une taille d'échantillon suffisante :**
La longueur de l'intervalle de confiance indique **la précision** obtenue.
Cette longueur est égale à $\frac{2}{\sqrt{n}}$.
Fixer une précision inférieure à t % permet de déterminer la taille minimale de l'échantillon.

EXEMPLE

Un intervalle de confiance de la proportion p des conducteurs prêts à utiliser des biocarburants au niveau de confiance 0,95 est :

$$\left[\frac{850}{980}-\frac{1}{\sqrt{980}}\,;\frac{850}{980}+\frac{1}{\sqrt{980}}\right]\approx[0{,}83\,;0{,}90].$$

Parmi un grand nombre d'intervalles de confiance obtenus à partir d'un grand nombre d'échantillons, 95 % environ contiennent p.
Au niveau de confiance 0,95, p est compris entre 83 % et 90 %.
Pour obtenir p avec une précision inférieure à 1 %, on résout :

$$\frac{2}{\sqrt{n}}\leqslant\frac{1}{100}\Leftrightarrow\sqrt{n}\geqslant 200\Leftrightarrow n\geqslant 40\,000.$$

Il faut donc interroger au moins 40 000 conducteurs.

Attention Ici, $n=900\geqslant 30$, $nf=850\geqslant 5$, et $n(1-f)=48\geqslant 5$. Les conditions d'approximation sont vérifiées.

Estimer la proportion inconnue d'un caractère

Exercice corrigé

Énoncé

La société SOAP veut estimer la taille du marché potentiel d'un nouveau produit de soin pour le corps auprès de femmes âgées de plus de 25 ans.
Un premier sondage est effectué auprès de 40 femmes âgées de plus de 25 ans. 24 sont satisfaites de ce nouveau produit. Lors d'un deuxième sondage auprès de 225 femmes âgées de plus de 25 ans, 150 sont satisfaites du nouveau produit.
Soit p la proportion de femmes de plus de 25 ans satisfaites de ce nouveau produit.

1 a. Calculer une **estimation ponctuelle** f de la proportion p, fournie par le premier échantillon.

b. À partir du 1er échantillon, déterminer une estimation de p par un intervalle de confiance au niveau de confiance 0,95. *Pour les bornes, on donnera les valeurs arrondies à 0,01 près.*

c. À l'issue de ce sondage, la société SOAP peut-elle prétendre que plus de la moitié des femmes de plus de 25 ans sont satisfaites de son nouveau produit ?

2 Répondre aux questions **1 a.** et **1 b.** pour le deuxième sondage.

3 Déterminer le nombre minimal de femmes de plus de 25 ans que l'on doit interroger pour obtenir une estimation de p avec une précision inférieure à 0,1.

Points méthode

1 a. L'estimation ponctuelle f de p est la fréquence observée dans un échantillon.

b. L'intervalle de confiance est $\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$.
On approche les bornes, en général, à 0,01 en prenant soin d'agrandir l'intervalle.

2 La **précision** de l'estimation de la proportion p est la longueur (ou **amplitude**) de l'intervalle de confiance, c'est-à-dire la différence de la borne supérieure et de la borne inférieure, soit $\frac{2}{\sqrt{n}}$.

Solution

1 a. Pour cet échantillon de taille $n=40$, une estimation ponctuelle de p est : $f=\frac{24}{40}=0{,}6$.

b. $n\geqslant 30$, $nf>5$ et $n(1-f)>5$. Un intervalle de confiance de p au niveau de confiance 0,95 correspondant à cet échantillon est $\left[0{,}6-\frac{1}{\sqrt{40}}\,;0{,}6+\frac{1}{\sqrt{40}}\right]$, soit, en arrondissant, $[0{,}44\,;0{,}76]$.

c. La borne inférieure de l'intervalle est 0,44. Il est donc plausible que la proportion des femmes satisfaites du produit soit inférieure à 0,5.
Donc la société ne peut pas raisonnablement prétendre que plus de la moitié des femmes âgées de plus de 25 ans sont satisfaites du nouveau produit.

2 a. Pour ce deuxième échantillon de taille $n=225$, une estimation ponctuelle de p est : $f=\frac{150}{225}=\frac{2}{3}$.

b. Une estimation par intervalle de confiance de p au niveau de confiance 0,95 de ce deuxième échantillon est $\left[\frac{2}{3}-\frac{1}{\sqrt{225}}\,;\frac{2}{3}+\frac{1}{\sqrt{225}}\right]$, soit environ $[0{,}60\,;0{,}74]$.

3 La précision de l'estimation de la proportion p est $\frac{2}{\sqrt{n}}$.
On résout l'inéquation : $\frac{2}{\sqrt{n}}<0{,}1\Leftrightarrow\sqrt{n}>\frac{2}{0{,}1}\Leftrightarrow n>400$.
Il faut donc interroger plus de 400 femmes de plus de 25 ans pour obtenir l'estimation de la proportion de femmes satisfaites du nouveau produit avec une précision inférieure à 0,1.

Exercices d'application

Voir exercices 26 à 30

5 Une enquête a été réalisée, auprès de 900 utilisateurs de tablettes numériques, sur leur consommation de presse en ligne. 270 utilisateurs sont prêts à télécharger la presse.
Soit p la proportion d'utilisateurs de tablettes numériques prêts à télécharger la presse en ligne. Calculer une estimation de p par un intervalle de confiance au niveau de confiance 0,95.

6 On interroge n possesseurs de GPS : 41 % habitent dans une grande agglomération.

a. Déterminer le nombre n de possesseurs de GPS que l'on doit interroger pour que la précision de l'intervalle de confiance de la proportion p de ceux qui habitent dans une grande agglomération soit inférieure à 1 %.

b. Calculer alors l'intervalle de confiance de p au niveau de confiance 0,95.

Faire le point

Capacités — Mise en œuvre

Savoir définir la variable aléatoire « fréquence » :
$$F_n = \frac{X}{n}.$$

Soit p la proportion d'un caractère dans une population.
Soit X_n la variable aléatoire qui, à cet échantillon, associe le nombre d'individus de cet échantillon ayant ce caractère.
La variable aléatoire F_n associe à cet échantillon la **fréquence** du caractère étudié dans cet échantillon.

Si X_n prend la valeur k, F_n prend la valeur $f = \frac{k}{n}$ et $P(X_n = k) = P\left(F_n = \frac{k}{n}\right)$.

On extrait un **échantillon aléatoire de taille n** de cette population sur lequel on observe une réalisation k de X_n et une réalisation f de F_n.

Savoir déterminer un intervalle de fluctuation asymptotique de F_n au seuil de 95 %.

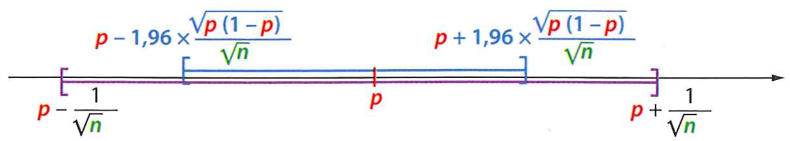
Soit p la proportion d'un caractère dans une population de taille n.
L'intervalle de fluctuation asymptotique de F_n au seuil de 95 % est :
$$\mathbf{I}_n = \left[p - 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \; ; \; p + 1{,}96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

Pour $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, alors $P(F_n \in \mathrm{I}_n) \approx 0{,}95$.

Savoir comparer les deux intervalles de fluctuation à 95%.

Soit l'intervalle de fluctuation vu en Seconde : $\mathbf{J}_n = \left[p - \frac{1}{\sqrt{n}} \; ; \; p + \frac{1}{\sqrt{n}}\right]$.

Alors : $\mathbf{I}_n \subset \mathbf{J}_n$ et $P(F_n \in \mathbf{J}_n) \geqslant P(F_n \in \mathbf{I}_n)$.



Savoir prendre une décision à partir de la fréquence d'un échantillon aléatoire.

On prélève dans la population **un échantillon de taille n** sur lequel on calcule **la fréquence observée f du caractère.**
- Si la fréquence observée **f n'appartient pas à l'intervalle I_n**, on **rejette l'hypothèse** selon laquelle la proportion est p dans l'ensemble de la population, avec **un risque d'erreur d'environ 5 %**.
- Si la fréquence observée **f appartient à l'intervalle I_n**, l'hypothèse selon laquelle la proportion est p dans la population **n'est pas remise en question.**

Savoir déterminer une estimation ponctuelle d'une proportion p inconnue et définir un intervalle de confiance pour une proportion p au niveau de confiance 0,95.

Une **estimation ponctuelle** d'une proportion p **inconnue** d'un caractère dans une population est la fréquence observée f du caractère dans cet échantillon.
Soit **f la fréquence observée** du caractère dans un **échantillon de taille n**.
Sous les conditions d'approximation $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 5$,
l'intervalle $\left[f - \frac{1}{\sqrt{n}} \; ; \; f + \frac{1}{\sqrt{n}}\right]$ est **un intervalle de confiance de p** au niveau de confiance 0,95.

