Statistique/Modèle linéaire

Corrigé commenté du partiel du 12 novembre 2002

Exercice 1

1. *Non, il n'est pas vrai que $f(x|\theta) > f(x|d)$ pour tout x !* Par contre, comme la fonction log est concave, par l'inégalité de Jensen,
$$\mathbb{E}_\theta\left[-\log h(X)\right] \geq -\log \mathbb{E}_\theta\left[h(X)\right] ,$$
et
$$\begin{aligned}\mathbb{E}_\theta\left[\log\frac{f(X|\theta)}{f(X|d)}\right] = \mathbb{E}_\theta\left[-\log\frac{f(X|d)}{f(X|\theta)}\right] &\geq -\log\mathbb{E}_\theta\left[\frac{f(X|d)}{f(X|\theta)}\right]\\ &= -\log\int\frac{f(x|d)}{f(x|\theta)}f(x|\theta)dx = 0\end{aligned}$$
L'égalité n'a lieu que si la variable aléatoire $f(X|d)/f(X|\theta)$ est presque sûrement constante, car log est strictement concave. Donc seulement si $f(\cdot|\theta) > f(\cdot|d)$.

2. Pour la loi $\mathcal{N}(\theta,1)$, $f(x|\theta)\propto \exp-(x-\theta)^2/2$ et
$$L^e(\theta,d) = \mathbb{E}_\theta\left[(X-d)^2-(X-\theta)^2\right]/2 = (\theta-d)^2/2\,.$$
Pour la loi $\mathcal{N}(0,e^{2\theta})$, $f(x|\theta)\propto e^{-\theta}\exp -x^2/2e^{2\theta}$, donc $\log f(x|\theta) = -\theta - e^{-2\theta}\,x^2/2 + c$ et
$$\begin{aligned}L^e(\theta,d) &= \mathbb{E}_\theta\left[d-\theta-(X^2/2)\left\{e^{-2\theta}-e^{-2d}\right\}\right]\\ &= d-\theta-e^{2\theta}\left\{e^{-2\theta}-e^{-2d}\right\}/2\\ &= d-\theta+\left\{e^{2(\theta-d)}-1\right\}/2\end{aligned}$$
Pour la loi $\mathcal{P}(e^\theta)$, $f(x|\theta) = e^{-e^\theta}\,e^{x\theta}/x!$, donc
$$L^e(\theta,d) = \mathbb{E}_\theta\left[e^{-d}-e^\theta+X(\theta-d)\right] = e^d - e^\theta + e^\theta(\theta-d)$$

3. Pour la distribution $\mathcal{N}(0,e^{2\theta})$, le coût à minimiser (en d) est
$$\mathbb{E}^\pi\left[d-\theta+\left\{e^{2(\theta-d)}-1\right\}/2|x\right] ,$$
donc en dérivant par rapport à d,
$$1-\mathbb{E}^\pi\left[e^{2(\theta-d)}|x\right] = 0$$
soit encore
$$d^\star = \log\left\{\mathbb{E}^\pi\left[e^{2\theta}|x\right]\right\}/2$$

4. Si $\mu=\exp(-2\theta)\sim\beta^\alpha\,\frac{\mu^{\alpha-1}}{\Gamma(\alpha)}\,\exp\left(-\beta\mu\right)$, la formule du Jacobien donne
$$\begin{aligned}\theta &\sim \beta^\alpha\,\frac{\exp\{-2(\alpha-1)\theta\}}{\Gamma(\alpha)}\,\exp\left\{-\beta\exp(-2\theta)\right\}\,2\exp\{-2\theta\}\\ &= 2\beta^\alpha\,\frac{\exp\{-2\alpha\theta\}}{\Gamma(\alpha)}\,\exp\left\{-\beta\exp(-2\theta)\right\}\end{aligned}$$
Cette loi est conjuguée si $X\sim\mathcal{N}(0,e^{2\theta})$ puisque
$$\pi(\theta|x)\propto\exp\{-2\alpha\theta\}\,\exp\left\{-\beta\exp(-2\theta)\right\}\times\exp\{-\theta\}\,\exp\left\{-(x^2/2)\exp(-2\theta)\right\} ;$$
α est donc changé en $\alpha+(1/2)$ et β en $\beta+(x^2/2)$. On retrouve simplement le résultat connu que, si $1/\mu\sim\mathcal{G}(\alpha,\beta)$,
$$1/\mu|x\sim\mathcal{G}(\alpha+1/2,\beta+(x^2/2))\,.$$
Alors
$$\delta^\pi(x) = \frac{1}{2}\,\log\mathbb{E}^\pi\left[e^{2\theta}|x\right] = \frac{1}{2}\,\log\mathbb{E}^\pi\left[\frac{1}{\mu}\middle|\,x\right] = \frac{1}{2}\,\log\frac{\alpha+1/2}{\beta+(x^2/2)} = \frac{1}{2}\,\log\frac{2\alpha+1}{2\beta+x}$$

5. Si $\alpha=\beta=0$, δ^π est bien défini pour $x>0$, à la fois dans la formule ci-dessus et comme estimateur associé à la loi a posteriori $\mathcal{G}(1/2,x^2/2)$.

