Statistique inférentielle

Intervalles de confiance

A. Godichon-Baggioni

I. Intervalles de confiance

Intervalles de confiance

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées.

Soit $\alpha \in (0,1)$, un intervalle de confiance pour le paramètre θ au niveau de confiance $1-\alpha$ est un intervalle de la forme

$$IC_{1-\alpha}(\theta) = \left[a\left(X_1, \dots, X_n\right); b\left(X_1, \dots, X_n\right)\right]$$

avec

$$\mathbb{P}\left[\theta \in \left[a\left(X_1, \dots, X_n\right); b\left(X_1, \dots, X_n\right)\right]\right] = 1-\alpha.$$

Attention ! Cela ne signifie pas que $\theta \in IC_{1-\alpha}(\theta)$.
Attention ! On ne peut pas dire que la probabilité que θ appartienne à la réalisation de $IC_{1-\alpha}(\theta)$ est de $1-\alpha$.

EXEMPLES

Exemple 1 : On considère des variables aléatoires i.i.d $X_1, \dots, X_n$ de densité

$$f_\theta(x) = \frac{\theta}{x^2} \mathbf{1}_{x^2 \geq \theta}$$

avec $\theta > 0$. Soit $\alpha \in (0,1)$, un intervalle de confiance de niveau $1-\alpha$ pour θ est donné par

$$IC_{1-\alpha}(\theta) = \left[X_{(1)} \alpha^{1/n} ; X_{(1)} \right].$$

Exemple 2 : loi uniforme. On considère des variables aléatoires i.i.d $X_1, \dots, X_n$ avec $X_1 \sim \mathcal{U}([0,\theta])$ et $\theta > 0$. Soit $\alpha \in (0,1)$, un intervalle de confiance de niveau $1-\alpha$ pour θ est donné par

$$IC_{1-\alpha}(\theta) = \left[X_{(n)} ; X_{(n)} \alpha^{-1/n} \right].$$

Remarque

Souvent, on cherche des intervalles tels que

$$\mathbb{P}\left[\theta \leq a\left(X_1, \ldots, X_n\right)\right] = \mathbb{P}\left[\theta \geq b\left(X_1, \ldots, X_n\right)\right] = \alpha/2.$$

Exemple 1 : On obtient un intervalle de la forme (si $\alpha < 1/2$)

$$IC_{1-\alpha}(\theta) = \left[X_{(1)}\left(\frac{\alpha}{2}\right)^{1/n} ; X_{(1)}\left(1 - \frac{\alpha}{2}\right)^{1/n}\right].$$

Exemple 2 : On obtient un intervalle de la forme (si $\alpha < 1/2$)

$$IC_{1-\alpha}(\theta) = \left[X_{(n)}\left(1 - \frac{\alpha}{2}\right)^{-1/n} ; X_{(n)}\left(\frac{\alpha}{2}\right)^{-1/n}\right]$$

Remarque

Un intervalle de confiance pour le paramètre θ au niveau de confiance au moins $1-\alpha$ est un intervalle de la forme

$$IC_{1-\alpha}(\theta) = [a(X_1,\ldots,X_n)\,;b(X_1,\ldots,X_n)]$$

avec

$$\mathbb{P}\left[\theta \in [a(X_1,\ldots,X_n)\,;b(X_1,\ldots,X_n)]\right] \geq 1-\alpha.$$

Exemple : loi de Bernoulli. Soit $X_1,\ldots,X_n$ des variables aléatoires i.i.d avec $X_1 \sim \mathcal{B}(\theta)$ et $\theta \in (0,1)$. Un intervalle de confiance de niveau au moins $1-\alpha$ est donné par

$$IC_{1-\alpha}(\theta) = \left[\overline{X}_n - \frac{1}{2\sqrt{\alpha n}}\,;\overline{X}_n + \frac{1}{2\sqrt{\alpha n}}\right].$$

BILATÈRE VS UNILATÈRE

Remarque : Pour les intervalles précédents, on parle d'intervalles de confiances bilatères.

