fluctuations d'échantillonnages
intervalle de fluctuations
intervalle de confiance

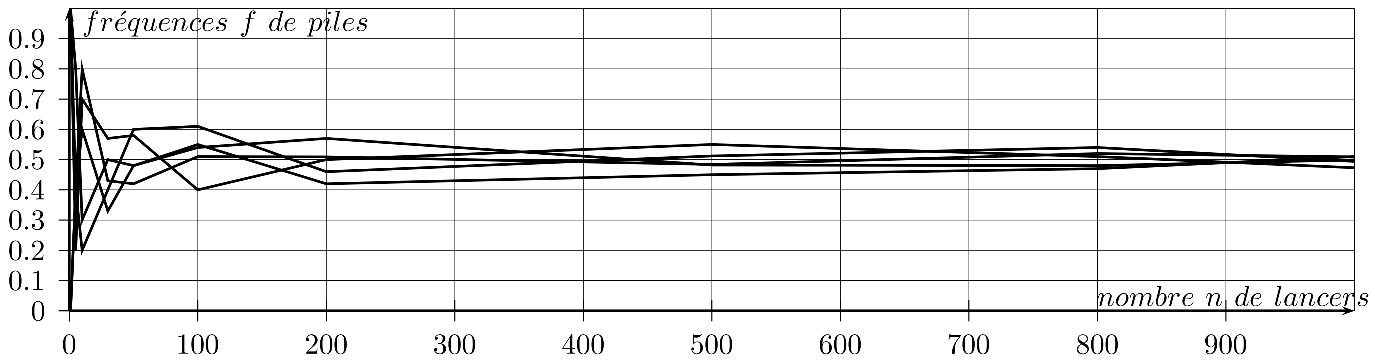
Table des matières

1 loi des grands nombres, intervalle de fluctuations

1.1 activité

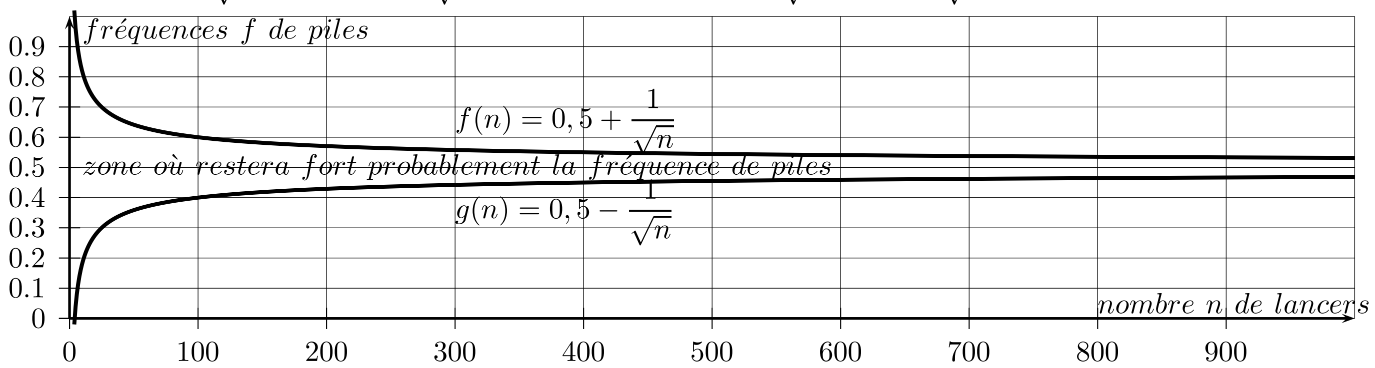
1.1.1 activité 1

on lance 1000 fois une pièce de monnaie équilibrée et on relève la fréquence f de piles au fur et à mesure que le nombre de lancers augmente
on recommence quatre fois de suite 1000 lancers et on obtient les 5 courbes ci dessous
on remarque qu'à chaque fois le pourcentage de piles s'est rapproché de $0,5$



un résultat remarquable est que, quel que soit le nombre n de lancers, la fréquence f de piles, restera "fort probablement" localisée entre $f(n) = 0,5 + \dfrac{1}{\sqrt{n}}$ et $g(n) = 0,5 - \dfrac{1}{\sqrt{n}}$

soit $0,5 - \dfrac{1}{\sqrt{n}} \leq f \leq 0,5 + \dfrac{1}{\sqrt{n}}$ ou encore $f \in [0,5 - \dfrac{1}{\sqrt{n}}\ ;\ 0,5 + \dfrac{1}{\sqrt{n}}]$ (intervalle de fluctuations)



le but de l'activité est de vérifier ce résultat au travers une simulation avec un tableur

1. **ouvrir et enregistrer dans votre dossier Mathématiques le fichier "tp (tableur)" de la ligne "fluctuations d'échantillonnages" de la page d'accueil de "secondes" du site "site.math.free.fr"**
2. **entrer la valeur 2 dans la cellule $B1$ pour fixer le nombre de résultats possibles à deux**
3. **simulation de 20 séries de 1000 lancers :**
   (a) **dans la cellule $B4$, entrer la formule $= ENT(\$B\$1 * ALEA())$**
   elle permet d'obtenir un nombre entier aléatoire égal à 0 (face) ou 1 (pile) (si $B1 = 2$)
   (b) **étirer la formule jusqu'à $U4$**
   (sont ainsi obtenus les premiers résultats de chacune des ... séries de ... tirages)
   (c) **sélectionner la plage de cellules $B4 : U4$ puis étirer la sélection vers le bas jusqu'à la ligne 1003 pour obtenir les 20 séries de 1000 tirages**
4. **obtention des valeurs de fréquences de piles pour chacune des 20 séries :**
   (a) **dans la cellule $W4$, entrer la formule : $= NB.SI(B\$4 : B4; 1)/\$A4$ puis étirer la formule jusqu'à $AP4$**
   à quoi sert le dollard devant le 4 ? : ...
   à quoi sert le dollard devant le A ? : ...
   (b) **sélectionner la plage de cellules $W4 : AP4$ puis étirer la sélection vers le bas jusqu'à la ligne 1003 pour obtenir 20 séries de fréquences de "piles" pour 1000 tirages**
   (c) **dans quelle cellule se trouve le pourcentage de "piles" que l'on a obtenu pour les 500 premiers lancers de la série 10 : ... et quel est ce pourcentage ? : ...**
   (d) **que signifie le résultat obtenu dans la cellule $AP1003$? : ...**

