STATISTIQUES PROBABILITÉS

3

Échantillonnage et estimation

Connaissances nécessaires à ce chapitre

- Connaître la définition d'un intervalle de fluctuation
- Déterminer un intervalle de fluctuation dans le cadre de la loi binomiale
- Tester une hypothèse à l'aide d'un intervalle de fluctuation
- Déterminer une estimation d'un paramètre inconnu avec un intervalle de confiance au seuil de 95 %

Auto-évaluation

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

1 On considère une variable aléatoire X qui suit une loi binomiale $\mathscr{B}(8\ ;\ 0{,}4)$ et on donne le tableau de probabilités (arrondies au millième) ci-dessous :

k	0	1	2	3	4
$P(X \leqslant k)$	0,017	0,106	0,315	0,594	0,826

k	5	6	7	8
$P(X \leqslant k)$	0,950	0,991	0,999	1

1) Déterminer une intervalle de fluctuation de X au seuil de 95 %.

2) En déduire un intervalle de fluctuation de la fréquence $\frac{X}{8}$ au seuil de 95 %.

2 Une machine produit des clous en série. Le fabricant de la machine affirme que 97 % des clous sont sans défaut. Un client teste ce pourcentage : il décide de compter le nombre X de clous défectueux dans un échantillon de 10 000 clous.

1) Quelle loi suit X sous l'affirmation du fabricant ? Préciser les paramètres.

2) Déterminer un intervalle de fluctuation de X au seuil de 95 %.

3) Le client compte 399 clous défectueux.
Peut-il remettre en cause l'affirmation du fabricant ?

3 L'an dernier, Jenny a remarqué que la probabilité qu'il pleuve dans son département était $p = 0{,}4$.
Cette année, elle a relevé la météo durant 20 jours pour comparer les deux années.
Sous l'hypothèse $p = 0{,}4$, un intervalle de fluctuation de la fréquence des jours pluvieux au seuil de 95 % est $[0{,}2\ ;\ 0{,}6]$.
Elle a remarqué que durant ces 20 jours, la pluie s'est manifestée 5 fois.

1) Peut-elle rejeter l'hypothèse que $p = 0{,}4$ au seuil de 95 % ?

2) Peut-elle affirmer que la probabilité qu'il pleuve est encore de 0,4 cette année ?

4 Un candidat à une élection commande un sondage portant sur 1 000 personnes.
Il est donné gagnant avec 51 % des voix.

1) Déterminer un intervalle de confiance sur la proportion des voix qu'il obtiendra lors de l'élection.

2) Est-il sûr de gagner, au seuil de 95 % ?

3) Quel devrait-être le nombre de personnes sondées pour qu'un sondage lui donnant 51 % d'intention de vote assure sa victoire au seuil de 95 % ?

Voir solutions p. ??

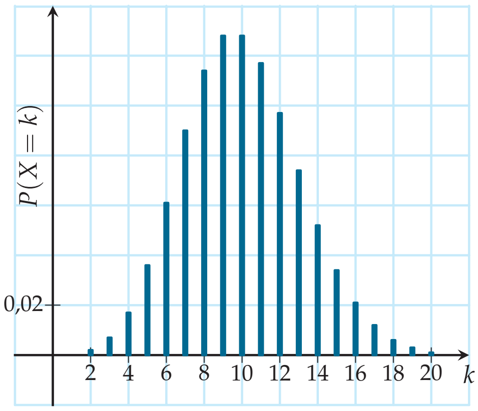
Activités d'approche

ACTIVITÉ 1 Contrôle de qualité

Une entreprise fabrique des vis en acier. Elle affirme que 5 % des pièces qu'elle produit ont un défaut à la fin de la chaîne de production.
La responsable du contrôle qualité prélève un échantillon de n vis (au vu du grand nombre de vis produites, on assimile ce prélèvement à un tirage avec remise) pour effectuer une analyse.

Partie 1 : Avec $n = 200$ et des logiciels de calcul

La contrôleuse prélève un échantillon de 200 vis au hasard.
On note X la variable aléatoire donnant le nombre de vis avec défaut qui suit donc une loi binomiale de paramètres $n = 200$ et $p = 0{,}05$ et $F = \dfrac{X}{200}$ la variable aléatoire donnant la fréquence des vis défectueuses dans l'échantillon.
On donne les éléments ci-dessous obtenus avec un tableur : un graphique représentant en partie cette loi et un tableau de valeurs (les probabilités sont arrondies à 10^{-3}).



k	$P(X \leqslant k)$	k	$P(X \leqslant k)$
2	0,002	11	0,700
3	0,009	12	0,796
4	0,026	13	0,870
5	0,062	14	0,922
6	0,124	15	0,956
7	0,213	16	0,976
8	0,327	17	0,988
9	0,455	18	0,994
10	0,583	19	0,997

1) Peut-on dire que la probabilité qu'elle relève exactement 3 pièces avec défaut est élevée ?
2) La probabilité qu'elle constate entre 6 et 10 vis avec défaut est-elle supérieure à 0,6 ?
3) **a)** Trouver la plus petite valeur de l'entier b tel que $P(4 \leqslant X \leqslant b) \geqslant 0{,}95$.
b) En déduire un intervalle $[f_1\ ;\ f_2]$ tel que $P(f_1 \leqslant F \leqslant f_2) \geqslant 0{,}95$.

> On dit que l'intervalle $[0{,}02\ ;\ 0{,}08]$ est un intervalle de fluctuation au seuil de 95 % de la variable aléatoire F donnant la fréquence : cela signifie qu'il y a plus de 95 % de chance que la fréquence des vis défectueuses (dans cet échantillon de 200 vis) soit dans cet intervalle.

4) **a)** Proposer un autre intervalle $[c\ ;\ d]$ (avec c et d entiers) tel que $P(c \leqslant X \leqslant d) \geqslant 0{,}95$.
b) En déduire un autre intervalle de fluctuation au seuil de 95 % de la variable aléatoire F.

5)

> En Première, pour déterminer un tel intervalle, on appliquait la méthode suivante :
> - on cherche le plus petit entier a tel que $P(X \leqslant a) > 0{,}025$;
> - on cherche le plus petit entier b tel que $P(X \leqslant b) \geqslant 0{,}975$;
> - on calcule l'intervalle de fluctuation au seuil de 95 % donné par $\left[\dfrac{a}{n}\ ;\ \dfrac{b}{n}\right]$ où n correspond au paramètre n de la loi binomiale utilisée.

Lequel des deux intervalles trouvés aux questions **3b** et **4b** obtient-on avec cette méthode ?

Partie 2 : Avec n quelconque et des résultats de Terminale

La contrôleuse prélève un échantillon de n vis au hasard et on note :

- X_n la variable aléatoire donnant le nombre de vis défectueuses, qui suit donc une loi binomiale de paramètres n et $p = 0{,}05$;
- $F_n = \dfrac{X_n}{n}$ la variable aléatoire donnant la fréquence des vis défectueuses dans l'échantillon ;
- $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ la variable aléatoire centrée réduite associée à X_n.

1) En utilisant le théorème de Moivre-Laplace, que peut-on dire de $P(a \leqslant Z_n \leqslant b)$ quand n tend vers $+\infty$?

2) Que peut-on en déduire pour $P(-1{,}96 \leqslant Z_n \leqslant 1{,}96)$ quand n tend vers $+\infty$?

3) En déduire un intervalle $[f_1\ ;\ f_2]$ dépendant de p et n tel que $P(f_1 \leqslant F_n \leqslant f_2) \approx 0{,}95$ et $P(f_1 \leqslant F_n \leqslant f_2) \geqslant 0{,}95$ quand n tend vers $+\infty$

> Cet intervalle obtenu grâce à une limite est dit **intervalle de fluctuation asymptotique** de F_n au seuil de 95 %.
> Lorsque l'on réalise n tirages avec remise (ou n tirages assimilables à des tirages avec remise, comme c'est le cas ici), cet intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence des succès est donné par $\left[p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où p est la probabilité d'un succès.

Partie 3 : Comparaison des intervalles obtenus

1) a) Calculer l'intervalle de fluctuation asymptotique au seuil de 95 % dans le cas $n = 200$ (arrondir à 10^{-3}).

b) Comparer cet intervalle avec celui obtenu par la méthode de Première de la partie **1** .

