Cours 8

6 Régions de Confiance

Dans l'approche classique, la construction de régions de confiance est rarement immédiate. On a besoin de stratégies particulières comme par exemple celle qui consiste à fabriquer une fonction pivotale.

Exemple – Considérons $(X_1,\cdots,X_n)$ un n-échantillon de v.a. de loi $\mathcal{N}(\mu,\sigma^2)$. Supposons σ^2 connu. Si on veut construire un intervalle de confiance symétrique pour μ de niveau $1-\alpha$, on s'appuie sur le fait que l'estimateur $\overline{X}_n$ de μ suit une loi normale de moyenne μ et de variance σ^2/n pour écrire :

$$P(a<\mu\leqslant b)=1-\alpha \Longleftrightarrow P\left(\frac{\overline{X}_n-b}{\sigma/\sqrt{n}}<\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\leqslant\frac{\bar{X}_n-a}{\sigma/\sqrt{n}}\right)=1-\alpha.$$

dont on déduit les bornes a et b, pour un intervalle symétrique :

$$\overline{X}_n\pm\frac{\sigma}{\sqrt{n}}\phi^{-1}(1-\alpha/2)$$

L'approche bayésienne présente l'avantage de permettre une construction directe d'une région de confiance. En effet, disposant de la loi a posteriori $\pi(\theta\mid x)$, le calcul de $[a,b]$ tel que :

$$P^{\pi(.|x)}([a,b]\ni\theta)=1-\alpha,$$

ne présente pas de difficultés. Cette région est dite de *plus forte densité a posteriori* et notée PFDP. D'une manière générale, on peut donner la définition suivante :

Définition 14 – *Soit un modèle statistique bayésien $(\mathscr{X},\mathscr{A},P_\theta,\pi(\theta))$, on appelle région de confiance de niveau α ou région α-crédible, tout ensemble C_x tel que :*

$$P^{\pi(.|x)}(\theta\in C_x)\geqslant 1-\alpha$$

où $P^{\pi(.|x)}$ est la loi de probabilité dont la densité a posteriori est $\pi(.\mid x)$.

On dira qu'une région est α-crédible PFDP si elle s'écrit :

$$C_x^\pi=\{\theta;\pi(\theta\mid x)\geqslant k_\alpha\}$$

où k_α est la plus grande valeur telle que :

$$P^{\pi(.|x)}(\theta\in C_x^\pi)\geqslant 1-\alpha$$

Exemple – Considérons une loi normale dont la moyenne θ suit a priori une loi $\mathcal{N}(0,\sigma^2)$. Rappelons que dans ce cas, la loi a posteriori est une loi normale $\mathcal{N}(\mu(x), 1/\rho^2)$ avec

$$\mu(x) = \frac{\tau^2 x}{\tau^2+\sigma^2} \quad et \quad \frac{1}{\rho^2} = \frac{1}{\tau^2} + \frac{1}{\sigma^2}.$$

On fabrique une région α-credible PFDP en écrivant :

$$\begin{aligned}
& P(|\theta| \leqslant a) = 1-\alpha \\
\Longleftrightarrow \quad & P\{|[\theta-\mu(x)]/\rho| \leqslant (a-\mu(x))/\rho\} = 1-\alpha \\
\Longleftrightarrow \quad & [a-\mu(x)]/\rho = \phi^{-1}(1-\alpha/2) \\
\Longleftrightarrow \quad & a = \mu(x) \pm \rho\ \phi^{-1}(1-\alpha/2).
\end{aligned}$$

Si $\tau^2 \to +\infty$, $\rho \longrightarrow 1/\sigma^2$ et on retrouve l'intervalle classique.
Notons que $\tau^2 \to +\infty$ correspond à une loi a priori uniforme sur $\mathbb{R}$.

7 Tests d'Hypothèses

7.1 Généralités

Une autre façon de mener l'inférence sur le paramètre θ est de proposer des hypothèses portant sur des régions de Θ succeptible de contenir ce paramètre. Autrement dit, on se pose la question de savoir s'il est possible d'admettre que θ appartienne à telle ou telle région de Θ et ce au vu de ce que l'on a observé.

