



### ROC

Les **ROC**, (**R**estitution **O**rganisée de **C**onnaissances), sont les démonstrations du cours à connaître indiquées explicitement dans le nouveau programme de terminale S entré en vigueur à la rentrée 2012.

Ce chapitre compte 1 ROC.

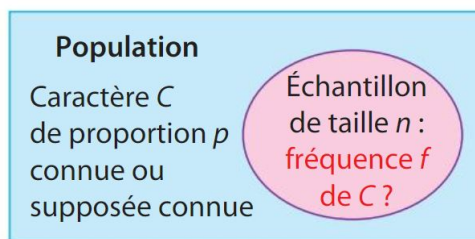
## I Intervalle de fluctuation au seuil de 95% et intervalle de fluctuation asymptotique au seuil de 95%

### I.1 La problématique

Dans une certaine population, on connaît la proportion d'individus présentant la caractère C.

On considère maintenant un échantillon de taille  $n$  d'individus de cette population.

Que peut-on dire de la fréquence  $f$  des individus de l'échantillon ayant la caractère C.



### I.2 A partir de la loi binomiale : un intervalle de fluctuation au seuil 0,95

On a vu en première que :

#### Propriété 1 (intervalle de fluctuation au seuil 0,95)

- Si la variable aléatoire  $X_n$  est le nombre d'individus d'un échantillon de taille  $n$ , présentant le caractère C;
- et si  $F_n = \frac{X_n}{n}$  est la variable aléatoire correspondant à la fréquence du caractère C dans l'échantillon de taille  $n$  alors on a :

1. La variable aléatoire  $X_n$  suit une loi binomiale  $\mathcal{B}(n; p)$ . Elle est définie par :

$$P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ pour tout entier } k \text{ tel que } 0 \leq k \leq n$$

2. On peut déterminer un intervalle de la forme  $\left[\frac{a}{n}; \frac{b}{n}\right]$ , où  $a$  et  $b$  sont les plus petit entiers tels que :

$$\begin{cases} P(X_n \leq a) > 0,025 \\ P(X_n \leq b) \geq 0,975 \end{cases}$$

3. Un tel intervalle est appelé **intervalle de fluctuation au seuil 0,95** de  $F_n$  et on a :

$$P\left(\frac{a}{n} \leq F_n \leq \frac{b}{n}\right) \geq 0,95$$



**Exemple**

Dans une urne contenant 3 boules rouges et 7 boules bleues, on effectue 100 tirages avec remise. On désigne par  $X$  la variable aléatoire correspondant au nombre de boules rouges obtenues et  $F$  la variable aléatoire fréquence associée.

Déterminer un intervalle de fluctuation au seuil 0,95 de la variable aléatoire  $F$ .



**Preuve**

• **Modélisation**

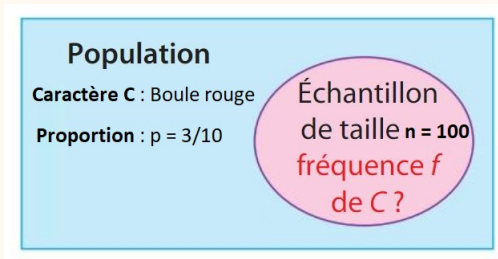
Il y a répétition de  $n = 100$  événements indépendants et identiques (on tire une boule rouge). Chaque tirage a deux issues possibles (épreuve de Bernoulli) :

- succès de probabilité  $p = 0,3$  quand une boule rouge est tirée;
- et échec de probabilité  $1 - p = 0,7$  sinon.