Déterminer une taille d'échantillon suffisante pour obtenir une estimation d'une proportion p avec une précision de t %.

La **précision** de l'estimation de la proportion p est la longueur de l'intervalle de confiance, soit $\frac{2}{\sqrt{n}}$.
Pour obtenir une précision inférieure à t %, on résout l'inéquation :
$$\frac{2}{\sqrt{n}} \leqslant \frac{1}{100} \Leftrightarrow n \geqslant \left(\frac{200}{t}\right)^2.$$

Faire le point

QCM

Voir corrigés en fin de manuel

Pour chaque question, indiquer la seule bonne réponse.

7 On interroge 5 élèves choisis au hasard. La variable aléatoire X égale au nombre d'élèves ayant réussi le Bac parmi les 5 élèves interrogés, suit la loi $\mathscr{B}(5\,;0{,}75)$. On pose $F = \frac{1}{5}X$.
a. $P(X=4) = P(F=0{,}8)$. **b.** $P(F \geqslant 1) = 1 - P(X=0)$. **c.** La loi de F est la loi $\mathscr{B}(5\,;0{,}75)$.

8 Si I est un intervalle de fluctuation à 95 % de F et f une réalisation de F, alors :
a. il y a 95 % de chance que f appartienne à I. **b.** l'intervalle I est centré en f. **c.** f peut ne pas être dans l'intervalle I.

9 En posant $J_n = \left[p - \frac{1}{\sqrt{n}} \,; p + \frac{1}{\sqrt{n}}\right]$, l'amplitude de l'intervalle de fluctuation asymptotique à 95 % de F est :
a. moins grande quand n est petit. **b.** moins grande si p est petit. **c.** moins grande que J_n.

10 Pour $n = 100$ et $p = 0{,}5$, on considère I_n, l'intervalle de fluctuation asymptotique à 95 % de F.
a. La longueur de I_n est maximale. **b.** $I_n = [0{,}4\,;0{,}6]$. **c.** $[0{,}4\,;0{,}6] \subset I_n$.

11 On fait l'hypothèse que la proportion du caractère dans une population est $p = 0{,}7$.
L'intervalle de fluctuation à 95 % de F est $[0{,}55\,;0{,}85]$. La fréquence observée dans un échantillon est $f = 0{,}8$.
a. On accepte l'hypothèse avec un risque d'erreur de 5 %. **b.** On rejette l'hypothèse avec un risque d'erreur de 5 %. **c.** On ne peut pas rejeter l'hypothèse $p = 0{,}7$.

12 Dans une population, on suppose que la proportion d'un caractère est $p = 0{,}7$. Dans deux échantillons A et B de même taille, on observe les fréquences $f_A = 0{,}6$ et $f_B = 0{,}5$ pour ce caractère.
a. A est plus représentatif de la population que B. **b.** Les résultats sur A sont compatibles avec l'hypothèse. **c.** Les résultats sur B sont compatibles avec l'hypothèse.

13 Une estimation ponctuelle de la proportion p d'un caractère d'une population est :
a. une fréquence observée dans un échantillon. **b.** une fréquence observée dans la population. **c.** un intervalle.

14 Après avoir interrogé un échantillon de 100 personnes, un institut de sondage publie que l'estimation ponctuelle de la proportion p d'un caractère de la population est 0,65.
a. On peut affirmer que $p = 0{,}65$. **b.** $[0{,}55\,;0{,}75]$ est un intervalle de confiance de p au seuil de 0,95. **c.** La probabilité $P(p \in [0{,}55\,;0{,}75])$ est 0,95.

15 L'intervalle de confiance sur un échantillon de taille n est d'autant plus grand que :
a. la précision est petite. **b.** n est grand. **c.** la précision est grande.

16 La fréquence observée dans un échantillon de taille 25 est 0,76.
Alors, la précision de l'intervalle de confiance est :
a. 95 %. **b.** 0,4. **c.** 0,76.

→Les exercices portant un numéro jaune sont corrigés à la fin du manuel.

1 Intervalle de fluctuation à 95 %

17 Vrai ou faux ?

Dans un journal, un test est proposé au lecteur.
Chaque question a 3 réponses possibles : A, B, C.
Le jeu comporte 25 questions. Maxence décide de répondre au hasard à ce questionnaire à l'aide d'un dé à 6 faces parfaitement équilibré : lorsque les faces 1 ou 2 apparaissent Maxence répond A.
Soit X la variable aléatoire égale au nombre de réponses A en répondant au hasard aux 25 questions et soit $F = \frac{X}{25}$.

Préciser si les affirmations suivantes sont vraies ou fausses. Justifier la réponse.

a. La loi suivie par la variable aléatoire X est une loi binomiale $\mathscr{B}(3\,;\frac{1}{25})$.

b. Un intervalle de fluctuation de F au seuil de 95 % obtenu à l'aide des probabilités cumulées de X est :
$$I = [\,0{,}16\,;0{,}52\,].$$

c. L'intervalle de fluctuation asymptotique au seuil de 95 % est égal à I.

d. Il est possible que Maxence donne 60 % de réponses A sur les 25 questions.

18 Vrai ou faux ?

a. Pour diviser par deux l'amplitude de l'intervalle de fluctuation asymptotique, il faut multiplier la taille de l'échantillon par 2.

b. Lorsque $p = 0{,}5$ et si $n \geqslant 10\,000$, l'amplitude de l'intervalle de fluctuation est inférieure à 1 %.

c. Si on remplace le nombre 1,96 par 3, on obtient un intervalle de fluctuation asymptotique à 99 %.

19 Noël pour les seniors

La municipalité d'une ville offre, pour Noël, un panier-cadeau aux seniors de plus de 60 ans.
17 % des habitants ont plus de 60 ans.
Un employé de la mairie est chargé de distribuer ces paniers dans un quartier comptant 550 habitants.
Il souhaite connaître une « fourchette » du nombre de seniors de ce quartier afin de prévoir le nombre de paniers-cadeaux à préparer.

1 Déterminer un intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence des habitants âgés de plus de 60 ans dans ce quartier.
Les bornes seront arrondies au centième près.

2 Combien de paniers-cadeaux l'employé municipal doit-il prévoir ?

20 Prestations familiales

Environ 47 % des Français sont couverts par une prestation versée par les Caisses d'allocations familiales.
Un candidat aux élections municipales souhaite connaître la proportion d'allocataires dans sa ville de 20 000 habitants.

1 Déterminer l'intervalle de fluctuation asymptotique à 95 % de la fréquence des allocataires dans cette ville.
Arrondir les bornes à 10^{-3} près.

2 Ce candidat souhaite connaître la proportion d'allocataires de sa ville avec une précision de 1 %.

a. Calculer l'amplitude $f(n)$ de l'intervalle de fluctuation asymptotique en fonction du nombre n de personnes à interroger.

b. Résoudre l'équation $f(n) = 0{,}01$. Conclure.

21 Amplitude maximale

La proportion d'un caractère dans une population est p.
On choisit, au hasard dans cette population, un échantillon de taille $n = 1\,000$.

1 Exprimer, en fonction de p, l'amplitude de l'intervalle de fluctuation asymptotique à 95 %.

2 Étudier les variations sur l'intervalle $[\,0\,;1\,]$ de la fonction f définie par : $f(x) = x(1-x)$.

3 Montrer que f admet un maximum dont on précisera les cordonnées.

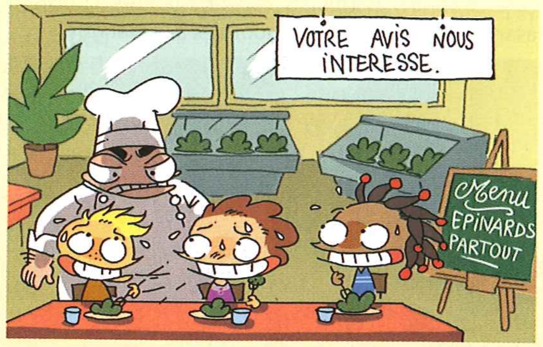
4 En déduire la valeur de la proportion p pour que l'amplitude de l'intervalle de fluctuation asymptotique soit maximale.

5 Un publicitaire interroge 1 000 adolescents au hasard afin de connaître la proportion d'entre eux préférant la barre chocolatée BonChoco. Calculer l'amplitude du plus grand intervalle de fluctuation asymptotique à 95 % de la proportion d'adolescents préférant cette barre chocolatée. Interpréter en termes de pourcentage.

2 Décision à partir de la fréquence d'un échantillon

22 Les élèves aiment les épinards

Lors de la commission « Menu » d'un lycée, le responsable du service de restauration affirme que 85 % des élèves ont apprécié le potage aux épinards.
On interroge au hasard les 35 élèves d'une classe.



