### Devoir maison – Modèles exponentiels

On considère un modèle paramétrique $\mathcal{M} = ((\mathbb{R}^{p\times n}), \mathcal{B}(\mathbb{R}^{p\times n}), (\mathbb{P}_\theta)_{\theta\in\Theta})$ avec $\Theta \subseteq \mathbb{R}^d$, où l'on observe $X = (X_1, \dots, X_n)$ avec $X_i$ vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^p$. Pour tout $\theta \in \Theta$, sous $\mathbb{P}_\theta$, $X_1$ a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^p$ donnée par :

$$f_\theta : x \mapsto b(x) \exp\left(\langle T(x), \theta\rangle - A(\theta)\right),$$

où $b : \mathbb{R}^p \to \mathbb{R}_+$ est mesurable et non nulle, $T : \mathbb{R}^p \to \mathbb{R}^d$ est mesurable, et $A : \Theta \to \mathbb{R}$.

On suppose que $b, T$ et $A$ sont des fonctions connues, et on note $T(x) = (T_1(x), \dots, T_d(x))$. On fait les hypothèses supplémentaires suivantes :

(H1) $\Theta$ est un ouvert convexe de $\mathbb{R}^d$,

(H2) Pour tout $\theta \in \Theta$, il existe un voisinage $U$ de $\theta$ dans $\mathbb{R}^p$, et $G_\theta, H_\theta : \mathbb{R}^p \to \mathbb{R}_+$ intégrables telles que pour tous $\theta' \in U$, $1 \leq i, j \leq d$, et $x \in \mathbb{R}^p$,

$$b(x)|T_i(x)| \exp\left(\langle T(x), \theta'\rangle\right) \leq G_\theta(x) \quad \text{et} \quad b(x)|T_i(x)T_j(x)| \exp\left(\langle T(x), \theta'\rangle\right) \leq H_\theta(x).$$

Une telle famille de lois $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est appelée *famille exponentielle*, et l'on parle de *modèle exponentiel.*

1. Montrer que les modèles suivants sont des modèles exponentiels, pour lesquels on précisera $b, T$ et $A$:

   1.(a). $p = 1$, $X_1 \sim \text{Exp}(\lambda)$, $\lambda > 0$.
   1.(b). $p = 1$, $X_1 \sim (\mathcal{U}([0, 1]))^{1/\alpha}$, $\alpha > 0$.
   1.(c). $p$ général, $X_1 \sim \mathcal{N}(\mu, I_p)$.

2. Montrer que pour tout $\theta \in \Theta$,

$$A(\theta) = \log \int_{\mathbb{R}^p} b(x) \exp\left(\langle T(x), \theta\rangle\right) dx.$$

3. En utilisant (H2), montrer que $T(X_1)$ a un moment d'ordre 2 fini sous tout $\mathbb{P}_\theta$, et que $A$ est de classe $C^2$ sur $\Theta$ avec

$$\mathbb{E}_\theta[T(X_1)] = \nabla A(\theta) \quad \text{et} \quad \text{Var}_\theta(T(X_1)) = \nabla^2 A(\theta),$$

   où $\nabla A$ et $\nabla^2 A$ désignent respectivement le gradient et la hessienne de $A$.

On dit qu'un vecteur aléatoire $Z$ à valeur dans $\mathbb{R}^p$ est de *rang plein* si pour tous $\theta \in \Theta$, $\lambda_1, \dots, \lambda_d \in \mathbb{R}$,

$$(\lambda_1 Z_1 + \ldots + \lambda_d Z_d = 0 \ \ \mathbb{P}_\theta - \text{p.s.}) \implies \lambda_1 = \ldots = \lambda_d = 0.$$

4. Montrer que $\nabla^2 A(\theta)$ est symétrique définie positive pour tout $\theta \in \Theta$ si et seulement si $T(X_1)$ est de rang plein.

Dans toute la suite on suppose $T(X_1)$ de rang plein. Par la question précédente, $A$ est donc de Hessienne toujours symétrique définie positive. On admet[1] que cela implique que $\nabla A : \Theta \to \nabla A(\Theta)$ est un $C^1$−difféomorphisme (i.e., une bijection $C^1$ de réciproque $C^1$).

5. Justifier que $\mathcal{M}$−p.s., pour $n$ assez grand, $\frac{1}{n}\sum_{i=1}^n T(X_i)$ appartient à $\nabla A(\Theta)$.

[1] On le démontre simplement en utilisant la stricte convexité de $A$ et le théorème d'inversion locale.

6. En utilisant ce qui précède, montrer que $\mathcal{M}-$p.s., pour $n$ assez grand, l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ est unique et qu'il est donné par

$$\widehat{\theta}_n = [\nabla A]^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right).$$

7. Montrer que $\widehat{\theta}_n$ est un estimateur fortement consistant de $\theta$.

On admet que sous nos hypothèses, le modèle $\mathcal{M}$ est régulier.

8. Montrer que l'information de Fisher du modèle à une seule observation s'écrit $I_1(\theta) = \nabla^2 A(\theta)$.

On rappelle le résultat de calcul différentiel suivant. Pour $U, V$ ouverts de $\mathbb{R}^d$, si $F : U \to V$ est un $C^1-$difféomorphisme, alors la matrice jacobienne de $F^{-1}$ est donnée pour tout $v \in V$ par

$$\forall u \in A(\Theta),\ \mathrm{Jac}(F^{-1})(u) = (\mathrm{Jac}(F)(F^{-1}(u)))^{-1}.$$

9. Montrer que pour tout $\theta \in \Theta$, sous $\mathbb{P}_\theta$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, (I_1(\theta))^{-1}).$$

*Le bilan de cette étude est le suivant. Dans un modèle exponentiel dont la statistique $T(X_1)$ est de rang plein, l'estimateur du maximum de vraisemblance est consistant et asymptotiquement efficace. En vérifiant que $\sum_{i=1}^n T(X_i)$ est suffisante et complète, on peut aussi montrer qu'il est UVMB pour $\theta$.*