Statistique Bayésienne

Tests et régions de confiance

Anna Simoni[2]

[2]CREST - Ensae and CNRS

Outline

- H_0 et H_1 peuvent être considerées : (1) soit comme deux régions (i.e. une partition) de l'espace paramétrique d'un unique modèle d'échantillonage, (2) soit comme deux modèles d'échantillonage différents.

- Débats entre approches classiques (fondé sur $P(\text{données}|\theta)$) et approches bayésien (fondé sur $P(\theta|\text{données})$). On peut considerer ces deux approches comme complémentaire plutôt qu'opposées.

- Deux points de vue :

  - Une *statistique de test* est un procédé statistique à valeurs dans un espace à deux points : "accepter" et "rejeter" une hypothèse.

  - les tests d'hypothèses peuvent aussi être considerés comme une façon pour les statisticien de gérer ses doutes relatifs à son modèle statistique.

Principes généraux des tests d'hypothèses. I

- Nous avons un espace de décisions avec deux points : $\mathscr{D} = \{\delta_0, \delta_1\}$ et une fonction de perte $L(\theta, \delta)$.

- On peut partitionner en deux classes l'ensemble des états de la nature :

$$\Theta = \Theta_0 \cup \Theta_1$$

où Θ_0 et Θ_1 sont définis :

$$\begin{aligned} \Theta_0 &= \{\theta;\, L(\theta, \delta_0) = 0\} \\ \Theta_1 &= \{\theta;\, L(\theta, \delta_1) = 0\}. \end{aligned}$$

- La specification de la fonction de perte est alors complétée comme suit :

$$L(\theta, \delta) = \begin{cases} L_1(\theta) & \text{si } \delta = \delta_1 \text{ et } \theta \in \Theta_0 \\ L_0(\theta) & \text{si } \delta = \delta_0 \text{ et } \theta \in \Theta_1, \end{cases}$$

c'est-à-dire $L(\theta, \delta_0) = \mathbb{1}_{\Theta_1}(\theta) L_0(\theta)$ et $L(\theta, \delta_1) = \mathbb{1}_{\Theta_0}(\theta) L_1(\theta)$.

Principes généraux des tests d'hypothèses. II

- On obtient donc

	Θ_0	Θ_1
δ_0	0	$L_0(\theta)$
δ_1	$L_1(\theta)$	0

- Cas particulier : $L_j(\theta) = L_j, j = 0, 1$ (fonction de perte constante sur les éléments de la partition des états de la nature).

- Lorsque Θ devient l'espace paramétrique d'un modèle statistique, les éléments de la partition $\Theta = \Theta_0 \cup \Theta_1$ s'appelleront des hypothèses statistiques.

- Approche de Neyman et Pearson : H_0 est choisie de telle sorte que l'erreur de première espèce soit la plus grave.

- L'analyse Bayésienne ne requiert pas une telle spécification.

Principes généraux des tests d'hypothèses. III

- Soit $x^{(n)} := (x_1, \ldots, x_n)$ l'observation d'un échantillon i.i.d. de $X \in \mathscr{X}$.
- La decision optimal a posteriori est définie par :

$$\begin{aligned} \delta^*(x^{(n)}) &= \arg\min_{\delta \in \mathscr{D}} \mathbf{E}[L(\theta, \delta) | x^{(n)}] \\ &= \arg\min_{\delta \in \mathscr{D}} \{\rho(\pi, \delta_0), \rho(\pi, \delta_1)\} \end{aligned}$$

where $\rho(\pi, \delta) := \mathbf{E}[L(\theta, \delta) | x^{(n)}]$ est le risque à posteriori de la décision δ.