5. **obtention des courbes des fréquences de piles pour chacune des 20 séries**
   (a) **sélectionner la plage de cellules** $W4 : AP1003$ ***(survoler les 20 colonnes des fréquences à la souris clic gauche maintenu)***
   (b) **insertion -> graphique -> courbes -> courbes avec marques affichées à chaque point -> terminer (supprimer la légende et placer le graphique à sa place)**
   (c) **observations :** ***(F9 pour relancer les tirages et changer les courbes)***
      i. **les fluctuations de la fréquence sont-elles de plus en plus grande ou de plus en plus petites quand le nombre de tirages augmente ? : ...**
      ii. **de quelle valeur les fréquences de piles semblent-elles se rapprocher ? : ...**
6. **construction des "courbes-enveloppes" correspondant à l'intervalle de confiance** $f(n) = 0,5 + \frac{1}{\sqrt{n}}$ **et** $g(n) = 0,5 - \frac{1}{\sqrt{n}}$
   (a) **dans la cellule** $AR4$**, entrer la formule** $= 1/\$B\$1 - 1/RACINE(A4)$
   (b) **dans la cellule** $AS4$**, entrer la formule** $= 1/\$B\$1 + 1/RACINE(A4)$
   (c) **sélectionner la plage de cellules** $AR4 : AS4$ ***(survoler à la souris clic gauche maintenu)*** **puis étirer la sélection vers le bas jusqu'à la ligne** 1003 **pour obtenir les tableaux de valeurs des fonctions** f **et** g
   (d) **ajouter la courbe de** f **dans le repère précédent comme suit :**
      i. **clic droit dans la zone de tracage -> données sources -> série -> Ajouter -> valeur -> sélectionner la plage de cellules** $AR4 : AR1003$ **-> ok**
      ii. **de même pour la courbe de** g ***(mettre des Marques rouges pour les deux courbes)***
   (e) **observations :**
      i. **les 20 courbes des fréquences restent-elles toujours strictement entre les courbes de** f **et** g **sans déborder ?** ***(relancer les calculs avec*** $F9$ ***pour voir)*** **: ...**
      ii. **restent-elles "relativement globalement" entre les courbes : ...**
7. **estimation du pourcentage de valeurs qui restent dans l'intervalle de fluctuations**
   (a) **dans la celulle** $AV4$ **entrer la formule :**
   $= SOMMEPROD((W4 : AP4 >= AR4) * (W4 : AP4 <= AS4))/20$
   (elle donne, pour chaque valeur de n***, le pourcentage de valeurs qui se trouvent dans l'intervalle de fluctuations parmi les*** 20 ***valeurs obtenues)***
   étirer cette formule jusqu'à $AV1003$
   (b) **expliquer ce que donne la cellule** $AY4$ **: ...**
   (c) **est-il raisonnable de dire que l'on a "pratiquement toujours"** 95% **des 20 valeurs dans l'intervalle de fluctuations ? : ...**
8. **pour le cas d'un dé équilibré à 6 faces, quelle valeur mettre dans la cellule** $B2$ **? : ...**
   (a) **si on met** 6 **dans la cellule** $B2$**, les courbes de fréquences restent-elles globalement entre les courbes de** f **et** g **? : ...**
   (b) **si on met** 10 **dans la cellule** $B2$**, les courbes de fréquences restent-elles globalement entre les courbes de** f **et** g **? : ...**
9. **conclusion :**

   si on répète n **fois une expérience aléatoire** ***(exemple : n = 1000 lancers d'une pièce)*** **ou la probabilité d'obtenir un certain résultat** ***(exemple : pile)*** **est égale à** p ***(exemple :*** $p = 0,5$***)***

   quelle estimation de probabilité a t-on d'obtenir entre $0,5 - \frac{1}{\sqrt{1000}} = 46,8\%$ **et** $0,5 + \frac{1}{\sqrt{1000}} = 53,1\%$ **de piles ? :** $\boxed{5\%}$, $\boxed{95\%}$ **ou** $\boxed{\text{on ne peut pas savoir}}$
10. **application :**

    si on lance 2000 fois une pièce équilibrée, on peut dire avec 95% **de chance d'avoir raison que le nombre de "piles" sera compris entre** x **et** y
    calculer les valeurs de x **et** y

1.1.2 activité 2

A. $\boxed{\text{Intervalle de fluctuations à 95\% et loi des grands nombres}}$

si $\left\{\begin{array}{l}\text{on répète } n \text{ fois } (\text{n} \geq 25) \text{ une même expérience aléatoire}\\ \text{les expériences sont indépendantes}\\ \text{pour chacune, la probabilité que l'événement } A \text{ (succès) se produise est } p \in]0,2;0,8[\end{array}\right.$

alors $\left\{\begin{array}{l}\text{la fréquence } f \text{ de succès parmi les } n \text{ expériences}\\ \text{est dans l'intervalle } [p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}] \text{ avec une probabilité de 95\%}\end{array}\right.$

1. Dans un certain pays, il est fait l'hypothèse que la proportion femmes est égale à 45%
   (a) Pour un échantillon de 5000 personnes choisies au hasard,
   justifier le calcul de l'intervalles de fluctuations à 95% pour la fréquence de femmes et interpréter le résultat
   (b) Pour l'échantillon de taille 5000, il y a eu 2400 femmes,
   l'hypothèse de 45% de femmes dans la population peut-elle être raisonnablement validée ? avec quelle risque d'erreur ? *(probabilité de la refuser sachant qu'elle est vraie)*
2. Dans un autre pays, il est fait l'hypothèse que la proportion femmes est égale à 48%
   (a) Pour un échantillon de 2000 personnes choisies au hasard,
   justifier le calcul de l'intervalles de fluctuations à 95% pour la fréquence de femmes et interpréter le résultat
   (b) Pour l'échantillon de taille 2000, il y a eu 920 femmes,
   l'hypothèse de 48% de femmes dans la population peut-elle être raisonnablement validée ?