Remarque : On peut également construire des intervalles de confiances de la forme

$$\left]-\infty, b\left(X_1, \ldots, X_n\right)\right] \qquad \text{et} \qquad \left[a\left(X_1, \ldots, X_n\right), +\infty\right[.$$

On parle alors d'intervalles de confiance unilatères.

Quantiles

On considère une variable aléatoire X et on note F sa fonction de répartition.

Définition

Pour tout $\alpha \in (0,1)$, on appelle quantile d'ordre α le réel q_α tel que

$$q_\alpha = \inf\left\{x \in \mathbb{R}, \quad F(x) \geq \alpha\right\}.$$

Si la fonction de répartition F est strictement croissante, elle est inversible et on a alors

$$F\left(q_\alpha\right) = \alpha \Leftrightarrow q_\alpha = F^{-1}(\alpha).$$

Exemples

Exemple 1 : la loi uniforme. Soit $X \sim \mathcal{U}([a,b])$. Soit $\alpha \in (0,1)$, le quantile q_α d'ordre α de X est donné par

$$q_\alpha = a + \alpha(b-a).$$

Exemple 2 : la loi exponentielle. Soit $X \sim \mathcal{E}(1)$. Soit $\alpha \in (0,1)$, le quantile q_α d'ordre α de X est donné par

$$q_\alpha = -\ln(1-\alpha).$$

Exemple 3 : la loi de Bernoulli. Soit $X \sim \mathcal{B}(\theta)$. On a

$$q_\alpha = \begin{cases} 0 & \text{si } \alpha \in (0, 1-\theta] \\ 1 & \text{sinon} \end{cases} \tag{1}$$

II. Rappels sur la loi normale

Rappels sur la loi normale

Soient $X_1, \ldots, X_n$ des variables aléatoires suivant des lois normales de moyennes $\mu_1, \ldots, \mu_n$ et de variances $\sigma_1^2, \ldots, \sigma_n^2$. On rappelle que la fonction caractéristique de X_i est définie pour tout $t \in \mathbb{R}$ par

$$\Phi_{X_i}(t) = \exp\left(\mu_i i t - \frac{t^2 \sigma_i^2}{2}\right)$$

Rappels sur la loi normale

Proposition

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivant des lois normales de moyennes $\mu_1, \ldots, \mu_n$ et de variances $\sigma_1^2, \ldots, \sigma_n^2$. Alors toute combinaison linéaire des X_i suit une loi normale. Plus précisément, soient $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, alors

$$\sum_{i=1}^{n} \lambda_i X_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

avec

$$\mu = \sum_{i=1}^{n} \lambda_i \mu_i$$

$$\sigma^2 = \sum_{i=1}^{n} \lambda_i^2 \sigma_i^2$$

Loi du Chi-deux

Définition

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant une loi normale centrée réduite. Alors la variable aléatoire

$$Z_n = \sum_{i=1}^{n} X_i^2$$

suit une loi du Chi-deux à n degrés de liberté (χ_n^2).

Loi du Chi-deux

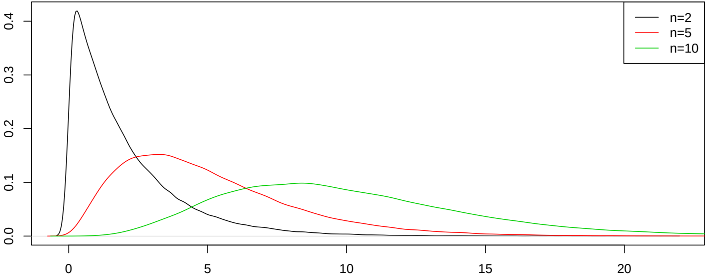

Figure – Densité d'une chi deux à $n = 2, 5, 10$ degrés de liberté

Loi de Student

Définition

Soient Z, U deux variables aléatoires indépendantes telles que $Z \sim \mathcal{N}(0,1)$ et $U \sim \chi_n^2$, alors

$$\frac{Z}{\sqrt{U/n}} \sim T_n$$

où T_n suit une loi de Student à n degrés de liberté.