Préciser si les affirmations suivantes sont vraies ou fausses. Justifier la réponse.

a. 26 élèves ont apprécié ce potage, donc on accepte l'hypothèse que 85 % des élèves apprécient ce potage.

b. 20 élèves ont apprécié ce potage, donc on rejette l'hypothèse que 85 % des élèves apprécient ce potage avec un risque de se tromper de 5 %.

c. Dans une autre classe de 35 élèves, seulement 40 % des élèves apprécient ce potage. L'affirmation du responsable est donc fausse.

23 Taux de chômage

Le taux de chômage en France en décembre 2011 a été évalué à 9,5 %. À la même période, on relève dans la ville A un taux de chômage de 10 % et de 9,8 % pour la ville B.
Un journaliste affirme, dans un article, que les taux de chômage de ces deux villes sont significativement différents de celui relevé dans la population française.
La population de la ville A est de 5 860 habitants et celle de la ville B de 35 000 habitants.
Déterminer les intervalles de fluctuation asymptotiques au seuil de 95 % des taux de chômage de chacune des villes.
Peut-on confirmer la remarque du journaliste ?

24 Résultats du Bac

En juin 2011, 77 % des candidats au baccalauréat de toutes les séries ont réussi l'examen au premier groupe d'épreuves.
Le nombre de candidats dans la série S était de 165 000 et de 105 300 en série ES. On note E_1 l'échantillon formé des candidats de la série S et E_2 celui de la série ES.

1 Déterminer les intervalles de fluctuation asymptotique à 95 % de la proportion de candidats admis à l'examen pour les échantillons E_1 et E_2.

2 79,7 % des candidats de la série S ont été admis.
Ce résultat est-il conforme à la proportion de candidats admis toutes séries confondues ?

3 70,8 % des candidats de la série ES ont été admis.
Doit-on mettre en défaut le résultat annoncé au baccalauréat de toutes les séries ?

25 Algorithmique

On reprend le programme donné dans « Savoir faire » page 237, qui permet d'obtenir les bornes de l'intervalle de fluctuation asymptotique à 95 %.

```
PROGRAM:FLUCTU
:Prompt N,P
:P-1.96*√(P(1-P))/√(N)→A
:P+1.96*√(P(1-P))/√(N)→B
:Disp A,B
```

PARTIE A Algorithme

On souhaite modifier ce programme afin de faire apparaître la règle de décision.

a. Ouvrir un nouveau programme appelé TEST puis recopier le programme précédent.

b. En ligne 5 du programme, ajouter une entrée supplémentaire correspondant à la fréquence F observée sur l'échantillon.

c. Poursuivre le programme par ce test.

```
Si A < F et F < B , alors
    Afficher « NON REJET »
Sinon
    Afficher « REJET »
```

Voir page 343

PARTIE B Application

L'office de tourisme d'une ville de la Côte d'Azur publie dans une brochure d'accueil que la ville bénéficie de 2 984 heures de soleil par an.

1 Calculer la proportion de jours de soleil dans cette ville en supposant qu'une année comporte 365 jours, un jour « de soleil » correspondant à une tranche de 12 h d'ensoleillement. *Arrondir à 1 % près.*

2 On choisit au hasard deux échantillons de 30 jours durant une année dans cette ville.
Dans le premier échantillon, on ne compte que 6 jours de soleil.
Dans le second échantillon on a 16 jours de soleil.

a. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la proportion de jours d'ensoleillement pour ces échantillons.

b. Quelle information sur l'affirmation de l'office de tourisme est fournie par chaque échantillon ? Commenter.

c. On considère l'échantillon formé par la réunion des deux échantillons précédents.
Que penser de l'affirmation de l'office de tourisme ?

3 Estimation et intervalle de confiance

26 Vrai ou faux ?

Le service de comptabilité d'un magasin étudie le montant des achats de la journée.
Pour un échantillon de 100 clients choisis au hasard, 37 ont dépensé plus de 125 euros.

a. La proportion p des clients du magasin dépensant plus de 125 euros est 37 %.

b. Un intervalle de confiance au niveau de confiance 0,95 de la proportion p des clients du magasin dépensant plus de 125 euros est $[0{,}27\,;0{,}47]$.

c. Dans 95 % des cas, la proportion p des clients du magasin dépensant plus de 125 euros est incluse dans l'intervalle $[0{,}27\,;0{,}47]$.

d. Soit p la proportion vue précédemment.
$$P(p\in[0{,}27\,;0{,}47])\approx 0{,}95.$$

27 Vrai ou faux ?

Une population présente un caractère avec une proportion p. On interroge au hasard 100 personnes et on détermine un intervalle de confiance I_1, au niveau de confiance 0,95.

a. Si l'on interroge 100 autres personnes, l'intervalle de confiance sera identique à I_1.

b. Les deux intervalles ont le même centre.

c. Ces deux intervalles peuvent n'avoir aucune valeur en commun.

d. Les amplitudes des deux intervalles de confiance sont égales.

28 Étude d'une part de marché

Un fabricant d'appareils photo numériques étudie la part du marché des appareils photo numériques de type reflex avant de décider l'arrêt éventuel de la fabrication.
Sur un échantillon de 1 500 acheteurs d'appareils photo numériques choisis au hasard, 210 achètent encore un appareil reflex.

1 Calculer une estimation ponctuelle de la part du marché des appareils reflex.

2 Déterminer un intervalle de confiance, au niveau de confiance 0,95, de la part du marché des appareils reflex parmi les appareils photo numériques.

3 Le fabricant décide d'arrêter la fabrication de ce type d'appareil si la borne supérieure de l'intervalle de confiance est inférieure à 15 %.
À l'issue de ce sondage, arrêtera-t-il la fabrication ?

29 Amélioration de la précision d'un sondage

Sur un marché, un stand propose des smoothies fabriqués sur place à différents parfums. Le vendeur souhaite connaître le parfum préféré de sa clientèle. Au cours d'une matinée, parmi 55 consommateurs choisis au hasard, 32 ont préféré les smoothies à la mangue.

1 Déterminer, à 10^{-4} près, les bornes de l'intervalle de confiance de la proportion p des clients préférant le smoothie à la mangue au niveau de confiance 0,95.

2 La précision de cet intervalle ne satisfait pas le vendeur. Elle ne lui permet pas de prévoir correctement son stock de mangues. Il souhaite réduire l'amplitude de l'intervalle de moitié.

a. Combien devra-t-il alors interroger de consommateurs ?

b. On suppose que la fréquence observée reste identique. Préciser le nouvel intervalle de confiance obtenu. Interpréter le résultat en termes de pourcentage.

30 Détermination de la taille d'un échantillon

Un laboratoire propose un nouveau type de vaccin pour lutter contre la grippe saisonnière.
Il souhaite connaître la taille minimale de l'échantillon qui testera ce vaccin pour obtenir une précision du taux de réussite du vaccin à moins de 1 %.

1 a. Calculer l'amplitude $f(n)$ de l'intervalle de confiance en fonction de la taille n d'un échantillon.

b. Représenter cette fonction sur l'écran de la calculatrice, ainsi que la droite d'équation $y=0{,}01$.
Déterminer graphiquement la valeur de n.

```
FENETRE
Xmin=0
Xmax=10000
Xgrad=1000
Ymin=-.01
Ymax=.05
Ygrad=.01
Xres=1
```

2 Résoudre, par le calcul, l'inéquation $f(n)\leqslant 0{,}01$. Conclure par une phrase.

31 Algorithmique

PARTIE A Détermination de l'intervalle de confiance

```
PROGRAM:FLUCTU
:Prompt N,P
:P-1.96*√(P(1-P)
)/√(N)→A
:P+1.96*√(P(1-P)
)/√(N)→B
:Disp A,B
```

Par analogie avec le programme ci-contre, on souhaite écrire un programme sur calculatrice permettant de déterminer l'intervalle de confiance, au niveau de confiance 0,95, de la proportion p d'un caractère d'une population.

a. Quelles doivent être les valeurs entrées en première ligne du programme ?

b. Quels sont les calculs qui devront être stockés en variable A et B ?

Application :
Pour connaître la part du marché p de pommes produites par la Chine annuellement, on choisit au hasard un échantillon de 500 pommes dont 150 sont produites en Chine.
Déterminer une estimation par intervalle de confiance au niveau de confiance 0,95 de p.

Exposition agricole à Nanning (Chine), 2011.

PARTIE B Détermination de la taille de l'échantillon

a. Soit T % la précision de l'estimation de la proportion p par intervalle de confiance au niveau de confiance 0,95 et soit n la taille de l'échantillon.
Écrire l'égalité vérifiée par T et n.

b. Écrire un programme, sur la calculatrice, permettant de déterminer la taille de l'échantillon, en suivant l'algorithme suivant :

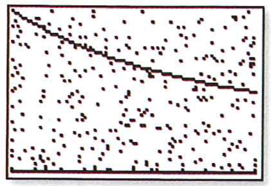
Entrer la précision demandée, T
Calculer $\left(\dfrac{200}{T}\right)^2$ et stocker ce résultat en N
Afficher N

Application :
On veut améliorer la précision de l'estimation de la part du marché de pommes produites par la Chine.
Calculer la taille minimale de l'échantillon pour que la précision de l'estimation soit inférieure à 1 %.

