BAT 2	Devoir maison de mathématiques n° 1	Année 2 013/2 014
À rendre, sur feuille, au plus tard le 15/05		
NOM, Prénom :		

Vous pouvez poser vos questions dans les commentaires de l'article sur le site.

Sommaire

I Introduction

I.1 Identification de la situation

On se situe, dans ce chapitre, dans deux domaines des statistiques, qui sont ceux de « l'échantillonnage » et de « l'estimation » .
Une ligne de pointillés :

....

Trois lignes de pointillés :

....
....
....

Une ligne de pointillés dans minipage :
....

Trois lignes de pointillés dans minipage :
....
....
....

Apprenons à reconnaître les différents contextes d'applications.

Exemples :

On considère deux urnes U_1 et U_2 contenant chacune un très grand nombre de boules, rouges ou bleues.

1. Dans l'urne U_1, **on connaît la proportion** p **de boules rouges**.
   On procède à des tirages avec remise de n boules, et on
   observe la fréquence d'apparition d'une boule rouge.
   Cette fréquence observée appartient « en général » à un « intervalle de fluctuation » de centre p, dont la longueur diminue lorsque n augmente.
   Cet intervalle est un « intervalle de fluctuation » .

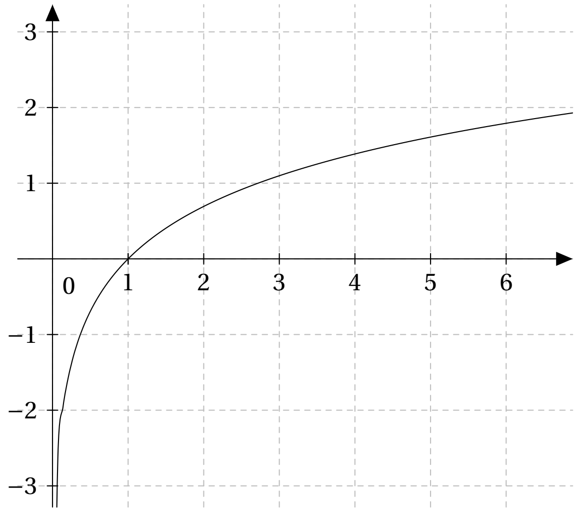
   On est ici dans le domaine de **l'échantillonnage** et de l'**intervalle de fluctuation**.

$$\left[\frac{5}{2};\sqrt{\frac{1}{3}}\right]$$

$$\left[\frac{5}{2};\sqrt{\frac{1}{3}}\right[$$

$$\left]\frac{5}{2};\sqrt{\frac{1}{3}}\right]$$

$$\left]\frac{5}{2};\sqrt{\frac{1}{3}}\right[$$



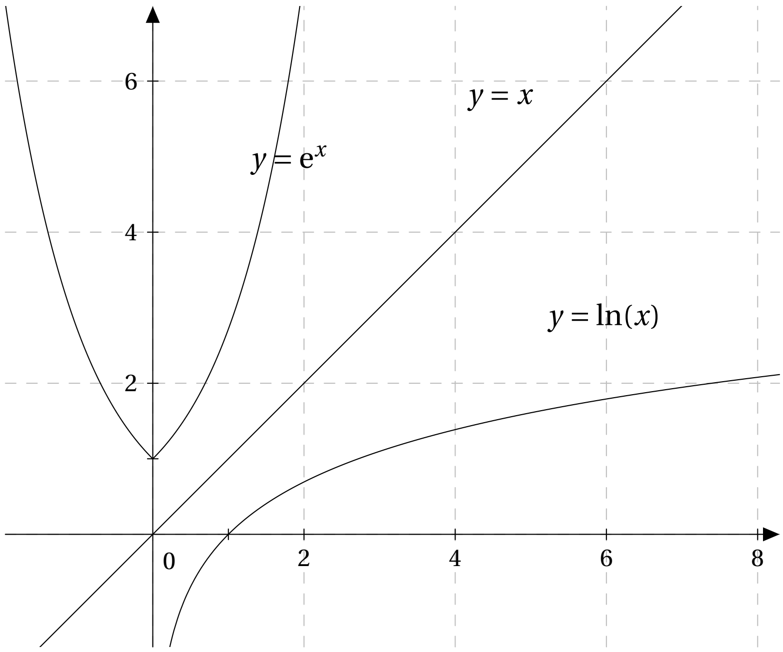
2. Dans l'urne U_2, **on ignore la proportion de boules rouges**.