B. $\boxed{\text{Intervalle de fluctuations asymptotique à 95\% et loi des grands nombres}}$

si $\left\{\begin{array}{l}\text{on répète } n \text{ fois } (\text{n} \geq 30) \text{ une même expérience aléatoire}\\ \text{les expériences sont indépendantes}\\ \text{pour chacune, la probabilité que l'événement } A \text{ (succès) se produise est}\\ \text{telle que } \left\{\begin{array}{l} np > 5 \\ n(1-p) > 5\end{array}\right.\end{array}\right.$

alors $\left\{\begin{array}{l}\text{la fréquence } f \text{ de succès parmi les } n \text{ expériences est dans l'intervalle}\\ [p-1,96\times\frac{\sqrt{p(1-p)}}{\sqrt{n}};p+1,96\times\frac{\sqrt{p(1-p)}}{\sqrt{n}}] \text{ avec une probabilité de 95\%}\end{array}\right.$

1. Une livraison de sacs de farines doit contenir 60% de sacs de 25kg (le reste de 20kg)
   (a) Pour un échantillon de 500 sacs choisis au hasard,
   justifier le calcul de l'intervalles de fluctuations asymptotique à 95% pour la fréquence de sacs de 25 kg et interpréter le résultat
   (b) Pour l'échantillon de taille 500, il y a eu 280 sacs de 25kg,
   l'hypothèse de 60% de sacs de 25kg dans la livraison peut-elle être raisonnablement validée ?
2. Une autre livraison de sucre doit contenir 70% de sacs de 25kg (le reste de 20kg)
   (a) Pour un échantillon de 200 sacs choisis au hasard,
   justifier le calcul de l'intervalles de fluctuations asymptotique à 95% pour la fréquence de sacs de 25 kg et interpréter le résultat
   (b) Pour l'échantillon de taille 200, il y a eu 125 sacs de 25kg,
   l'hypothèse de 70% de sacs de 25kg dans la livraison peut-elle être raisonnablement validée ?

1.2 à retenir

propriété 1 : *(intervalle de fluctuations)*

Etant donnée une expérience aléatoire
Soit un événement quelconque A de probabilité p connue
On répète n fois cette expérience aléatoire avec indépendance
Soit f la proportion de fois où l'événement A se réalise parmi les n expériences
si $\boxed{n \geq 25 \text{ et } 0,2 \leq p \leq 0,8}$
alors $\boxed{f \text{ est dans l'intervalle } [p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}] \text{ avec une probabilité de } 95\%}$

Remarques :

i. f **n'est pas dans l'intervalle** $[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}]$ **avec une probabilité de** 5%

ii. **l'intervalle ci dessus est appelé** $\boxed{\text{intervalle de fluctuations}}$

Exemple :

pièce équilibrée : $p = 0,5$ **avec** $n = 100$ **lancers,**
on a bien :
$n \geq 25$ **et** $0,2 \leq p \leq 0,8$
donc
la fréquence de piles sera dans l'intervalle $[0,5 - \frac{1}{\sqrt{100}}\ ;\ 0,5 + \frac{1}{\sqrt{100}}]$
soit $[0,4\ ;\ 0,6]$ **avec une probabilité de** 95%

propriété 2 : *(intervalle de fluctuations asymptotique)*

Etant donnée une expérience aléatoire
Soit un événement quelconque A de probabilité p connue
On répète n fois cette expérience aléatoire avec indépendance
Soit f la proportion de fois où l'événement A se réalise parmi les n expériences
si $\boxed{n \geq 30 \text{ et } np \geq 5 \text{ et } n(1-p) \geq 5}$
alors
$\boxed{f \text{ est dans } [p - 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}}] \text{ avec une probabilité de } 95\%}$

Remarques :

i. f **n'est pas dans cet l'intervalle avec une probabilité de** 5%

ii. **l'intervalle ci dessus est appelé** $\boxed{\text{intervalle de fluctuations asymptotique}}$

iii. **l'intervalle de fluctuations asymptotique est "plus précis" que l'intervalle de fluctuations**

Exemple :

pièce équilibrée : $p = 0,5$ **pour** $n = 100$ **lancers,**
on a bien :
$n \geq 30$ **et** $np = 100 \times 0,5 = 50$ **soit np > 5 et** $n(1-p) = 100 \times (1-0,5) = 50$
soit n(1-p)>5
donc
la fréquence de piles sera dans l'intervalle
$[0,5 - 1,96 \times \frac{\sqrt{0,5(1-0,5)}}{\sqrt{100}}\ ;\ 0,5 + 1,96 \times \frac{\sqrt{0,5(1-0,5)}}{\sqrt{100}}]$ **soit** $[0,402\ ;\ 0,598]$ **avec une probabilité de** 95%

1.3 exercices

exercice 1 :

Dans une usine, une machine automatisée fabrique un certain objet

Quand elle est réglée au mieux, cette machine produit 20% d'objets qui ont un petit travail à terminer à la main

On souhaite savoir si cette machine est encore réglée au mieux ou si elle doit être à nouveau réglée

pour cela on controle 1000 objets prélevés au hasard dans le stock des objets produits par cette machine

Il a ainsi été obtenu 840 objets conformes et 160 non conformes

1. quelle est la probabilité p de tomber sur un objet non conforme si on choisit un objet au hasard dans la production ?
2. pour l'échantillon des 1000 objets prélevés, quelle est la fréquence f d'objets non conformes ?
   (a) les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de fluctuations pour f ? *(justifier)*
   (b) déterminer l'intervalle de fluctuations pour la fréquence f
   (c) faut-il régler la machine ou bien est-elle réglée au mieux *(justifier)*
   (d) pour une autre machine du même type, le controle de 500 objets prélevés au hasard dans le stock des objets produits par cette machine a donné 400 objets conformes, faut-il régler cette machine ? *(justifier)*
3. reprendre les questions précédentes avec l'intervalle de fluctuations asymptotique

exercice 2 :

Une entreprise est composée de 200 hommes et de 300 femmes

Chaque jour, un des 500 noms d'employés est choisit pour donner son avis sur le menu du jour à la cantine

