Statistique/Modèle linéaire

Partiel du 9 novembre 2004

Avec documents, durée deux heures. Trois exercices indépendants.

Exercice 1

Soient $Y_1, \ldots, Y_n$ des variables aléatoires indépendantes associées à des variables explicatives unidimensionnelles $x_1, \ldots, x_n$ telles que Y_i suit la loi (conditionnellement à x_i)

$$Y_i \sim \mathcal{P}\mathrm{oi}(\exp\{a + bx_i\}) .$$

1. Donner la vraisemblance et la matrice d'information de Fisher associée au paramètre (a, b) et à l'échantillon $(Y_1, \ldots, Y_n)$ conditionnellement à $(x_1, \ldots, x_n)$.
2. En admettant les hypothèses de régularité usuelles, donner la loi asymptotique de l'estimateur du maximum de vraisemblance de (a, b). (On ne calculera pas cet estimateur.)
3. On cherche à tester l'hypothèse nulle $H_0 : b = 0$. Donner la forme des statistiques de test de Wald, du score et du rapport de vraisemblance.
4. Montrer que la statistique de test du score est calculable explicitement et donner la décision sur H_0 au niveau $\alpha = .05$ lorsque

$$n = 25, \quad \bar{y} = 3.27, \quad \sum_i x_i(y_i - \bar{y}) = 5.52, \quad \sum x_i^2 = 23.87 .$$

5. On suppose à présent que l'estimateur du maximum de vraisemblance vaut $(\hat{a}, \hat{b}) = (1.18, -0.42)$. Si $\sum_i x_i y_i = -6.73$, montrer que la statistique du rapport de vraisemblance pour le test de $H_0 : b = 0$ vaut

$$2\left[n\bar{y}(\hat{a} - \hat{a}_0) + \hat{b}\sum_{i=1}^n x_i y_i\right]$$

et déduire en utilisant les valeurs numériques de la question précédente si H_0 est alors acceptée au niveau $\alpha = .05$.

Exercice 2

Soit $X_1, \ldots, X_n$ un échantillon de loi de densité sur $\mathbb{R}$

$$f(x|\theta) = \frac{1}{4\theta\sqrt{|x|}}\, e^{-\sqrt{|x|}/\theta}\,, \quad \theta > 0\,.$$

1. Montrer que $Y_i = \sqrt{|X_i|}$ suit une loi exponentielle de moyenne θ.
2. En déduire l'estimateur du maximum de vraisemblance de θ et l'information de Fisher pour ce modèle.
3. On cherche à tester l'hypothèse $H_0 : \theta = 1$. Donner les statistiques de test de Wald et du rapport de vraisemblance.
4. On se place à présent dans une perspective bayésienne en supposant que θ suit la loi a priori $\pi(\theta) = \theta^{-3} \exp(-2/\theta)$, soit une loi $\mathcal{IG}a(2, 2)$. Donner la loi a posteriori associée à l'échantillon $(x_1, \ldots, x_n)$ et en déduire la moyenne a posteriori correspondante.
5. Donner la loi marginale de $(x_1, \ldots, x_n)$ associée à cette loi et en déduire la forme du facteur de Bayes pour le test de $H_0 : \theta = 1$.

Exercice 3

On considère deux échantillons normaux de taille n,

$$\begin{aligned} x_{11}, \ldots, x_{1n} &\sim \mathcal{N}(\mu_1, \sigma^2), \\ x_{21}, \ldots, x_{2n} &\sim \mathcal{N}(\mu_2, \sigma^2)\,. \end{aligned}$$

a. On s'intéresse d'abord au paramètre $\xi = (\xi_1, \xi_2) = (\mu_1/\sigma, \mu_2/\sigma)$ et la loi a priori associée est impropre,

$$\pi(\mu_1, \mu_2, \sigma) = \sigma^{-p}.$$

Donner (sous forme intégrale) la densité de la loi a posteriori **marginale** $\pi(\xi|x)$ et montrer qu'elle ne dépend que de $z = (z_1, z_2) = (\bar{x}_1/s, \bar{x}_2/s)$, si $\bar{x}_1, \bar{x}_2$ sont les moyennes empiriques des deux échantillons et

$$s^2 = \sum_{i=1}^{n} \{(x_{1i} - \bar{x}_1)^2 + (x_{2i} - \bar{x}_2)^2\}\,.$$

b. Calculer la densité de la loi marginale de z, $f(z|\theta)$, et montrer qu'elle ne dépend que de ξ.