Master professionnel II : Ingénierie mathématique : Option Statistique

Statistique Bayésienne.

Anne Philippe
Université de Nantes
Laboratoire de Mathématiques Jean Leray

Fiche 3. Inférence Bayésienne.

Exercice 1.

On dispose de n observations $X_1, ..., X_n$ iid suivant une loi de Bernoulli de paramètre $\theta \in]0,1[$.

1) On suppose que la loi a priori sur le paramètre θ est une loi finie sur $]0,1[$ de support $\{b_1, ..., b_k\}$ telle que pour tout $i = 1, ..., k$: $P(\theta = b_i) = \pi_i$ et $\sum_{i=1}^k \pi_i = 1$. Quelle est la loi a posteriori du paramètre θ ?
2) On suppose que la loi a priori est un melange de lois uniformes. Elle admet pour densité

$$\pi(\theta) = \sum_{i=0}^{k} \pi_i \frac{1}{a_{i+1} - a_i} \mathbb{I}_{[a_i, a_{i+1}[}(\theta) \tag{1}$$

avec $0 = a_0 < a_1 < \cdots < a_k < a_{k+1} = 1$ et $\sum_{i=0}^k \pi_i = 1$. Donner la densité de la loi a posteriori de θ à une constante multiplicative près.
3) Quelle est la loi a posteriori lorsque la loi a priori est la loi beta de paramètres $(a, b) \in \mathbb{R}_+^2$

Application

On veut estimer p la proportion des étudiants qui dorment plus de 8 heures par nuit. Les observations sur un échantillon de 27 étudiants sont :

11 étudiants dorment plus de 8 heures
16 étudiants dorment moins de 8 heures.

On envisage trois lois a priori différentes sur p :

A- La loi discrète définie dans la Table 1

B- Une loi a priori de type (1) défini par

i	0	1	2	3	4	5	6	7	8	9
q_i	2	4	8	8	4	2	1	1	1	1

où

$$a_i = i/10$$

et

$$\pi_i = \frac{q_i}{\sum_{j=0}^9 q_j}$$

C- La loi Beta de paramètres $a = 3.4$ and $b = 7.4$.

i	b_i	$P(p = b_i)$
1	0.05	0.03
2	0.15	0.18
3	0.25	0.28
4	0.35	0.25
5	0.45	0.16
6	0.55	0.07
7	0.65	0.02
8	0.75	0.00
9	0.85	0.00
10	0.95	0.00

Table 1.

Commande R .

— La fonction `stepfun` retourne une fonction constante par morceaux

— La fonction `integrate` permet d'approcher une intégrale de Riemann. La syntaxe est par exemple

```
f = function(x) {x^2}
integrate(f,0,1)
```

pour approcher $\int_0^1 x^2 \, \mathrm{d}x$

4) Représenter graphiquement les trois lois a priori, calculer la moyenne et la variance de ces lois a priori.
5) Pour les trois lois a priori, représenter la densité de la loi a posteriori. Puis calculer la moyenne et la variance des trois lois a posteriori
6) Commenter les résultats obtenus

Exercice 2. Modèle Gaussien

On dispose de n observations $X_1, ..., X_n$ iid suivant une loi gaussienne $\mathcal{N}(\theta, 1)$.
On choisit comme loi a priori sur θ la loi gaussienne $\mathcal{N}(0, \tau^{-2})$, $\tau > 0$

1) Montrer que la loi a posteriori est une loi Gaussienne
$$\mathcal{N}(\frac{\bar{X}_n}{1+\tau^2/n}, \frac{1}{n+\tau^2})$$
2) Montrer que les régions HPD de niveau $1-\alpha(= .95)$ sont de la forme
$$\theta \in \left[\frac{\bar{X}_n}{1+\tau^2/n} - \frac{u_{1-\alpha/2}}{\sqrt{n+\tau^2}}; \frac{\bar{X}_n}{1+\tau^2/n} + \frac{u_{1-\alpha/2}}{\sqrt{n+\tau^2}}\right] = I^{HPD}(\tau, \bar{X}_n)$$
où u_α est le quartile d'ordre α de la loi gaussienne standard.
3) Montrer que
$$P_\theta(\theta \in I^{HPD}(\tau, \bar{X}_n)) = F\left(\frac{\theta\tau^2}{\sqrt{n}} + u_{1-\alpha/2}\sqrt{\frac{n+\tau^2}{n}}\right) - F\left(\frac{\theta\tau^2}{\sqrt{n}} - u_{1-\alpha/2}\sqrt{\frac{n+\tau^2}{n}}\right)$$
où F est la fonction de répartition de la loi gaussienne standard.
4) Quelle est la limite de cette probabilité quand $n \to \infty$. Commenter.
5) Quelle est la limite de cette probabilité quand $\tau \to 0$. Commenter.
6) Comment choisir la loi a priori pour que les régions HPD de niveau $1-\alpha$ soient aussi des régions de confiance au sens classique de niveau $1-\alpha$.