1. quelle est la probabilité p de tomber sur un homme si on choisit un employé de l'entreprise au hasard ?
2. Le chef d'entreprise dit que pour chacun des 180 jours des 6 derniers mois, "le nom de l'employé a été choisit au hasard"
   Il a ainsi été obtenu 50 hommes et 130 femmes
   (a) calculer la fréquence f d'hommes obtenue
   (b) les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de fluctuations pour f ? *(justifier)*
   (c) déterminer l'intervalle de fluctuations pour la fréquence f
   (d) le chef d'entreprise dit que pour l'avis sur le menu du jour, "le nom de l'employé a été choisit au hasard"
   pensez vous que cela ait de grandes chances d'être vrai ? *(préciser)*
   (e) les 6 mois précédents, pour l'avis sur la cantine, il avait été obtenu 60 hommes et 120 femmes
   les noms d'employés peuvent-ils raisonnablement avoir été choisit au hasard ? *(justifier)*
3. reprendre les questions précédentes avec l'intervalle de fluctuations asymptotique

exercice 3 :

1. quelle doit être la taille de l'échantillon pour que l'amplitude de l'intervalle de fluctuations soit inférieure ou égale à 0,01 ?
2. est - il vrai que si on double la taille de l'échantillon alors l'amplitude de l'intervalle est divisée par deux ?

1.4 corrigés exercices

corrigé exercice 1 :

Dans une usine, une machine automatisée fabrique un certain objet
Quand elle est réglée au mieux, cette machine produit 20% d'objets qui ont un petit défaut à corriger à la main
On souhaite savoir si cette machine est encore réglée au mieux ou si elle doit être à nouveau réglée
pour cela on controle 1000 objets prélevés au hasard dans le stock des objets produits par cette machine
Il a ainsi été obtenu 840 objets conformes et 160 non conformes

1. **probabilité p de tomber sur un objet non conforme si on choisit un objet au hasard dans la production :** $\boxed{p = 0,2 = 20\%}$
2. **pour l'échantillon des 1000 objets prélevés, quelle est la fréquence f d'objets non conformes ? :** $\boxed{f = \dfrac{160}{1000} = 0,16 = 16\%}$
   (a) **les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de fluctuations pour f ?**
   oui car : $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq p \leq 0,8 \end{cases}}$
   (b) **déterminer l'intervalle de fluctuations pour la fréquence f**
   $\boxed{0,2 - \dfrac{1}{\sqrt{1000}} \simeq 0,17}$ $\boxed{0,2 + \dfrac{1}{\sqrt{1000}} \simeq 0,23}$ **soit :** $\boxed{[0,17\ ;\ 0,23]}$
   (c) **faut-il régler la machine ou bien est-elle réglée au mieux** ***(justifier)***
   $f = 0,16$ n'est pas dans l'intervalle $[0,17\ ;\ 0,23]$ donc la machine n'est pas bien réglée avec un risque de se tromper de 5%
   (d) **pour une autre machine du même type, le controle de 500 objets prélevés au hasard dans le stock des objets produits par cette machine a donné 400 objets conformes, faut-il régler cette machine ?** ***(justifier)***
   $\boxed{f = \dfrac{100}{500} = 0,2 = 20\%}$
   $f = 0,2$ est dans l'intervalle $[0,17\ ;\ 0,23]$ donc la machine est bien réglée avec un risque inconnu de se tromper

3. **avec l'intervalle de fluctuations asymptotique :**
   (a) **première machine :** $\begin{cases} n \geq 30 \\ np = 1000 \times 0,2 = 200 \geq 5 \\ nq = n(1-p) = 1000 \times 0,8 = 800 \geq 5 \end{cases}$
   l'intervalle de fluctuations asymptotisque est : $I = [0,175; 0,225]$
   $f = \dfrac{160}{1000} = 0,16 = 16\%$
   $0,16 \notin I$
   donc la macchine n'est pas bien réglée avec un risque de se tromper de 5 %
   (b) **seconde machine :** $\begin{cases} n \geq 30 \\ np = 500 \times 0,2 = 100 \geq 5 \\ nq = n(1-p) = 500 \times 0,8 = 400 \geq 5 \end{cases}$
   l'intervalle de fluctuations asymptotisque est : $I = [0,165; 0,235]$
   $f = \dfrac{100}{500} = 0,2 = 20\%$
   $0,2 \in I$
   donc la machine est bien réglée avec un risque inconnu de se tromper

corrigé exercice 2 :

Une entreprise est composée de 200 hommes et de 300 femmes
Chaque jour, un des 500 noms d'employés est choisit pour donner son avis sur le menu du jour à la cantine

1. **la probabilité p de tomber sur un homme si on choisit un employé de l'entreprise au hasard est :** $\boxed{p = \dfrac{200}{200+300} = 0,4 = 40\%}$
2. **Le chef d'entreprise dit que pour chacun des 180 jours des 6 derniers mois, "le nom de l'employé a été choisit au hasard"**
   Il a ainsi été obtenu 50 hommes et 130 femmes
   (a) **fréquence f d'hommes obtenue :** $\boxed{f = \dfrac{50}{50+130} \simeq 0,28 = 28\%}$
   (b) **les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de fluctuations pour f ?**
   oui car : $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq p \leq 0,8 \end{cases}}$
   (c) **intervalle de fluctuations pour la fréquence f :**
   $\boxed{0,4 - \dfrac{1}{\sqrt{180}} \simeq 0,33}$ $\boxed{0,4 + \dfrac{1}{\sqrt{180}} \simeq 0,47}$ **soit :** $\boxed{[0,33\ ;\ 0,47]}$
   (d) **le chef d'entreprise dit que pour l'avis sur le menu du jour, "le nom de l'employé a été choisit au hasard"**
   pensez vous que cela ait de grandes chances d'être vrai ? ***(préciser)***
   non, car $\boxed{0,28 \text{ n'appartient pas à } [0,33\ ;\ 0,47]}$ **avec un risque de se tromper de 5%**
   (e) **les 6 mois précédents, pour l'avis sur la cantine, il avait été obtenu 60 hommes et 120 femmes**
   les noms d'employés peuvent-ils raisonnablement avoir été choisit au hasard ?
   oui, car : $\boxed{f = \dfrac{60}{60+120} \simeq 0,33 = 33\%}$ **et** $\boxed{0,33 \in [0,33\ ;\ 0,47]}$
   avec une probabilité inconnue d'avoir raison

corrigé exercice 3 :