Loi de Student

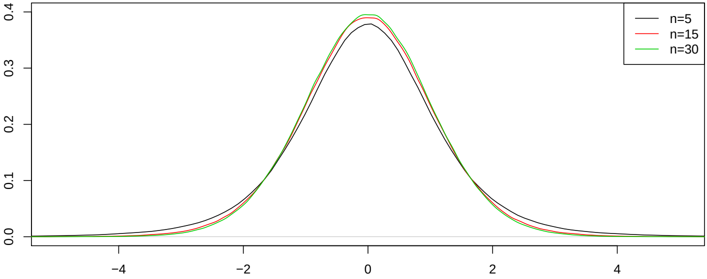

FIGURE – Densité d'une loi de Student à $n = 5, 15, 30$ degrés de liberté.

Loi de Student

Proposition

Soit T_n une variable aléatoire suivant une loi de Student à n degrés de liberté. Si $n \geq 2$, alors :

- *T_n admet un moment d'ordre 1*
- $\mathbb{E}[T_n] = 0$.
- *La loi de Student est symétrique en 0.*
- *On a la convergence en loi*

$$T_n \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

III. Cas Gaussien

Cas gaussien

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivant une loi normale d'espérance μ et de variance σ^2.

Proposition

On a

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Cas ou la variance est connue

Proposition

Pour tout $\alpha \in (0,1)$,

$$\mathbb{P}\left[\overline{X}_n - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha,$$

où $q_{1-\alpha/2}$ *est le quantile d'ordre* $1-\alpha/2$ *de la loi normale centrée réduite, i.e si* $Z \sim \mathcal{N}(0,1)$,

$$\mathbb{P}\left[Z \leq q_{1-\alpha/2}\right] = 1-\alpha/2.$$

On obtient donc l'intervalle de confiance de niveau $1-\alpha$

$$IC_{1-\alpha}(\mu) = \left[\overline{X}_n - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} ; \overline{X}_n + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Intervalles unilatères

On peut également obtenir les intervalles de confiances unilatères suivants :

$$IC_{1-\alpha}(\mu) = \left]-\infty; \overline{X}_n + q_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right]$$

$$IC_{1-\alpha}(\mu) = \left[\overline{X}_n - q_{1-\alpha}\frac{\sigma}{\sqrt{n}}; +\infty\right[$$

Cas où la variance est inconnue

Proposition

Soient $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ *et* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$*, alors*

1. $\frac{n-1}{\sigma^2}S_n^2 \sim \chi^2_{n-1}$.
2. S_n *et* $\overline{X}_n$ *sont indépendants.*

Corollaire

On a

$$\sqrt{n}\frac{\overline{X}_n - \mu}{S_n} \sim T_{n-1},$$

où T_{n-1} *suit une loi de Student à* $n-1$ *degrés de liberté.*

Cas où la variance est inconnue

Corollaire (Intervalles de confiance)

Soit $\alpha \in (0,1)$*, alors*

$$\mathbb{P}\left[\overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right] = 1-\alpha$$

où $t_{n-1,1-\alpha/2}$ *est le quantile d'ordre* $1-\alpha/2$ *de la loi de Student à* $n-1$ *degrés de liberté, i.e si* $T \sim T_{n-1}$*,*

$$\mathbb{P}\left[T \leq t_{n-1,1-\alpha/2}\right] = 1-\alpha/2.$$

On obtient donc l'intervalle de confiance au niveau $1-\alpha$

$$IC_{1-\alpha}(\mu) = \left[\overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} ; \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right]$$