On va donc émettre une hypothèse notée H_0 dite *hypothèse nulle* que l'on va chercher à tester. L'observation nous autorise-t-elle à considérer cette hypothèse comme vraie ou nous conduit-elle à la rejeter ?

En général, on teste l'hypothèse H_0 contre une hypothèse dite *hypothèse alternative* notée H_1. Le test s'écrit :

$$H_0 : \theta \in \Theta_0 \quad \text{contre} \quad H_1 : \theta \in \Theta_1$$

où Θ_0 et Θ_1 sont des sous-ensembles de Θ.
Remarquons que ces hypothèses ne sont pas nécessairement complémentaires. Par exemple, on peut tester l'hypothèse que la moyenne d'une gaussienne est égale à une valeur donnée contre l'hypothèse que cette moyenne est plus grande que cette valeur. Ceci se formulera de la manière suivante :

$$H_0 : \mu = \mu_0 \quad \text{contre} \quad H_1 : \mu > \mu_0.$$

Notons que H_0 s'écrit encore $\mu - \mu_0 = 0$ de là, la terminologie *hypothèse nulle*, pour *il n'y a pas de différence* dans les tests de comparaison.

Les deux hypothèses d'un test n'ont pas exactement le même statut. En effet, H_0 est l'hypothèse privilégiée. On pourra penser à l'analogie du procés. La justice priviligie l'hypothèse de non culpabilité (H_0 : non coupable) ; c'est la *présomption d'innocence*.

La démarche de test permet de rejeter H_0 avec un certain degré de certitude. L'accusation (ce qu'on a observé) vise à prouver que l'accusé est coupable ou plus précisement qu'il n'est pas possible de considérer ce dernier non coupable. Cependant, si elle y parvient, cela ne signifie pas qu'il l'est. Si on rejette H_0, le test nous dit que les informations dont on dispose, ne permettent pas de considérer qu'il est non coupable.

Cette analogie permet de saisir l'importance de deux erreurs que l'on peut commettre :

— Condamner un innoncent : rejeter H_0 alors qu'elle est vraie

— Libérer un coupable : ne pas rejeter H_0 alors qu'elle est fausse.

Dans la construction du test, on cherchera à faire en sorte que ces deux erreurs soient les plus faibles possibles.

La première de ces erreurs : *rejeter H_0 alors qu'elle est vraie* est appelée *erreur de $1^{ère}$ espèce* ou *erreur de type I* ou parfois *risque de $1^{ère}$ espèce*. On la note : α.

La seconde : *ne pas rejeter H_0 alors qu'elle est fausse* est appelée *erreur de $2^{ème}$ espèce* ou *erreur de type II* ou parfois *risque de $2^{ème}$ espèce*. On la note : β.

Un test est un problème de décision. Suivant ce que l'on aura observé, on décidera de rejeter ou non H_0. On va chercher à explorer les sous-ensembles de l'espace des observations $\mathcal{X}$ qui contredise l'hypothèse nulle. Ils formeront une région dite *région de rejet* ou *région critique*.

$$W = \{(x_1, x_2, \cdots, x_n) | \text{ on rejette } H_0\}.$$

Les deux erreurs décrites précédemment s'expriment alors par :

$\alpha = P_\theta(W \mid \theta \in \Theta_0)$, mesure du sous-espace de l'observation dans lequel on rejette à tort et

$\beta = 1 - P_\theta(W \mid \theta \in \Theta_1))$, mesure du sous-espace de l'observation dans lequel on ne rejette pas H_0 alors qu'on devrait le faire puisque θ est dans Θ_1.

Tester les hypothèses consistera donc à définir une région de rejet telle que les probabilités définies ci-dessus soient les plus petites possibles.

On décidera alors, à travers ce que l'on a observé effectivement, de rejeter ou non l'hypothèse nulle.