32 Estimation de ln(2) par la méthode de Monte Carlo

1 Justifier que $\ln(2)=\displaystyle\int_1^2\frac{1}{x}\,dx$. Interpréter.

2 Tracer la courbe de la fonction inverse sur l'écran de la calculatrice.
Pour le réglage de la fenêtre, choisir :
$1\leqslant X\leqslant 2$ et $0\leqslant Y\leqslant 1$.



3 Représenter un nuage de 300 points aléatoires obtenus à partir des listes L1 et L2 ainsi définies :

TI™	Casio
suite(1+NbrAleat, x, 1, 300, 1) → L1 Suite(NbrAleat, x, 1, 300, 1) → L2	Seq(1+Ran#, x, 1, 300, 1) → List 1 Seq(Ran#, x, 1, 300,1) → List 2

4 Calculer la proportion des points situés sous la courbe de la fonction inverse.

TI™	Casio
L3 = 1/L1 Somme(L2 ≤ L3)/300	1/List 1 → List 3 Sum(List 2 ≤ List 3)/300

De quelle aire obtient-on ainsi une estimation ponctuelle ?
Déterminer l'intervalle de confiance de la valeur de ln(2) au niveau de confiance 0,95.

5 On veut améliorer la précision de l'estimation à l'aide d'un tableur. Combien doit-on choisir de points aléatoires, si l'on souhaite obtenir une précision de 1% ?

33 Intervalle de confiance sur calculatrice

Les calculatrices utilisent un intervalle de confiance non simplifié (que l'on ne peut pas justifier), où f est la fréquence observée sur l'échantillon de taille n :
$$\left[f-1{,}96\sqrt{\frac{f(1-f)}{n}}\,;\,f+1{,}96\sqrt{\frac{f(1-f)}{n}}\right].$$

Sur 500 employés interrogés, 190 approuvent les réformes proposées par la direction. Déterminer un intervalle de confiance de la proportion p d'employés dans l'entreprise, approuvant les réformes de la direction.

TI™	Casio
stats TEST A:1-PropZint 1-PropZInt x:190 n:500 Niveau-C:.95 Calculs	STAT INTR Z 1-Prop ZInterval C-Level :0.95 x :190 n :500 Save Res:None Execute

Dans l'énoncé	Comment faire ou rédiger ?
Déterminer un intervalle de fluctuation asymptotique à 95 % de la fréquence.	Repérer dans l'énoncé la **proportion p connue** du caractère étudié, puis la taille n de l'échantillon étudié. Vérifier ensuite les **critères d'approximation** (page 236), puis calculer les bornes de l'intervalle à l'aide de la formule du cours.
Arrondir les bornes à 0,01 près.	Arrondir les bornes de manière à ce que l'intervalle trouvé **contienne** l'intervalle de fluctuation asymptotique.
Interpréter le résultat.	Préciser que, dans **environ 95 % des cas**, la fréquence appartient à l'intervalle trouvé. On peut revenir aux effectifs en multipliant les bornes par la taille n de l'échantillon.
Déterminer un intervalle de confiance au niveau de confiance 0,95 .	Repérer dans l'énoncé la taille n de l'échantillon, puis calculer la **fréquence observée f** du caractère dans l'échantillon. Vérifier les critères d'approximation, puis calculer les bornes de l'intervalle en retranchant ou en ajoutant $\frac{1}{\sqrt{n}}$ à la fréquence observée f.
Déterminer la taille minimale de l'échantillon pour une précision donnée.	La longueur de l'intervalle de confiance est $\frac{2}{\sqrt{n}}$. Cette longueur doit être inférieure à la précision demandée. On résout alors une inéquation d'inconnue n.

Exercice guidé

34 Un questionnaire a été distribué à chacun des participants de plusieurs stages de formation.

1 L'organisme de formation a constaté que dans l'ensemble de tous ces stages, 75 % des stagiaires sont prêts à poursuivre le stage.
Un formateur a animé un des stages dans lequel étaient inscrits 80 personnes.
Déterminer l'intervalle de fluctuation asymptotique à 95 % de la fréquence du nombre de stagiaires prêts à poursuivre le stage de ce formateur.
Donner un intervalle dont les bornes sont arrondies au centième près.
Interpréter le résultat.

2 L'organisme de formation désire connaître la satisfaction des stagiaires sur la situation du lieu du stage.
Il interroge 120 stagiaires : 48 ne sont pas satisfaits.
a. Déterminer un intervalle de confiance au niveau de confiance 0,95 de la proportion de stagiaires non satisfaits du lieu du stage.
On donnera un intervalle dont les bornes sont arrondies à 10^{-4} près.
b. Combien de stagiaires au moins le formateur doit-il interroger pour que l'amplitude de l'intervalle de confiance diminue de moitié ?

Aide

1 p est la proportion des stagiaires prêts à poursuivre le stage, dans l'ensemble de tous les stages. Un stage est un échantillon.
Le nombre de stagiaires inscrits à un stage est n.
Appliquer la formule du cours pour calculer les bornes de l'intervalle.
Prendre, pour la borne inférieure, une valeur approchée par défaut et pour la borne supérieure, une valeur approchée par excès.

2 **a.** La taille n de l'échantillon est le nombre de stagiaires interrogés.
On calcule une estimation de la proportion p de stagiaires non satisfaits, en calculant la fréquence des stagiaires non satisfaits dans l'échantillon. Après avoir vérifié les conditions d'approximation, on calcule les bornes de l'intervalle de confiance en les arrondissant au centième près.
b. Pour diminuer l'amplitude de l'intervalle de confiance de moitié, il faut multiplier l'effectif par $2^2 = 4$.

Revoir les outils de base

Savoir résoudre des équations ou inéquations contenant des racines carrées

35 **1** Démontrer que :
$$b-\frac{1}{\sqrt{n}} \leqslant a \leqslant b+\frac{1}{\sqrt{n}} \Leftrightarrow a-\frac{1}{\sqrt{n}} \leqslant b \leqslant a+\frac{1}{\sqrt{n}}.$$
2 Calculer la valeur de n pour laquelle :
a. $\left[0{,}8-\frac{1}{\sqrt{n}} ; 0{,}8+\frac{1}{\sqrt{n}}\right]=[0{,}75 ; 0{,}85]$;
b. $\left[0{,}6-\frac{1}{\sqrt{n}} ; 0{,}6+\frac{1}{\sqrt{n}}\right]=[0{,}58 ; 0{,}62]$;
c. $\left[0{,}8-\frac{0{,}4}{\sqrt{n}} ; 0{,}8+\frac{0{,}4}{\sqrt{n}}\right]=[0{,}78 ; 0{,}82]$.
3 À partir de quelle valeur de n a-t-on :
a. $\frac{2}{\sqrt{n}} \leqslant 0{,}1$? **b.** $\frac{2}{\sqrt{n}} \leqslant 0{,}01$?

Aide

1 Séparer cet encadrement en deux inégalités.
2 L'équation $0{,}8-\frac{1}{\sqrt{n}}=0{,}75$ est équivalente à l'équation $0{,}8+\frac{1}{\sqrt{n}}=0{,}85$.
Pour que les deux intervalles soient égaux, il suffit donc de résoudre une seule équation.
Pour résoudre cette équation, isoler d'abord $\frac{1}{\sqrt{n}}$, puis montrer que pour tout $k>0$, $\frac{1}{\sqrt{n}}=k \Leftrightarrow \sqrt{n}=\frac{1}{k}$ et élever au carré.
3 Isoler $\sqrt{n}$ par multiplication de nombres positifs.

Savoir différencier population et échantillon

36 Une enquête révèle que 38 % des Français possédant une maison individuelle déclarent être prêts à équiper leur résidence en panneaux solaires thermiques. On interroge au hasard 150 propriétaires de maison individuelle : 66 sont prêts à équiper leur résidence en panneaux solaires thermiques.
a. Préciser la population et le caractère étudiés.
b. Quelle est la taille de l'échantillon considéré ?
c. Calculer la fréquence observée.

Aide

a. On observe un caractère dans toute la population dont la proportion est p. Ici, $p = 0{,}38$.
b. On choisit un échantillon au hasard, c'est-à-dire une partie de la population. La taille de l'échantillon n est le nombre de personnes choisies au hasard.
c. $f = \frac{\text{nb de personnes ayant le caractère étudié dans l'échantillon}}{\text{taille de l'échantillon}}$.