Donc la variable aléatoire  $X$  qui est égale au nombre de succès au cours de ces  $n = 100$  épreuves indépendantes de Bernoulli de paramètre  $p = 0,3$  suit une loi binomiale de paramètres  $n = 100$  et  $p = 0,3$ . On peut écrire :

$$X \text{ suit } \mathcal{B}(100; 0,3) \text{ ou } X \sim \mathcal{B}(100; 0,3).$$

• **Intervalle de fluctuation au seuil 0,95**



D'après la cours, on sait que l'on peut déterminer un intervalle de la forme  $\left[\frac{a}{n}; \frac{b}{n}\right]$ , où  $a$  et  $b$  sont les plus petit entiers tels que :

$$\begin{cases} P(X_n \leq a) > 0,025 \\ P(X_n \leq b) \geq 0,975 \end{cases}$$

• **Utilisation de la calculatrice**

La calculatrice nous donne  $a$  est le plus petit entier tel que  $P(X_n \leq a) > 0,025$  :

$$\begin{cases} P(X_n \leq 20) \approx 0,0165 < 0,025 \\ P(X_n \leq 21) \approx 0,0288 > 0,025 \end{cases} \implies a = 21$$

La calculatrice nous donne  $b$  est le plus petit entier tel que  $P(X_n \leq b) \geq 0,975$  soit :

$$\begin{cases} P(X_n \leq 38) \approx 0,966 < 0,975 \\ P(X_n \leq 39) \approx 0,979 > 0,975 \end{cases} \implies b = 39$$

• **Conclusion.**

Un intervalle de fluctuation au seuil 0,95 de  $F_n$ , la fréquence de boules rouges dans un échantillon de taille 100 est donc :

$$\left[\frac{a}{n}; \frac{b}{n}\right] = [0,21; 0,39] \implies P(0,21 \leq F_n \leq 0,39) \geq 0,95$$

### I.3 Intervalle de fluctuation asymptotique au seuil de 95%

#### I.3.1 ROC : Application du Théorème de Moivre-Laplace

##### Théorème 1 (Intervalle de fluctuation asymptotique)

Soit  $X_n$  une variable aléatoire qui suit la loi binomiale  $\mathcal{B}(n; p)$  et  $F_n = \frac{X_n}{n}$ .  
Soit  $\alpha \in ]0; 1[$  et  $u_\alpha$  le réel tel que  $P(-u_\alpha \leq Z_n \leq u_\alpha) = 1 - \alpha$ .  
où  $Z$  suit la loi normale centrée réduite  $\mathcal{N}(0; 1)$ .

1. Alors

$$\lim_{n \rightarrow +\infty} P(F_n \in I_n) = 1 - \alpha \quad \text{avec} \quad I_n = \left[ p - u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

2. L'intervalle  $I_n$  est un intervalle de fluctuation asymptotique au seuil  $1 - \alpha$  de la fréquence  $F_n = \frac{X_n}{n}$ .

3. Un intervalle de fluctuation asymptotique de la variable aléatoire  $F_n$  au seuil  $1 - \alpha$  est un intervalle déterminé à partir de  $p$  et de  $n$  et qui contient  $F_n$  avec une probabilité d'autant plus proche de  $1 - \alpha$  que  $n$  est grand.



#### ROC 1 : ROC Exigible

Soit  $\alpha \in ]0; 1[$  et  $u_\alpha$  le réel tel que  $P(-u_\alpha \leq Z_n \leq u_\alpha) = 1 - \alpha$ .  
où  $Z$  suit la loi normale centrée réduite  $\mathcal{N}(0; 1)$ .

- Si  $X_n$  est une variable aléatoire de loi  $\mathcal{B}(n; p)$ , alors son espérance est  $E(X_n) = np$  et son écart-type  $\sigma = \sqrt{np(1-p)}$ .
- D'après le théorème de Moivre-Laplace, on a :

$$\lim_{n \rightarrow +\infty} P\left(-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha\right) = P(-u_\alpha \leq Z_n \leq u_\alpha) = 1 - \alpha$$

• Or

$$\begin{aligned} P\left(-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha\right) &= P\left(-u_\alpha \sqrt{np(1-p)} \leq X_n - np \leq u_\alpha \sqrt{np(1-p)}\right) \\ &= P\left(np - u_\alpha \sqrt{np(1-p)} \leq X_n \leq u_\alpha \sqrt{np(1-p)} + np\right) \\ &= P\left(\frac{np - u_\alpha \sqrt{np(1-p)}}{n} \leq \frac{X_n}{n} \leq \frac{u_\alpha \sqrt{np(1-p)} + np}{n}\right) \\ &= P\left(p - u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \end{aligned}$$

• Donc

$$\lim_{n \rightarrow +\infty} P(F_n \in I_n) = 1 - \alpha \quad \text{avec} \quad I_n = \left[ p - u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

**I.3.2 Cas particulier : le seuil 0,95**

Au seuil  $1 - \alpha = 0,95$  on cherche à écrire cet intervalle de fluctuation asymptotique. Il nous faut donc déterminer  $u_\alpha$ .