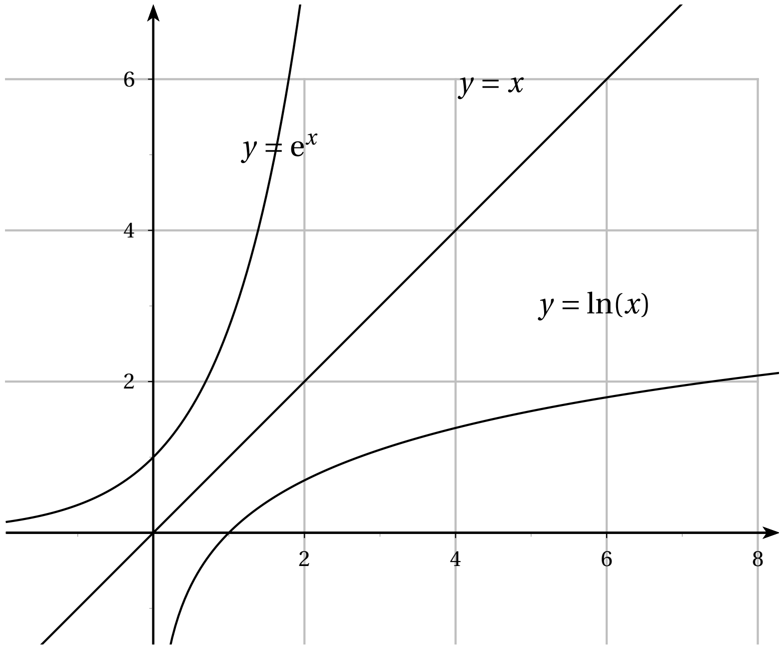
En procédant à des tirages avec rmise de n boules, on va essayer d'estimer la proportion p de boules rouges dans l'urne.

Cette estimation se fait au moyen d'un « intervalle de confiance » .

Cet intervalle dépend d'un coefficient, le « niveau de confiance » , que l'on attribue à l'estimation.

On est ici dans le domaine de l'**estimation**, et de l'**intervalle de confiance**.





Questions	**Réponses**
1. Question	☐ Proposition 1 ☐ Proposition 2 ☐ Proposition 3
2. Question	☐ Proposition 1 ☐ Proposition 2 ☐ Proposition 3
3. Question	☐ Proposition 1 ☐ Proposition 2 ☐ Proposition 3
4. Question	☐ Proposition 1 ☐ Proposition 2 ☐ Proposition 3

Questions	Réponses
1. Pour tout $x \in]-3\,;\,2]$, $f'(x) \geqslant 0$.	☑ **V** ❐ **F**
2. La fonction F présente un maximum en 2	❐ **V** ☑ **F**
3. $\displaystyle\int_0^2 f'(x)\,\mathrm{d}x = -2$	❐ **V** ☑ **F**

I.2 Intervalle de fluctuation ou intervalle de confiance ?

On s'intéresse à une population, dont on étudie un caractère particulier.

Deux cas

- **Échantillonnage**

  On utilise un **intervalle de fluctuation** quand :

  – on **connaît la proportion** p de présence du caractère étudié dans la population,

  – OU, on **fait une hypothèse sur la valeur de cette proportion** (on est alors dans le cas de la « prise de décision » .

- **Estimation**

  On utilise un **intervalle de confiance** quand :

  on **ignore la valeur de la proportion** p de présence du caractère dans la population, et on ne formule pas d'hypothèse sur cette valeur.

En appliquant l'agorithme de Dijsktra, on obtient :

A	B	C	D	E	F	G	H	Sommet
0	∞	∞	∞	∞	∞	∞	∞	A
	0+2 2(A)	0+1 1(A)	0+4 4(A)	∞	∞	∞	∞	C
	2(A)		4(A)	1+5 6(C)	∞	∞	∞	B
			4(A)	6(C)	∞	∞	∞	D
				6(C)	4+1 5(D)	∞	∞	F
				6(C)		5+4 9(F)	5+1 6(F)	E
						6+2 8(E)	6(F)	H
						6+4 8(E)		G

$$\left[\frac{5}{2};-2\right]$$

Exemples :

1. On dispose d'une pièce de monnaie.

   Comment décider qu'elle est « équilibrée » ou pas ?