1. **quelle doit être la taille de l'échantillon pour que l'amplitude de l'intervalle de fluctuations soit inférieure ou égale à 0,01 ?**

   Amplitude : A $= 1+\sqrt{n}-(1-\sqrt{n}) = \dfrac{2}{\sqrt{n}}$

   on cherche $n \geq 0$ **solution de** $\dfrac{2}{\sqrt{n}} \leq 0,01$

   $\iff 0,01\sqrt{n} \geq 2$

   $\iff \sqrt{n} \geq \dfrac{2}{0,01}$

   $\iff \sqrt{n} \geq 200$

   $\iff n \geq 200^2$

   $\iff n \geq 40000$

2. **est - il vrai que si on double la taille de l'échantillon alors l'amplitude de l'intervalle est divisée par deux ?**

   Amplitude pour un échantillon de taille n : A $= \dfrac{2}{\sqrt{n}}$

   Amplitude pour un échantillon de taille 2n : : A' $= \dfrac{2}{\sqrt{2n}}$

   $$\frac{A'}{A} = \frac{\dfrac{2}{\sqrt{2n}}}{\dfrac{2}{\sqrt{n}}} = \frac{\sqrt{n}}{\sqrt{2n}} = \sqrt{\frac{n}{2n}} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}}$$

   $$A' = \frac{A}{\sqrt{2}}$$

   l'amplitude est donc divisée par $\sqrt{2} \simeq 1,414$
   mais pas par 2

2 intervalle de confiance

2.1 activité

2.1.1 activité 1

Intervalle de confiance

Etant donnée une expérience aléatoire
Soit un événement quelconque A de probabilité p inconnue
On répète n fois cette expérience aléatoire
Soit f la proportion de fois où l'événement A se réalise parmi les n expériences

si $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}}$

alors $\boxed{p \text{ est dans l'intervalle } [f - \frac{1}{\sqrt{n}} \; ; \; f + \frac{1}{\sqrt{n}}] \text{ avec une probabilité de } 95\%}$

1. **Un maire désire savoir s'il va être réélu**
   il commande un sondage avec un échantillon de taille $n = 1024$ sur les 42821 personnes de sa commune
   532 personnes ont déclaré qu'elles voterait pour le maire actuel
   (a) **justifier le calcul de l'intervalle de confiance à 95%**
   (b) **le maire peut - il espérer être réélu à coup sur ?**
2. **le maire commande alors un second sondage**
   (a) **quelle taille d'échantillon n_0 faut-il au moins pour avoir une amplitude de l'intervalle de confiance de moins de 2% ?**
   (b) **avec un tel échantillon de taille n_0 le maire obtient 5195 avis favorables**
   peut-il craindre de ne pas être réélu ?

2.2 à retenir

propriété 3 : *(intervalle de confiance)*

Etant donnée une expérience aléatoire
Soit un événement quelconque A de probabilité p inconnue
On répète n fois cette expérience aléatoire
Soit f la proportion de fois où l'événement A se réalise parmi les n expériences

si $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}}$

alors $\boxed{p \text{ est dans l'intervalle } [f - \frac{1}{\sqrt{n}} \ ; \ f + \frac{1}{\sqrt{n}}] \text{ avec une probabilité de } 95\%}$

Remarques :

i. p **n'est pas dans l'intervalle** $[f - \frac{1}{\sqrt{n}} \ ; \ f + \frac{1}{\sqrt{n}}]$ **avec une probabilité de** 5%

ii. **l'intervalle ci dessus est appelé** $\boxed{\textbf{intervalle de confiance}}$

Exemples :

a. **pièce équilibrée ou non ? :** $p(pile) = ?$
pour $n = 100$ **lancers, on obtient** 45 **piles et** 55 **faces**
la fréquence de piles est donc $f = 0,45$
on a bien : $\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}$
donc $p(piles) = p$ **sera dans l'intervalle** $[0,45 - \frac{1}{\sqrt{100}} \ ; \ 0,45 + \frac{1}{\sqrt{100}}]$ **soit** $[0,35 \ ; \ 0,55]$ **avec une probabilité de** 95%
donc $p = 0,5$ **peut-être accepté avec un risque non connu de se tromper**

b. **autre pièce équilibrée ou non ? :** $p(pile) = ?$
pour $n = 200$ **lancers, on obtient** 70 **piles et** 130 **faces**
la fréquence de piles est donc $f = \frac{70}{200} = 0,35$
on a bien : $\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}$
donc
$p(piles) = p$ **sera dans l'intervalle** $[0,35 - \frac{1}{\sqrt{200}} \ ; \ 0,35 + \frac{1}{\sqrt{200}}]$ **soit** $[0,279 \ ; \ 0,421]$ **avec une probabilité de** 95%
donc $p = 0,5$ **n'est pas acceptée avec un risque de** 5% **de se tromper**

2.3 exercices

exercice 4 :

Un vote va être organisé dans un lycée de 2000 élèves concernant une certaine mesure
On souhaite au préalable, savoir si cette mesure a des chances d'être acceptée et avec quelle proportion p de "pour". Cette évaluation sera faite par sondage
pour cela on tire 200 nom élèves au hasard et on leur demande leur avis
Il a ainsi été obtenu 80 "pour" et 120 "contres"

1. quelle est la fréquence f de "pour" obtenue suite au sondage ?
2. les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de confiance pour p ? *(justifier)*
3. déterminer l'intervalle de confiance pour p
4. sachant que la mesure est acceptée si au moins 50% des élèves sont "pour"
   pensez vous qu'elle va être acceptée ? *(justifier)*

exercice 5 :

Lors d'une élection, un sondage sur un échantillon de taille 1000 a donné pour trois candidats, les résultats suivants

candidat	A	B	C
résultat	32%	30%	26%

1. les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de confiance ?
2. déterminer l'intervalle de confiance pour chaque candidat
3. les résultats du premier tour ont donné les résultats suivants

candidat	A	B	C
résultat	$29,9\%$	$27,2\%$	$26,9\%$

   Faut-il remettre en cause la validité du sondage effectué avant l'élection ? (justifier)

exercice 6 :