Savoir interpréter un intervalle de fluctuation à 95 % avec la loi binomiale

37 Une enquête récente publie que la population de l'Inde compte 48 % de femmes.
Dans une province, on choisit 500 personnes au hasard : 400 d'entre elles sont des hommes.
On assimile ce choix à un tirage successif avec remise.
Soit X la variable aléatoire égale au nombre de femmes choisies.
a. Quelle est la loi suivie par la variable aléatoire X ?
b. Déterminer un intervalle de fluctuation à 95 % de la fréquence correspondant à X.
c. La province étudiée est-elle représentative de l'Inde ?

k	$P(X \leqslant k)$
217	0,02185
218	0,02699

k	$P(X \leqslant k)$
261	0,9728
262	0,9780

Aide

a. On fait l'hypothèse que la proportion de femmes est $p = 0{,}48$. On choisit 500 personnes au hasard.
La variable aléatoire X suit la loi binomiale $\mathscr{B}(500 ; 0{,}48)$.
b. À l'aide de la table fournie, on détermine le plus petit entier a tel que : $P(X \leqslant a) > 0{,}025$.
On recherche le plus petit entier b tel que :
$P(X \leqslant b) \geqslant 0{,}975$.
Comme la taille de l'échantillon est $n = 500$, l'intervalle de fluctuation à 95 % de la fréquence correspondant à X, est obtenu en divisant les bornes a et b par 500.
c. On calcule la **fréquence observée f** dans l'échantillon.
Si f n'appartient pas à l'intervalle de fluctuation à 95 %, on rejette l'hypothèse $p = 0{,}48$ et on conclut, avec un risque d'erreur de 5 %, que la répartition des femmes dans cette province n'est pas représentative de celle de la population de l'Inde.

Accompagnement personnalisé

Savoir interpréter un intervalle de fluctuation de la forme $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$

38 On reprend l'enquête précédente auprès des Français possédant une maison individuelle (exercice **36**).

L'enquête affirme que 38 % des propriétaires de maison individuelle déclarent être prêts à équiper leur résidence en panneaux solaires.
La fréquence observée dans l'échantillon de taille 150, choisi au hasard, est égale à 0,44 .
En utilisant l'intervalle de fluctuation vu en Seconde, doit-on remettre en question les résultats de l'enquête ?

Aide

On vérifie que les critères d'approximation sont vérifiés, puis on détermine l'intervalle de fluctuation vu en Seconde :

$$\left[p-\frac{1}{\sqrt{n}}\,;\,p+\frac{1}{\sqrt{n}}\right].$$

Si f appartient à l'intervalle de fluctuation, on ne peut pas remettre en question les résultats de l'enquête.

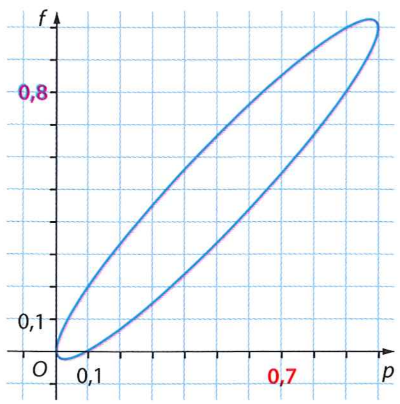
Pour aller plus loin

39 Détermination d'un autre intervalle de confiance

1 La proportion d'un caractère dans une population est p. On extrait de cette population un échantillon de taille $n=36$.
On a construit dans un repère orthonormé, les courbes des fonctions g et h, correspondant aux bornes de l'intervalle de fluctuation, définies sur $[0\,;1]$ par :

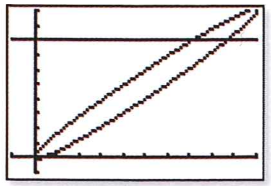
$$g(p)=p-1{,}96\frac{\sqrt{p(1-p)}}{6}$$

et $$h(p)=p+1{,}96\frac{\sqrt{p(1-p)}}{6}.$$



a. Si $p=0{,}95$, a-t-on $h(p)\leqslant 1$?
b. Déterminer graphiquement les valeurs de p pour lesquelles l'intervalle de fluctuation à 95 % peut être déterminé.
c. Déterminer graphiquement un **intervalle de fluctuation** de la fréquence pour $p=0{,}70$.
d. On suppose que la fréquence observée est $f=0{,}8$. Comment déterminer graphiquement un **intervalle de confiance** de p au niveau de confiance 0,95 ?

2 On suppose maintenant que $n=100$.
La fréquence observée sur cet échantillon est $f=0{,}8$.
a. Tracer sur l'écran de la calculatrice les courbes correspondant aux bornes de l'intervalle de fluctuation à 95 % puis la droite d'équation $y=0{,}8$.



b. En connaissant la fréquence f, comment déterminer graphiquement un autre intervalle de confiance de p au niveau de confiance 0,95 ?
c. Montrer que $p-1{,}96\frac{\sqrt{p(1-p)}}{6}=0{,}8$ et $p\geqslant 0{,}8$

$\Leftrightarrow\ p^2(1+0{,}196^2)-p(1{,}6+0{,}196^2)+0{,}64=0$.

Résoudre cette équation. *Arrondir les solutions à 10^{-3} près.*
d. Par calcul, déterminer les bornes de l'intervalle de la question **b.**, approchées à 10^{-2} près.

3 Un fabricant de produits cosmétiques a réalisé un sondage auprès de sa clientèle pour connaître sa satisfaction pour la nouvelle crème amincissante qu'il vient de commercialiser.
Il interroge 100 personnes : 80 % d'entre elles sont satisfaites de sa nouvelle crème.
a. Les conditions d'approximation sont-elles vérifiées ?
b. Utiliser les résultats vus en **2** pour obtenir un intervalle de confiance au niveau de confiance 0,95 de la proportion de clients satisfaits.
Déterminer l'intervalle de confiance à l'aide des formules du cours. Les comparer.

Réussir au Bac

40 Vrai ou faux ?

À l'entrée d'un stade accueillant une rencontre de football entre M et P , on interroge 80 spectateurs sur le pronostic de victoire pour le match à venir.
47 personnes interrogées prédisent la victoire de l'équipe M.
Préciser si les affirmations suivantes sont vraies ou fausses. Justifier la réponse.
a. Les critères d'approximation ne sont pas vérifiés.
b. $\frac{47}{80}$ est la proportion de spectateurs du stade prédisant la victoire de l'équipe M.
c. $\frac{47}{80}$ est l'estimation ponctuelle de la proportion p de spectateurs du stade prédisant la victoire de l'équipe M.
d. Au niveau de confiance 0,95 , l'intervalle de confiance de la proportion de spectateurs prédisant une victoire de l'équipe M est $[0{,}476\,;0{,}699]$, lorsque l'on arrondit les bornes à 0,001 près.

41 QCM

Indiquer la seule bonne réponse.
Avant le deuxième tour d'une élection municipale, un candidat veut réaliser un sondage pour estimer, à l'aide d'un intervalle de confiance au niveau de confiance 0,95 , la proportion p des électeurs prêts à lui apporter leur soutien.
1 Pour que la précision de l'estimation soit de 0,8 % , le nombre d'électeurs qu'il doit interroger est :
a. 15 625. **b.** 625. **c.** 62 500.
2 Le candidat a demandé un sondage auprès de 1 000 électeurs : 53 % lui sont favorables.
a. Le candidat est assuré de remporter l'élection.
b. Le candidat sait que, au seuil de confiance 0,95, son score serait compris entre 49,8 % et 56,2 % .
c. Le candidat a 95 % de chance que son score soit compris entre 49,8 % et 56,2 % .

42 Vrai ou faux ?

Une enquête réalisée dans un hôpital, indique que 60 % des patients et du personnel sont insatisfaits des horaires des repas.

On interroge à ce sujet 44 patients et 124 membres du personnel : 22 patients et 92 membres du personnel sont insatisfaits des horaires.
Préciser si les affirmations suivantes sont vraies ou fausses. Justifier la réponse.
a. L'intervalle de fluctuation asymptotique à 95 % de la fréquence des patients insatisfaits est l'intervalle $[0{,}45\,;0{,}75]$.
b. L'intervalle de fluctuation asymptotique à 95 % de la fréquence des membres du personnel insatisfaits est l'intervalle $[0{,}51\,;0{,}69]$.
c. L'échantillon des patients choisi est conforme à l'enquête réalisée dans l'hôpital.
d. L'échantillon des membres du personnel choisi est conforme à l'enquête réalisée dans l'hôpital.

43 Compte épargne

Un établissement bancaire réalise une étude sur les montants déposés sur les « comptes épargne », afin de proposer de nouveaux types de placement aux propriétaires de ces comptes.
Ils constatent que 15 % des montants d'épargne de ces comptes sont supérieurs à 2 500 €.
1 Soit p la proportion des comptes épargne, au plan national, dont le montant d'épargne dépasse 2 500 €.
a. Dans une ville, cet établissement bancaire étudie 100 comptes pris au hasard.
En prenant comme estimation ponctuelle de p la valeur $f=0{,}15$, déterminer un intervalle de confiance de p au niveau de confiance 0,95 . *Les bornes de cet intervalle seront arrondies au centième près.*
b. Quel doit être le nombre de comptes étudiés par l'établissement bancaire pour que l'intervalle de confiance soit inclus dans $[0{,}14\,;0{,}16]$?
c. On étudie un grand nombre d'échantillons ayant pour taille celle trouvée en **1 b.** et on suppose que $p=0{,}5$.
Estimer la proportion d'échantillons pour lesquels l'intervalle de confiance contient p .
2 On admet qu'en France la proportion de comptes épargne dont le montant est supérieur à 2 500 € est 15 % .
Dans une autre ville, l'établissement bancaire étudie 340 comptes épargne choisis au hasard, avec remise et indépendamment les uns des autres.
a. Calculer, à l'aide de la calculatrice, la probabilité d'obtenir entre 39 et 64 comptes épargne de montant supérieur à 2 500 €.
b. Pour l'intervalle de fluctuation asymptotique à 95 % de la fréquence du nombre de titulaires de compte épargne, dont le montant est supérieur à 2 500 €, les critères d'approximation sont-ils vérifiés ?
Si oui, déterminer cet intervalle.
Arrondir les bornes à 10^{-3} près.
c. En déduire une « fourchette » du nombre de titulaires de compte épargne auxquels l'établissement bancaire pourra proposer son nouveau type de placement.