2) On donne le tableau ci-dessous regroupant des intervalles de fluctuation au seuil de 95 % obtenus avec la méthode de Première pour $p = 0{,}05$ et différentes valeurs de n :

Pour $n =$	30	500	1000
Intervalle de fluctuation	$[0\ ;\ 0{,}133]$	$[0{,}032\ ;\ 0{,}07]$	$[0{,}037\ ;\ 0{,}064]$
Intervalle de fluctuation asymptotique			

a) Recopier et compléter le tableau en calculant les intervalles de fluctuation asymptotiques au seuil de 95 % correspondant à chaque valeur de n.

b) Que peut-on dire des intervalles présents dans chaque colonne quand n augmente ?

> Cet intervalle de fluctuation asymptotique est plus facile à déterminer que l'intervalle de fluctuation obtenu avec la méthode de Première. On estime qu'il en donne une approximation satisfaisante lorsque $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

Partie 4 : Prise de décision

La contrôleuse a finalement choisi de prélever 400 vis et 26 d'entre elles ont un défaut.
Elle demandera un nouveau réglage des machines si la fréquence observée n'est pas dans l'intervalle de fluctuation asymptotique au seuil de 95 %.
Que va t-elle décider ?

1. Intervalle de fluctuation

DÉFINITION

Soit I un intervalle, s un réel de $]0\,;\,1[$ et X une variable aléatoire.
I est un intervalle de fluctuation de X au seuil de s si $P(X \in I) \geqslant s$.

PROPRIÉTÉ

Soit X_n une variable aléatoire qui suit une loi binomiale $\mathscr{B}(n\,;\,p)$ avec $p \in]0\,;\,1[$ et $\alpha \in]0\,;\,1[$.
Alors, d'après le théorème de Moivre-Laplace,

$$\lim_{n \mapsto +\infty} P\left(\frac{X_n}{n} \in I_n\right) \geqslant 1-\alpha$$

où $I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ et u_α est le nombre tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$ lorsque Z suit la loi normale centré réduite $\mathcal{N}(0\,;\,1)$.

I_n est appelé **intervalle de fluctuation asymptotique** de $\frac{X_n}{n}$ au seuil de $1-\alpha$.

PREUVE

- Soit $\alpha \in]0\,;\,1[$ et Y suivant la loi $\mathcal{N}(0\,;\,1)$: il existe u_α tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$.
- D'autre part, comme $E(X_n) = np$ et $\sigma(X_n) = \sqrt{np(1-p)}$, d'après Moivre-laplace :

$$\lim_{n \to +\infty} P\left(-u_\alpha \leqslant \frac{X_n - np}{\sqrt{np(1-p)}} \leqslant u_\alpha\right) = P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha \text{ or}$$

$$\begin{aligned}
-u_\alpha \leqslant \frac{X_n - np}{\sqrt{np(1-p)}} \leqslant u_\alpha \quad &\Leftrightarrow \quad -u_\alpha\sqrt{np(1-p)} \leqslant X_n - np \leqslant u_\alpha\sqrt{np(1-p)} \\
&\Leftrightarrow \quad np - u_\alpha\sqrt{n}\sqrt{p(1-p)} \leqslant X_n \leqslant np + u_\alpha\sqrt{n}\sqrt{p(1-p)} \\
&\Leftrightarrow \quad p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{X_n}{n} \leqslant p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}
\end{aligned}$$

donc $\displaystyle\lim_{n \to +\infty} P\left(p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{X_n}{n} \leqslant p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1-\alpha.$

Valeurs particulières : On obtient comme intervalle de fluctuation asymptotique :

- $I_n = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ au seuil de 95 % ;
- $I_n = \left[p - 2{,}58\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 2{,}58\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ au seuil de 99 %.

Remarque : En pratique, ces deux intervalles permettent des **prises de décisions** au seuil de 95 % ou de 99 % sous les conditions suivantes : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

PROPRIÉTÉ

L'intervalle de fluctuation asymptotique au seuil de 95 % est inclus dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$.

PREUVE

Voir TP 1 page 18.

2. Prise de décision

MÉTHODE 1 Tester une hypothèse en étudiant un échantillon

► Ex. 17 p. 9

On considère une population dans laquelle on souhaite savoir si la proportion d'individus vérifiant une certaine propriété est p : c'est l'hypothèse à tester.
Pour cela, on détermine un intervalle de fluctuation asymptotique I (à un certain seuil) de la fréquence du caractère dans un échantillon de taille n prélevé dans la population (en admettant que ce prélèvement est assimilable à des tirages avec remise) puis on observe effectivement cette fréquence f dans un échantillon donné et :

- si $f \notin I$ alors on rejette l'hypothèse que la proportion est p au seuil considéré ;
- si $f \in I$ alors on ne rejette pas l'hypothèse que la proportion est p au seuil considéré.

Exercice d'application

Le pourcentage de personnes du groupe sanguin O dans la population française est de 43 %.
On souhaite déterminer si l'on peut faire la même hypothèse pour d'autre populations, en étudiant des échantillons de 250 personnes dans ces populations (dont on suppose qu'elles sont suffisamment grandes pour assimiler ces prélèvements d'échantillons à des tirages avec remise).
On note X la variable aléatoire donnant le nombre de personnes du groupe O dans un échantillon de 250 personnes issu d'une population dont 43 % des individus sont du groupe O.

1) a) Quelle loi suit X ?
   b) Déterminer l'intervalle de fluctuation asymptotique de la fréquence des individus du groupe O au seuil de 95 % dans un tel échantillon. Arrondir à 10^{-3} près.
2) a) On observe pour un échantillon de la population canadienne une proportion de 47 % d'individus du groupe O.
   Peut-on rejeter l'hypothèse que 43 % des Canadiens sont du groupe O ?
   b) On observe pour un échantillon de Basques 138 individus du groupe O parmi les 250 personnes de l'échantillon.
   Peut-on rejeter l'hypothèse que 43 % des Basques sont du groupe O ?

Correction

1) a) X suit la loi binomiale de paramètres $n = 250$ et $p = 0{,}43$.
   b) On a $n = 250 \geqslant 30$, $np = 250 \times 0{,}43 \geqslant 5$ et $n(1-p) = 250 \times (1 - 0{,}43) \geqslant 5$: les conditions sont bien vérifiées. On a alors :
   - $p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} = 0{,}43 - 1{,}96 \times \dfrac{\sqrt{0{,}43 \times (1-0{,}43)}}{\sqrt{n}} \approx 0{,}368$ (arrondi par défaut) ;
   - $p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} = 0{,}43 + 1{,}96 \times \dfrac{\sqrt{0{,}43 \times (1-0{,}43)}}{\sqrt{n}} \approx 0{,}492$ (arrondi par excès).

   L'intervalle de fluctuation asymptotique au seuil de 95 % est donc $I = [0{,}368\ ;\ 0{,}492]$.
2) a) La fréquence observée est de 0,47 et appartient donc à l'intervalle de fluctuation asymptotique I : au seuil de 95 %, on ne peut pas rejeter l'hypothèse que la proportion de Canadiens du groupe O est de 43 %.
   b) La fréquence observée est de $\dfrac{138}{250} = 0{,}552$ et n'appartient donc pas à l'intervalle de fluctuation asymptotique I : au seuil de 95 %, on peut rejeter l'hypothèse que la proposition de Basques du groupe O est de 43 %.

Cours - Méthodes

3. Intervalle de confiance

On considère une population dans laquelle on souhaite estimer une proportion p inconnue d'individus vérifiant une certaine propriété. Pour cela, on prélève un échantillon de taille n dans cette population et on appelle f la fréquence des individus vérifiant la propriété dans cet échantillon.

PROPRIÉTÉ

La proportion p inconnue est telle que, pour $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 5$, on a :

$$P\left(f - \frac{1}{\sqrt{n}} \leqslant p \leqslant f + \frac{1}{\sqrt{n}}\right) \geqslant 0{,}95.$$

PREUVE Pour $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 5$, on a $P\left(p - \frac{1}{\sqrt{n}} \leqslant f \leqslant p + \frac{1}{\sqrt{n}}\right) \geqslant 0{,}95$.