1. L'intervalle de fluctuations au seuil de 95% est $[0,55\ ;\ 0,65]$
   déterminer la proportion p ainsi que la taille de l'échantillon n
2. L'intervalle de confiance au seuil de 95% est $[0,72\ ;\ 0,82]$
   déterminer la fréquence f ainsi que la taille de l'échantillon n

exercice 7 :

On souhaite estimer à 0,1 près la proportion p de personnes atteintes de la grippe dans la population française au niveau de confiance de 95%
Pour cela on considère un échantillon de taille n pour lequel on obtiendra un pourcentage f de personnes atteintes de la grippe

1. rappeler l'intervalle de confiance au niveau de confiance de 95% en fonction de n et f
2. quelle valeur faut-il choisir pour la taille n de l'échantillon pour que l'amplitude de l'intervalle soit d'au plus 0,1 ? Conclure
3. même question si on souhaite une précision de 0,01

exercice 8 :

Ecrire un algorithme qui donne les bornes de l'intervalles de fluctuations si on entre la proportion p ainsi que la taille de l'échantillon n

2.4 corrigés exercices

corrigé exercice 4 :

Un vote va être organisé dans un lycée de 2000 élèves concernant une certaine mesure
On souhaite au préalable, savoir si cette mesure a des chances d'être acceptée et avec quelle proportion p de "pour". Cette évaluation sera faite par sondage
pour cela on tire 200 nom élèves au hasard et on leur demande leur avis
Il a ainsi été obtenu 80 "pour" et 120 "contres"

1. **fréquence f de "pour" obtenue suite au sondage :** $\boxed{f = \dfrac{80}{80+120} = 0,4 = 40\%}$
2. **les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de confiance pour p ?** ***(justifier)***
   oui car : $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}}$
3. **intervalle de confiance pour p**
   $\boxed{0,40 - \dfrac{1}{\sqrt{200}} \simeq 0,33}$ $\boxed{0,4 + \dfrac{1}{\sqrt{200}} \simeq 0,47}$ **soit :** $\boxed{[0,33\ ;\ 0,47]}$
4. **sachant que la mesure est acceptée si au moins 50% des élèves sont "pour"**
   pensez vous qu'elle va être acceptée ? ***(justifier)***
   $p = 0,5$ **n'est pas dans l'intervalle** $[0,33\ ;\ 0,47]$ **donc la mesure ne sera acceptée avec un risque de se tromper de** 5%

corrigé exercice 5 :

Lors d'une élection, un sondage sur un échantillon de taille 1000 a donné pour trois candidats, les résultats suivants

candidat	A	B	C
résultat	32%	30%	26%

1. **les hypothèses sont-elles réunies pour appliquer la propriété de l'intervalle de confiance** ***(justifier)*** **oui car :** $\boxed{\begin{cases} n \geq 25 \\ 0,2 \leq f \leq 0,8 \end{cases}}$ **pour chacun des candidats**

2. **déterminer l'intervalle de confiance pour chaque candidat :**

   pour A : $\boxed{0,32 - \dfrac{1}{\sqrt{1000}} \simeq 0,29}$ $\boxed{0,32 + \dfrac{1}{\sqrt{1000}} \simeq 0,35}$ **soit :** $\boxed{[0,29 \ ; \ 0,35]}$

   pour B : $\boxed{0,3 - \dfrac{1}{\sqrt{1000}} \simeq 0,27}$ $\boxed{0,3 + \dfrac{1}{\sqrt{1000}} \simeq 0,33}$ **soit :** $\boxed{[0,27 \ ; \ 0,33]}$

   pour C : $\boxed{0,26 - \dfrac{1}{\sqrt{1000}} \simeq 0,23}$ $\boxed{0,26 + \dfrac{1}{\sqrt{1000}} \simeq 0,29}$ **soit :** $\boxed{[0,23 \ ; \ 0,29]}$

3. **les résultats du premier tour ont donné les résultats suivants**

candidat	A	B	C
résultat	$29,9\%$	$27,2\%$	$26,9\%$

   Faut-il remettre en cause la validité du sondage effectué avant l'élection ? (justifier)

   non, le sondage a été réalisé au hasard car :
   $\boxed{0,299 \in [0,29 \ ; \ 0,35]}$ **et** $\boxed{0,272 \in [0,27 \ ; \ 0,33]}$ **et** $\boxed{0,269 \in [0,23 \ ; \ 0,29]}$

corrigé exercice 6 :

1. **L'intervalle de fluctuations au seuil de 95% est $[0,55\ ;\ 0,65]$**
déterminer la proportion p ainsi que la taille de l'échantillon n

$$\boxed{p = \frac{0,55+0,65}{2} = 0,6}$$

$$0,6 + \frac{1}{\sqrt{n}} = 0,65$$

$$\iff \frac{1}{\sqrt{n}} = 0,65 - 0,6 = 0,05$$

$$\iff \sqrt{n} \times 0,05 = 1$$

$$\iff \sqrt{n} = \frac{1}{0,05} = 20$$

$$\iff \boxed{n = 20^2 = 400}$$

2. **L'intervalle de confiance au seuil de 95% est $[0,62\ ;\ 0,92]$**
déterminer la fréquence f ainsi que la taille de l'échantillon n

$$\boxed{f = \frac{0,62+0,92}{2} = 0,77}$$

$$0,77 + \frac{1}{\sqrt{n}} = 0,92$$

$$\iff \frac{1}{\sqrt{n}} = 0,92 - 0,77 = 0,15$$

$$\iff \sqrt{n} \times 0,15 = 1$$

$$\iff \sqrt{n} = \frac{1}{0,15} \simeq 6,67$$

$$\iff \boxed{n \simeq 6,67^2 \simeq 44}$$

corrigé exercice 7 :

Ecrire un algorithme qui donne les bornes de l'intervalles de fluctuation si on entre la proportion p ainsi que la taille de l'échantillon n

2.5 travaux pratiques

TP : $\boxed{\textbf{Instituts de sondages et intervalle de fluctuations}}$

Des instituts de sondages vont essayer de connaître à l'avance les résultats d'un référendum où chaque personne de la population doit choisir entre "pour" et "contre" ***(un vote nul ou blanc sera comptabilisé "contre")***
Pour cela, chaque institut va "sonder" un certain nombre de personnes ***(un sous ensemble)*** **de la population choisies au hasard de façons anonymes et recueillir leur avis**
Chaque institut va alors calculer les pourcentages de "pour" et "contre" obtenus dans leurs "sous population" et tirer une conclusion quand au résultat possible du référendum à venir.