Soit  $\alpha \in ]0 ; 1[$  et  $u_\alpha$  le réel tel que  $P(-u_\alpha \leq Z_n \leq u_\alpha) = 1 - \alpha = 0,95$  avec  $Z \sim \mathcal{N}(0 ; 1)$ .

On applique alors la propriété suivante :

**Propriété 2**

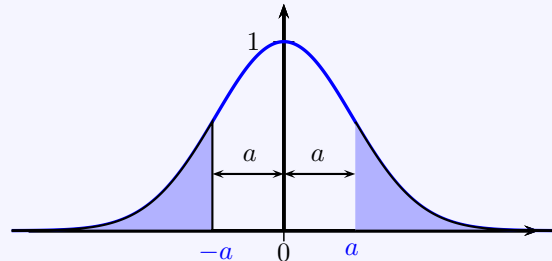
Soit  $X$  une variable aléatoire qui suit la loi normale centrée réduite  $\mathcal{N}(0 ; 1)$ .

1. La fonction  $\Phi$  est définie sur  $\mathbb{R}$  par

$$\Phi(t) = P(X \leq t)$$

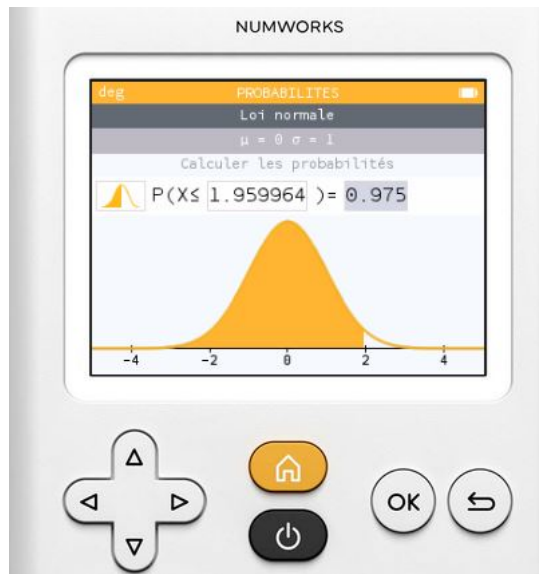
2. Pour tout réel  $a$  on a :

- (1) :  $P(X \leq -a) = P(X \geq a)$
- (2) :  $\Phi(-a) = 1 - \Phi(a)$
- (3) :  $P(-a \leq X \leq a) = 2\Phi(a) - 1$



Donc puisque  $Z \sim \mathcal{N}(0 ; 1)$  on a :

$$\begin{aligned} P(-u_\alpha \leq Z_n \leq u_\alpha) = 1 - \alpha = 0,95 &\iff 2\Phi(u) - 1 = 0,95 \\ &\iff \Phi(u) = \frac{1,95}{2} = 0,975 \\ &\iff \underline{u \approx 1,96} \end{aligned}$$



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

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

I.3.3 Intervalle de fluctuation asymptotique au seuil de 95%

**Théorème 2** (Intervalle de fluctuation asymptotique au seuil 0,95)

Si les conditions suivantes sont remplies :

$$\begin{cases} \checkmark & n \geq 30 \\ \checkmark & np \geq 5 \\ \checkmark & n(1-p) \geq 5 \end{cases}$$

Alors un **intervalle de fluctuation asymptotique** au seuil de confiance de 95% de la fréquence  $F_n$  d'un caractère dans un échantillon de taille  $n$  est si  $p$  désigne la proportion de ce caractère dans la population :

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

**Remarque**

Un intervalle de fluctuation asymptotique de la variable aléatoire  $F_n$  au seuil 0,95 est un intervalle déterminé à partir de  $p$  et de  $n$  et qui contient  $F_n$  avec une probabilité d'autant plus proche de 0,95 que  $n$  est grand.