   On va ici faire l'hypothèse que la fréquence d'apparition de « Pile » , par exemple, est égale à 0,5, et on va tester cette hypothèse.

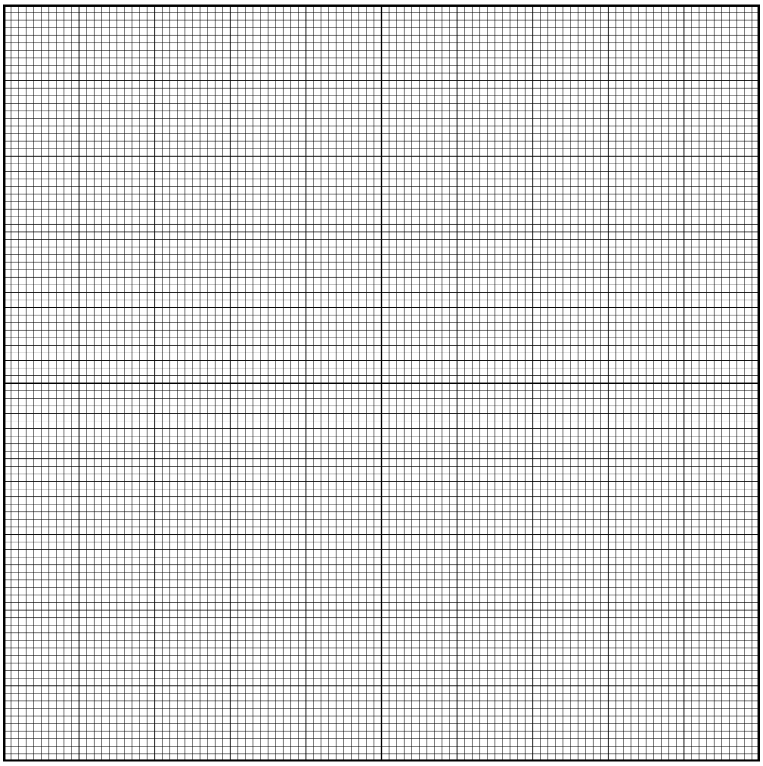
   On est dans une **situation d'échantillonnage**.

2. Une usine fabrique des fusées de feux d'artifice.

   Sur 100 fusées choisies au hasard à l'issue du processus de fabrication et mises à feu, on trouve 12 fusées qui ne fonctionnent pas.

   Comment se faire une idée de la proportion de fusées défectueuses dans la production ?

   On est dans une **situation d'estimation** : on n'a, au départ, aucune idée de la valeur de la proportion étudiée dans la production.



Exercice :

Texte à trous :

Un enfant a réalisé 4 040 d'une pièce de monnaie, et il a obtenu 2 048 fois le résultat « Pile » .

Nous allons nous intéresser au problème suivant : « Peut-on considérer que la pièce utilisée est équilibrée ? »

1. Prendre une décision en utilisant l'intervalle de fluctuation au seuil de 95% sous la forme étudiée en classe de 2^{nde}.
2. Reprendre le même travail, en utilisant un intervalle de fluctuation au seuil 95% obtenu à partir de la loi binomiale vue en classe de $1^{\text{ère}}$.

Solution :

1. En classe de 2^{nde}, on a vu que pour $n > 25$ et $0{,}2 < p < 0{,}8$, l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ constitue un intervalle de fluctuation au seuil 95%.

Ici, si on appelle p la « proportion théorique » de résultats « Pile » obtenus sur un très grand nombre de lancers, on veut tester l'hypothèse selon laquelle la pièce est équilibrée, c'est-à-dire l'hypothèse $p = 0{,}5$.