1. **quelques questions d'intuition**
   (a) **pensez vous qu'un sondage puisse donner à chaque fois le bon résultat à l'avance ? : ...**
   (b) **pensez vous qu'un sondage "bien fait" puisse donner avec de "fortes chances" le bon résultat à l'avance ? : ...** ***(il reste à savoir ce qu'est un sondage "bien fait")***
2. **ouvrir et enregistrer dans votre dossier Mathématiques le fichier "tp2 (tableur)" de la ligne "fluctuations d'échantillonnages" de la page d'accueil de "secondes" du site "site.math.free.fr"**
3. **simulations des sondages**
   (a) **entrer la valeur** 20 **dans la cellule** $A2$, 80 **dans** $B2$ **et la formule** $= A2 + B2$ **dans la cellule** $C2$
   (b) **dans la cellule** $B6$, **entrer la formule =SI(ENT(C2*ALEA())+1 <=A2 ;"P" ;"C")**
   (cette formule écrit P ***avec une probabilité de*** 20% ***et*** C ***avec une probabilité de*** 80% ***et s'adapte aux valeurs de "pour" et de "contre" entrées précédemment)***
   (c) **étirer la formule jusqu'à** $D6$
   (sont ainsi obtenus les premiers avis de chacune des ... séries)
   (d) **sélectionner la plage de cellules** $B6 : D6$ **puis étirer la sélection vers le bas jusqu'à la ligne** 1005 **pour obtenir** 3 **séries de** 1000 **avis**
   (e) **mettre** 0 **dans la cellule** $A2$, **qu'obtient-on alors pour les avis ? : ...**
   mettre 10000 **dans la cellule** $A2$, **n'obtient-on alors toujours que des "pour" ? : ...**
   (f) **remettre** 20 **dans la cellule** $A2$
4. **obtention de la fréquence de "pour" en fonction du nombre d'avis**
   (a) **dans la cellule** $F6$, **entrer la formule :** $= NB.SI(B\$6 : B6; "P")/\$A6$ **puis étirer la formule jusqu'à** $H6$
   (remarquer les dollards devant le 6 ***et le*** A ***pour empécher les décalages des "coordonnées" des cellules verticalement ou horizontalement)***
   (b) **sélectionner la plage de cellules** $F6 : H6$ **puis étirer la sélection vers le bas jusqu'à la ligne** 1005 **pour obtenir les séries de fréquences de "pour" jusqu'a** 1000 **avis**
   (c) **dans quelle cellule se trouve le pourcentage de "pour" que l'on a obtenu pour les 10 premiers avis de la série I2 ? : ... quel est ce pourcentage ? : ...**
   pour la série I1 le pourcentage est de : ... pour la série I3 : ...
   (d) **que signifie le résultat obtenu dans la cellule** $H1005$ **? : ...**

5. **obtention des courbes des fréquences de "pour" pour chacune des** 3 **séries**
   (a) **sélectionner la plage de cellules** $F6 : H1005$ ***(survoler les 3 colonnes des fréquences à la souris, clic gauche maintenu)***
   (b) **insertion -> graphique -> courbes -> courbes avec marques affichées à chaque point -> terminer** ***(supprimer la légende et placer le graphique à sa place)***
   (c) **mettre une légende à chaque axe du graphique :**
   clic droit sur la zone de traçage-> Options du graphique -> ...
   (d) **observations :** ***(F9 pour relancer les tirages et changer les courbes)***
      i. **les fluctuations des fréquences de "pour" sont-elles de plus en plus grande ou de plus en plus petites quand le nombre d'avis augmente ? : ...**

ii. de quelle valeur les fréquences de "pour" semblent-elles se rapprocher ? : ...

iii. Même question si on remplace le 20 de la cellule $A2$ par un 50 :
les fréquences semblent se rapprocher de (valeur exacte puis approchées à 1% près) : ...

iv. conclusions partielles (mettre 500 dans $A2$ et 2000 dans $B2$)

A. le pourcentage de "pour" de la population entière est donc de : ...

B. si chaque institut de sondage ne recueille au hasard que 5 avis
ont-ils de fortes chances que les fréquences de "pour" obtenues soient proches de la fréquence de "pour" de la population entière ? : ...
par quelle plage de cellules pouvez vous illustrer votre avis ?
plage de cellules : ...   pourcentages obtenus : ...

C. si chaque institut de sondage recueille au hasard 1000 avis
ont-ils de fortes chances que les fréquences de "pour" obtenues soient proches de la fréquence de "pour" de la population entière ? : ...
par quelle plage de cellules pouvez vous illustrer votre avis ?
plage de cellules : ...   pourcentages obtenus : ...

D. pour que la fréquence de "pour" sur la sous population choisie au hasard soit relativement proche de la "vraie" proportion de "pour" de la "vraie" population il faut que le nombre d'avis recueilli soit assez ...

