

Estimation par intervalle de confiance

I - Intervalle de confiance

I.1 - Définition

Définition 1

Soit X une variable aléatoire suivant une loi de probabilité de paramètre inconnu θ et soit $(X_1, \dots, X_n)$ un n -échantillon de X (indépendantes et de même loi que X).

On considère U_n et V_n deux estimateurs de θ .

Pour tout $\alpha \in]0; 1[$, on dit que $[U_n, V_n]$ est un **intervalle de confiance de θ au niveau de confiance $1 - \alpha$** (ou au risque α) ssi :

$$P([U_n \leq \theta \leq V_n]) \geq 1 - \alpha$$

I.2 - Utilisation de l'inégalité de Bienaymé-Tchebychev

Méthode 1

- On suppose que T_n est un estimateur sans biais de θ . Ainsi, on a : $E(T_n) = \theta$.

L'inégalité de Bienaymé-Tchebychev donne :

$$\begin{aligned} \forall \epsilon > 0, \quad P(|T_n - \theta| \geq \epsilon) &\leq \frac{V(T_n)}{\epsilon^2} \iff 1 - P(|T_n - \theta| < \epsilon) \leq \frac{V(T_n)}{\epsilon^2} \\ &\iff P(|T_n - \theta| < \epsilon) \geq 1 - \frac{V(T_n)}{\epsilon^2} \\ &\iff P(-\epsilon < T_n - \theta < \epsilon) \geq 1 - \frac{V(T_n)}{\epsilon^2} \\ &\iff P(T_n - \epsilon < \theta < T_n + \epsilon) \geq 1 - \frac{V(T_n)}{\epsilon^2} \\ &\iff P(T_n - \epsilon \leq \theta \leq T_n + \epsilon) \geq 1 - \frac{V(T_n)}{\epsilon^2} \\ &\text{car } (T_n - \epsilon < \theta < T_n + \epsilon) \subset (T_n - \epsilon \leq \theta \leq T_n + \epsilon) \end{aligned}$$

- On cherche ensuite une valeur $\tilde{\epsilon}$ de ϵ telle que $\frac{V(T_n)}{\tilde{\epsilon}^2} = \alpha$ de telle sorte que :

$$P(T_n - \tilde{\epsilon} \leq \theta \leq T_n + \tilde{\epsilon}) \geq 1 - \alpha \quad \text{soit} \quad P(\theta \in [T_n - \tilde{\epsilon}; T_n + \tilde{\epsilon}]) \geq 1 - \alpha$$

- Si $\tilde{\epsilon}$ ne dépend pas de θ** , l'intervalle $[T_n - \tilde{\epsilon}; T_n + \tilde{\epsilon}]$ est donc un intervalle de confiance de θ au niveau de confiance $1 - \alpha$.

En effet, les variables $U_n = T_n - \tilde{\epsilon}$ et $V_n = T_n + \tilde{\epsilon}$ sont deux estimateurs de θ qui vérifient

$$P(U_n \leq \theta \leq V_n) \geq 1 - \alpha$$

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.

- Rappeler sans démonstration les valeurs de $E(\overline{X}_n)$ et de $V(\overline{X}_n)$.
- Montrer à l'aide de l'inégalité de BT que : $\forall \epsilon > 0, \quad P\left(\left|\overline{X}_n - p\right| \leq \epsilon\right) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$.
- En déduire que pour tout $\alpha \in]0; 1[$: $P\left(p \in \left[\overline{X}_n - \sqrt{\frac{p(1-p)}{\alpha n}}, \overline{X}_n + \sqrt{\frac{p(1-p)}{\alpha n}}\right]\right) \geq 1 - \alpha$.
- 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 $1 - \alpha$.
- En déduire un intervalle de confiance de p au niveau de confiance 0,95.
- 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?

II - Intervalle de confiance asymptotique

II.1 - Définition

Définition 2

Soit X une variable aléatoire suivant une loi de probabilité de paramètre inconnu θ et soit $(X_1, \dots, X_n)$ un n -échantillon de X (indépendantes et de même loi que X).

On dit que $[U_n, V_n]$ est un **intervalle de confiance asymptotique de θ au niveau de confiance $1 - \alpha$** (ou au risque α) ssi :

$$\lim_{n \rightarrow +\infty} P([U_n \leq \theta \leq V_n]) \geq 1 - \alpha$$

II.2 - Utilisation du théorème limite central : étude d'un exemple

Méthode 2

- On considère un n -échantillon i.i.d $(X_1, \dots, X_n)$ d'une loi d'**espérance m inconnue de variance σ^2 connue**. On cherche un intervalle de confiance de m au niveau de confiance $1 - \alpha$.

On considère l'estimateur $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ et on rappelle que : $E(\bar{X}_n) = m$ et $V(\bar{X}_n) = \frac{\sigma^2}{n}$.

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

$$\bar{X}_n^* = \sqrt{\frac{n}{\sigma^2}} (\bar{X}_n - m) \xrightarrow{\text{loi}} N \quad \text{avec } N \hookrightarrow \mathcal{N}(0, 1)$$

et en particulier pour tout $t > 0$:

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

- On cherche alors une valeur de t , notée t_α telle que :

$$2\Phi(t_\alpha) - 1 = 1 - \alpha \quad \text{soit} \quad \Phi(t_\alpha) = 1 - \frac{\alpha}{2}$$

Par exemple, pour $\alpha = 0,05$, on a : $1 - \frac{\alpha}{2} = 0,975$, et la table de la loi normale donne $t_\alpha \approx 1,96$.

- On peut donc écrire pour cette valeur de t_α :

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

soit

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

Ainsi, l'intervalle $\left[\bar{X}_n - t_\alpha \sqrt{\frac{\sigma^2}{n}}; \bar{X}_n + t_\alpha \sqrt{\frac{\sigma^2}{n}}\right]$ est un intervalle de confiance asymptotique de m au niveau de confiance $1 - \alpha$.

Exercice 2

On reprend les notations de l'exercice 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 que : $\lim_{n \rightarrow +\infty} P\left(-t \leq \sqrt{\frac{n}{p(1-p)}} (\bar{X}_n - p) \leq t\right) = 2\Phi(t) - 1$.
- 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 que : $\lim_{n \rightarrow +\infty} P\left(\bar{X}_n - t_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + t_\alpha \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$.
- 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 $1 - \alpha$.
- 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.
- 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?