On donne la définition suivante :

Définition 15 – *Un test pur est une application ϕ de $\mathcal{X}$ dans $\{0,1\}$ telle que $\phi(x)$ est la probabilité de rejeter H_0 si x est observé.*

L'ensemble des x tels que $\phi(x) = 1$ est la région de rejet du test. ϕ est donc tout simplement la fonction indicatrice de la région de rejet. Autrement dit, écrire un test revient à estimer une fonction indicatrice.

Ce problème peut être plongé dans le cadre de la théorie de la décision. L'espace des actions contient deux éléments :

d_0 : on ne rejette pas H_0,

d_1 : on rejette H_0.

A ces décisions vont être associées des coûts $L(\theta, d_0(x))$ et $L(\theta, d_1(x))$. Par définition, les décisions bayésiennes sont celles qui minimisent le coût a posteriori : $E^{\pi(.|x)}[L(\theta, d_i)]$.
Considérons le coût 0-1 :

$$L(\theta, d_i(x)) = \begin{cases} 1 & \text{si } \theta \in \Theta_i \\ 0 & \text{sinon.} \end{cases} \quad \text{soit} \quad L(\theta, d_i(x)) = \mathbb{1}(\theta \in \Theta_i),\ i = 0,1.$$

Si on décide de ne pas rejeter H_0 alors qu'effectivement $\theta \in \Theta_0$, la pénalité est nulle. Elle sera encore nulle si on décide de rejeter H_0 alors que $\theta \in \Theta_1$.
Calculons le coût a posteriori de chaque décision :

$$E^{\pi(.|x)}[L(\theta, d_i(x))] = \int_\Theta L(\theta, d_i)\pi(\theta \mid x)d\theta = \int_{\Theta_i} \pi(\theta \mid x)d\theta = P(\theta \in \Theta_i \mid x)\ ,\quad i = 0,1.$$

Autrement dit, si $P(\Theta_1 \mid x) < P(\Theta_0 \mid x)$, alors on choisit de rejeter H_0.

La décision bayésienne est donc l'hypothèse ayant la plus grande probabilité a posteriori.
Considérons maintenant une fonction de coût de la forme :

$$L(\theta, d_i(x)) = \begin{cases} 0 & \text{si } \theta \in \Theta_i \\ K_i & \text{si } \theta \notin \Theta_i \end{cases} \quad i = 0,1.$$

Comme précédemment, on calcule le coût a posteriori :

$$E^{\pi(.|x)}[L(\theta, d_i)] = K_i P(\theta \in \Theta_j \mid x)\ ,\quad i = 0,1; i \neq j.$$

La décision bayésienne est obtenue en comparant : $K_0 P(\theta \in \Theta_1 \mid x)$ et $K_1 P(\theta \in \Theta_0 \mid x)$.
On rejettera H_0, si le coût a posteriori de d_1 est plus faible i.e. :

$$K_1 P(\theta \in \Theta_0 \mid x) < K_0 P(\theta \in \Theta_1 \mid x) \Longleftrightarrow \frac{K_0}{K_1} > \frac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)}.$$

Mais $P(\theta \in \Theta_0 \mid x) + P(\theta \in \Theta_1 \mid x) = 1$, on a donc

$$\frac{K_0}{K_1} > \frac{1}{P(\theta \in \Theta_1 \mid x)} - 1 \Longleftrightarrow P(\theta \in \Theta_1 \mid x) > \frac{K_1}{K_0 + K_1}$$

Ainsi, en utilisant la terminologie classique, on peut écrire que la région critique du test bayésien est de la forme :

$$C = \left\{ x \mid P(\theta \in \Theta_1 \mid x) > \frac{K_1}{K_0 + K_1} \right\}$$

Exemple : Soit $X \sim \mathcal{N}(\theta, \sigma^2)$. On souhaite effectuer le test suivant :

$$\mathrm{H}_0 : \theta \geqslant \theta_0 \;\text{ contre }\; \mathrm{H}_1 : \theta < \theta_0.$$