6. Intervalle de fluctuations
Des mathématiciens ont montré que :

$$\text{si} \begin{cases} \text{la vraie proportion de "pour" est égale à } p \text{ avec : } 0,2 < p < 0,8 \\ \\ \text{le nombre d'avis recueillis au hasard est } n \text{ avec } n > 25 \end{cases}$$

alors la proportion f de "pour" dans la sous population sera avec une probabilité de 95% dans "l'intervalle de fluctuations" suivant : $\boxed{[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}}]}$

(a) dans le cas où : $\begin{cases} \text{vraie population de 2500 personnes dont 500 "pour"} \\ \\ \text{le nombre d'avis recueillis au hasard est } n = 1000 \end{cases}$

i. la valeur de p est égale à : ...

ii. la valeur de $p - \frac{1}{\sqrt{n}}$ est (à 0,01 près) : ...

iii. la valeur de $p + \frac{1}{\sqrt{n}}$ est (à 0,01 près) : ...

iv. l'intervalle de fluctuations est : ...

v. pour 1000 avis recueillis au hasard, la probabilité que la fréquence de "pour" soit dans l'intervalle : ...   est de ...   %

(b) le résultat précédent est-il en accord avec les résultats obtenus avec la tableur ?
(justifier clairement en évoquant les valeurs de trois fréquences ainsi que les noms des cellules dans lesquelles se trouvent ces valeurs)

7. application :
si on lance 2000 fois un dé équilibré, on peut dire avec 95% de chance d'avoir raison que le nombre de "6" sera compris entre x et y
calculer les valeurs de x et y et donner l'intervalle de fluctuations

2.6 devoir maison

corrigé devoir maison

1. (a) **pourcentage de femmes dans l'entreprise B :** $\dfrac{1150}{2500} = 0,46 = 46\%$

   pourcentage de femmes dans l'entreprise A : $\dfrac{43}{100} = 0,43 = 43\%$

   (b) **le pourcentage de femmes est plus élevé dans l'entreprise B car $46\% > 43\%$, cette entreprise B est plus proche de la parité que l'autre.**

2. (a) **intervalles de fluctuations à 95%**

   pour A : $100 \geq 25$ **et** $0,2 \leq 0,5 \leq 0,8$ **donc les hypothèses sont validées pour calculer l'intervalle de fluctuations :**

   $\boxed{0,5 - \dfrac{1}{\sqrt{100}} = 0,4}$ $\boxed{0,5 + \dfrac{1}{\sqrt{100}} \simeq 0,6}$ **soit :** $\boxed{[0,4\ ;\ 0,6]}$

   pour B : $2500 \geq 25$ **et** $0,2 \leq 0,5 \leq 0,8$ **donc les hypothèses sont validées pour calculer l'intervalle de fluctuations :**

   $\boxed{0,5 - \dfrac{1}{\sqrt{2500}} = 0,48}$ $\boxed{0,5 + \dfrac{1}{\sqrt{2500}} = 0,52}$ **soit :** $\boxed{[0,48\ ;\ 0,52]}$

   (b) **pour A :** $0,43 \in [0,4\ ;\ 0,6]$ **donc on peut considérer que la parité est respectée au risque de 5% de se tromper**

   pour B : $0,46 \notin [0,48\ ;\ 0,52]$ **donc on peut considérer que la parité n'est pas respectée au risque de 5% de se tromper**

   Ce qui ne conforte pas la première impression

3 compléments *(pour prof)*

Justification du $\boxed{f \in [\ p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\ ] \text{ avec une probabilité de } 95\%}$

1. **Soit** $X_i \in \{0;1\}$ **une variable aléatoire de loi de Bernoulli de paramètre** $p \in [0;1]$
2. **Soit** $S_n = X_1 + ... + X_n$ **somme de** n **variables aléatoires indépendantes du type précédent**
   (a) S_n **suit une loi binomiale** $B(n;p)$
   (b) $p(na \leq S_n \leq nb) = p(a \leq \frac{S_n}{n} \leq b)$ **où** $\frac{S_n}{n}$ **correspond au** f **ci dessus**
   (c) **on cherche** a **et** b **tels que** $p(na \leq S_n \leq nb) \simeq 0,95$
   (d) **sous certaines conditions** ***(np assez grand)*** **la loi de** S_n **est approchée par la loi de** Y **où** Y **suit une loi normale** $N(np; \sqrt{np(1-p)}$
   (e) **on cherche** a **et** b **tels que** $p(na \leq Y \leq nb) \simeq 0,95$
   ce qui équivaut à
   on cherche a **et** b **tels que** $p(a \leq \frac{Y}{n} \leq b) \simeq 0,95$
   où $Z = \frac{Y}{n}$ **suit une loi** $N(p; \frac{\sqrt{np(1-p)}}{n})$
   donc
   on cherche a **et** b **tels que** $p(a \leq Z \leq b) \simeq 0,95$ **où** Z **suit une loi** $N(p; \frac{1}{\sqrt{n}}\sqrt{p(1-p)})$
3. **soit** $T = \frac{Z-p}{\sigma}$ **où** $\sigma = \frac{1}{\sqrt{n}}\sqrt{p(1-p)}$
   Z **suit une loi normale centrée réduite** $N(0;1)$
   (a) **on cherche** a **et** b **tels que** $p(\frac{a-p}{\sigma} \leq T \leq \frac{b-p}{\sigma}) \simeq 0,95$
   (b) **or on sait que** $p(-2 \leq T \leq 2) \simeq 0,95$
   (c) **il suffit alor s de prendre** $\frac{b-p}{\sigma} = 2$ **et** $\frac{a-p}{\sigma} = -2$
   (d) **ce qui donne** $a = p - 2\sigma$ **et** $b = p + 2\sigma$
   (e) **soit** $a = p - 2\frac{1}{\sqrt{n}}\sqrt{p(1-p)}$ **et** $b = p + 2\frac{1}{\sqrt{n}}\sqrt{p(1-p)}$
   (f) **or pour** $p \in [0,2\ ;\ 0,8]$ **on a** $2\sqrt{p(1-p)} \in [0,8\ ;\ 1]$
4. **on peut donc prendre les approximations** $a \simeq p - \frac{1}{\sqrt{n}}$ **et** $b \simeq p + \frac{1}{\sqrt{n}}$
5. **un intervalle de confiance plus précis est avec :**
   $a = p - \frac{1}{\sqrt{n}} \times 2\sqrt{p(1-p)}$ **et** $b = p + \frac{1}{\sqrt{n}} \times 2\sqrt{p(1-p)}$
   soit
   $\boxed{f \in [\ p - \frac{1}{\sqrt{n}} \times 2\sqrt{p(1-p)}\ ;\ p + \frac{1}{\sqrt{n}} \times 2\sqrt{p(1-p)}\ ] \text{ avec une probabilité de } 95\%}$