44 Chien ou chat ?


Julie Manet (ou l'enfant au chat), Auguste Renoir (1841-1919).

En 2010, une enquête portant sur 14 000 foyers a été réalisée par la Chambre syndicale des fabricants d'aliments pour animaux. Elle a révélé que 48,7 % des foyers possèdent un animal de compagnie.
On s'intéresse aux foyers possédant soit un chat, soit un chien. Dans ces foyers, 41 % possèdent un chien, dont 47,7 % un chien mâle.
Pour 55,5 % des foyers possédant un chat, cet animal est une femelle.

1 On interroge au hasard un foyer possédant un chat ou un chien. On note :
C l'évènement : « le foyer possède un chien » ;
M l'évènement : « le foyer possède un mâle ».

a. Montrer que la probabilité qu'un foyer ayant un chat ou un chien possède un animal mâle est égale à 0,45812 .

b. Sachant qu'un foyer possède une chienne ou une chatte, calculer la probabilité que ce soit une chienne.

c. On interroge au hasard, de manière indépendante, 10 foyers possédant soit un chat soit un chien, en assimilant ce choix à un tirage avec remise.
Soit X la variable aléatoire égale au nombre de foyers possédant un animal mâle. Calculer la probabilité qu'au moins un foyer possède un animal mâle.

2 On interroge au hasard 250 foyers français :
130 déclarent posséder un animal de compagnie.

a. Donner un intervalle de fluctuation asymptotique à 95 % de la fréquence des foyers ayant un animal de compagnie. *On donnera un intervalle dont les bornes sont approchées au millième près.*

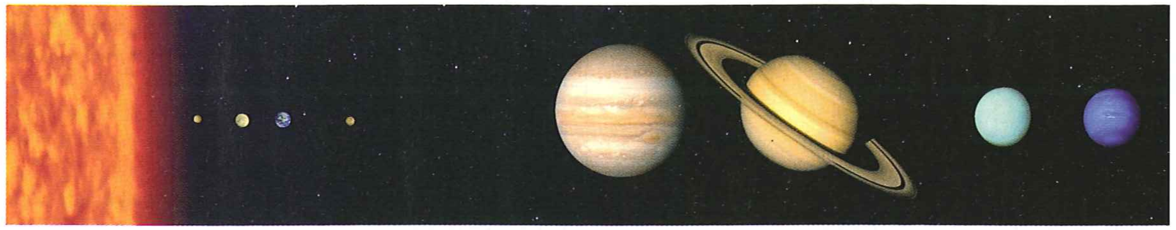
b. L'échantillon choisi est-il conforme à la population des foyers français selon l'enquête réalisée ?

3 Soit p la proportion de foyers, ayant un animal de compagnie, qui possèdent un poisson.
On interroge 1 600 foyers possédant un animal de compagnie : 512 déclarent posséder un poisson.

a. Calculer une estimation ponctuelle f de p .

b. Déterminer un intervalle de confiance au niveau de confiance 0,95 de p . Interpréter.

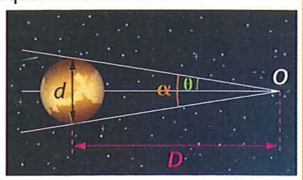
45 Diamètre d'une planète



Au cours d'une séance d'observation astronomique d'une heure, on a réalisé plusieurs mesures du diamètre apparent d'une planète, à l'aide d'un télescope.
On a obtenu les résultats suivants, en seconde d'angle :

61,60	63,01	62,86	63,68	61,70	61,85
63,28	63,10	62,66	63,68	63,60	62,87
63,68	63,31	63,69			

Pour info Le diamètre apparent α d'un corps céleste correspond à **l'angle** sous lequel il est observé.
Ce « diamètre » n'est pas absolu puisqu'il dépend de la distance D entre l'observateur O et l'objet. Ici, pendant la durée de l'observation, la planète s'est peu déplacée.



On peut donc dire que les écarts entre les valeurs sont dus aux seules incertitudes de la mesure.

1 **a.** Calculer la moyenne μ et l'écart type σ de cette série. *Arrondir les résultats au centième près.*

b. Déterminer les quartiles et calculer l'écart interquartile.

c. Calculer la fréquence des mesures dans l'intervalle :
$$E = [\mu - \sigma ; \mu + \sigma].$$

d. Peut-on penser à une distribution normale des mesures ?

2 On effectue un grand nombre de mesures et on considère que la variable aléatoire X égale au diamètre apparent de la planète suit la loi normale :
$$\mathcal{N}(62{,}97 ; 0{,}5041).$$

a. Calculer la probabilité $P(62{,}26 \leqslant X \leqslant 63{,}68)$.

b. Déterminer le nombre a tel que :
$$P(62{,}97 - a \leqslant X \leqslant 62{,}97 + a) \approx 0{,}68 .$$

3 Dans l'ensemble de toutes les mesures prises, on note p la proportion des mesures dans l'intervalle E.
On effectue, cette fois-ci, 40 mesures dont 21 dans E.

a. Déterminer une estimation de p par intervalle de confiance au niveau de confiance 0,95 .

b. Combien doit-on faire de mesures pour réduire la longueur de l'intervalle de confiance de p à 0,1 ?

46 Fabrication de clés USB

On considère un stock important de clés USB fabriquées par une entreprise pendant un mois.

1 On note E l'évènement : « la clé est défectueuse ».
On suppose que $P(E) = 0{,}05$. On prélève au hasard 400 clés USB dans le stock afin de les vérifier.
Le stock est suffisamment important pour que l'on puisse assimiler ce prélèvement à un tirage de 400 clés avec remise.
On considère la variable aléatoire X qui, à tout prélèvement ainsi défini, associe le nombre de clés défectueuses de ce prélèvement.

a. Quelle est la loi suivie par X ? Donner ses paramètres.

b. Calculer la probabilité $P(X \leqslant 30)$.

2 On pose $Z = \dfrac{X - 20}{\sqrt{19}}$. Pourquoi peut-on considérer que Z suit la loi normale $\mathcal{N}(0 ; 1)$?
En utilisant Z, déterminer le nombre a de clés USB tel que $P(X \leqslant a) = 0{,}9$. Interpréter.

3 On s'intéresse, dans cette partie, à la proportion inconnue p de clés du stock présentant une erreur de capacité de mémoire. Pour cela, on prélève au hasard, et avec remise, 2 000 clés dans le stock : 90 sont défectueuses.

a. Déterminer une estimation de la proportion p par un intervalle de confiance au niveau de confiance 0,95 .
Arrondir les bornes de l'intervalle au centième près.
Interpréter en termes de pourcentage.

b. Si la borne supérieure de l'intervalle de confiance est supérieure à 0,05 , la fabrication de ce type de clé USB doit être arrêtée.
L'entreprise devra-t-elle arrêter la fabrication ?

47 Recrudescence de la rougeole en France

En 2010, en France, une enquête indique que 38 % des cas de rougeole déclarés étaient des enfants de moins d'un an. Dans une ville, le nombre d'enfants de moins d'un an atteints de la rougeole semble beaucoup plus important. Les services sanitaires se demandent si les résultats de l'enquête ne sont pas à remettre en question. Ils sélectionnent, de manière aléatoire, 100 enfants de moins d'un an de cette ville.

1 Déterminer un intervalle de fluctuation asymptotique au seuil de 95 % de la proportion d'enfants de moins d'un an atteints de la rougeole parmi les enfants sélectionnés. *Arrondir les bornes au centième près.*

2 L'étude, réalisée sur les 100 enfants choisis, a dénombré 44 enfants atteints de la rougeole. L'échantillon choisi permet-il de remettre en cause la proportion $p = 0{,}38$?