- Dans le cas particulier $L_j(\theta) = L_j, j = 0, 1$, on définit :
$\pi(\theta \in \Theta_0 | x^{(n)}) = p(x^{(n)})$ et on peut alors écrire :

$$\begin{aligned} \mathbf{E}[L(\theta, \delta_0) | x^{(n)}] &= L_0 \times (1 - p(x^{(n)})) \\ \mathbf{E}[L(\theta, \delta_1) | x^{(n)}] &= L_1 \times p(x^{(n)}). \end{aligned}$$

La règle optimale de decision devient :

$$\delta^*(x^{(n)}) = \delta_0 \quad \Leftrightarrow \quad L_0 \times (1 - p(x^{(n)})) < L_1 p(x^{(n)}).$$

On peut aussi écrire la règle optimale de décision en termes des *quotients d'enjeux* (*odds ratio*) :

$$\delta^*(x^{(n)}) = \delta_0 \quad \Leftrightarrow \quad \frac{p(x^{(n)})}{1 - p(x^{(n)})} > \frac{L_0}{L_1}.$$

Principes généraux des tests d'hypothèses. IV

- Par exemple : si $L_1 = 19L_0$, alors $\delta^*(x^{(n)}) = \delta_1 \quad \Leftrightarrow \quad \frac{p(x^{(n)})}{1-p(x^{(n)})} < \frac{1}{19}$.
  Ceci est une façon de formaliser l'idée que l'erreur de type I est beaucoup plus grave que l'erreur de type II.

- En général donc, en analyse Bayésienne, on calcule tout simplement $\pi(\Theta_0|x^{(n)})$ et $\pi(\Theta_1|x^{(n)})$ et on decide en conséquence. Ces probabilités sont les probabilités subjectives des hypothèses au vu des données et de l'a priori.

- La règle de Bayes pour une perte $0-1$ consiste à choisir l'hypothèses avec la probabilité a posteriori plus haute.

- Un autre outil utilisé dans des problèmes de test est le *Facteur de Bayes* :

Définition

Soit $\pi(\Theta_0|x^{(n)})/\pi(\Theta_1|x^{(n)})$ *le odds ratio à posteriori et* $\pi(\Theta_0)/\pi(\Theta_1)$ *le odds ratio à priori. La quantité*

$$B_{01} = \frac{\textit{posterior odds ratio}}{\textit{prior odds ratio}} = \frac{\pi(\Theta_0|x^{(n)})\pi(\Theta_1)}{\pi(\Theta_1|x^{(n)})\pi(\Theta_0)}$$

est appelée le facteur de Bayes en faveur de Θ_0.

Principes généraux des tests d'hypothèses. IV

- Plus la valeur de BF_{01} est petite et plus forte est l'evidence contre H_0.

- Lorsque les hypothèses en présence sont des hypothèses simples (i.e. $\Theta_j = \{\theta_j\}, j = 0, 1$) et donc $\Theta = \{\theta_0, \theta_1\}$, le facteur de Bayes est exactement égal au quotient des vraisemblances et est donc indépendant de l'à priori.

- En général, B_{01} depend de l'a priori. De plus, $B_{10} = 1/B_{01}$.

- Problème avec cet approche : si l'a priori est impropre alors $\pi(\Theta_0)$ et $\pi(\Theta_1)$ peuvent être indéfinies.

- Si notre vue de H_0 est comme dans l'approche fréquentiste (i.e. H_0 ne devrait pas être rejetée sauf s'il y a suffisamment d'evidence pour le contraire) alors il est raisonable d'assigner plus de probabilité a priori à H_0 que à H_1. Un choix objectif serait d'assigner probabilités a priori égaux.

- Tout ça peut être mieux fait avec la specification de l'a priori suivante.