Or $p - \frac{1}{\sqrt{n}} \leqslant f \leqslant p + \frac{1}{\sqrt{n}} \Leftrightarrow -f - \frac{1}{\sqrt{n}} \leqslant -p \leqslant -f + \frac{1}{\sqrt{n}} \Leftrightarrow f + \frac{1}{\sqrt{n}} \geqslant p \geqslant f - \frac{1}{\sqrt{n}}$, on a donc bien $P\left(f - \frac{1}{\sqrt{n}} \leqslant p \leqslant f + \frac{1}{\sqrt{n}}\right) \geqslant 0{,}95$.

DÉFINITION

L'intervalle $\left[f - \frac{1}{\sqrt{n}} \,;\, f + \frac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance de la proportion p au seuil de confiance de 95 %.

MÉTHODE 2 Déterminer un intervalle de confiance

► Ex. 30 p. 12

Exercice d'application

Dans une école, on cherche à estimer la proportion d'élèves malades durant une épidémie de grippe. Pour cela, on choisit au hasard 50 élèves ; parmi eux 13 sont malades.
Donner un intervalle de confiance de la proportion d'élèves malades dans l'école.

Correction

La fréquence vaut $\frac{37}{50} = 0{,}74$, on a donc bien $n = 50 \geqslant 30$, $nf = 50 \times 0{,}74 = 37 \geqslant 5$ et $n(1-f) = 50 \times (1 - 0{,}74) = 13 \geqslant 5$: les conditions sont bien vérifiées. On a alors :

- $f - \frac{1}{\sqrt{n}} = 0{,}74 - \frac{1}{\sqrt{50}} \approx 0{,}598$ (arrondi par défaut) ;
- $f + \frac{1}{\sqrt{n}} = 0{,}74 + \frac{1}{\sqrt{50}} \approx 0{,}882$ (arrondi par excès).

On en déduit, au niveau de confiance de 95 %, que la proportion d'élèves malades dans l'école est dans l'intervalle $[0{,}598 \,;\, 0{,}882]$.

REMARQUES :

- Dans les deux méthodes du chapitre, on a arrondi la borne inférieure de l'intervalle par défaut et la borne supérieure de l'intervalle par excès : on procédera toujours de cette manière pour s'assurer que l'on a bien un intervalle au seuil souhaité (que ce soit un intervalle de confiance ou de fluctuation).
- Les propriétés et méthodes du chapitre s'appliquent également dans le cas où f désigne la fréquence de succès lorsque l'on réalise n tirages avec remise (ou assimilables à des tirages avec remise) pour une expérience de Bernoulli de probabilité de succès p.

Activités mentales

1 Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % si $n = 100$ et $p = 0{,}5$.

2 Déterminer l'intervalle de fluctuation asymptotique au seuil de 99 % si $n = 10\,000$ et $p = 0{,}2$.

3 Mélanie s'intéresse au nombre de spams reçu dans ses emails. Une de ses connaissances affirme que les spams représentent 10 % des mails échangés, ce dont doute Mélanie.
Elle décide d'étudier un échantillon de 100 mails pour tester cette hypothèse.
On note X le nombre de mails qui sont des spams dans un échantillon de 100 mails.

1) Préciser la loi suivie par X.
2) Calculer de tête $\dfrac{\sqrt{0{,}1 \times 0{,}9}}{\sqrt{100}}$.
3) En déduire l'intervalle de fluctuation asymptotique au seuil de 95 %.
4) Mélanie a compté 6 spams parmi 100 mails reçus. Que penser de l'hypothèse de départ ?

4 Une bûcheronne travaillant dans un bois a compté 3 320 chênes parmi les 10 000 arbres rencontrés.
Sans calculatrice, déterminer un intervalle de confiance au seuil de 95 % de la proportion de chênes dans ce bois.

5 Un maire souhaite lancer un sondage pour déterminer si les habitants de sa commune approuvent un projet immobilier.

1) Combien de personnes doit-il sonder s'il veut avoir une estimation à 1 % près de la proportion de personnes favorables ?
2) Combien de personnes doit-il sonder s'il veut avoir une estimation à 0,1 % près de la proportion de personnes défavorables ?

6 On lance n fois une pièce équilibrée.

1) Déterminer l'intervalle de fluctuation asymptotique de la fréquence de « face » obtenus au seuil de 95 % pour $n = 100$.
2) Même question pour $n = 10\,000$.
3) Combien de lancers devrait-on effectuer pour avoir l'intervalle de fluctuation asymptotique d'amplitude $1{,}96 \times 10^{-3}$?

Intervalles de fluctuation

Dans les exercices **7** à **9**, on vérifiera que les conditions d'utilisation de l'intervalle de fluctuation asymptotique sont vérifiées et on arrondira les bornes au millième.

7 Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % pour $n = 100$ et $p = 0{,}4$.

8 Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % pour $n = 4\,000$ et $p = \dfrac{1}{3}$.

9 Déterminer l'intervalle de fluctuation asymptotique au seuil de 99 % pour $n = 77$ et $p = 0{,}89$.

10 Selon une enquête de la DREES, 70 % des plus de 20 ans de la population française portent des lunettes ou des lentilles de contact.
On considère un échantillon de 400 personnes tirées au sort dans la population française et on admet que la population française est suffisamment grande pour l'assimiler à un tirage au sort avec remise.
On note X le nombre de personnes qui portent des lunettes ou des lentilles dans l'échantillon.

1) Quelle loi suit X ?
2) Vérifier que n et p vérifient bien les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.
3) En déduire l'intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence des porteurs de lunettes ou de lentilles dans cet échantillon.
4) Donner une interprétation concrète du résultat précédent.

11 On lance 50 fois de suite une pièce équilibrée.
On note X le nombre de « pile » obtenus.

1) a) Déterminer l'intervalle de fluctuation asymptotique de la fréquence de « pile » obtenus au seuil de 95 %.
   b) En déduire l'intervalle de fluctuation asymptotique de X au seuil de 95 %.
2) Déterminer l'intervalle de fluctuation asymptotique de X au seuil de 99 %.

12 ALGO

Écrire un algorithme :

- demandant en entrée les valeurs de p et n ;
- donnant en sortie les bornes de l'intervalle de fluctuation asymptotique au seuil de 95 %.

S'entraîner

13 Mickaël possède un dé dont il souhaite vérifier l'équilibre.

Il lance ce dé 100 fois, et note le nombre de 6 obtenus.

1) Déterminer l'intervalle de fluctuation asymptotique de la fréquence de 6 obtenus au seuil de 95 %.
2) Déterminer l'intervalle de fluctuation asymptotique de la fréquence de 6 obtenus au seuil de 99 %.
3) a) Quel intervalle est inclus dans l'autre ?
   b) Pourquoi était-ce prévisible ?

14 Dans une fabrique de chocolat, une machine met en forme les tablettes. Elle fabrique des tablettes imparfaites avec une probabilité 0,025.

Quand une tablette est parfaitement formée, elle est vendue 2 €. Lorsqu'elle est imparfaite, elle est vendue en vrac à 0,75 € dans le magasin d'usine. Chaque jour, l'usine produit 20 000 tablettes de chocolat.

1) Déterminer un intervalle de fluctuation de la fréquence de tablettes imparfaites au seuil de 95 %.
2) On suppose que toute la production est vendue. Déterminer un intervalle de fluctuation du chiffre d'affaires quotidien réalisé au seuil de 95 %.

15 Samuel veut acheter des fusées pour un feux d'artifice qu'il souhaite grandiose.

Le vendeur affirme que 15 % des mèches de fusées s'éteignent, empêchant le départ des fusées. Samuel souhaite qu'au moins 100 fusées soient fonctionnelles ; ne pouvant être sûr ce rien, il souhaite avoir au moins 95 % de chance d'avoir 100 fusées opérationnelles.

On note n le nombre de fusées achetées par Samuel. On suppose que le stock de fusées est suffisamment grand pour assimiler le choix des fusées à un tirage au sort avec remise. On note X le nombre de fusées opérationnelles parmi celles achetées par Samuel.

1) Quelle loi suit X ? Préciser ses paramètres.
2) Déterminer l'intervalle de fluctuation asymptotique de la fréquence de fusées opérationnelles au seuil de 95 % en fonction de n.
3) En déduire l'intervalle de fluctuation asymptotique du nombre de fusées opérationnelles au seuil de 95 % en fonction de n.
4) Déterminer la quantité de fusées que Samuel doit acheter pour avoir au moins 95 % de chance d'avoir 100 fusées opérationnelles.