Pour l'échantillon considéré, on a $n = 4\,040$; les conditions d'application sont donc réalisées, et on obtient l'intervalle de fluctuation

$$\left[0{,}5 - \frac{1}{\sqrt{4\,040}}; 0{,}5 + \frac{1}{\sqrt{4\,040}}\right]$$

En arrondissant à 10^{-4} près, on obtient l'intervalle de fluctuation $\mathrm{I}_1 = [0{,}4843; 0{,}5157]$.

La fréquence observée dans l'échantillon est donnée par :

$$f = \frac{2\,048}{4\,040} \approx 0{,}5069 \text{ à } 10^{-4} \text{ près}$$

Ainsi, on a $f \in \mathrm{I}_1$, donc on accepte l'hypothèse de pièce équilibrée au seuil de confiance de 95%.

2. En supposant que la pièce est équilibrée, la variable aléatoire X qui dénombre les résultat « Pile » obtenus parmi les $4\,040$ lancers réalisés suit une loi binomiale de paramètres $n = 4\,040$ et $p = 0{,}5$.

On a vu en classe de $1^{\text{ère}}$ que l'ntervalle de fluctuation à 95% associé à la variable aléatoire X est l'intervalle $\mathrm{I}_2 = \left[\frac{k_1}{n}; \frac{k_2}{n}\right]$, où k_1 est le plus petit entier k vérifiant $\mathrm{P}(\mathrm{X} \leq k) > 0{,}025$ et k_2 le plus petit entier k vérifiant $\mathrm{P}(\mathrm{X} \leq k) \geq 0{,}975$.

On détermine k_1 et k_2 à l'aide d'un tableur (ou d'une calculatrice...mais le tableau est très long, ici).

On obtient $k_1 = 1\,958$ et $k_2 = 2\,082$, qui donnent comme intervalle de fluctuation :

$$\mathrm{I}_2 = \left[\frac{1\,958}{4\,040}; \frac{2\,082}{4\,040}\right]$$

soit en arrondissant à 10^{-4} près $\mathrm{I}_2 = [0{,}4847; 0{,}5153]$.

La fréquence observée $f \approx 0{,}5069$ vérifie $f \in \mathrm{I}_2$.

On est ici aussi conduit à accepter l'hypothèse de pièce équilibrée au seuil de confiance de 95%.

EXERCICE 2 : **8 points**

Une agence de voyages propose un circuit touristique pour visiter trois villes A, B et C.
Le client peut choisir la durée de son séjour dans chacune des villes.
L'agence propose des tarifs qui diffèrent selon la période. Il existe trois périodes touristiques : une période haute, une période moyenne et une période basse.
Les prix journaliers, en **centaines d'euros** par personne, dans les différents lieux sont donnés dans le tableau suivant :

	Ville A	**Ville B**	**Ville C**
Période haute	2,5	3,5	1,5
Période moyenne	2	2,2	1,5
Période basse	1,5	1,2	1

On considère la matrice $M = \begin{pmatrix} 2{,}5 & 3{,}5 & 1{,}5 \\ 2 & 2{,}2 & 1{,}5 \\ 1{,}5 & 1{,}2 & 1 \end{pmatrix}$

1. M. Gibbs choisit un circuit de trois jours : deux jours dans la ville A et un jour dans la ville C.
Calculer, à l'aide d'un **produit de matrices**, le prix de son séjour pour chaque période.

2. La matrice ci-dessous donne toutes les combinaisons d'un circuit de trois jours :

$$D = \begin{pmatrix} 3 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & 0 & 2 & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Ainsi, la troisième colonne indique que M. Gibbs choisit un circuit de deux jours dans la ville A et un jour dans la ville C.

a. Compléter la matrice D.

b. Calculer la matrice $M \times D$.

3. En déduire quels sont les circuits de trois jours possibles pour que le séjour coûte moins de :

a. 700 € en période haute ;

b. 560 € en période moyenne ;

c. 330 € en période basse.

4. Dans une publicité, l'agence de voyages affirme qu'un circuit complet de 10 jours est possible au prix de 2 800 € en période haute, 2 000 € en période moyenne et 1 250 € en période basse.
Comment se compose ce voyage ?