Principes généraux des tests d'hypothèses. V I

- Écrivons l'a priori comme :

$$\pi(\theta) = \begin{cases} \pi_0 g_0(\theta) & \text{if } \theta \in \Theta_0 \\ \pi_1 g_1(\theta) & \text{if } \theta \in \Theta_1 \end{cases} \tag{1}$$

où $\pi_j = \pi(\Theta_j), j = 0, 1$ et g_0 et g_1 sont des densités propres. Donc,

$$\pi(\theta) = \pi_0 g_0 \mathbb{1}_{\Theta_0}(\theta) + (1 - \pi_0) g_1(\theta) \mathbb{1}_{\Theta_1}(\theta).$$

- Alors, on peut écrire le posterior odds ratio :

$$\begin{aligned} \frac{\pi(\Theta_0|x^{(n)})}{\pi(\Theta_1|x^{(n)})} &= \frac{\int_{\Theta_0} \pi(\theta|x^{(n)})d\theta}{\int_{\Theta_1} \pi(\theta|x^{(n)})d\theta} = \frac{\int_{\Theta_0} f(x^{(n)}|\theta)\pi_0 g_0(\theta)d\theta/m(x^{(n)})}{\int_{\Theta_1} f(x^{(n)}|\theta)\pi_1 g_1(\theta)d\theta/m(x^{(n)})} \\ &= \frac{\pi_0 \int_{\Theta_0} f(x^{(n)}|\theta)g_0(\theta)d\theta}{\pi_1 \int_{\Theta_0} f(x^{(n)}|\theta)g_0(\theta)d\theta} \end{aligned}$$

et le facteur de Bayes :

Principes généraux des tests d'hypothèses. V II

$$B = \frac{\int_{\Theta_0} f(x^{(n)}|\theta) g_0(\theta) d\theta}{\int_{\Theta_0} f(x^{(n)}|\theta) g_0(\theta) d\theta}$$

qui est le ratio des vraisemblances ponderées (par g_0 et g_1) de Θ_0 et Θ_1.

- On a que le posterior odds ratio est égale à :

$$\frac{\pi_0}{1-\pi_0} BF_{01}$$

et il devient égale à BF_{01} si $\pi_0 = 1/2$.

Exemple A I

- Consider a blood test conducted for determining the sugar level of a person with diabetes two hours after he had his breakfast.
- We want to see if his medication has controlled his blood sugar levels.
- Assume that the test result X is $\mathcal{N}(\theta, 100)$, where θ is the true level.
- In the appropriate population (diabetic but under this treatment), $\theta \sim \mathcal{N}(100, 900)$,
- Then, marginally $X \sim \mathcal{N}(100, 1000)$, and the posterior distribution is

$$\theta | X = x \sim \mathcal{N}(0.9x + 10, 90).$$

- We want to test :

$$H_0 : \theta \leq 130$$
$$H_1 : \theta > 130.$$

- If the blood test shows a sugar level of 130, what can be concluded ?

Exemple A II

- Given this test result, the posterior is $\mathcal{N}(127, 90)$. Consequently :

$$\begin{aligned}\pi(\theta \leq 130|X=130) &= \Phi\left(\frac{130-127}{\sqrt{90}}\right) = \Phi(.316) = 0.624 \\ \pi(\theta > 130|X=130) &= 0.376.\end{aligned}$$

  Therefore, the posterior odds ratio is : $0.624/0.376 = 1.66$.

- Because $\pi_0 = \Phi\left(\frac{130-100}{30}\right) = \Phi(1)$, the prior odds ratio is $\Phi(1)/(1-\Phi(1)) = 0.8413/0.1587 = 5.3$ and thus the Bayes factor is

$$BF_{01} = \frac{1.66}{5.3} = 0.313.$$

- It can also be noted here that in one-sided testing situations when a continuous prior π can be specified readily for the entire parameter space, there is no need to express it in the form of $\pi(\theta) = \pi_0 g_0 \mathbb{1}_{\Theta_0}(\theta) + (1-\pi_0) g_1(\theta) \mathbb{1}_{\Theta_1}(\theta)$. However, the problem of testing a point null hypothesis turns out to be quite different.