Prise de décision

16 On cherche à savoir si un dé cubique est équilibré : pour cela, on le lance 1 000 fois et on s'intéresse au nombre de 1 obtenus.

Dans l'hypothèse où l'on lance 1 000 fois un dé équilibré, un intervalle de fluctuation de la fréquence de 1 obtenus au seuil de 95 % est $[0{,}144\ ;\ 0{,}190]$.

Que peut-on dire du dé si l'on obtient :

1) 200 fois le nombre 1 ? 2) 150 fois le nombre 1 ?

17 ► MÉTHODE 1 p. 5

Un producteur de jus de pomme a constaté que 4 % de sa production n'avait pas pu être commercialisée l'an dernier à cause d'une teneur en sucre trop élevée.

Il décide de tester un échantillon de sa nouvelle production pour savoir si la proportion de bouteilles non commercialisables est différente de celle de l'année dernière.

Il choisit au hasard dans sa production 598 bouteilles, et compte le nombre X de bouteilles non commercialisables (on suppose que le volume de sa production est tel que l'on peut assimiler le choix de cet échantillon à un tirage au sort avec remise).

1) Quelle loi suit X sous l'hypothèse où la proportion de bouteilles non commercialisables n'aurait pas évolué d'une année sur l'autre ?
2) Déterminer un intervalle de fluctuation de la fréquence de bouteilles non commercialisables au seuil de 95 % dans cet échantillon.
3) Le producteur trouve finalement 19 bouteilles non commercialisables.
   Peut-il affirmer qu'il a fait mieux que l'an dernier ?

18 D'après une étude de l'INSEE en 2006, la moitié des bébés français sont nés hors mariage.

Sur un échantillon de 1 000 naissances au cours de l'année 2010, on a observé que 556 ont eu lieu hors mariage.

On fait l'hypothèse que la proportion de naissances hors mariage en 2010 est la même qu'en 2006.

1) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 %.
2) Que peut-on en déduire concernant l'hypothèse émise ?

19 Une compagnie de train annonce que 90 % de ses trains arrivent à l'heure.

Un usager mécontent conteste ce nombre et décide de compter pendant 60 jours le nombre de fois où le train arrive en retard sur leur trajet.

1) Déterminer un intervalle de fluctuation au seuil de 95 % de la fréquence des trains arrivés à l'heure.
2) Cet usager a relevé que son train avait eu 12 fois du retard.
   a) Déterminer la fréquence des trains arrivés à l'heure. Cette fréquence appartient-elle à l'intervalle de fluctuation ?
   b) Proposer deux hypothèses permettant d'expliquer le résultat de la question précédente.

20 Lors d'une campagne électorale, le maire sortant, élu lors des précédentes élections avec 52 % des voix, est donné vainqueur par un sondage avec 50,5 % d'intention de vote.

Si le maire souhaite ne pas rejeter l'hypothèse que sa côte de popularité est la même que lors de l'élection précédente, doit-il privilégier un intervalle de fluctuation au seuil de 95 % ou au seuil de 99 % ?

21 ALGO

1. *Liste des variables utilisées*
2. `n` : *entier*
3. `a,b,f,p` : *réels*
4. *Traitement et affichage*
5. Demander `p`
6. Demander `n`
7. Demander `f`
8. Donner à `a` la valeur de $p - 1{,}96 * \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$
9. Donner à `b` la valeur de $p + 1{,}96 * \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$
10. **Si** f<a ou f>b **Alors**
11. Afficher "On peut rejeter cette hypothèse au seuil de . . . "
12. **Sinon**
13. Afficher ". . . "
14. **Fin Si**

1) Compléter les lignes 11 et 12 de l'algorithme.
2) Que fait-il ?
3) Modifier l'algorithme pour qu'il demande d'abord à l'utilisateur s'il souhaite un seuil de 95 % ou de 99 %.

22 **D'après Bac (Pondichéry–2014) Quest. ouv.**

Une entreprise annonce que le pourcentage de moteurs défectueux dans sa production est égal à 1 %.

Afin de vérifier cette affirmation, 800 moteurs sont prélevés au hasard. On constate que 15 moteurs sont défectueux.

Le résultat de ce test remet-il en question l'annonce de l'entreprise ? Justifier. On pourra s'aider d'un intervalle de fluctuation.

23 **Question ouverte**

Le gérant d'un service de transports urbains annonce fièrement : « Il y a suffisamment de bus en circulation pour que les usagers soient assis 95 % du temps ».

Sébastiana pense que ce n'est pas possible. Elle fait un relevé sur 120 trajets, et note qu'elle n'a pas pu s'asseoir durant 27 trajets.

Que peut-elle penser de l'affirmation du gérant ?

24 **Question ouverte**

Une généticienne souhaite tester une hypothèse sur la transmission d'un caractère chez les drosophiles. Elle s'attend à trouver, après une génération, un quart de la population avec des yeux bruns.

Pour tester son hypothèse, elle observe un échantillon de 2 500 mouches. Elle en compte 633 avec des yeux bruns. Que peut-elle penser de son hypothèse ?

25 Un client d'un supermarché achète une mousse à raser, attiré par l'étiquette qui indique : « 97 % des utilisateurs sont satisfaits de cette mousse à raser. N'hésitez plus ! ».

En retrant chez lui, il trouve cette mousse assez irritante. Furieux, il décide de vérifier l'affirmation du producteur de mousse.

Patient, il demande l'avis de clients à la sorte du supermarché. Au bout de quelques heures, 177 clients ayant utilisé cette mousse à raser ont répondu à ses questions, et 166 d'entre eux ont révélé apprécier la mousse.

1) a) Déterminer un intervalle de fluctuation de la fréquence des personnes satisfaites au seuil de 99 %.
   b) Que peut penser le client de l'affirmation du producteur ?
2) Reprendre la question précédente avec un intervalle de fluctuation au seuil de 95 %.

26 Élise reçoit sa copie de philosophie avec la remarque suivante : « Copie bourrée de fautes d'orthographes ! Il y en a une tous les trois mots ! J'espère que vous vous débrouillez mieux en mathématiques ! ».

Élise regarde la première phrase qui contient 25 mots mais seulement 5 fautes.

1) Peut-elle remettre en cause l'affirmation du professeur ?
2) Pour une même fréquence de fautes observées, combien de mots doit-elle considérer pour pouvoir le faire ?

27 On donne dans le tableau ci-dessous les intervalles de fluctuation asymptotiques (I.F.A) d'une fréquence F à différents seuils, obtenus avec la formule du cours pour $p = 0{,}2$ et $n = 200$.

Seuil	75 %	90 %
I.F.A	[0,167 ; 0,233]	[0,153 ; 0,247]

Seuil	95 %	99 %
I.F.A	[0,144 ; 0,256]	[0,127 ; 0,273]

1) Indiquer les différentes inclusions d'intervalles.
2) Pour $a \leqslant b$, on souhaite comparer les deux intervalles de fluctuation asymptotiques suivants obtenus avec la formule du cours :
   - l'intervalle $I_{a\,\%}$, au seuil de a % ;
   - l'intervalle $I_{b\,\%}$, au seuil de b %.

   a) Traduire les intervalles précédents par des probabilités faisant intervenir F.
   b) Que peut-on en conclure concernant les intervalles $I_{a\,\%}$ et $I_{b\,\%}$?
3) Discuter de l'intérêt de choisir tel ou tel niveau de confiance dans chacun des cas suivants :
   a) Une usine produit des pièces automobiles. Réparer une pièce défectueuse est très coûteux.
   Un contrôleur qualité de l'usine souhaite montrer que la proportion de pièces défectueuses n'est que de 2 %.
   b) Un chef d'entreprise qui emploie 45 % de femmes souhaite expliquer qu'il ne pratique pas de discrimination à l'embauche.
   c) Un vendeur de dé cabossé souhaite faire passer son dé pour équilibré, alors qu'un client veut le tester. Que va choisir le vendeur ? Et le client ?