3 Les services sanitaires, insatisfaits de cette réponse, pensent que le nombre d'enfants interrogés était insuffisant pour mettre en évidence qu'il y avait plus d'enfants de moins d'un an atteints de la rougeole dans cette ville qu'en France.

a. Soit x la borne supérieure de l'intervalle de fluctuation asymptotique à 95 % . Montrer que l'expression de la taille n de l'échantillon en fonction de x est :
$$\frac{1{,}96^2 \times 0{,}38 \times 0{,}62}{(x - 0{,}38)^2}.$$

b. Étudier les variations de f sur $]\,0{,}38 ; 0{,}5\,]$.

c. Tracer la courbe représentative de f dans un repère orthogonal. (*Unités graphiques : 2 cm pour 0,01 sur l'axe des abscisses ; 1 cm pour 100 sur l'axe des ordonnées ; l'origine en (0,38 ; 0)*)

d. En déduire le nombre d'enfants que l'on doit sélectionner pour que la fréquence 0,44 soit à l'extérieur de l'intervalle de fluctuation. Comparer à 100.

48 Incidents sur les routes de montagne

Une entreprise d'autocars dessert une région montagneuse. En chemin, les véhicules peuvent être bloqués par des incidents extérieurs comme des chutes de pierres, des troupeaux sur la route, etc.
On note D la variable aléatoire qui mesure la distance parcourue, en kilomètre, du départ de son entrepôt au lieu où un incident survient.
La loi de probabilité de D est définie par :
$$P(D \leqslant d) = \int_0^d \frac{1}{80}\, e^{\frac{-x}{80}}\, dx .$$

1 **a.** Calculer la probabilité pour que la distance parcourue sans incident soit :
- comprise entre 50 et 100 km ;
- supérieure à 300 km.

b. Sachant que l'autocar a déjà parcouru 350 km sans incident, quelle est la probabilité qu'il n'en subisse pas non plus au cours des prochains 25 km ?

2 L'entreprise possède 250 autocars. La proportion d'autocars n'ayant subi aucun incident après avoir parcouru 300 km est p . Le directeur de l'entreprise interroge 80 chauffeurs : 10 n'ont eu aucun incident.

a. Calculer l'estimation ponctuelle f de p .

b. Déterminer un intervalle de confiance au niveau de confiance 0,95 de p . *Approcher les bornes à 0,01 près.*

c. Est-il probable qu'au moins un quart des chauffeurs ne rencontrent pas d'incidents après 300 km ?

d. Déterminer une fourchette du nombre d'autocars n'ayant subi aucun incident après 300 km.

49 Sondages politiques

M. W est un candidat aux élections présidentielles. Il est arrivé au second tour.
La proportion p des personnes inscrites sur les listes électorales prêtes à voter pour le candidat W au second tour est inconnue.
Le tableau ci-dessous donne les résultats publiés par trois instituts de sondage :

Institut	Sofrop	Ifres	Tnp
Nombre de personnes interrogées	798	955	934
Nombre de votants pour le candidat W	462	582	568

1 Calculer les estimations ponctuelles de la proportion p, suivant les trois sondages.
Arrondir les résultats au centième près.

2 **a.** Pour chacun des trois sondages, déterminer l'intervalle de confiance au niveau de confiance 0,95 .
On les notera I_1, I_2 et I_3.
Approcher les bornes à 10^{-2} près.
b. Calculer les longueurs des trois intervalles de confiance. Comparer ces longueurs.
c. Traduire les intervalles I_1, I_2 et I_3 en termes de pourcentage.
d. D'après ces résultats que peut-on conclure quant aux chances de M. W de remporter les élections ?

3 En réalité l'institut Sofrop a fourni comme intervalle de confiance $[0{,}55 ; 0{,}60]$.
Combien de personnes a-t-il interrogées ?

4 On admet que M. W peut espérer 60,5 % des votes des électeurs.
M. Z est le deuxième candidat à ces élections présidentielles au second tour.
Ni M. W, ni M. Z ne sont candidats de la liste Écologie.
Parmi les électeurs prêts à voter pour M. W, 8 % avaient soutenu le candidat de la liste Écologie au premier tour.
6,8 % de tous les électeurs ont voté au premier tour pour le candidat de la liste Écologie.
On interroge un électeur au hasard.
On note :
W l'événement : « L'électeur a voté pour M. W » ;
V l'événement : « L'électeur a voté pour le candidat de la liste Écologie ».
a. Construire un arbre de probabilité représentant la situation décrite ci-dessus. Indiquer les pondérations connues.
b. Calculer la probabilité d'interroger, parmi les électeurs ayant voté pour le candidat Z, un électeur de la liste Écologie.
c. On interroge un électeur de la liste Écologie.
Quelle est la probabilité qu'il ait voté pour M. W ?

50 Contraintes budgétaires

Un gérant de multiplexe cinématographique souhaite estimer, par intervalle de confiance, la proportion p de spectateurs achetant, dans son établissement, des friandises : bonbons, glaces, pop corn, etc.

1 150 spectateurs sont interrogés et, parmi eux, 60 déclarent acheter des friandises. Déterminer un intervalle de confiance de la proportion p au niveau de confiance 0,95 . *Approcher les bornes au millième près.*
Interpréter le résultat en termes de pourcentage.

2 Le gérant choisit, pour l'estimation de p, un niveau de confiance de 0,95 et souhaite une précision de l'estimation de p inférieure à 5 % .
Combien de spectateurs, au minimum, le gérant doit-il prévoir d'interroger ?

3 Pour réaliser son enquête, le gérant possède un budget de 20 000 euros.
Les coûts fixes pour la préparation du questionnaire et la formation des enquêteurs s'élèvent à 8 000 euros.
Les coûts variables, en euro, pour le traitement informatique des données s'expriment en fonction du nombre x de spectateurs interrogés par $C(x) = 0{,}5\,e^{0{,}01x}$.
a. Exprimer, en fonction de x, le coût total à prévoir pour cette enquête.
b. Calculer le coût total lorsque 150 spectateurs sont interrogés.
c. Par résolution d'une inéquation, déterminer la contrainte sur la taille de l'échantillon afin que le budget ne dépasse pas 40 000 €.
d. Est-il alors possible de mener cette enquête en respectant la contrainte de budget et en obtenant un intervalle de confiance satisfaisant le critère de la question **2** ?

4 Pour respecter les contraintes budgétaires, le gérant réalise le sondage auprès de 1 000 spectateurs.
Parmi eux, 400 achètent des friandises.
a. Déterminer alors un intervalle de confiance de la proportion p au niveau de confiance 0,95 .
Approcher les bornes au millième près.
b. Le multiplexe réalise en moyenne 9 000 entrées par mois. Le gérant prévoit un stock de friandises pour environ 3 000 spectateurs.
Doit-il diminuer ou augmenter ses prévisions ?

51 Les centenaires en France

En France, comme dans beaucoup de pays, le nombre de centenaires n'est pas bien connu.
Pour en étudier l'évolution, on combine les résultats des recensements et les données d'état civil pour procéder à des estimations.
En 2011, en France, on estime que la proportion de centenaires est de 258 centenaires pour un million d'habitants.
Une société spécialisée dans le service à la personne souhaite connaître le nombre de centenaires de sa région parmi les 2 550 000 habitants.

1 **a.** Déterminer l'intervalle de fluctuation à 95 % de la fréquence des centenaires dans cette région.
Arrondir les bornes de l'intervalle à 0,00001 près.
b. Déterminer une « fourchette » du nombre de centenaires de cette région.

2 On admet que cette région compte 650 centenaires. Parmi eux, 520 ont moins de 105 ans.
a. Calculer une estimation ponctuelle de la proportion des centenaires âgés de moins de 105 ans en France.
b. Déterminer un intervalle de confiance de la proportion p de centenaires de moins de 105 ans, parmi les centenaires, en France, au niveau de confiance 0,95 .
Approcher les bornes à 10^{-3} près.
c. Interpréter le résultat obtenu ci-dessus, en termes de pourcentage.

3 Le tableau suivant donne l'évolution, en France, du nombre de centenaires entre 2007 et 2011 :

Année	Rang x_i	Nombre de centenaires y_i
2007	1	13 981
2008	2	14 432
2009	3	14 941
2010	4	14 944
2011	5	16 269

a. Calculer le taux global d'évolution du nombre de centenaires entre 2007 et 2011.
Arrondir à 0,1 point de pourcentage près.
b. Calculer le taux d'évolution annuel moyen du nombre de centenaires.
En supposant que ce taux d'évolution reste le même, calculer le nombre de centenaires prévu en 2013.
c. On prévoit que l'évolution du nombre de centenaires ralentira dans les années à venir. On propose donc de modéliser cette évolution par la fonction f telle que :
$$f(x) = 1\,190 \ln x + 13\,775 ,$$
où x est le rang de l'année, avec $x \geqslant 1$.
D'après ce modèle, à partir de quelle année le nombre de centenaires sera-t-il supérieur à 17 000 ?
On sera ammené à résoudre une inéquation.

52 Équipement automobile en France

On propose de modéliser l'évolution du taux, en pourcentage, d'équipement des ménages français en automobile à partir de 1980 par la fonction f définie sur l'intervalle $[0 ; +\infty[$ par : $f(x) = \dfrac{82}{1 + 0{,}15\,e^{-0{,}08x}}$,
où x désigne le rang de l'année, avec $x = 0$ en 1980.

