1 Exercices

Exercice 1. Nous considérons la suite d'expériences statistiques

$$\left(\{0,1\}^n, \mathscr{P}(\{0,1\}^n), \left\{\mathbb{P}_{n,\theta} = \bigotimes_{i=1}^n \mathrm{Ber}\left(\varphi(\theta^T \mathbf{x}_i)\right) \,:\, \theta \in \Theta = \mathbb{R}^p\right\}\right)$$

où $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ sont des variables explicatives quantitatives ($\mathbf{x}_i \in \mathbb{R}^p$) et

$$\varphi(t) = (1 + \mathrm{e}^{-t})^{-1} \, .$$

Nous suppose que, pour tout $n \geq p$, la matrice $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ est de rang p. On note $\mathrm{L}_n(\theta)$ la fonction de log-vraisemblance des observations $(Y_1, \cdots, Y_n)$.

1. Déterminer $\theta \mapsto \mathrm{L}_n(\theta)$.
2. Montrer que

$$\nabla \mathrm{L}_n(\theta) = \sum_{i=1}^n \{Y_i - p_i(\theta)\} \mathbf{x}_i$$
$$-\nabla^2 \mathrm{L}_n(\theta) = \sum_{i=1}^n h(\theta^T \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T \qquad \text{où } h = \varphi(1-\varphi)$$

   En déduire que la fonction $\theta \mapsto \mathrm{L}_n(\theta)$ est strictement concave.
3. Montrer que l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est défini comme la solution d'un système d'équations que l'on précisera.
4. Proposer un algorithme pour déterminer l'estimateur du maximum de vraisemblance.

On souhaite établir la normalité asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta)$ sous $\mathbb{P}_\theta$. On fait les hypothèses suivantes :

— La suite $\{\hat{\theta}_n,\, n \in \mathbb{N}\}$ est consistante.

— Pour tout $\theta \in \Theta$ il existe une matrice Q_θ inversible telle que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n h(\theta^T \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T = Q_\theta$$

— $\sup_{n\geq 0} n^{-1} \sum_{i=1}^n \|\mathbf{x}_i\|^3 < \infty$.

5. Montrer que

$$\left(\frac{\nabla^2 \mathrm{L}_n(\theta)}{n} + R_n\right) \sqrt{n}(\hat{\theta}_n - \theta) = -\frac{\nabla \mathrm{L}_n(\theta)}{\sqrt{n}}$$

   où pour tout $\theta \in \Theta$ et $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}_{n,\theta}(\|R_n\| \geq \varepsilon) = 0$.
6. Montrer que $n^{-1/2} \nabla \mathrm{L}_n(\theta)$ est asymptotiquement normal.
7. Conclure que $\sqrt{n}(\hat{\theta}_n - \theta) \overset{\mathbb{P}_{n,\theta}}{\Longrightarrow} \mathrm{N}(0, Q_\theta^{-1})\,.$

8. On note $\beta_{n,k}$ le $k$ème coefficient de la diagonale de $(-\nabla^2 \mathrm{L}_n(\hat{\theta}_n)/n)^{-1}$. Montrer que
$$\sqrt{\frac{n}{\beta_{n,k}}}\,(\hat{\theta}_{n,k}-\theta_k) \stackrel{\mathbb{P}_\theta}{\Longrightarrow} \mathrm{N}(0,1)\,.$$
9. Construire un intervalle de confiance asymptotique pour le paramètre θ_k pour $k \neq p$.
10. En déduire un test de niveau asymptotique α pour $H_0 \ : \ \theta_k = 0$ contre $H_0 \ : \ \theta_k \neq 0$.
11. Calculer la p-valeur asymptotique de ce test.

Exercice 2. Soit $\{X_k,\ k \in \mathbb{N}\}$ un échantillon de loi de densité
$$p(\theta,x) = \begin{cases} (\theta_1+\theta_2)^{-1}\mathrm{e}^{-x/\theta_1} & x>0 \\ (\theta_1+\theta_2)^{-1}\mathrm{e}^{x/\theta_2} & x\leq 0 \end{cases} \quad \theta=(\theta_1,\theta_2)\in\Theta := \mathbb{R}_*^+ \times \mathbb{R}_*^+ .$$
On considère le test d'hypothèses
$$H_0 :\ \theta_1 = \theta_2, \qquad \text{contre} \qquad H_1 :\ \theta_1 \neq \theta_2$$
1. Construire le test de Wald.
2. Construire le test de Rao.
3. Construire le test du rapport de vraisemblance généralisé.

Exercice 3. On rappelle qu'une loi Beta(α,β), $(\alpha,\beta)\in\mathbb{R}_+^*\times\mathbb{R}_+^*$ est une loi de densité sur $[0,1]$ donnée par
$$f(\alpha,\beta,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}_{]0,1[}(x)\,.$$
Soit $(X_1,\ldots,X_n)$ des variables aléatoires indépendantes distribuées suivant une loi Beta de paramètre $(\mu,1)$ et $(Y_1,\ldots,Y_m)$ des variables aléatoires indépendantes de loi Beta $(\nu,1)$. On suppose de plus que $(X_1,\ldots,X_n)$ et $(Y_1,\ldots,Y_m)$ sont indépendantes. On effectue le test
$$H_0 :\ \mu = \nu, \qquad \text{contre} \qquad H_1 :\ \mu \neq \nu\,.$$
1. Déterminer le test de rapport de vraisemblance généralisé.
2. Montrer que le test du rapport de vraisemblance généralisé est équivalent au test de fonction critique
$$\phi(X_1,\ldots,X_n,Y_1,\ldots,Y_m) = \mathbb{1}\{T_{n,m}\notin[c_1,c_2]\}$$
où $c_1 < c_2$ et
$$T_{n,m} := \frac{\sum_{i=1}^n \log(X_i)}{\sum_{i=1}^n \log(X_i) + \sum_{i=1}^m \log(Y_i)}$$
3. Montrer que $-\log(X_i)$ et $-\log(Y_i)$ sont distribuées suivant des lois exponentielles et donc que $-\sum_{i=1}^n \log(X_i)$ et $-\sum_{j=1}^m \log(Y_j)$ sont distribuées suivant des lois Γ que l'on déterminera. Déterminer la distribution sous l'hypothèse nulle de la statistique de test, puis déterminer c_1 et c_2 permettant d'obtenir un test de niveau α pour $\alpha\in]0,1[$.
4. Supposons que $m=m_n$ est une suite telle que $\lim_{n\to\infty} m_n/n = \gamma > 0$ et $\gamma - m_n/n = o(n^{-1/2})$. Montrer que la suite $\{T_{n,m_n}, n\in\mathbb{N}\}$ admet une limite λ que l'on déterminera puis déterminer la distribution limite de $\sqrt{n}(T_{n,m_n}-\lambda)$. Proposer une construction du test de niveau asymptotique α pour $\alpha\in]0,1[$.