28 Matthieu et Hélène observent l'évolution de leur nouveau-né Camille.

Ils ont remarqué qu'en moyenne, Camille s'endormait 30 minutes après chaque biberon.

Ils modélisent le temps T (en heure) nécessaire à un endormissement de Camille par une variable aléatoire suivant une loi exponentielle de paramètre λ.

1) Déterminer la valeur de λ.
2) a) Selon ce modèle, quelle est la probabilité que Camille mette plus d'une heure à s'endormir ? Arrondir au millième.
   b) Hélène et Matthieu doutent un peu du modèle.
   Ils ont remarqué que durant le mois de janvier, après 150 repas, Camille a mis plus d'une heure à s'endormir seulement 10 fois.
   Que peuvent-ils en conclure (on pourra utiliser l'intervalle de fluctuation asymptotique) ?
3) Trouver une bonne raison en défaveur d'une loi exponentielle pour ce type de modélisation.

29 On souhaite mener une étude sur les salariés travaillant dans des fast-food.

En interrogeant 532 employés, on obtient les résultats suivants :

	Femmes	Hommes	Total
Surpoids	98	82	180
Pas de surpoids	232	120	352
Total	330	202	532

On souhaite tester la représentativité de cet échantillon parmi l'ensemble de la population française.

Dans la population française, on compte :

- 51,3 % de femmes ;
- 32 % de personnes en surpoids.

1) a) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la fréquence des femmes dans un échantillon de 532 personnes tirées au sort dans la population française.
   b) L'échantillon de l'énoncé est-il représentatif de la population française sur ce critère ?
2) Reprendre la question précédente en s'intéressant maintenant aux personnes en surpoids.
3) D'après cette étude, que peut-on dire des salariés qui travaillent en fast-food par rapport à l'ensemble de la population française ?

Intervalles de confiance

30 ► MÉTHODE 2 p. 6

On considère une population de très grand effectif, dont certains individus présentent un caractère particulier. On observe la fréquence d'apparition de ce caractère dans un échantillon de $n = 400$ individus. On a relevé ce caractère pour 135 individus de l'échantillon.

1) Déterminer la fréquence d'apparition du caractère dans l'échantillon.

2) Peut-on affirmer qu'un individu de la population a une probabilité de 0,337 5 de présenter ce caractère ?

3) Déterminer une estimation de cette probabilité à l'aide d'un intervalle de confiance au niveau 95 %.

31 Quelle taille devrait avoir un échantillon pour obtenir un intervalle de confiance :

1) d'amplitude 0,01 ? **2)** de rayon 0,001 ?

32 On souhaite déterminer la proportion p d'oliviers affectés par la bactérie *Xylella fastidiosa* dans une région. Pour cela, on effectue un test de dépistage sur 530 oliviers et on note le nombre X d'arbres infectés.

Après examens cliniques, on trouve que 77 arbres sont malades.

1) Déterminer une estimation de la proportion p à l'aide d'un intervalle de confiance au seuil de 95 %.

2) Combien d'arbres devrait-on tester pour avoir un intervalle de taille inférieure ou égale à 0,01 ?

3) Par combien doit-on multiplier la taille de l'échantillon de la question **1** pour avoir une précision dix fois plus grande de l'intervalle de confiance ?

33 Un laboratoire pharmaceutique souhaite tester l'efficacité d'un médicament destiné à soulager les maux de tête.

Il est administré à 579 patients volontaires. Parmi eux, 370 ont noté une amélioration de leur état de santé.

1) Déterminer une estimation du taux d'efficacité de ce médicament sous la forme d'un intervalle de confiance au seuil de 95 %.

2) Que penser de l'efficacité de ce médicament ?

3) On souhaite comparer le médicament avec un placebo. Dans le cas de migraines, l'administration d'un placebo (sans agent actif) soulage les patients dans 60 % des cas. Peut-on affirmer, au seuil de 95 %, que le médicament a une utilité ?

4) Avec le même taux de réussite pour le médicament, combien de patients devraient être testés pour distinguer son effet d'un placebo ?

34 Juste Bienbir, chanteur sur le déclin, a recueilli 43 % d'opinion favorable lors d'un sondage portant sur 10 000 personnes.

1) Déterminer une estimation de sa côte de popularité à l'aide d'un intervalle de confiance.

2) Quelle devrait être la taille de l'échantillon pour qu'avec une même fréquence d'opinions favorables, Juste puisse encore penser être apprécié par une majorité de personnes ?

35 **D'après Bac (Antilles-Guyane–2013)** ROC

Soient n un entier naturel, p un nombre réel compris entre 0 et 1, et X_n une variable aléatoire suivant une loi binomiale de paramètres n et p.

On note $F_n = \dfrac{X_n}{n}$ et f une valeur prise par F_n.

On rappelle que, pour n assez grand, l'intervalle $\left[p - \dfrac{1}{\sqrt{n}}\,;\, p + \dfrac{1}{\sqrt{n}}\right]$ contient la fréquence f avec une probabilité au moins égale à 0,95.

En déduire que l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}\,;\, f + \dfrac{1}{\sqrt{n}}\right]$ contient p avec une probabilité au moins égale à 0,95.

36 **Question ouverte**

Un lycée compte 1 200 élèves. Un élève de terminale souhaite offrir une rose à chacun des garçons du lycée lors du bal de fin d'année. Il ne peut pas compter tous les élèves du lycée, mais il a remarqué que sa classe contenait 16 filles pour 13 garçons.

Quelle quantité de roses devra-t-il prévoir pour le bal de fin d'année, au niveau de confiance de 95 % ?

37 **Question ouverte**

D'après une réglementation sur les cultures OGM, le seuil acceptable de contamination des cultures biologiques par des cultures OGM pour être qualifiées de « sans OGM » est de 0,9 %.

Après une analyse faite par un laboratoire, Bernard, producteur de produits biologiques, a été informé que sur les 1 000 plants de maïs testés dans son champ, 93 provenaient d'une culture OGM.

Préparer sa défense.

Préparer le bac

38 D'après Bac (Métropole–2013)

Dans une usine, on utilise une machine pour fabriquer des pièces. On estime que la machine est convenablement réglée si 90 % des pièces qu'elle fabrique sont conformes.

On décide de contrôler cette machine en examinant n pièces choisies au hasard (n entier naturel) dans la production. On assimile ces n tirages à des tirages successifs indépendants et avec remise.

On note X_n le nombre de pièces qui sont conformes dans l'échantillon de n pièces, et $F_n = \dfrac{X_n}{n}$ la proportion correspondante.

1) Justifier que la variable aléatoire X_n suit une loi binomiale et préciser ses paramètres.
2) Dans cette question, on prend $n = 150$.
   Déterminer l'intervalle de fluctuation asymptotique I au seuil de 95 % de la variable aléatoire F_{150}.
3) Un test qualité permet de dénombrer 21 pièces non conformes sur un échantillon de 150 pièces produites.
   Cela remet-il en cause le réglage de la machine ? Justifier la réponse.

39 D'après Bac (Centres étrangers–2015)

Un fournisseur produit deux sortes de cadenas. Les uns sont *premier prix*, et les autres sont *haut de gamme*. Un magasin de bricolage dispose d'un stock de cadenas provenant de ce fournisseur ; ce stock comprend un grand nombre de cadenas de chaque type.

1) Le fournisseur affirme que, parmi les cadenas *haut de gamme*, il n'y a pas plus de 3 % de cadenas défectueux dans sa production. Le responsable du magasin de bricolage désire vérifier la validité de cette affirmation dans son stock ; à cet effet, il prélève un échantillon aléatoire de 500 cadenas *haut de gamme*, et en trouve 19 qui sont défectueux.
   Ce contrôle remet-il en cause le fait que le stock ne comprenne pas plus de 3 % de cadenas défectueux ?
   On pourra pour cela utiliser un intervalle de fluctuation asymptotique au seuil de 95 %.
2) Le responsable du magasin souhaite estimer la proportion de cadenas défectueux dans son stock de cadenas *premier prix*. Pour cela il prélève un échantillon aléatoire de 500 cadenas *premier prix*, parmi lesquels 39 se révèlent défectueux.
   Donner un intervalle de confiance de cette proportion au niveau de confiance 95 %.

40 D'après Bac (Amérique du nord–2014)

Une association de consommateurs décide d'estimer la proportion de personnes satisfaites par l'utilisation d'une crème fraîche.

