Tuto intervalles de confiance

Colin 'Alaska' Coërchon

2024

Table des matières

1	Intervalles de confiance	2
1.1	Intervalle de confiance de loi normales $\mathcal{N}(\mu, \sigma^2)$	2
1.1.1	Estimation de l'espérance μ lorsque la variance σ^2 est connue	2
1.1.2	Estimation de l'espérance μ lorsque la variance σ^2 est inconnue	2
1.1.3	Estimation de la variance σ^2 lorsque l'espérance μ est inconnue	3
1.1.4	Estimation de la variance σ^2 lorsque l'espérance μ est connue	3
1.2	Intervalle de confiance pour une proportion ($\mathcal{B}(\theta)$)	4
1.3	Intervalle de confiance asymptotique pour n'importe quelle loi	5
1.3.1	Exemples	5
1.4	Intervalle de confiance avec information de Fisher	5
1.5	Quelques valeurs des quantiles	6
1.5.1	Quantiles de $\mathcal{N}(0, 1)$	6
1.5.2	Quantiles de $\mathcal{T}_{n-1}$	6
1.5.3	Quantiles de χ_{n-1}^2 (ou χ_n^2)	6

1 Intervalles de confiance

On note α la précision de notre intervalle de confiance. On cherche un intervalle tel que la proba que θ soit à l'intérieur soit égale à $1 - \alpha$. C'est-à-dire que :

$$\mathbb{P}(\theta \in \text{IC}) = 1 - \alpha$$

1.1 Intervalle de confiance de loi normales $\mathcal{N}(\mu, \sigma^2)$

1.1.1 Estimation de l'espérance μ lorsque la variance σ^2 est connue

On souhaite approximer μ . Donc notre estimateur ici est : $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$.

On sait que : $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ **SANS TCL**

Donc :

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

On en déduit par définition que :

$$\begin{aligned} \mathbb{P}\left(-u_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq u_{1-\alpha/2}\right) &= 1 - \alpha \\ \implies \mathbb{P}\left(\bar{X}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

Et donc, voici notre intervalle de confiance :

$$\boxed{\text{IC}_{1-\alpha}(\mu) = \left[\bar{x}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]}$$

Avec $u_{1-\alpha/2}$ le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$.

1.1.2 Estimation de l'espérance μ lorsque la variance σ^2 est inconnue

Ici, σ^2 est inconnue, donc on ne peut plus considérer $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$.

On doit alors utiliser un estimateur de la variance !

On pose alors :

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

On considère alors :

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$$

Et, il se trouve que cette variable aléatoire suit une loi de student à $n - 1$ degrés de liberté, que l'on note $\mathcal{T}_{n-1}$. On en déduit alors :

$$\begin{aligned} \mathbb{P}\left(-t_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq t_{1-\alpha/2}\right) &= 1 - \alpha \\ \implies \mathbb{P}\left(\bar{X}_n - t_{1-\alpha/2} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{1-\alpha/2} \frac{S_n}{\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

Et donc, voici notre intervalle de confiance :

$$\text{IC}_{1-\alpha}(\mu) = \left[\overline{x}_n - t_{1-\alpha/2} \frac{s_n}{\sqrt{n}}, \overline{x}_n + t_{1-\alpha/2} \frac{s_n}{\sqrt{n}} \right]$$

Avec $t_{1-\alpha/2}$ le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{T}_{n-1}$.

1.1.3 Estimation de la variance σ^2 lorsque l'espérance μ est inconnue

On estime la variance σ^2 supposée inconnue par la variance empirique :

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2$$

Et on sait que $S_n^2 \sim \frac{\sigma^2}{n-1} \chi_{(n-1)}^2$, et que S_n^2 est un estimateur sans biais de σ^2 .

On peut alors écrire par définition :

$$\begin{aligned} \mathbb{P} \left(v_{\alpha/2} \leq \frac{n-1}{\sigma^2} S_n^2 \leq v_{1-\alpha/2} \right) &= 1 - \alpha \\ \implies \mathbb{P} \left(\frac{(n-1)S_n^2}{v_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{v_{\alpha/2}} \right) &= 1 - \alpha \end{aligned}$$

Et donc, voici notre intervalle de confiance :

$$\text{IC}_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s_n^2}{v_{1-\alpha/2}}, \frac{(n-1)s_n^2}{v_{\alpha/2}} \right]$$

Avec $v_{1-\alpha/2}$ et $v_{\alpha/2}$ les quantiles d'ordre $1 - \alpha/2$ et $\alpha/2$ de la loi χ_{n-1}^2 .