Exercice 2

1. *La loi exponentielle est une loi sur $\mathbb{R}_+$, pas sur $\mathbb{N}$. Et la loi de Y a comme support $\{0,1\}$...* Il vient $Y_i\sim B\left(p\right)$, où
$$p = P\left(X_i\geq 2\right) = \int_2^\infty \theta\;e^{-\theta x}\;dx = \left.-e^{-\theta x}\right|_2^\infty = e^{-2\theta},$$

2. $\widehat{\theta} = -\dfrac{\log \overline{Y}}{2}$ converge p.s. vers $-\dfrac{\log \mathbb{E}(Y_i)}{2}$. Donc, comme la transformation est régulière,

$$\widehat{\theta} \overset{n\to\infty}{\longrightarrow} -\frac{\log p}{2} = -\frac{\log e^{-2\theta}}{2} = \theta. \qquad (p.s.)$$

3. Par le TCL,

$$\sqrt{n}\left(\overline{Y} - p\right) \overset{n\to\infty}{\approx} \mathcal{N}\left(0, p\left(1-p\right)\right).$$

Posons $g(x) = -\log(x)/2$. Alors, par le théroème de Slutsky,

$$\begin{aligned} & \sqrt{n}\left(g\left(\overline{Y}\right) - g\left(p\right)\right) \overset{n\to\infty}{\approx} \mathcal{N}\left(0, p\left(1-p\right) g'^2\left(p\right)\right) \\ \Leftrightarrow \quad & \sqrt{n}\left(\widehat{\theta} - \theta\right) \overset{n\to\infty}{\approx} \mathcal{N}\left(0, \frac{p\left(1-p\right)}{4p^2}\right) \\ \Leftrightarrow \quad & \sqrt{n}\left(\widehat{\theta} - \theta\right) \overset{n\to\infty}{\approx} \mathcal{N}\left(0, \frac{e^{2\theta}-1}{4}\right) \qquad \left[p = e^{-2\theta}\right] \end{aligned}$$

4. Dans le cas où l'alternative est $\theta = \theta_1$, le test de Neyman-Pearson est fondé sur le rapport des vraisemblances en θ_0 et en θ_1, qui ne dépend que de

$$\exp\left\{(\theta_1 - \theta_0) n\overline{x}\right\}.$$

Donc, si $\theta_1 > \theta_0$, ce test rejette H_0 pour $\overline{x} > k$, tandis que, pour $\theta_1 < \theta_0$, ce test rejette H_0 pour $\overline{x} < k'$. Il n'y a donc pas de test de Neyman-Pearson pour l'ensemble des alternatives possibles. On prendra à la place un test de la forme

$$\varphi(\overline{x}) = \begin{cases} 1 & \text{si } \overline{x} \in [k', k] \\ 0 & \text{sinon} \end{cases}$$

Le problème étant la détermination des bornes, on simplifie en prenant (k, k') tel que

$$k^{n-1} e^{-\theta_0 k} = k'^{n-1} e^{-\theta_0 k'},$$

avec $\Pr_{\theta_0}(\overline{X} \in [k', k]) = \alpha$.

5. Comme $\tilde{\theta} = 1/\overline{X}$ est l'emv dans ce cas, que l'information de Fisher est $1/\theta^2$ pour la loi exponentielle, on a l'approximation

$$\sqrt{n}\,\frac{\tilde{\theta} - \theta}{\theta} \approx \mathcal{N}(0,1)$$

et la statistique de Wald vaut

$$\xi_n^W = n\,\frac{(\tilde{\theta} - \theta_0)^2}{\tilde{\theta}^2} = n\overline{X}^2\left\{(1/\overline{X}) - \theta_0\right\}^2 \sim \chi_1^2 \qquad \text{sous } H_0.$$

6. Suivant le résultat de la question 3,

$$\sqrt{n}\,\sqrt{\frac{4}{e^{2\theta}-1}}\left(\widehat{\theta} - \theta\right) \overset{n\to\infty}{\approx} \mathcal{N}\left(0, 1\right),$$

et la statistique de Wald vaut

$$\xi_n^W = \frac{4n}{e^{2\theta}-1}(\widehat{\theta} - \theta)^2 = \frac{4n}{(1/\overline{Y}) - 1}\left\{-\frac{\log \overline{Y}}{2} - \theta_0\right\}^2$$

7. On accepte H_0 si

$$n\overline{x}^2((1/\overline{x}) - \theta_0)^2 = 143.143\,(\theta_0 - 0.473)^2 < 3.84$$

pour les X_i, et si

$$\frac{4n}{(1/\overline{y}) - 1}(\widehat{\theta} - \theta_0)^2 = 112.601\,(\theta_0 - 0.380)^2 < 3.84,$$

pour les Y_i. Les réponses au test pourront donc différer suivant les valeurs de θ_0, comme le montre le graphe ci-dessous (où la courbe marron correspond aux Y_i et celle en noir aux X_i).