Elle réalise un sondage parmi les personnes utilisant ce produit. Sur 140 personnes interrogées, 99 se déclarent satisfaites.

Estimer, par intervalle de confiance au seuil de 95 %, la proportion de personnes satisfaites parmi les utilisateurs de la crème.

41 D'après Bac (Polynésie–2015)Quest. ouv.

En étudiant une maladie dans la population d'un pays, on a constaté que le taux, en nanogrammes par millilitre $\left(\text{ng.ml}^{-1}\right)$, d'une substance Gamma présente dans le sang est plus élevé chez les personnes atteintes de cette maladie que chez les personnes qui n'en sont pas atteintes. Pour dépister chez une personne la maladie étudiée, on effectue une prise de sang à jeun. Les données montrent que 82 % des patients malades ont un dépistage positif.

Pour améliorer le confort des personnes susceptibles de subir cet examen sanguin, on souhaite vérifier si le fait d'être à jeun est une condition indispensable dans le protocole.

On considère un groupe de 300 personnes malades sur lesquelles la prise de sang n'est pas effectuée à jeun.

Le dépistage se révèle positif pour 74 % d'entre elles.

Ce dépistage peut-il être effectué sur des personnes qui ne sont pas à jeun ?

42 D'après Bac (Nouvelle-Calédonie–2014)

Les trois parties A, B et C sont indépendantes.

Une fabrique de desserts glacés dispose d'une chaîne automatisée pour remplir des cônes de glace.

Partie A

Les cônes de glace sont emballés individuellement puis conditionnés en lots de 2 000 pour la vente en gros.

On considère que la probabilité qu'un cône présente un défaut quelconque avant son conditionnement en gros est égale à 0,003.

On nomme X la variable aléatoire qui, à chaque lot de 2 000 cônes prélevés au hasard dans la production, associe le nombre de cônes défectueux présents dans ce lot.

On suppose que la production est suffisamment importante pour que les tirages puissent être supposés indépendants les uns des autres.

1) Quelle est la loi suivie par X ? Justifier la réponse et préciser les paramètres de cette loi.
2) Si un client reçoit un lot contenant au moins 12 cônes défectueux, l'entreprise procède alors à un échange de celui-ci.
   Déterminer la probabilité qu'un lot ne soit pas échangé ; le résultat sera arrondi au millième.

Partie B

Chaque cône est rempli avec de la glace à la vanille. On désigne par Y la variable aléatoire qui, à chaque cône, associe la masse (exprimée en grammes) de crème glacée qu'il contient.

On suppose que Y suit une loi normale $\mathcal{N}\left(110\ ;\ \sigma^2\right)$, d'espérance $\mu = 110$ et d'écart-type σ.

Une glace est considérée comme commercialisable lorsque la masse de crème glacée qu'elle contient appartient à l'intervalle $[104\ ;\ 116]$.

Déterminer une valeur approchée à 10^{-1} près du paramètre σ telle que la probabilité de l'évènement « la glace est commercialisable » soit égale à 0,98.

Partie C

Une étude réalisée en l'an 2000 a permis de montrer que le pourcentage de Français consommant régulièrement des glaces était de 84 %.

En 2010, sur 900 personnes interrogées, 795 d'entre elles déclarent consommer des glaces.

Peut-on affirmer, au niveau de confiance de 95 % et à partir de l'étude de cet échantillon, que le pourcentage de Français consommant régulièrement des glaces est resté stable entre les années 2000 et 2010 ?

43 D'après Bac (Liban–2015)

En prévision d'une élection entre deux candidats A et B, un institut de sondage recueille les intentions de vote de futurs électeurs.

Parmi les 1 200 personnes qui ont répondu au sondage, 47 % affirment vouloir voter pour le candidat A et les autres pour le candidat B.

Compte-tenu du profil des candidats, l'institut de sondage estime que 10 % des personnes déclarant vouloir voter pour le candidat A ne disent pas la vérité et votent en réalité pour le candidat B, tandis que 20 % des personnes déclarant vouloir voter pour le candidat B ne disent pas la vérité et votent en réalité pour le candidat A.

On choisit au hasard une personne ayant répondu au sondage et on note :

- A l'évènement « La personne interrogée affirme vouloir voter pour le candidat A » ;
- B l'évènement « La personne interrogée affirme vouloir voter pour le candidat B » ;
- V l'évènement « La personne interrogée dit la vérité ».

1) Construire un arbre de probabilités traduisant la situation.
2) a) Calculer la probabilité que la personne interrogée dise la vérité.
   b) Sachant que la personne interrogée dit la vérité, calculer la probabilité qu'elle affirme vouloir voter pour le candidat A.
3) Démontrer que la probabilité que la personne choisie vote effectivement pour le candidat A est 0,529.
4) L'institut de sondage publie alors les résultats suivants :

   > 52,9 % des électeurs* voteraient pour le candidat A.
   >
   > *estimation après redressement, fondée sur un sondage d'un échantillon représentatif de 1 200 personnes.

   Au seuil de confiance de 95 %, le candidat A peut-il croire en sa victoire ?
5) Pour effectuer ce sondage, l'institut a réalisé une enquête téléphonique à raison de 10 communications par demi-heure. La probabilité qu'une personne contactée accepte de répondre à cette enquête est 0,4.
   L'institut de sondage souhaite obtenir un échantillon de 1 200 réponses.
   Quel temps moyen, exprimé en heures, l'institut doit-il prévoir pour parvenir à cet objectif ?

44 Malcolm et Robert jouent aux fléchettes.
Sur 120 lancers, Malcolm a atteint la cible 37 fois, alors que Robert a réussi 320 de ses 1 002 lancers.
Robert fanfaronne : « Je suis à coup sûr un meilleur tireur que toi ! ». Que peut-on en penser ?

45 En 2014, on compte 27 % de femmes élues à l'Assemblée nationale, et 25 % de femmes au Sénat.
On rappelle que l'Assemblée nationale compte 577 députés, alors que le Sénat est constitué de 348 sénateurs.
1) Peut-on penser qu'il y a une discrimination envers les femmes ? Justifier votre réponse.
2) De l'Assemblée et du Sénat, une des deux institutions semble-t-elle moins inégalitaire ?

46 Une scierie produit des planches à partir de tronc d'arbres.
Celles-ci doivent faire entre 2 cm et 3 cm d'épaisseur pour pouvoir être commercialisées et :
- lorsqu'une planche produite a une épaisseur supérieur à 3 cm, un coup de rabot est nécessaire pour la rendre commercialisable ;
- lorsqu'une planche produite a une épaisseur inférieure à 2 cm, elle est inutilisable.

La gérante a relevé que sur un échantillon de 4 000 planches produites, 3 816 étaient directement commercialisables. Elle a aussi remarqué qu'à cause du processus de découpage, il y avait autant de planches dont l'épaisseur dépassait 3 cm que de planches de moins de 2 cm d'épaisseur.
Elle veut commercialiser au moins 100 000 planches.
Déterminer une fourchette du coût de production, au seuil de 95 % sachant que produire une planche coûte 7 € et qu'un coût de rabot coûte 2 € supplémentaires.

47 ROC
Dans certains cas, en fonction de l'hypothèse à tester, il peut-être plus intéressant de s'intéresser à des intervalles de fluctuation dits unilatéraux, de la forme $[0\,;\,a]$ ou $[a\,;\,1]$.
Le but de cet exercice est de déterminer d'autres types d'intervalles de fluctuation asymptotiques, non centrés sur la valeur p, mais toujours en utilisant le théorème Moivre-Laplace.
Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p et $s \in]0\,;\,1[$.
On cherche à déterminer, à l'aide du théorème de Moivre-Laplace, un intervalle $I = [0\,;\,a]$ tel que $P\left(\dfrac{X}{n} \in I\right) \geqslant s$.