1.1.4 Estimation de la variance σ^2 lorsque l'espérance μ est connue

Ce cas est très similaire au précédent. On estime la variance σ^2 supposée inconnue par une variance empirique plus simple :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

Et on sait que $S_n^2 \sim \frac{\sigma^2}{n} \chi_n^2$, et que S_n^2 est un estimateur sans biais de σ^2 .

On peut alors écrire par définition :

$$\begin{aligned} \mathbb{P} \left(v_{\alpha/2} \leq \frac{n}{\sigma^2} S_n^2 \leq v_{1-\alpha/2} \right) &= 1 - \alpha \\ \implies \mathbb{P} \left(\frac{n S_n^2}{v_{1-\alpha/2}} \leq \sigma^2 \leq \frac{n S_n^2}{v_{\alpha/2}} \right) &= 1 - \alpha \end{aligned}$$

Et donc, voici notre intervalle de confiance :

$$\text{IC}_{1-\alpha}(\sigma^2) = \left[\frac{n s_n^2}{v_{1-\alpha/2}}, \frac{n s_n^2}{v_{\alpha/2}} \right]$$

Avec $v_{1-\alpha/2}$ et $v_{\alpha/2}$ les quantiles d'ordre $1 - \alpha/2$ et $\alpha/2$ de la loi χ_n^2 .

1.2 Intervalle de confiance pour une proportion ($\mathcal{B}(\theta)$)

Prenons un exemple. On cherche à estimer la proportion θ de droitiers dans la population. On prends n personnes aléatoirement et on note X_i qui vaut 1 si la personne est droitère, 0 sinon.

On estime alors θ par la moyenne empirique

$$T_n = \overline{X}_n = \sum_{k=1}^n X_i$$

Et donc, comme $X_i \sim \mathcal{B}(1, \theta)$, on en déduit que :

$$\forall i \in \llbracket 1, n \rrbracket, \quad X_i \sim \mathcal{B}(1, \theta) \implies \begin{cases} \mathbb{E}[X_i] = \theta < \infty \\ \text{Var}(X_i) = \theta(1 - \theta) < \infty \end{cases}$$

Finalement, $\overline{X}_n$ est la moyenne empirique de n variables i.i.d de Bernoulli (les X_i). Donc, d'après le **Théorème Central Limite**, on a :

$$\boxed{\sqrt{n} \frac{\overline{X}_n - \theta}{\sqrt{\theta(1 - \theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)}$$

Seulement, nous voulons déterminer un intervalle de confiance à $\alpha\%$ de notre estimation, pour ensuite vérifier si θ se trouve dans cet intervalle.

Si on ne change rien, notre intervalle de confiance va dépendre de θ , ce qui n'est pas du tout souhaitable. On va donc utiliser le **théorème de Slutsky** pour se ramener à un intervalle de confiance indépendant de θ .

Premièrement, comme $X_1, X_2, \dots, X_M$ sont des variables aléatoires i.i.d, le **la loi forte des grands nombres** nous assure que :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^M X_i \xrightarrow{p.s.} \mathbb{E}[X_1] = \theta$$

Et donc, on a nécessairement le résultat plus faible suivant :

$$\overline{X}_n \xrightarrow{\mathbb{P}} \theta$$

Ainsi, en appliquant le théorème de continuité avec la fonction $x \mapsto \frac{1}{\sqrt{x(1-x)}}$ sur $]0, 1[$, on a :

$$\boxed{\frac{1}{\sqrt{\overline{X}_n(1 - \overline{X}_n)}} \xrightarrow{\mathbb{P}} \frac{1}{\sqrt{\theta(1 - \theta)}}$$

Par conséquent, d'après le **théorème de Slutsky**, on en déduit que :

$$\boxed{\sqrt{n} \frac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1 - \overline{X}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)}$$

On en déduit donc, en notant u_α les quantiles d'ordre α pour la loi $\mathcal{N}(0, 1)$, que :

$$\mathbb{P} \left(u_{\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1 - \overline{X}_n)}} \leq u_{1-\frac{\alpha}{2}} \right) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$$

$$\Rightarrow \mathbb{P} \left(\sqrt{\frac{\overline{X}_n (1 - \overline{X}_n)}{n}} u_{\frac{\alpha}{2}} \leq \overline{X}_n - \theta \leq \sqrt{\frac{\overline{X}_n (1 - \overline{X}_n)}{n}} u_{1-\frac{\alpha}{2}} \right) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$$

Et donc, voici notre intervalle de confiance **asymptotique** :

$$\boxed{\text{IC}_{1-\alpha}(\theta) = \left[\overline{x}_n - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}_n (1 - \overline{x}_n)}{n}}; \overline{x}_n + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}_n (1 - \overline{x}_n)}{n}} \right]}$$

1.3 Intervalle de confiance asymptotique pour n'importe quelle loi

Dans le cas où on ne traite pas des estimations de loi normales $\mathcal{N}(\mu, \sigma^2)$, il faut impérativement s'y ramener quand même par le **TCL**!!!!