Exemple B I

- Mister A is interested in determining his true weight from a variable bathroom scale.
- Assume the measurements are $X_i \sim \mathcal{N}(\mu, 9)$.
- Sample (measurements in pounds) :
  182, 172, 173, 176, 176, 180, 173, 174, 179, 175.
- μ=Mister A's true weight
- Suppose Mister A is interested in assessing if his true weight is more than 175 pounds. He wishes to test the hypotheses

$$
\begin{aligned}
H_0 &: \mu \leq 175 \\
H_1 &: \mu > 175.
\end{aligned}
$$

- Prior : $\mu \sim \mathcal{N}(170, 5)$.

- The prior odds of H_0 is given by

$$\frac{\pi_0}{\pi_1} = \frac{P(\mu \leq 175)}{P(\mu > 175)}.$$

```
> pmean=170; pvar=25
> probH=pnorm(175,pmean,sqrt(pvar))
> probA=1-probH
> prior.odds=probH/probA
> prior.odds
[1] 5.302974
```

- So, a priori, H_0 is five times more likely than H_1.
- We enter the ten weight measurements into R and compute the sample mean $\bar{y}$ and the associated sampling variance σ^2/n :

```
> weights=c(182, 172, 173, 176, 176, 180, 173, 174, 179, 175)
> ybar=mean(weights)
> sigma2 = 3^2/length(weights)
```

- The posterior precision of μ is the sum of the precisions of the data and the prior :

```
> post.precision=1/sigma2+1/pvar
> post.var=1/post.precision
```

- The posterior mean of μ is the weighted average of the sample mean and the prior mean, where the weights are proportional to the respective precisions :

```
>
post.mean=(ybar/sigma2+pmean/pvar)/post.precision
> c(post.mean,sqrt(post.var))
[1]  175.7915058  0.9320547
```

- The posterior density of μ is $\mathcal{N}(175.79, 0.93)$.

Exemple B IV

- Using this normal posterior density, we calculate the odds of H_0 :

```
>
post.odds=pnorm(175,post.mean,sqrt(post.var))/
+ (1-pnorm(175,post.mean,sqrt(post.var)))
> post.odds
[1]  0.2467017
```

- So, the BF_{01} in support of H_0 is

```
> BF = post.odds/prior.odds
> BF
[1]  0.04652139
```

- From the prior probabilities and the Bayes factor, we can compute the posterior probability of H_0 :

```
> postH=probH*BF/(probH*BF+probA)
> postH
[1]  0.1978835
```

- Based on this calculation, we can conclude that it is unlikely that Mister A's weight is at most 175 pounds.

Test d'une hypothèses nulle ponctuelle I

La loi a priori définie en (1) est utile si on veut tester une hypothèse nulle ponctuelle.

- Une hypothèse nulle ponctuelle $H_0 : \theta = \theta_0$ (contre $H_1 : \theta \neq \theta_0$)ne peut pas être testée sous une loi a priori continue.

- De plus, le facteur de Bayes n'est défini que lorsque $\pi_0 \neq 0$ et $\pi_1 \neq 0$. Cela implique que, si H_0 ou H_1 sont a priori impossibles, les observations ne vont pas modifier cette information absolue : des probabilités nulles a priori le restent a posteriori.

Cette modification de la loi a priori est surprenante, puisqu'elle revient à mettre un poids a priori sur un ensemble de mesure 0 :

- Une probabilité $\pi_0 > 0$ doit être assignée au point θ_0 et $(1 - \pi_0)$ doit être répandue sur $\{\theta \neq \theta_0\}$ utilisant une densité g_1.
- g_0 est alors prise égale à un point masse sur θ_0.