1) Soit Z une variable aléatoire qui suit la loi normale centrée réduite.
   a) Soit f la fonction définie sur $\mathbb{R}$ par
   $$f(x) = P(Z \leqslant x).$$
   Montrer que f est une fonction continue, strictement croissante, et déterminer les limites de f en $-\infty$ et en $+\infty$.
   b) En déduire qu'il existe un unique réel u tel que $P(Z \leqslant u) = s$.
   c) Montrer que :
   $$\frac{X-np}{\sqrt{np(1-p)}} \leqslant u \Leftrightarrow \frac{X}{n} \leqslant p + u\frac{\sqrt{p(1-p)}}{\sqrt{n}}.$$
   d) On sait d'après le théorème de Moivre-Laplace que $\lim\limits_{n\mapsto+\infty} P\left(\dfrac{X-np}{\sqrt{np(1-p)}} \leqslant u\right) = P(Z \leqslant u)$.
   Conclure en trouvant un intervalle de fluctuation asymptotique (dépendant de u) de la fréquence $\dfrac{X}{n}$ au seuil s.
2) **Application** : Un fabricant d'assiettes « un peu louche » affirme que ses produits présentent très peu de défaut. Il affirme que cela ne concerne que 3 % de sa production.
   On va tester son hypothèse sur un échantillon de 3 000 assiettes.
   a) À l'aide de la question **1d**, déterminer un intervalle de fluctuation asymptotique de la forme $[0\,;\,a]$ au seuil de 95 %.
   b) On a trouvé 267 assiettes fêlées dans l'échantillon. Que peut-on en conclure ?
3) Comment trouver un intervalle de fluctuation asymptotique de $\dfrac{X}{n}$ de la forme $[a\,;\,1]$ au seuil de 95 % ?
4) Dans chacun des cas suivants, préciser s'il faut privilégier un intervalle de fluctuation asymptotique « classique » ou un intervalle de fluctuation unilatéral (à préciser).
   - On chercher à tester la côte de popularité d'un homme politique.
   - On chercher à tester la proportion de ticket gagnants dans une tombola.
   - On chercher à tester le taux de guérison en une semaine pour un médicament par rapport à celui, connu, d'un placebo.

Je teste mes connaissances

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

- Déterminer un intervalle de fluctuation
- Savoir rejeter ou non une hypothèse à l'aide d'un intervalle de fluctuation
- Faire preuve d'esprit critique lors d'un test d'hypothèse
- Déterminer une estimation d'un paramètre à l'aide d'un intervalle de confiance
- Déterminer la taille d'un échantillon pour obtenir un intervalle de confiance d'amplitude ou de rayon donné

QCM d'auto-évaluation

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

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

48 Pour pouvoir utiliser un intervalle de fluctuation asymptotique, il faut que les paramètres n et p vérifient :

(a) $p \geqslant 5$ (b) $(1-p)n \geqslant 5$ (c) $np < 5$ (d) $np \geqslant 30$

49 Pour $n = 45$ et $p = 0{,}01$, l'intervalle de fluctuation asymptotique est :

(a) inutilisable (b) utilisable

50 Un intervalle de fluctuation au seuil de 95 % :

(a) est forcément un intervalle de fluctuation au seuil de 90 %
(b) est forcément un intervalle de fluctuation au seuil de 99 %

Dans une usine, une machine fabrique des tiges métalliques. L'ingénieur chargé du réglage affirme que les tiges fabriquées présentent un défaut dans 0,8 % des cas.
On s'intéresse à une échantillon de 800 tiges prélevées au hasard dans le stock. On suppose que le stock est suffisamment grand pour assimiler cela à un tirage au sort avec remise. On note X le nombre de tiges sans défaut.

51 X suit une loi binomiale de paramètres :

(a) $n = 800$ et $p = 0{,}8$ (b) $n = 640$ et $p = 0{,}008$ (c) $n = 800$ et $p = 0{,}008$ (d) $n = 800$ et $p = 0{,}992$

52 L'intervalle de fluctuation asymptotique de la fréquence des tiges sans défaut au seuil de 95 % est

(a) $[0{,}985\ ;\ 0{,}999]$ (b) $[0{,}983\ ;\ 1]$ (c) $[0\ ;\ 0{,}95]$

53 Un ouvrier trouve 12 tiges défectueuses dans l'échantillon. Il peut en conclure que :

(a) Au seuil de 95 %, l'hypothèse de l'ingénieur est à rejeter
(b) Au seuil de 95 %, on ne peut pas rejeter l'hypothèse de l'ingénieur
(c) Il faut recommencer l'expérience

Florient affirme que 15 % des êtres humains sont gauchers.
Marjolaine trouve ce pourcentage très important ; elle souhaite tester cette hypothèse sur un échantillon de 79 personnes.

54 Un intervalle de fluctuation asymptotique au seuil de 99 % est :

ⓐ $[0\ ;\ 0{,}99]$ ⓑ $[0{,}071\ ;\ 0{,}229]$ ⓒ $[0{,}99\ ;\ 1]$ ⓓ $[0{,}046\ ;\ 0{,}254]$

55 Elle trouve finalement 19 gauchers parmi les 79 personnes étudiées.

ⓐ Au seuil de 99 %, l'hypothèse est à rejeter
ⓑ Au seuil de 99 %, on ne peut pas rejeter l'hypothèse
ⓒ Il faut recommencer l'expérience

56 Elle cherche ensuite à tester l'hypothèse au seuil de 95 %.

ⓐ Au seuil de 95 %, l'hypothèse est à rejeter
ⓑ Au seuil de 95 %, on ne peut pas rejeter l'hypothèse
ⓒ Il faut recommencer l'expérience

Dans un club de sport, 65 % des inscrits sont des hommes.

57 Lors d'une réunion de 55 personnes de cette association :

ⓐ il y a 35,75 hommes
ⓑ il y a entre 28 et 43 hommes
ⓒ il peut y avoir moins de 15 hommes

Un client désœuvré à la terrasse d'un café décide de compter le nombre de voitures rouges qui roulent dans la ville.

58 Sur 504 voitures, il en a compté 63 rouges. La proportion de voitures rouges roulant dans la ville est :

ⓐ Exactement 0,125
ⓑ Comprise entre 0,08 et 0,17 avec une probabilité supérieure à 0,95
ⓒ Comprise entre 0,05 et 0,2 avec une probabilité supérieure à 0,95
ⓓ Comprise entre 0,13 et 0,17 avec une probabilité supérieure à 0,95

59 Pour avoir un intervalle de confiance d'amplitude 0,02 au seuil de 95 %, le client aurait dû compter :

ⓐ 50 voitures ⓑ 100 voitures ⓒ 250 voitures ⓓ 10 000 voitures

60 Pour avoir un intervalle de confiance de rayon 0,05 au seuil de 95 %, le client aurait dû compter :

ⓐ 100 voitures ⓑ 400 voitures ⓒ 1 000 voitures ⓓ 4 000 voitures

TP 1 Comparaison avec un intervalle de seconde

On considère une population dans laquelle la proportion d'individus vérifiant un certain caractère est p est dans laquelle on prélève un échantillon de taille n.

Le but de ce TP est d'étudier le lien entre :

- l'intervalle de fluctuation de la fréquence du caractère dans l'échantillon au seuil de 95 % vu en seconde (pour $0{,}2 \leqslant p \leqslant 0{,}8$ et $n \geqslant 25$) :
$$\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$$
- l'intervalle de fluctuation asymptotique de cette même fréquence au seuil de 95 % (pour $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$) :
$$\left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

1 Quelques cas particuliers, avec un tableur

1) a) Dans un logiciel de tableur, créer la feuille de calcul ci-dessous :

	A	B	C	D	E	F
1					n	100
2					p	0,3
3		Borne inférieure	Borne supérieure			
4	Intervalle de fluctuation de seconde					
5	Intervalle de fluctuation asymptotique					

b) Saisir des formules dans les cellules B4, C4, B5 et C5 permettant d'obtenir les bornes inférieure et supérieure des intervalles de fluctuation pour les valeurs de n et p écrites respectivement dans les cellules F1 et F2.

2) À l'aide du tableur, déterminer l'intervalle de fluctuation au seuil de 95 % vu en seconde et l'intervalle de fluctuation asymptotique au seuil de 95 % vu en terminale dans les cas suivants :

a) pour $n = 100$ et $p = 0{,}3$

b) pour $n = 500$ et $p = 0{,}8$

c) pour $n = 1\,000$ et $p = 0{,}3$

d) pour $n = 10\,000$ et $p = 0{,}25$

3) Dans chacun des cas précédents, que peut-on dire en termes d'inclusion de ces deux intervalles de fluctuation ?