C'est notamment le cas où on nous parle d'intervalle de confiance **asymptotique**.

1.3.1 Exemples

Soit $X_1, \dots, X_n$ des VAR qui suivent une certaine loi. On cherche à estimer le paramètre θ . Deux cas s'offrent alors à nous :

- **1er cas (majorité des cas)** : On a $\boxed{\mathbb{E}[X] = \theta}$.

On peut donc prendre par la méthode des moments l'estimateur $\overline{X}_n$, et alors :

- ▷ Si on peut calculer la variance σ^2 , on se ramène au paragraphe "Estimation de l'espérance μ lorsque la variance σ^2 est connue".
- ▷ Si on ne peut PAS calculer la variance, on se ramène au cas "Estimation de l'espérance μ lorsque la variance σ^2 est inconnue".

- **2ème cas (pas de chance)** : On a $\boxed{\mathbb{E}[X] \neq \theta}$.

On doit quand même se ramener au **TCL**. Il faut souvent utiliser la **méthode Delta**, se débrouiller pour ce ramener au **1er cas**.

1.4 Intervalle de confiance avec information de Fisher

Pour lire cette partie, il y a de très très grandes chances qu'il soit apparu dans l'exercice :

1. L'estimateur du maximum de vraisemblance $\hat{\theta}^{\text{MV}}$.
2. L'information de Fisher $I_1(\theta)$.
3. Et enfin, une question sur un intervalle de confiance de θ .

Si cela réunit tous ces cas, il faut très certainement utiliser un intervalle de confiance asymptotique avec l'information de Fisher.

Pour cela, il y a un gros théorème dans le cours (se servant du **TCL**) qui nous donne :

$$\sqrt{n} \left(\hat{\theta}^{\text{MV}} - \theta \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, I_1^{-1}(\theta) \right)$$

Cependant, $I_1^{-1/2}(\theta)$ dépend du paramètre θ , on doit donc l'estimer lui aussi. Si $\theta \mapsto I_1(\theta)$ est continue (*c'est vrai dans la plupart des cas, il faut l'admettre au pire*), d'après le **théorème de Slutsky**, nous en déduisons que :

$$\mathbb{P} \left(u_{\alpha/2} \leq \sqrt{n} I_1^{1/2}(\hat{\theta}^{\text{MV}}) \left(\hat{\theta}^{\text{MV}} - \theta \right) \leq u_{1-\alpha/2} \right) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$$

Et donc, voici notre intervalle de confiance **asymptotique** :

$$\text{IC}_{1-\alpha}(\theta) = \left[\hat{\theta}^{\text{MV}} - u_{1-\frac{\alpha}{2}} \frac{I_1^{-1/2}(\hat{\theta}^{\text{MV}})}{\sqrt{n}}, \hat{\theta}^{\text{MV}} + u_{1-\frac{\alpha}{2}} \frac{I_1^{-1/2}(\hat{\theta}^{\text{MV}})}{\sqrt{n}} \right]$$

NOTE : $I_1^{-1/2}(\hat{\theta}^{\text{MV}})$ signifie la racine carrée de $I_1^{-1}(\hat{\theta}^{\text{MV}})$.

Donc :

- S'il n'y a qu'un paramètre $\theta \in \mathbb{R}$ à estimer, on fait juste la racine carrée de la valeur qu'on a obtenue pour I_1^{-1} . i.e :

$$I_1^{-1/2} = \frac{1}{\sqrt{I_1}}$$

- S'il y a plus de 2 paramètres à estimer, exemple $\theta = (\alpha, \beta)$, c'est la "racine carrée" de la **matrice** I_1 . C'est plus complexe, donc IMPOSSIBLE QUE ÇA TOMBE.

1.5 Quelques valeurs des quantiles

1.5.1 Quantiles de $\mathcal{N}(0, 1)$

- $\alpha = 10\%$: $u_{1-\frac{\alpha}{2}} \approx 1.65$
- $\alpha = 5\%$: $u_{1-\frac{\alpha}{2}} \approx 1.96$
- $\alpha = 1\%$: $u_{1-\frac{\alpha}{2}} \approx 2.58$

1.5.2 Quantiles de $\mathcal{T}_{n-1}$

Ça dépend du degré de liberté $n - 1$, donc ça sera donné s'il y a une question sur ça !

1.5.3 Quantiles de χ_{n-1}^2 (ou χ_n^2)

Ça dépend du degré de liberté $n - 1$ (ou n), donc ça sera donné s'il y a une question sur ça !