II Intervalles de fluctuation

II.1 Intervalle de fluctuation asymptotique

On dispose d'un urne contenant un très grand nombre de boules rouges et bleues.
On sait que la proportion de boules rouges dans l'urne est égale à $p = 0{,}4$.
Si on tire, successivement avec remise, n boules dans l'urne (n entier naturel non nul), et si on appelle X_n la variable aléatoire dénombrant les boules rouges tirées, alors X_n suit une loi binomiale $\mathscr{B}(n; p)$.

Théorème et définition

Soit X_n une variable aléatoire suivant une loi binomiale $\mathscr{B}(n; p)$, et α un réel tel que $0 < \alpha < 1$.
Si X est une variable aléatoire suivant la loi normale centrée réduite $\mathscr{N}(0; 1)$, on appelle u_α l'unique réel tel que :

$$P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$$

- On appelle I_n l'intervalle :

$$I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p}}{\sqrt{n}} \right]$$

  Alors,

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

- L'intervalle I_n contient la fréquence $F_n = \frac{X_n}{n}$ avec une probabilité qui se rapproche de $1 - \alpha$ lorsque n augmente :
  on dit que c'est l'**intervalle de fluctuation asymptotique de** F_n **au seuil** $1 - \alpha$.

Démonstration (exigible) :

On pose $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ et on applique le théorème de Moivre-Laplace.
Si X suit la loi normale $\mathscr{N}(0; 1)$: $\lim\limits_{n \to +\infty} P(Z_n \in [-u_\alpha; u_\alpha]) = P(X \in [-u_\alpha; u_\alpha] = 1 - \alpha$.
Or, $Z_n \in [-u_\alpha; u_\alpha] \Leftrightarrow -u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha \Leftrightarrow np - u_\alpha\sqrt{np(1-p} \leq X_n \leq np + u_\alpha\sqrt{np(1-p)}$.
$Z_n \in [-u_\alpha; u_\alpha] \Leftrightarrow p - u_\alpha \frac{\sqrt{p(1-p}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}$.
Donc, $Z_n \in [-u_\alpha; u_\alpha] \Leftrightarrow \frac{X_n}{n} \in I_n$.

Exemple :
On tire 50 boules de l'urne décrite ci-dessus, et on souhaite déterminer un intervalle de fluctuation au seuil 0,9 (c'est-à-dire $\alpha = 0{,}1$).
À l'aide de la calculatrice, on trouve 1,645 comme valeur approchée à 10^{-3} près de $u_{0,1}$.
On obtient donc comme intervalle de fluctuation :

$$I_{50} = \left[0{,}4 - u_{0,1} \frac{\sqrt{0{,}4 \times 0{,}6}}{\sqrt{50}} ; 0{,}4 + u_{0,1} \frac{\sqrt{0{,}4 \times 0{,}6}}{\sqrt{50}} \right]$$

soit $I_{50} = [0{,}286; 0{,}514]$.
Ainsi, en effectuant 50 tirages dans cette urne, la fréquence d'apparition d'une boule rouge est comprise entre 0,28 et 0,514 avec une probabilité d'environ 0,9.
Pour 500 tirages, au même seuil de 0,9, on obtient $I_{500} = [0{,}364; 0{,}436]$.

Remarque :
La longueur de l'intervalle a, pour un même seuil, été divisée par plus de 3 en passant de 50 à 500 tirages !!

Propriété
L'intervalle de fluctuation asymptotique au seuil de 95% pour une variable aléatoire X_n suivant une loi binomiale $\mathscr{B}(n; p)$ est l'intervalle :

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

Démonstration :
Cette propriété découle de celle vue au chapitre 11 concernant la loi normale.
On a vu en effet que $u_{0,05} \approx 1{,}96$.

Remarque :
En majorant $1{,}96\sqrt{p(1-p)}$ par 1, on retrouve l'intervalle de fluctuation présenté en classe de 2^nde^.
La majoration précédente résulta de ce qui suit :
$p(1-p) = -p^2 + p$ est un trinôme du second degré.
Il admet un maximum en $p = 0{,}5$, qui est égal à 0,25.
Ainsi, pour tout $p \in [0; 1], 0 \leq p(1-p) \leq 0{,}25$.
La fonction racine carrée étant croissante sur $\mathbb{R}_+$, on en déduit : $0 \leq \sqrt{p(1-p)} \leq 0{,}5$,
puis $0 \leq 1{,}96\sqrt{p(1-p)} \leq 1{,}95 \times 0{,}5 \leq 2 \times 0{,}5 = 1$.