1 **a.** Étudier les variations de f sur $[0 ; +\infty[$.
b. Justifier que l'équation $f(x) = 80$ admet une solution unique sur l'intervalle $[0 ; 40[$.
c. En déduire à partir de quelle année le taux d'équipement en automobile a dépassé 80 % .
d. Si l'évolution se poursuit suivant le même modèle, estimer le taux d'équipement des Français en 2015.

2 Parmi les ménages français possédant une automobile en 2010, 36,6 % possèdent deux véhicules.
Calculer la part des ménages français possédant deux véhicules en 2010. *Arrondir à 0,1 % près.*

3 On admet désormais que 30 % des ménages français possèdent deux véhicules.
Une compagnie d'assurance mène une enquête dans une ville comportant 5 000 ménages, afin de proposer un nouveau contrat aux propriétaires de deux véhicules.
1 400 ménages affirment posséder deux véhicules.
a. Déterminer un intervalle de fluctuation asymptotique à 95 % de la fréquence des ménages de cette ville possédant deux véhicules, dont les bornes sont arrondies à 0,01 près.
b. Déterminer une fourchette du nombre de nouveaux contrats que la compagnie d'assurance peut prévoir.
c. Peut-on considérer que la répartition des ménages possédant deux véhicules dans cette ville est conforme à celle de la population française ?

4 La part des ménages français possédant trois véhicules est p . Une enquête réalisée dans une ville comportant 6 000 ménages montre que 6 % de ces ménages possèdent trois véhicules. Déterminer un intervalle de confiance de la proportion p de ménages français possédant trois véhicules au niveau de confiance 0,95 .
Interpréter en termes de pourcentage.

53 ... en ornithologie

Estimation d'une population

On désire évaluer le nombre N de cormorans installés dans une région de la Bretagne. Pour cela, on capture au hasard 900 cormorans. Les captures sont indépendantes les unes des autres. Ces cormorans sont bagués, puis relâchés.

a. Exprimer, en fonction de N, la proportion p de cormorans bagués dans la région.

b. On capture ultérieurement 1 500 cormorans au hasard dans la région, parmi lesquels on dénombre 321 cormorans bagués. Déterminer un intervalle de confiance de la proportion p de cormorans bagués parmi tous les cormorans de la région, pour un niveau de confiance de 0,95 .
Approcher les bornes au millième près.
Interpréter le résultat en termes de pourcentage.

c. En déduire une fourchette du nombre de cormorans de cette région, au niveau de confiance 0,95 .

54 ... en médecine

Test d'un médicament

Une race de souris présente des tumeurs. Le taux est toujours de 18 %.
On veut tester l'action d'un nouveau médicament sur ces tumeurs. On applique un traitement à l'ensemble de la population constituée par les souris de cette race.

Plus tard, on effectue les deux tests suivants :
Test A : on prélève 100 souris au hasard et on observe que 25 d'entre elles présentent encore une tumeur.
Test B : on prélève 400 souris au hasard et on observe que 56 d'entre elles présentent encore une tumeur.

a. Déterminer, pour chacun des échantillons prélevés, un intervalle de fluctuation asymptotique à 95 % de la proportion des souris présentant une tumeur.

b. Les échantillons prélevés sont-ils conformes à la répartition des souris présentant une tumeur dans la population des souris de cette race ?

c. Peut-on conclure que la substance étudiée a eu une action ? *On pourra réunir les deux échantillons en un seul.*

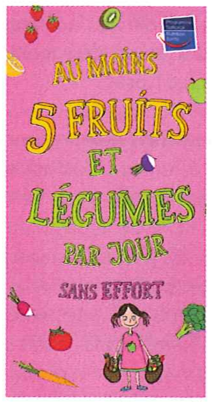
55 ... en SVT

5 fruits et légumes

Afin de mieux connaître leurs goûts et habitudes alimentaires, une entreprise agro-alimentaire a lancé une enquête auprès de 300 enfants âgés de 5 à 11 ans.

Selon cette étude, les enfants entre 5 et 11 ans ont un bon rapport avec les fruits et légumes. 186 sont prêts à en manger cinq par jour.

Cependant, ils sont encore 57 % à en manger moins de cinq par jour.



www.mangerbouger.fr

On arrondirra les bornes des intervalles à 10^{-3} près.

a. Déterminer un intervalle de confiance de la proportion p d'enfants de 5 à 11 ans prêts à manger au moins cinq fruits ou légumes par jour, au niveau de confiance 0,95 .

b. Déterminer un intervalle de confiance de la proportion p' d'enfants de 5 à 11 ans mangeant moins de cinq fruits ou légumes par jour, au niveau de confiance 0,95 .

c. Combien l'entreprise doit-elle interroger d'enfants pour que l'amplitude de l'intervalle de confiance de la proportion p' soit inférieure à 10 % ?

d. L'entreprise peut-elle publier dans le rapport de son enquête que 57 % des enfants de 5 à 11 ans mangent moins de 5 fruits et légumes par jour ?
Comment doit-elle présenter le résultat de son enquête pour que l'affirmation soit exacte ?

Réinvestir...

56 ... en SES

Précision d'un sondage

PARTIE A Marge d'erreur

Un institut de sondage publie dans un journal, en accompagnement des résultats d'une enquête, le tableau ci-dessous intitulé : **« Marge d'erreur ou intervalle de confiance à 95 % ».**

Et si l'effectif est… / *Si le pourcentage trouvé est…*	5 ou 95 %	10 ou 90 %	20 ou 80 %	30 ou 70 %	40 ou 60 %	50 %
50	6,2	8,5	11,3	13,0	13,9	14,1
100	4,4	6,0	8,0	9,2	9,8	10,0
200	3,1	4,2	5,7	6,5	6,9	7,1
250	2,8	3,8	5,1	5,8	6,2	6,3
300	2,5	3,5	4,6	5,3	5,7	5,8
350	2,3	3,2	4,3	4,9	5,2	5,3
400	2,2	3,0	4,0	4,6	4,9	5,0
450	2,1	2,8	3,8	4,3	4,6	4,7
500	1,9	2,7	3,6	4,1	4,4	4,5
600	1,8	2,4	3,3	3,7	4,0	4,1
700	1,6	2,3	3,0	6,5	3,7	3,8
800	1,5	2,1	2,8	3,2	3,5	3,5
900	1,4	2,0	2,6	3,0	3,2	3,3
1 000	1,4	1,8	**2,5**	2,8	3,0	3,1
2 000	1,0	1,3	**1,8**	2,1	2,2	2,2
4 000	0,7	0,9	1,3	1,5	1,6	1,6
6 000	0,6	0,8	1,1	1,3	1,4	1,4
10 000	0,4	0,6	0,8	0,9	0,9	1,0

a. Cet institut de sondage définit la marge d'erreur d'un résultat, comme la demi-amplitude en % d'un intervalle de confiance. En utilisant les formules du cours, calculer la marge d'erreur pour $n = 1\ 000$, puis $n = 2\ 000$. Dépend-elle de la fréquence observée ? Les résultats obtenus correspondent-ils aux résultats sur fond rose dans le tableau ?

b. L'institut de sondage a utilisé, pour calculer les valeurs de ce tableau, une forme non simplifiée de l'intervalle de confiance donné par : $\mathrm{I} = \left[f - 1{,}96\sqrt{\dfrac{f(1-f)}{n}}\ ;\ f + 1{,}96\sqrt{\dfrac{f(1-f)}{n}}\right]$.

Calculer alors, arrondie au millième près, la marge d'erreur lorsque $n = 1\ 000$, puis $n = 2\ 000$, pour $f = 0{,}20$.
Comparer aux résultats du tableau.

c. Avant le premier tour d'une élection présidentielle, un sondage publié dans un quotidien a été réalisé auprès de 800 personnes. Les résultats des différents candidats étaient les suivants :

A : 24 % B : 23 % C : 21 % D : 19 % E : 13 %.

Estimer, en utilisant le tableau précédent, les marges d'erreur pour ces trois résultats, ainsi que les intervalles de confiance correspondants. *On prendra les valeurs approchées des fréquences.*
Peut-on prévoir quel candidat sera en tête à l'issue du premier tour ?

d. Pour les candidats A et D, déterminer les intervalles de confiance I_A et I_B à l'aide de la formule donnée en **b.**

PARTIE B Élections 2002

Jacques Chirac	19 %
Lionel Jospin	18 %
Jean-Marie Le Pen	14 %

Lors du premier tour des élections présidentielles de 2002, le dernier sondage publié effectué sur 1 000 personnes prévoyait les résultats ci-contre.
L'absence de publication des marges d'erreur a joué. Beaucoup d'électeurs ont cru que Lionel Jospin serait qualifié au premier tour, donc présent au second.
Estimer les marges d'erreur de chacun de ces résultats, puis les intervalles de confiance correspondants.
Pouvait-on prévoir l'ordre des candidats au premier tour de l'élection ?