4) Quelle conjecture peut-on émettre ? La mettre à l'épreuve avec d'autres valeurs de n et p.

2 Démonstration de la conjecture

On souhaite maintenant démontrer la conjecture émise à la question **4** de la partie **1**.

1) a) Étudier les variations de $p \mapsto \sqrt{p(1-p)}$ sur $[0\ ;\ 1]$.

b) En déduire que $1{,}96\sqrt{p(1-p)} \leqslant 1$ sur $[0\ ;\ 1]$.

2) Conclure.

Travaux pratiques

TP 2 Le bon seuil

1 Trouver u_α

1) On considère l'algorithme ci-dessous écrit avec le logiciel AlgoBox où :

- a désigne un réel dans l'intervalle $]0\,;\,1[$;
- u désigne le nombre vérifiant $P(-u \leqslant X \leqslant u) = a$ où X suit la loi normale centrée réduite et a est le nombre défini au point précédent ;
- la commande `ALGOBOX_INVERSE_LOI_NORMALE_CR(p)` donne le nombre x tel que $P(X < x) = p$ où X suit la loi normale centrée réduite.

```
1.  VARIABLES
2.     a EST_DU_TYPE NOMBRE
3.     u EST_DU_TYPE NOMBRE
4.  DEBUT_ALGORITHME
5.     LIRE a
6.     u PREND_LA_VALEUR ALGOBOX_INVERSE_LOI_NORMALE_CR(......)
7.     AFFICHER u
8.  FIN_ALGORITHME
```

Compléter la ligne 6 pour que l'algorithme affiche le nombre u souhaité.

2) Sans utiliser l'algorithme ou une calculatrice, donner la valeur affichée par l'algorithme (à 10^{-2} près) si l'utilisateur rentre $a = 0{,}95$.

2 Intervalle de fluctuation asymptotique

En s'appuyant sur l'algorithme précédent, en écrire un nouveau qui :

- demande à l'utilisateur la proportion p d'individus vérifiant une certaine propriété dans une population, la taille n d'un échantillon et un seuil a (sous forme décimale) ;
- affiche les bornes inférieure et supérieure de l'intervalle de fluctuation asymptotique au seuil a de la fréquence d'individus vérifiant la propriété dans un échantillon de taille n.

3 Application

Une confiserie a lancé une nouvelle gamme de bonbons, les « chanceux ».

Quand on ouvre le papier autour d'un chanceux, soit le bonbon est rose et il ne se passe rien, soit il est vert et l'on obtient gratuitement un autre bonbon.

La confiserie l'affirme dans une publicité :

15 % des bonbons sont gagnants !

1) On admet que les services de la répression des fraudes mènent leurs études sur des échantillons de taille 50 et au seuil de 90 %.

Déterminer alors, en utilisant un intervalle de fluctuation asymptotique, dans quel intervalle doit se trouver le nombre de bonbons gagnants pour que l'on n'accuse pas la confiserie de publicité mensongère.

2) En réalité, cette confiserie « triche » et ne produit que 14,5 % de bonbons gagnants.

Risque-t-elle plus de se faire accuser de publicité mensongère que si elle ne trichait pas ?

TP 3 Sondages

1 Avant le premier tour

Pour le premier tour d'une élection, on compte un candidat et deux candidates, que l'on note 1, 2 et 3.

Un journal commande à un institut un sondage, de taille « habituelle », soit sur un échantillon de 1 000 personnes dites représentatives.

Les résultats du sondage sont donnés dans le tableau si-dessous :

Vote	Candidat 1	Candidate 2	Candidate 3	Abstention	Blanc ou nul
Effectif	191	243	176	237	153

La candidate 2 pavoise, certaine d'être en tête au deuxième tour.

1) Déterminer un intervalle de confiance au seuil de 95 % des résultats pour chacun des candidat et candidates.

2) Que peut-on en tirer comme conclusion ?

3) Combien de personnes faudrait-il interroger pour pouvoir classer les candidat et candidates au seuil de 95 % en gardant les mêmes pourcentages d'intentions de vote ?

2 Après le premier tour

Le soir du premier tour, les résultats sont publiés :

Vote	Candidat 1	Candidate 2	Candidate 3	Abstention ou Blanc ou nul
Pourcentage	21 %	19 %	18 %	42 %

On souhaite simuler différents sondages sur 1 000 personnes.

1) Réaliser la feuille de calcul ci-dessous :

	A	B	C	D	E	F
1	Vote				Nombre de votes	Poucentage
2				Candidat 1		
3				Candidate 2		
4				Candidate 3		
5				Autre		

2) Dans la cellule A2, entrer la formule `=ALEA()` pour simuler un nombre entre 0 et 1.

3) Dans la cellule B2, entrer la formule `=SI(A2<0,21;1;SI(A2<0,4;2;SI(A2<0,58;3;4)))`.

4) Recopier ces formules vers le bas jusqu'à la ligne 1 001 pour simuler les choix des sondés.

5) Dans la cellule E2, entrer la formule `=NB.SI(B2:B1001;1)` pour compter le nombre d'intentions de votes pour le candidat 1.

6) Compléter de même les cellules de la colonne E puis la colonne F.

7) Relancer l'expérience avec la touche F9 ou CTRL+MAJ+F9.
Le candidat 1 est-il toujours donné gagnant ?

8) Recommencer l'expérience avec un sondage portant sur 10 000 personnes.
Qu'observe-t-on ?

9) a) Pour les trois candidats précédents, déterminer les intervalles de fluctuation asymptotiques au seuil de 95 % pour $n = 1\,000$ puis pour $n = 10\,000$.

b) Commenter les réponses aux questions **7** et **8** à l'aide de la question **9a**.

Travaux pratiques

TP 4 Surréservation

1 Intervalle de fluctuation

Une compagnie aérienne fait voler des avions pouvant accueillir 300 personnes.
Pour chaque vol, elle vend un nombre n de billets qui excède le nombre de places disponibles, comptant sur le fait que des voyageurs ne se présenteront pas pour prendre l'avion.
La probabilité qu'une personne se présente bien pour prendre son vol est p.
On suppose que les actes des passagers sont indépendants les uns des autres et on note X le nombre de personnes qui ne se présentent pas à l'embarquement.

1) Quelle loi suit X ? Préciser ses paramètres.

2) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % en fonction de n et p.

3) La compagnie aérienne veut limiter les risques : elle souhaite que la probabilité que tous les passagers se présentant à l'aéroport aient une place soit supérieure à 0,95.
Expliquer pourquoi cela sera vérifié si :

$$p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} - \frac{300}{n} \leqslant 0$$

2 Avec le tableur

À l'aide d'une feuille de calcul, déterminer la valeur maximale de n qui est solution de l'inéquation précédente dans chacun des cas suivants :

a) $p = 0{,}95$ b) $p = 0{,}98$ c) $p = 0{,}99$

3 Pour aller plus loin

Pour $p = 0{,}95$, résoudre l'inéquation (d'inconnue n) de la partie 1 en posant $x = \sqrt{n}$ et retrouver l'un des résultats obtenus à la partie 2.

Récréation, énigmes

Trop beau pour être vrai !

Lucas affirme : « j'ai lancé 1000 fois cette pièce . Il est parfaitement équilibré, la preuve, j'ai obtenu 500 fois face ».
Lucie n'y croit pas. Comment peut-elle se convaincre que Lucas a truqué ses résultats ?

Partie A : Des intervalles contre les trucages

Les intervalles de fluctuation asymptotiques vus dans ce chapitre sont des intervalles bilatéraux : les valeurs refusées sont celles soit trop grandes, soit trop petites. Parfois, pour tester une hypothèse, on peut vouloir refuser des valeurs de fréquences trop « parfaites », qui seraient le résultat de trucages embellissant la théorie.

1) En vous appuyant sur le théorème de Moivre-Laplace, déterminer ainsi un intervalle de fluctuation asymptotique au seuil de 5 %, c'est-à-dire tel que la probabilité d'être en dehors soit proche de 95 %.

2) Discuter les résultats de Lucas.

Partie B : Un exemple célèbre

Pour un exemple de résultats pour appuyer une théorie, chercher des informations concernant Gregor Mendel, moine botaniste, et son expérience sur les petits pois.