Intervalles unilatères

On peut également obtenir les intervalles de confiances unilatères suivants :

$$IC_{1-\alpha}(\mu) = \left]-\infty; \overline{X}_n + t_{n-1,1-\alpha}\frac{S_n}{\sqrt{n}}\right]$$

$$IC_{1-\alpha}(\mu) = \left[\overline{X}_n - t_{n-1,1-\alpha}\frac{S_n}{\sqrt{n}}; +\infty\right[$$

Estimation de la variance

Proposition

Soit $\alpha \in (0,1)$, alors

$$\mathbb{P}\left[\frac{(n-1)S_n^2}{k_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{k_{\alpha/2}}\right] = 1-\alpha$$

où $k_{\alpha/2}$ et $k_{1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi du Chi-deux à $n-1$ degrés de liberté, i.e si $Z \sim \chi^2_{n-1}$,

$$\mathbb{P}\left[Z \leq k_{\alpha/2}\right] = \alpha/2 \qquad \mathbb{P}\left[Z \leq k_{1-\alpha/2}\right] = 1-\alpha/2.$$

On obtient donc l'intervalle de confiance au niveau $1-\alpha$

$$IC_{1-\alpha}\left(\sigma^2\right) = \left[\frac{(n-1)S_n^2}{k_{1-\alpha/2}}; \frac{(n-1)S_n^2}{k_{\alpha/2}}\right].$$

IV. Intervalles de confiance asymptotiques

Intervalles de confiance asymptotiques

On s'intéresse à l'estimation d'une caractéristique ou d'un paramètre θ d'une variable aléatoire X. On dispose d'un estimateur $\hat{\theta}_n$ asymptotiquement normal, i.e il existe $\sigma^2 > 0$ tel que

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right).$$

On supposera également que l'on a un estimateur consistant $\hat{\sigma}_n^2$ de σ^2.

Intervalles de confiance asymptotiques

Proposition

Soit $\alpha \in (0,1)$*,*

$$\mathbb{P}\left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}\right] \xrightarrow[n\to+\infty]{} 1-\alpha,$$

où $q_{1-\alpha/2}$ *est le quantile d'ordre* $1-\alpha/2$ *de la loi normale centrée réduite.*

On obtient donc l'intervalle de confiance asymptotique de niveau $1-\alpha$

$$IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}; \hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}}\right].$$

Exemples

Exemple 1 : le lancer de pièce. On considère une variable aléatoire X suivant une loi de Bernoulli de paramètre $\theta \in (0, 1)$. Pour tout $\alpha \in (0, 1)$, un intervalle de confiance asymptotique de niveau $1 - \alpha$ de θ

$$IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n - q_{1-\alpha/2} \frac{\sqrt{\hat{\theta}_n \left(1 - \hat{\theta}_n\right)}}{\sqrt{n}} ; \hat{\theta}_n + q_{1-\alpha/2} \frac{\sqrt{\hat{\theta}_n \left(1 - \hat{\theta}_n\right)}}{\sqrt{n}} \right].$$

Exemples

Exemple 2 : la loi exponentielle. On considère une variable aléatoire X suivant une loi exponentielle de paramètre $\theta > 0$. Pour tout $\alpha \in (0, 1)$, un intervalle de confiance asymptotique de niveau $1 - \alpha$ de θ

$$IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{\sqrt{n}}; \hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{\sqrt{n}}\right].$$

EXEMPLES

Exemple 2bis : la loi exponentielle. En réalité, pour la loi exponentielle, on peut être malin et obtenir un intervalle de confiance asymptotique de niveau $1 - \alpha$ de θ

$$IC_{1-\alpha}(\theta) = \left[\frac{\hat{\theta}_n}{1 + \frac{q_{1-\alpha/2}}{\sqrt{n}}} ; \frac{\hat{\theta}_n}{1 - \frac{q_{1-\alpha/2}}{\sqrt{n}}} \right].$$