Considérons une loi $\mathcal{N}(\mu, \tau^2)$ comme loi a priori sur θ, nous avions calculé que la loi a posteriori était dans ce cas, une loi normale de paramètres :

$$\mu(x) = \frac{\sigma^2\mu + \tau^2 x}{\sigma^2 + \tau^2} \quad et \quad \rho^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

On calcule alors :

$$\begin{aligned} P^{\pi(.|x)}(\Theta_1) &= P(U \leqslant \theta_0) \text{ avec } U \sim \mathcal{N}(\mu(x), \rho^2) \\ &= \Phi\left(\frac{\theta_0 - \mu(x)}{\rho}\right) \end{aligned}$$

On rejettera donc l'hypothèse nulle si

$$\Phi\left(\frac{\theta_0 - \mu(x)}{\rho}\right) > \frac{K_1}{K_0 + K_1} \Longleftrightarrow \mu(x) < \theta_0 - \rho\ \Phi^{-1}\left(\frac{K_1}{K_0 + K_1}\right)$$

Or $\mu(x) = \rho^2\ (x/\sigma^2 + \mu/\tau^2)$.
On a donc :

$$\begin{aligned} & \rho^2\ (x/\sigma^2 + \mu/\tau^2) < \theta_0 - \rho\ \Phi^{-1}\left(\frac{K_1}{K_0 + K_1}\right) \\ \Longleftrightarrow \quad & x < \frac{\sigma^2}{\rho^2}(\theta_0 - \tau) - \rho\Phi^{-1}\big(K_1/(K_1 + K_2)\big) \end{aligned}$$

Le test bayésien rejettera donc l'hypothèse nulle si :

$$\sqrt{\rho}(\theta_0 - \mu(x)) > \phi^{-1}\left(\frac{(K_1}{K_1 + K_2}\right).$$

La région critique est donc de la forme :

$$C = \left\{x < \left(\frac{1}{\rho}\left[\theta_0 - \frac{N_\alpha}{\sqrt{\rho}}\phi^{-1}(K_1/(K_1 + K_2))\right] - \frac{\mu}{\tau^2}\right)\sigma^2\right\}.$$

□

7.2 Facteur de Bayes

Dans le cadre bayésien, disposant d'une loi a priori et ayant calculé une loi a posteriori, la mesure des régions Θ_0 et Θ_1 ne présente pas de difficulté et est directe. On calculera deux probabilités que l'on comparera pour prendre une décision : rejeter ou accepter H_0. Ces probabilités sont :

$$\alpha_0 = P(\theta \in \Theta_0 \mid x) = \int_{\Theta_0} \pi(\theta \mid x) d\theta \quad \text{et} \quad \alpha_1 = P(\theta \in \Theta_1 \mid x) = \int_{\Theta_1} \pi(\theta \mid x) d\theta.$$

La décision repose donc sur la nature du rapport : α_0/α_1.
On donne la définition suivante :

Définition 16 *Le facteur de Bayes est le rapport :*

$$B = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}$$

où

$$\pi_i = Pr(\theta \in \Theta_i) = \int_{\Theta_i} \pi(\theta) d\theta \ , \quad i = 0, 1.$$

Il s'agit de comparer le rapport des lois a posteriori et le rapport des lois a priori.

Remarquons que lorsque $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$, le facteur de Bayes n'est autre que le rapport de vraisemblance classique. En effet,

$$B = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}$$

puisque $\int_{\Theta_i} f(x \mid \theta)\pi(\theta) d\theta = f(x \mid \theta_i)\pi(\theta_i)\ , \quad i = 0, 1.$

α_0/α_1 est appelé *odds ratin a posteriori* et π_0/π_1 est appelé *odds ratio a priori*.
Supposons que l'on veuille tester : $H_0 : \theta = \theta_0$.
Soit π_0 la probabilité a priori que $\theta = \theta_0$ et g_1 densité a priori sur $\Theta_1 = \{\theta \neq \theta_0\}$.
La loi a priori $\pi(\theta)$ s'écrit : $\pi(\theta) = \pi_0 \mathbb{1}_{\{\theta=\theta_0\}} + (1-\pi_0) g_1(\theta) \mathbb{1}_{\{\theta\neq\theta_0\}}$.
Calculons la loi a posteriori

