

- EXERCICE 1 -

On considère un n -échantillon $(X_1, \dots, X_n)$ i.i.d (indépendants et de même loi) d'une loi de Bernoulli de paramètre p .

On note $\overline{X}_n$ la moyenne empirique associée à l'échantillon.

1. **Rappeler sans démonstration les valeurs de $E(\overline{X}_n)$ et de $V(\overline{X}_n)$.**

On a (calcul classique) : $E(\overline{X}_n) = p$ et $V(\overline{X}_n) = \frac{p(1-p)}{n}$.

2. **Montrer à l'aide de l'inégalité de BT que :** $\forall \epsilon > 0, \quad P\left(\left|\overline{X}_n - p\right| \leq \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$.

Comme $\overline{X}_n$ admet un moment d'ordre 2, on a d'après l'inégalité de BT : $\forall \epsilon > 0$,

$$\begin{aligned} P\left(\left|\overline{X}_n - E(\overline{X}_n)\right| \geq \epsilon\right) &\leq \frac{V(\overline{X}_n)}{\epsilon^2} \\ \Leftrightarrow P\left(\left|\overline{X}_n - p\right| \geq \epsilon\right) &\leq \frac{p(1-p)}{n\epsilon^2} \\ \Leftrightarrow 1 - P\left(\left|\overline{X}_n - p\right| < \epsilon\right) &\leq \frac{p(1-p)}{n\epsilon^2} \quad \text{en considérant l'évènement contraire} \\ \Leftrightarrow P\left(\left|\overline{X}_n - p\right| < \epsilon\right) &\geq 1 - \frac{p(1-p)}{n\epsilon^2} \end{aligned}$$

Enfin, comme l'évènement $\left[\left|\overline{X}_n - p\right| \leq \epsilon\right]$ contient l'évènement $\left[\left|\overline{X}_n - p\right| < \epsilon\right]$, (i.e. $\left[\left|\overline{X}_n - p\right| < \epsilon\right] \subset \left[\left|\overline{X}_n - p\right| \leq \epsilon\right]$),

on a : $P\left(\left|\overline{X}_n - p\right| \leq \epsilon\right) \geq P\left(\left|\overline{X}_n - p\right| < \epsilon\right)$ et donc :

$$\forall \epsilon > 0, \quad P\left(\left|\overline{X}_n - p\right| \leq \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

3. **En déduire que pour tout $\alpha \in]0; 1[$:** $P\left(p \in \left[\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}}, \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right]\right) \geq 1 - \alpha$.

On commence par enlever les valeurs absolues dans le résultat de la question précédente qui donne alors :

$$\forall \epsilon > 0, \quad P\left(-\epsilon \leq \overline{X}_n - p \leq \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

puis on isole p :

$$\forall \epsilon > 0, \quad P\left(-\overline{X}_n - \epsilon \leq -p \leq -\overline{X}_n + \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

soit

$$\forall \epsilon > 0, \quad P\left(\overline{X}_n - \epsilon \leq p \leq \overline{X}_n + \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

Comme cette relation est valable pour tout $\epsilon > 0$, on va choisir une valeur de ϵ qui permet de répondre à la question, donc telle que :

$$1 - \frac{p(1-p)}{n\epsilon^2} = 1 - \alpha \Leftrightarrow \frac{p(1-p)}{n\epsilon^2} = \alpha \Leftrightarrow n\epsilon^2 = \frac{p(1-p)}{\alpha} \Leftrightarrow \epsilon = \sqrt{\frac{p(1-p)}{\alpha n}}$$

Pour cette valeur de ϵ , on a bien en remplaçant dans la relation précédente :

$$P\left(\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}} \leq p \leq \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right) \geq 1 - \alpha$$

soit

$$P\left(p \in \left[\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}}, \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right]\right) \geq 1 - \alpha$$

4. **En majorant la fonction $p \mapsto p(1-p)$ sur $[0, 1]$, montrer que l'intervalle $\left[\overline{X}_n - \frac{1}{\sqrt{4\alpha n}}, \overline{X}_n + \frac{1}{\sqrt{4\alpha n}}\right]$ est un intervalle de confiance de p au niveau de confiance $1 - \alpha$.**

L'intervalle précédent n'est pas satisfaisant car il dépend de p .