Test d'une hypothèses nulle ponctuelle II

- Alors on a que $\pi(\theta)$ a une partie continue et une parti discrète :

$$\pi(\theta) = \pi_0 \mathbb{1}_{\theta_0}(\theta) + (1 - \pi_0) g_1(\theta) \mathbb{1}_{\theta \neq \theta_0}(\theta).$$

- Puisque :

$$\begin{aligned}
\pi(\theta_0 | x^{(n)}) &= \frac{\pi_0 f(x^{(n)} | \theta_0)}{\pi_0 f(x^{(n)} | \theta_0) + (1 - \pi_0) \underbrace{\int_{\theta \neq \theta_0} f(x^{(n)} | \theta) g_1(\theta) d\theta}_{=: m_1(x^{(n)})}} \\
&= \left(1 + \frac{1 - \pi_0}{\pi_0} \frac{m_1(x^{(n)})}{f(x^{(n)} | \theta_0)} \right)^{-1}
\end{aligned}$$

le posterior odds ratio devient

$$\frac{\pi(\theta_0 | x^{(n)})}{1 - \pi(\theta_0 | x^{(n)})} = \frac{\pi_0 f(x^{(n)} | \theta_0)}{(1 - \pi_0) m_1(x^{(n)})}$$

Test d'une hypothèses nulle ponctuelle III

et le facteur de Bayes est :

$$B_{01} = \frac{f(x^{(n)}|\theta_0)}{m_1(x^{(n)})}.$$

Outline

L'approche classique de la théorie des tests est la théorie de Neyman-Pearson (see e.g. Lehmann, 1986). Sous le coût $0-1$, noté L ci-dessous, la notion fréquentiste d'optimalité est fondée sur la puissance d'un test, définie comme :

Définition

La puissance d'une procédure de test φ *est la probabilité de rejeter* H_0 *sous l'hypothèse alternative :* $1-\beta(\theta)=1-\mathbf{E}_\theta[\varphi(x)]$ *lorsque* $\theta \in \Theta_1$. *La quantité* $\beta(\theta)$ *est appelée erreur de deuxième espèce, tandis que l'erreur de première espèce est* $\mathbf{E}_\theta[\varphi(x)]$ *lorsque* $\theta \in \Theta_0$.

Les tests fréquentistes optimaux sont ceux qui minimisent le risque $\mathbf{E}_\theta[L(\theta,\varphi(x))]$ sous H_1 seulement :

Définition

Si $\alpha \in (0,1)$ *et* $\mathscr{C}_\alpha$ *est la classe des procédures* φ *satisfaisant la contrainte suivante sur l'erreur de première espèce :*

$$\sup_{\theta\in\Theta_0} \mathbf{E}_\theta[L(\theta,\varphi(x))] = \sup_{\theta\in\Theta_0} P_\theta(\varphi(x)=1) < \alpha, \tag{2}$$

une procédure de test φ *est dite uniformément plus puissante (UPP) au niveau* α *si elle minimise dans* $\mathscr{C}_\alpha$ *le risque* $\mathbf{E}_\theta[L(\theta,\varphi(x))]$ *uniformément sur* Θ_1.

Tests UPP et UPPS. II

- Cette optimalité entraîne une asymétrie entre les hypothèses nulle et alternative.
- Elle implique la sélection d'un niveau de confiance α par le décideur, en plus du choix de la fonction de coût L, ce qui entraîne généralement le recours à des niveaux standard, comme 0.05 ou 0.01.
- Elle ne suggère pas nécessairement une réduction suffisante de la classe des procédures de test et ne permet pas toujours la sélection d'une procédure unique optimale.
- Si les hypothèses nulle et alternative sont ponctuelles, $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, le lemme de Neyman-Person établit l'existence de procédures de test UPP, de la forme :

$$\varphi(x) = \begin{cases} 0 & \text{si } f(x|\theta_1) < kf(x|\theta_0) \\ 1 & \text{sinon,} \end{cases}$$

k étant donné par le niveau de confiance choisi α.

Tests UPP et UPPS. III

- Soit $T(x)$ une statistique.

