

Feuille de Travaux Dirigés 5

Estimation par intervalles de confiance

Définition 1. Soient Y une v.a. réelle et $\alpha \in]0, 1[$. On appelle quantile d'ordre α de Y le nombre q_α tel que

$$q_\alpha = \inf\{y \in \mathbb{R} : F_Y(y) \geq \alpha\}.$$

Propriétés :

1. On a $\mathbb{P}(Y \leq q_\alpha) = F_Y(q_\alpha) = \alpha$.
2. Si Y est une v.a. continue la fonction $q_\alpha :]0, 1[\rightarrow \{x : f_Y(x) > 0\}$ est bijective et continue.
3. Si Y est une v.a. continue alors pour tout $0 \leq \beta \leq \gamma \leq 1$:

$$\mathbb{P}(q_\beta < Y \leq q_\gamma) = \mathbb{P}(Y \leq q_\gamma) - \mathbb{P}(Y \leq q_\beta) = \gamma - \beta.$$

4. Si f_Y est une fonction paire (= la loi de Y est symétrique autour de zéro, $-Y$ a la même loi de Y) alors $q_{1-\alpha} = -q_\alpha$.
5. $q_{1/2}$ est la médiane. $q_{1/4}$ le premier quartile.

Exercice 1. Montrer que si f_Y est une fonction paire alors $q_{1-\alpha} = -q_\alpha$.

Exercice 2. Calculer la fonction q_α dans les cas suivantes

1. $Y \sim B(2, 1/2)$;
2. $Y \sim \mathcal{U}([0, 1])$;
3. $Y \sim \mathcal{E}(\lambda)$, $\lambda > 0$;
4. $Y \sim \mathcal{G}eom(p)$, $p \in]0, 1[$;

Problème

Une entreprise reçoit d'un de ses fournisseurs un lot de pièces qui doit "normalement" contenir une proportion $\theta \leq 10\%$ de pièces défectueuses. L'entreprise voudrait, par examen d'un échantillon de taille n , décider entre $\theta \leq 10\%$ et $\theta > 10\%$, sachant qu'elle acceptera le lot dans le premier cas et le rejettera dans le deuxième cas.

On définit

$$X_i = \begin{cases} 1 & \text{si la pièce prélevée est défectueuse;} \\ 0 & \text{sinon.} \end{cases}$$

X_1, \dots, X_n sont n variables iid de loi de Bernoulli de paramètre θ qui composent l'échantillon \mathbf{X} . L'EMV est $\hat{\theta}_n = \bar{X}_n$ (égale à l'estimateur par méthode des moments).

Supposons $n = 100$ et que on observe $\bar{X}_n = 0.195$.

Question : Quelle décision l'entreprise doit prendre? Accepter ou rejeter le lot? Et, sur quel critère l'entreprise doit se baser pour prendre sa décision?

Définition 2. Soit $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ un modèle paramétrique. On dispose d'un échantillon $\mathbb{X} = (X_1, \dots, X_n)$ de n v.a. iid $\sim \mathbb{P}_\theta$. Soient A_n et B_n deux statistiques. On dira que $[A_n, B_n]$ est un intervalle de confiance de niveau $1 - \alpha$ pour θ si

$$\mathbb{P}_\theta(A_n \leq \theta \leq B_n) = 1 - \alpha$$

pour tout $\theta \in \Theta$.

On dira que $[A_n, B_n]$ est un intervalle de confiance de niveau asymptotiquement égal à $1 - \alpha$ pour θ si

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta(A_n \leq \theta \leq B_n) = 1 - \alpha$$

pour tout $\theta \in \Theta$.

Remarque : Dans les applications on utilise souvent les valeurs $\alpha = 0.05, 0.01$.

Exercice 3. Soit $X \sim \mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$ notre modèle paramétrique. Soient ζ_α les quantiles de la v.a. Gaussienne standard (centrée et réduite). On pose $A_n = \bar{X}_n - \zeta_\gamma / \sqrt{n}$ et $B_n = \bar{X}_n + \zeta_\beta / \sqrt{n}$. Déterminer β et γ dans $[0, 1]$ tels que $[A_n, B_n]$ soit un intervalle de confiance de niveau $1 - \alpha$ pour μ .

Remarques :

- Il existe un nombre infini des intervalles de confiance de niveau $1 - \alpha$.
- Si $\beta \neq 0$ et $\gamma \neq 1$ on parlera d'un intervalle de confiance bilatérale.
- Si $\beta = 0$ ($\gamma = 1 - \alpha$) ou si $\gamma = 1$ ($\beta = \alpha$) on parlera d'un intervalle de confiance unilatéral.
- Si $\beta = \alpha/2$ et $\gamma = 1 - \alpha/2$ on parlera d'un intervalle de confiance bilatéral symétrique.
- Valeurs utiles de ζ_α : $\zeta_{1/2} = 0$, $\zeta_{0.9} = 1.28$, $\zeta_{0.9} = 1.645$, $\zeta_{0.975} = 1.96$, $\zeta_{0.995} = 2.58$.

Exercice 4. Reprenons le problème introductif. $X \sim \text{Ber}(\theta)$. L'EMV pour θ est \bar{X}_n .

1. Montrer que

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

(utiliser le TCL, la LGN, le théorème de continuité et le lemme de Slutsky).

2. Donner un intervalle de confiance asymptotique et bilatéral symétrique de niveau $1 - \alpha$ pour θ .

Application : Si on fixe $\alpha = 0.05$. Pour la valeur observé de $\bar{X}_n = 0.195$ ($n = 100$) on a que l'intervalle de confiance trouvé dans l'exercice précédent est

$$\theta \in [0.117, 0.273]$$

(vérifier). Ce qui permet de rejeter le lot avec niveau de confiance 95%.

Exercice 5. Soit $X \sim \mathcal{E}(\lambda)$, $\lambda > 0$. Déterminer un intervalle de confiance de la forme $\{\lambda > a\}$ de niveau $1 - \alpha$ pour λ .