I.4 Exemple type bac : Intervalle de fluctuation asymptotique au seuil de 95%

**Exemple 1 : Point Bac**

La proportion de chômeurs dans la population active est de 10%. Déterminer un intervalle de fluctuation au seuil de 95% de la fréquence des chômeurs dans les échantillons de taille 400. (Arrondir au millième).

On a pour le cas étudié,  $n = 400$ ,  $p = 10\%$ . Vérifions les conditions d'application du théorème :

$$\begin{cases} \checkmark & n = 400 \geq 30 \\ \checkmark & np = 400 \times 0,1 = 40 \geq 5 \\ \checkmark & n(1-p) = 400 \times 0,9 = 360 \geq 5 \end{cases}$$

Un intervalle fluctuation asymptotique au seuil de confiance de 95% de la fréquence  $f$  dans un échantillon de taille  $n = 400$  : est alors :

$$I_n = \left[ p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] = \left[ 0,1 - 1,96 \frac{\sqrt{0,1 \times 0,9}}{\sqrt{400}} ; 0,1 + 1,96 \frac{\sqrt{0,1 \times 0,9}}{\sqrt{400}} \right]$$

Soit puisque les borne sont :

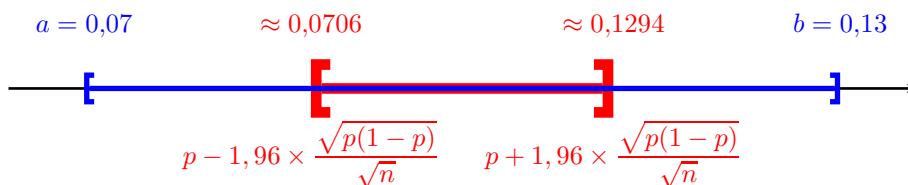
- $p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0,0706$  . On arrondit la borne inférieure par défaut  $10^{-3}$  près soit 0,07.
- $p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0,1294$  . On arrondit la borne supérieure par excès  $10^{-3}$  près soit 0,13.

$$I_{400} \approx [0,07 ; 0,13]$$

**Conclusion** : En considérant un échantillon de taille 400 dans la population, la probabilité que la fréquence de chômeurs dans cet échantillon appartienne à l'intervalle  $I_{400}$  est voisine de 95%. Autrement dit, il y a 95% de chances que la fréquence de chômeurs de cet échantillon soit comprise entre 7% et 13%.

**Remarque**

Il faut toujours proposer un intervalle en valeurs approchées qui contient l'intervalle en valeurs exactes.



## II Décision à partir de la fréquence d'un échantillon, au seuil de 5%

Quand les critères d'approximation sont vérifiés, l'intervalle de fluctuation asymptotique  $I_n$  permet de déterminer des seuils de décision :

- pour accepter ou rejeter l'hypothèse selon laquelle  $p$  est la proportion d'un caractère dans la population ;
- pour déterminer si un échantillon issu de la population est représentatif.

### Méthode 1

On formule l'hypothèse que la proportion d'un caractère dans la population est  $p$ . On prélève dans la population un échantillon de taille  $n$  et on note  $f$  la fréquence observée du caractère étudié. Lorsque  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$  on détermine  $I_n$  l'intervalle de fluctuation asymptotique au seuil 95% de la fréquence  $f$ .

- Si la fréquence observée  $f$  n'appartient pas à l'intervalle  $I_n$ , alors on rejette l'hypothèse selon laquelle  $p$  est la proportion du caractère étudié dans la population avec un risque d'erreur de 5 %.
- Si la fréquence observée  $f$  appartient à l'intervalle  $I_n$ , alors l'hypothèse selon laquelle  $p$  est la proportion du caractère étudié dans la population est acceptée.