Proposition

Soit $f(x|\theta)$ à rapport de vraisemblance monotone dans $T(x)$. Pour $H_0 : \theta \leq \theta_0$ et $H_1 : \theta > \theta_0$ il existe un test UPP tel que

$$\varphi(x) = \begin{cases} 0 & \text{si } T(x) < c \\ \gamma & \text{si } T(x) = c \\ 1 & \text{sinon,} \end{cases}$$

γ et c étant déterminés par la contrainte

$$\mathbf{E}_{\theta_0}[\varphi(x)] = \alpha$$

- Cependant, il n'existe pas de test UPP correspondant au cas : $H_0 : \theta_{\leq}\theta \leq \theta_2$. Ce paradoxe montre l'absence de symétrie du critère UPP et jette un doute sur la validité de l'analyse de Neyman-Pearson ou sur la pertinence d'un coût asymétrique comme le coût $0 - 1$.

Lois a priori les moins favorables. I

Lorsque aucun test UPPS n'existe, il devient assez difficile de défendre et de construire une procédure de test dans un cadre fréquentiste.

- Considérons le rapport de vraisemblance

$$\frac{\sup_{\theta \in \Theta_0} f(x|\theta)}{\sup_{\theta \in \Theta_1} f(x|\theta)}$$

- Ce rapport illustre un lien avec l'approche bayésienne, car il s'agit d'un facteur de Bayes pour une loi a priori π de support réduit aux points $\widehat{\theta}_0$ et $\widehat{\theta}_1$, estimateurs du maximum de vraisemblance de θ sur Θ_0 et Θ_1.
- Soient $H_0 : \theta \in \Theta_0$, $H_1 : \theta = \theta_1$ avec π une loi a priori sur Θ_0. D'un point de vue bayésien, ce problème de test peut être représenté comme le test de $H_\pi : x \sim m_\pi$ contre $H_1 : x \sim f(x|\theta_1)$, où $m_\pi(x) = \int_{\Theta_0} f(x|\theta)\pi(\theta)d\theta$.
- Puisque H_0 et H_1 sont des hypothèses ponctuelles, le lemme de Neyman-Pearson assure l'existence d'un test UPP φ_π à un niveau de signification α et de puissance $1 - \beta_\pi = P_{\theta_1}(\varphi_\pi = 1)$ de la forme :

$$\varphi_\pi(x) = \begin{cases} 0 & \text{si } m_\pi(x) > kf(x|\theta_1) \\ 1 & \text{sinon,} \end{cases}$$

Lois a priori les moins favorables. II

Définition

Une loi la moins favorable est une loi a priori π *qui maximise la puissance* $1-\beta_\pi$.

Théorème

Soit $H_0 : \theta \in \Theta_0$ *et* $H_1 : \theta = \theta_1$. *Si le test UPP* φ_π *au niveau* α *pour* H_π *contre* H_1 *satisfait*

$$\sup_{\theta \in \Theta_0} \mathbf{E}_\theta[L(\theta, \varphi_\pi)] \leq \alpha$$

alors

(i) φ_π *est UPP au niveau* α *;*

(ii) *si* φ_π *est le seul test de niveau* α *de* H_π *contre* H_1, φ_π *est le seul test UPP au niveau* α *pour tester* H_0 *contre* H_1 *; et*

(iii) π *est une loi la moins favorable.*

Outline

Intervalles de crédibilité. I

L'équivalent Bayésien des intervalles de confiance fréquentist est l'intervalle de crédibilité.

Définition

Un ensemble $100(1-\alpha)\%$ *crédible pour* θ *est un sousensemble* $C_x \subset \Theta$ *tel que*

$$1-\alpha \leq \pi(C_x|x).$$

On peut donc parler de la probabilité que θ est dans C_x.

Définition

Une région $100(1-\alpha)\%$ *crédible HPD (Highest Posterior Density) est un sousensemble* $C_x \subset \Theta$ *de la forme*

$$C_x = \{\theta \in \Theta; \pi(\theta|x) \geq k(\alpha)\}$$

où k est la plus grande constante telle que

$$P(C|x) \geq 1-\alpha.$$