U. Dauphine/MD4/Ch. Robert 2002/2003

Statistique/Modèle linéaire

Partiel du 12 novembre 2002

Avec documents (durée deux heures). Deux exercices indépendants.

Exercice 1

Etant donnée une densité de probabilité $f(\cdot|\theta)$ paramétrée par le paramètre θ, on définit la fonction de coût entropique associée à l'estimation de θ par d comme

$$L^e(\theta, d) = \mathbb{E}_\theta \left[\log \frac{f(X|\theta)}{f(X|d)} \right] = \int \log \frac{f(x|\theta)}{f(x|d)} f(x|\theta) dx \,.$$

1. Montrer que $L^e(\theta, d)$ est toujours positive ou nulle et qu'elle ne s'annule que si $f(\cdot|\theta) = f(\cdot|d)$.

2. Donner $L^e(\theta, d)$ pour les distributions suivantes

$$\mathcal{N}(\theta, 1), \qquad \mathcal{N}(0, e^{2\theta}), \qquad \mathcal{P}(e^\theta) \,.$$

3. Dans le cas de la distribution $\mathcal{N}(0, e^{2\theta})$, si θ est muni d'une loi *a priori* de densité $\pi(\theta)$, donner la forme intégrale (ou espérance *a posteriori*) de l'estimateur de Bayes associé à π et au coût entropique $L^e(\theta, d)$.

4. Si $\mu = \exp(-2\theta)$ est muni d'une loi *a priori* gamma, de densité

$$\beta^\alpha \frac{\mu^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\beta\mu\right) \,,$$

donner la loi a priori de θ. Montrer que cette loi est conjuguée lorsque la loi des observations est $\mathcal{N}(0, e^{2\theta})$. Dans ce cas, donner la forme explicite de l'estimateur de Bayes de la question 3.

5. Examiner si le cas limite où $\alpha = \beta = 0$ est bien défini.

Exercice 2

Soit $X_1, \ldots, X_n$ un échantillon iid de loi $\mathcal{E}xp\,(\theta)$. Pour $i = 1, \ldots, n$, on n'observe que

$$Y_i = \mathbb{I}_{X_i \geq 2} = \begin{cases} 1 & \text{si } X_i \geq 2 \\ 0 & \text{sinon} \end{cases}.$$

1. Donner la loi des Y_i en fonction de θ.
2. Montrer que $\widehat{\theta} = -\dfrac{\log \overline{Y}}{2}$ converge presque sûrement vers θ.
3. Calculer la loi asymptotique de $\widehat{\theta}$ (en partant de la loi asymptotique de $\overline{Y}$).

On considère le problème de test de l'hypothèse $\theta = \theta_0$.

4. Quel est le test de Neyman Pearson (fondé sur les X_i) pour cette hypothèse ?
5. Donner les statistiques des tests de Walk et du score (fondés sur les X_i) pour cette hypothèse.
6. Même question pour les statistiques fondées sur les Y_i.
7. Pour les données x_i suivantes, comparer les résultats obtenus en utilisant ces différents tests :

0,60	0,19	6,44	1,74	0,02	2,34	4,08	0,17	1,16	1,23	1,74
0,55	0,25	3,43	1,64	5,32	3,80	1,27	0,68	2,22	2,77	2,85
2,85	3,67	0,26	2,48	4,08	1,23	0,35	5,05	3,22	0,00.	

Table 1: Températures relevées au mois de février.

Aide : La moyenne des x_i vaut $2,115$ et celle des y_i $0,4687$.