### Exemple 2 : point bac

Dans un cabinet d'assurance, une étude est réalisée sur la fréquence des sinistres déclarés par les clients ainsi que leur coût. Une enquête affirme que 30 % des clients ont déclaré un sinistre au cours de l'année. Un expert indépendant interroge un échantillon de 100 clients choisis au hasard dans l'ensemble des clients du cabinet d'assurance.

L'expert constate que 19 clients ont déclaré un sinistre au cours de l'année. Déterminer, en justifiant, si l'affirmation du cabinet d'assurance : « 30 % des clients ont déclaré un sinistre au cours de l'année » peut être validée par l'expert.

#### • Analyse des données :

- « Sur un échantillon de  $n = 100$  clients. Il est constaté que 19 d'entre eux ont déclaré un sinistre. ».  
Donc la fréquence observée clients qui ont déclaré un sinistre est

$$f = 19 \div 100 = 0,19 \text{ soit } \underline{f = 0,19}$$

- On veut tester l'hypothèse : « la proportion de clients qui ont déclaré un sinistre est  $p = 30\%$  ».

#### • Intervalle de fluctuation :

On a pour le cas étudié,  $n = 100$ ,  $p = 30\%$ . Vérifions les conditions d'application du théorème :

$$\left\{ \begin{array}{l} \checkmark \quad n = 100 \geq 30 \\ \checkmark \quad np = 100 \times 0,3 = 30 \geq 5 \\ \checkmark \quad n(1 - p) = 100 \times 0,7 = 70 \geq 5 \end{array} \right.$$

Un intervalle fluctuation asymptotique au seuil de confiance de 95% de la fréquence  $f$  dans un échantillon de taille  $n = 100$  : est alors :

$$I_n = \left[ p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] = \left[ 0,3 - 1,96 \frac{\sqrt{0,3 \times 0,7}}{\sqrt{100}} ; 0,3 + 1,96 \frac{\sqrt{0,3 \times 0,7}}{\sqrt{100}} \right]$$

Soit puisque les borne sont :

$$\left| \begin{array}{l} \blacksquare \quad p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0,21018. \text{ On arrondit la borne inférieure par défaut } 10^{-3} \text{ près soit } \underline{0,21}. \\ \blacksquare \quad p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0,38982. \text{ On arrondit la borne supérieure par excès } 10^{-3} \text{ près soit } \underline{0,39}. \end{array} \right.$$

$$I_{100} \approx [0,21 ; 0,39]$$

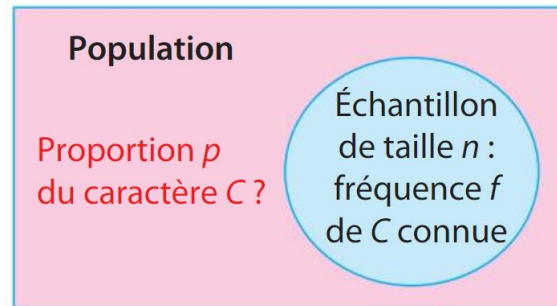
#### • Conclusion

La fréquence observée  $f = 0,19$  n'appartient pas à l'intervalle de fluctuation asymptotique  $I_{100}$ , donc le résultat du contrôle remet en question l'hypothèse, avec un risque d'erreur de 5%.

### III Intervalle de confiance

#### III.1 La problématique

Dans une certaine population, on connaît la fréquence d'individus présentant la caractère C sur un échantillon de taille  $n$  d'individus de cette population.  
Que peut-t-on dire de la proportion  $p$  des individus de ayant la caractère C dans la population ?



#### III.2 Théorème

##### Théorème 3

Si les conditions sont remplies :  $\left\{ \begin{array}{l} \checkmark \quad n \geq 30 \\ \checkmark \quad nf \geq 5 \\ \checkmark \quad n(1 - f) \geq 5 \end{array} \right.$ .

En notant  $f$  la fréquence observée d'un caractère dans un échantillon de taille  $n$  extrait d'une population dans laquelle la proportion de ce caractère est  $p$ . Alors l'intervalle de confiance au seuil 95% de la proportion inconnue  $p$  de ce caractère dans la population est :

$$\left[ f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$$

- Interprétation : la proportion  $p$  est dans l'intervalle de confiance au niveau de confiance 0,95.