Une étude rapide montre que $\forall p \in [0, 1], \quad p(1-p) \leq 1/4$.

On a alors :

$$p(1-p) \leq \frac{1}{4} \Rightarrow \frac{p(1-p)}{\alpha n} \leq \frac{1}{4\alpha n} \Rightarrow \sqrt{\frac{p(1-p)}{\alpha n}} \leq \frac{1}{\sqrt{4\alpha n}} \Rightarrow \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}} \leq \overline{X}_n + \frac{1}{\sqrt{4\alpha n}}$$

et de même

$$p(1-p) \leq \frac{1}{4} \Rightarrow \overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}} \geq \overline{X}_n - \frac{1}{\sqrt{4\alpha n}}$$

On en déduit alors que l'intervalle $\left[\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}}, \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right]$ est inclus dans l'intervalle

$\left[\overline{X}_n - \frac{1}{\sqrt{4\alpha n}}, \overline{X}_n + \frac{1}{\sqrt{4\alpha n}}\right]$, ce qui donne :

$$P\left(p \in \left[\overline{X}_n - \frac{1}{\sqrt{4\alpha n}}, \overline{X}_n + \frac{1}{\sqrt{4\alpha n}}\right]\right) \geq P\left(p \in \left[\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}}, \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right]\right) \geq 1 - \alpha$$

d'où le résultat.

5. **En déduire un intervalle de confiance de p au niveau de confiance 0,95.**

Pour obtenir un intervalle de confiance de p au niveau de confiance 0,95, c'est-à-dire $1 - \alpha = 0,95$, il faut

choisir $\alpha = 0,05$, ce qui donne avec la question précédente : $IC_{0,05} = \left[\overline{X}_n - \frac{1}{\sqrt{0,2n}}, \overline{X}_n + \frac{1}{\sqrt{0,2n}}\right]$

6. **On souhaite réaliser un classement des plus grandes légendes du football de tous les temps. Deux génies du ballon rond sont pressentis pour la première place : l'immense Z.Z et le modeste mais non moins grand S.A.**

On réalise un sondage auprès de 500 membres de la communauté sportive mondiale. 280 affirment vouloir placer S.A en première position, soit 56% des sondés.

Peut-on affirmer avec un risque d'erreur de 5% que S.A sera bien sur la première place du podium ?

Ici, $n = 500$ et $\overline{X}_n = 0,56$, ce qui donne : $IC_{0,05} = \left[0,56 - \frac{1}{\sqrt{100}}, 0,56 + \frac{1}{\sqrt{100}}\right] = [0,46, 0,66]$.

On ne peut donc pas affirmer avec un risque d'erreur de 5% que Mr S.A sera bien classé premier, la valeur de p pouvant être inférieure à 0,5.

- EXERCICE 2 -

On reprend les notations de l'exercice 1.

1. **Montrer que** $\sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p)$ **converge en loi vers une loi normale centrée réduite puis en déduire**

$$\text{que : } \lim_{n \rightarrow +\infty} P\left(-t \leq \sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p) \leq t\right) = 2\Phi(t) - 1.$$

On rappelle que : $E(\bar{X}_n) = p$ et $V(\bar{X}_n) = \frac{p(1-p)}{n}$.