II.2 Exercice

Énoncé
Dans un casino, il a été décidé que les « machines à sous » doivent être réglées sur une fréquence de gain du joueur de $g = 0{,}06$.
Une fréquence inférieure est supposée faire « fuir » le client, et une fréquence supérieure est susceptible de ruiner le casino.
Trois contrôleurs différents vérifient une même machine.
Le premier a joué 50 fois et gagné 2 fois,
le second a joué 120 fois et gagné 14 fois,
le troisième a joué 400 fois et gagné 30 fois.
En utilisant des intervalles de fluctuation asymptotiques au seuil 95%, examiner dans chaque cas la décision à prendre par le contrôleur, à savoir accepter ou rejeter l'hypothèse $g = 0{,}06$.

Solution :

- **Premier contrôleur** Le contrôle a porté sur 50 parties, donc $n = 50$.
  On a bien $n \geq 30$, mais $np = 50 \times 0{,}06 = 3$ et donc $np < 5$.
  On n'est pas dans les conditions d'application d'un intervalle de fluctuation asymptotique.
  Ce contrôle ne peut rien donner de probant en termes de prise de décision.
- **Deuxième contrôleur**
  Le contrôle a porté sur 120 parties, donc $n = 120, np = 7{,}2$ et $n(1-p) = 112{,}8$.
  Les conditions d'utilisation d'un intervalle de fluctuation asymptotique sont donc réunies.
  L'intervalle de fluctuation asymptotique au seuil 95% pour un échantillon de taille $n = 120$ est égal à :
  $$\mathrm{I}_{120} = \left[0{,}06 - 1{,}96\frac{\sqrt{0{,}06 \times 0{,}94}}{\sqrt{120}} ; 0{,}06 + 1{,}96\frac{\sqrt{0{,}06 \times 0{,}94}}{\sqrt{120}}\right]$$
  soit en arrondissant les bornes à 10^{-4} près : $\mathrm{I}_{120} = [0{,}0175 ; 0{,}1025]$.
  La fréquence observée par le deuxième contrôleur est $f_2 = \frac{14}{120} \approx 0{,}1167$.
  Cette fréquence est à l'extérieur de l'intervalle de fluctuation I_{120}, et le second contrôleur est conduit à rejeter l'hypothèse $g = 0{,}06$.
- **Troisième contrôleur**
  La contrôle a porté sur 400 parties, donc $n = 400, np = 24$ et $n(1-p) = 376$.
  Les conditions d'utilisation d'un intervalle de fluctuation asymptotique sont donc réunies.
  L'intervalle de fluctuation asymptotique au seuil 95% pour un échantillon de taille $n = 400$ est égal à :
  $$\mathrm{I}_{400} = \left[0{,}06 - 1{,}96\frac{\sqrt{0{,}06 \times 0{,}94}}{\sqrt{400}} ; 0{,}06 + 1{,}96\frac{\sqrt{0{,}06 \times 0{,}94}}{\sqrt{400}}\right]$$
  soit en arrondissant les bornes à 10^{-4} près : $\mathrm{I}_{400} = [0{,}0367 ; 0{,}0833]$.
  La fréquence observée par le deuxième contrôleur est $f_3 = \frac{30}{400} = 0{,}075 \in \mathrm{I}_{400}$.
  Ce dernier est donc amené à accepter l'hypothèse $g = 0{,}06$.

Remarques

1. Pour tester une hypothèse à l'aide d'un intervalle de fluctuation asymptotique, on doit vérifier ses conditions d'utilisation :
   $$n \geq 30, \quad np \geq 5, \quad n(1-p) \geq 5$$
   Ces conditions vérifiées, on détermine l'intervalle de fluctuation I_n au seuil demandé (souvent 95%).
   Si la fréquence observée f n'appartient pas à I_n, on rejette l'hypothèse.
2. L'intervalle de fluctuation étudié est « bilatéral » : on est conduit à rejeter l'hypothèse dans le cas où la fréquence observée est « trop grande » , ou « trop petite » .