##### Remarques

1. La précision d'un intervalle de confiance centré sur la fréquence  $f$  est donnée par son amplitude soit  $\frac{2}{\sqrt{n}}$ .
2. Plus la taille de l'échantillon est grande, plus les intervalles de confiance obtenus sont précis.
3. La **différence entre deux fréquences  $f_1$  et  $f_2$  observées** sur deux échantillons est considérée comme significative quand les intervalles de confiance correspondants sont disjoints.  
Dans ce cas, on considère que les deux proportions  $p_1$  et  $p_2$  sont différentes.  
Dans le cas contraire, on ne peut pas conclure.
4. Un **intervalle de confiance** encadre une valeur réelle que l'on cherche à estimer à l'aide de mesures prises par un procédé aléatoire.  
En particulier, cette notion permet de définir une marge d'erreur entre les résultats d'un sondage et un relevé exhaustif de la population totale.

## III.3 Exemple type Bac : intervalle de confiance

**Exemple 3 : Point bac**

On interroge au hasard 300 clients ayant effectué des achats sur un site internet et s'étant fait livrer le produit à domicile. Le temps de livraison a été jugé raisonnable par 160 personnes interrogées. Donner un intervalle de confiance de la proportion  $p$  de clients satisfaits.

- **Analyse des données :**

- « Sur un échantillon de  $n = 300$  clients. Il est constaté que 160 sont satisfaits du temps de livraison. ». Donc la fréquence observée de clients satisfaits du temps de livraison est

$$f = 160 \div 300 \approx 0,533333333 \text{ soit } f \approx \underline{0,533}$$

- **Intervalle de confiance :**

On a pour le cas étudié,  $n = 300$ ,  $f \approx 0,533$ . Vérifions les conditions d'application du théorème :

$$\left\{ \begin{array}{l} \checkmark \quad n = 300 \geq 30 \\ \checkmark \quad nf = 300 \times \frac{160}{300} = 160 \geq 5 \\ \checkmark \quad n(1 - f) = 300 \times \frac{140}{300} = 140 \geq 5 \end{array} \right.$$

Un intervalle de confiance au seuil de confiance de 95% de la proportion  $p$  est alors :

$$I_n = \left[ f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right] = \left[ \frac{160}{300} - \frac{1}{\sqrt{300}} ; \frac{160}{300} + \frac{1}{\sqrt{300}} \right]$$

Soit puisque les borne sont :

$$\left| \begin{array}{l} \blacksquare f - \frac{1}{\sqrt{n}} = \frac{160}{300} - \frac{1}{\sqrt{300}} \approx 0,4756. \text{ On arrondit la borne inférieure par défaut } 10^{-3} \text{ près soit } \underline{0,475}. \\ \blacksquare f + \frac{1}{\sqrt{n}} = \frac{160}{300} + \frac{1}{\sqrt{300}} \approx 0,59107. \text{ On arrondit la borne supérieure par excès } 10^{-3} \text{ près soit } \underline{0,592}. \end{array} \right.$$

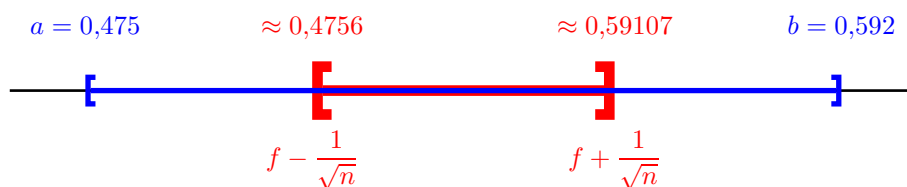
$$I_{300} \approx [0,475 ; 0,592]$$

- **Conclusion**

Cet intervalle contient la proportion  $p$  au niveau de confiance 95% (ou au risque d'erreur de 5%). La proportion de clients satisfaits se situe donc, au niveau de confiance 95%, entre 47,5% et 59,2%.

- **Remarque**

Il faut toujours proposer un intervalle en valeurs approchées qui contient l'intervalle en valeurs exactes.



↔ **Fin du cours** ↔