D'après le théorème limite central, $\bar{X}_n^*$ converge en loi vers la loi normale centrée réduite, soit :

$$\bar{X}_n^* = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p) \xrightarrow{\text{loi}} N \quad \text{avec } N \hookrightarrow \mathcal{N}(0, 1)$$

et en particulier pour tout $t > 0$:

$$\lim_{n \rightarrow +\infty} P\left(-t \leq \sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p) \leq t\right) = P(-t \leq N \leq t) = 2\Phi(t) - 1$$

2. Soit $\alpha \in]0, 1/2[$. Justifier qu'il existe un réel t_α tel que : $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$. puis déduire de la question précédente

$$\text{dente que : } \lim_{n \rightarrow +\infty} P\left(\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha.$$

La fonction Φ est continue sur $\mathbb{R}$ (fonction de répartition d'une variable à densité) et strictement croissante sur $\mathbb{R}$ (car $\phi(x) > 0$ sur $\mathbb{R}$) donc réalise une bijection de $\mathbb{R}$ sur $]0, 1[$ (limites d'une fonction de répartition).

Comme $\alpha \in]0, 1/2[$, on a $1 - \frac{\alpha}{2} > 0$ donc il existe un réel t_α tel que : $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$.

On a alors : $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$ donc $2\Phi(t_\alpha) - 1 = 1 - \alpha$.

On peut donc écrire pour cette valeur de t_α :

$$\lim_{n \rightarrow +\infty} P\left(-t_\alpha \leq \sqrt{\frac{n}{p(1-p)}}(\bar{X}_n - p) \leq t_\alpha\right) = 2\Phi(t_\alpha) - 1 = 1 - \alpha$$

soit après quelques étapes de calcul :

$$\lim_{n \rightarrow +\infty} P\left(\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

3. En déduire que l'intervalle $\left[\bar{X}_n - \frac{t_\alpha}{\sqrt{4n}}, \bar{X}_n + \frac{t_\alpha}{\sqrt{4n}}\right]$ est un intervalle de confiance asymptotique de p au niveau de confiance $1 - \alpha$.

Comme $\forall p \in [0, 1]$, $p(1-p) \leq 1/4$, le même raisonnement que dans l'exercice précédent (inclusion des intervalles) montre que :

$$\lim_{n \rightarrow +\infty} P\left(\bar{X}_n - t_\alpha \frac{1}{\sqrt{4n}} \leq p \leq \bar{X}_n + t_\alpha \frac{1}{\sqrt{4n}}\right) \geq \lim_{n \rightarrow +\infty} P\left(\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

Ainsi, l'intervalle $\left[\bar{X}_n - \frac{t_\alpha}{\sqrt{4n}}, \bar{X}_n + \frac{t_\alpha}{\sqrt{4n}}\right]$ est un intervalle de confiance asymptotique de p au niveau de confiance $1 - \alpha$.

4. Pour $\alpha = 0,05$, on donne $t_\alpha = t_{0,05} = 1,96$. En déduire un intervalle de confiance asymptotique de p au niveau de confiance 0,95.

Pour obtenir un intervalle de confiance de p au niveau de confiance 0,95, il faut choisir $\alpha = 0,05$, et $t_\alpha = t_{0,05} = 1,96$ (d'après l'énoncé).

On obtient alors avec la question précédente : $IC_{0,05} = \left[\bar{X}_n - \frac{1,96}{2\sqrt{n}}, \bar{X}_n + \frac{1,96}{2\sqrt{n}}\right]$

5. On reprend notre sondage de l'exercice précédent.

Peut-on cette fois affirmer avec un risque d'erreur de 5% que S.A sera bien sur la première place du podium ?

Ici, $n = 500$ et $\bar{X}_n = 0,56$, ce qui donne : $IC_{0,05} = \left[0,56 - \frac{1,96}{20\sqrt{5}}, 0,56 + \frac{1,96}{20\sqrt{5}}\right] \approx [0,516, 0,603]$.

On peut donc affirmer avec un risque d'erreur de 5% que Mr S.A sera bien classé premier.

Le contraire eut été étonnant!

Ce résultat n'est pas contradictoire avec celui de l'exercice précédent. Le théorème central limite fournit simplement un intervalle de confiance plus précis que l'inégalité de BT.