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{f(x)}$$

$$\text{avec } f(x) = \int_{\Theta} f(x \mid \theta)\pi(\theta) d\theta = \pi_0 f(x \mid \theta) + (1-\pi_0) \underbrace{\int_{\{\theta\neq\theta_0\}} f(x \mid \theta) g_1(\theta) d\theta}_{f_1(x)}.$$

Et la probabilité que $\theta = \theta_0$ s'écrit :

$$\frac{\pi_0 f(x \mid \theta_0)}{\pi_0 f(x \mid \theta_0) + (1-\pi_0) f_1(x)} = \left[1 + \frac{1-\pi_0}{\pi_0} \cdot \frac{f_1(x)}{f(x \mid \theta_0)}\right]^{-1}.$$

Il s'agit de α_0 et

$$\alpha_1 = \int_{\{\theta\neq\theta_0\}} \frac{(1-\pi_0) g_1(\theta) f(x \mid \theta)}{\pi_0 f(x \mid \theta_0) + (1-\pi_0) f_1(x)} d\theta = \frac{(1-\pi_0) f_1(x)}{\pi_0 f(x \mid \theta_0) + (1-\pi_0) f_1(x)}$$

Ainsi le facteur de Bayes a pour expression :

$$B = \frac{\pi_0 f(x \mid \theta_0)}{(1-\pi_0) f_1(x)} / \frac{\pi_0}{1-\pi_0} = \frac{f(x \mid \theta_0)}{f_1(x)}.$$

On obtient donc une expression similaire à un rapport de vraisemblance modifié et on remarque la relation avec la loi a posteriori :

$$\pi(\theta = \theta_0 \mid x) = \left[1 + \frac{1-\pi_0}{\pi_0} \frac{1}{B}\right]^{-1}$$

Exemple : Reprenons le cas normal, on avait : $x \sim \mathcal{N}(\theta, \sigma^2)$, $\theta \sim \mathcal{N}(\mu, \tau^2)$ et $\theta \mid x \sim \mathcal{N}(\mu(x), \rho^2)$. On veut tester : $H_0 : \theta = 0$. Il semble raisonnable de considérer une loi a priori de la forme $\mathcal{N}(0, \tau^2)$ et on calcule :

$$\frac{1}{B} = \frac{f_1(x)}{f(x \mid 0)}.$$

On a :

$$g_1(\theta) \propto \frac{1}{\tau} \exp\left\{-\left(\frac{\theta}{2\tau}\right)^2\right\} \quad et \quad f(x \mid 0) \propto \frac{1}{\sigma} \exp\left\{-\left(\frac{x}{2\sigma}\right)^2\right\}$$

$$\begin{aligned} f_1(x) &= \int_{\{\theta \neq 0\}} f(x \mid \theta) g_1(\theta) d\theta \\ &\propto \frac{1}{\tau\sigma} \int_{\{\theta \neq \theta\}} \exp\left\{-\left[\left(\frac{\theta}{\tau}\right)^2 + \left(\frac{x-\theta}{\sigma}\right)^2\right]\right\} d\theta \\ &\propto \frac{1}{\tau\sigma} \int_{\{\theta \neq 0\}} \exp\left\{-\left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\right)^2 \left[\theta - \left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\right)^{-1} \frac{x}{\sigma^2}\right]^2\right\} \exp\left\{\frac{\tau^2}{2(\sigma^2 + \tau^2)} \cdot \frac{x^2}{\sigma^2}\right\}. \end{aligned}$$

On calcule alors directement la probabilité a posteriori :

$$\pi(\theta = 0 \mid x) = \left[1 + \frac{1 - \pi_0}{\pi_0} \sqrt{\frac{\sigma^2}{\sigma^2 + \tau^2}} \exp\left(\frac{\tau^2 x^2}{2\sigma^2(\sigma^2 + \tau^2)}\right)\right]^{-1}.$$

□