Exercices, 5

EXERCICE 1 (exemples d'EMV). — Calculer l'estimateur du maximum de vraisemblance et étudier ses propriétés (lien avec un estimateur par moments, risque quadratique, loi limite et intervalle de confiance asymptotique) dans les cas suivants :

1. On observe $X_1, \dots, X_n$ i.i.d. de loi de Poisson de paramètre $\lambda > 0$.
2. On observe $X \sim \mathcal{B}(n, p)$, où $p \in [0, 1]$. Pour la loi limite et l'intervalle de confiance, on considérera l'hypothèse supplémentaire $p \in]0, 1[$.
3. On observe $X_1, \dots, X_n$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma > 0$.

EXERCICE 2. — On considère l'expérience statistique

$$\mathcal{E} = (\{0, 1\}, \mathcal{P}(\{0, 1\}), (\mathbb{P}_\theta)_{\theta \in \{\frac{1}{10}, \frac{8}{10}\}}),$$

$$\text{où } \mathbb{P}_\theta(dx) = \theta \delta_{\{0\}}(dx) + (1 - \theta) \delta_{\{1\}}(dx).$$

1. Montrer que le modèle est dominé, et calculer sa fonction de vraisemblance.
2. En déduire l'estimateur du maximum de vraisemblance.

EXERCICE 3 (loi Gamma). — On rappelle que la loi Gamma $\Gamma(\alpha, \beta)$, $\alpha, \beta > 0$ est une v.a. réelle de densité

$$x \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0.$$

On se donne un échantillon $(X_1, \dots, X_n)$ de loi $\Gamma(\alpha, \beta)$.

1. On suppose le paramètre β connu. Proposer un estimateur de α par la méthode des moments.
2. On suppose à présent que les deux paramètres α, β sont inconnus. Proposer un estimateur de (α, β) par la méthode des moments.
3. Toujours dans le cas où α, β sont inconnus, donner le système d'équation que satisfont les estimateurs de (α, β) par la méthode du maximum de vraisemblance.

EXERCICE 4. — Soit $(X_1, \dots, X_n)$ un n -échantillon de la loi uniforme sur $[-\theta, \theta]$, avec $\theta > 0$.

1. Décrire le modèle statistique associé.
2. Proposer un estimateur $\hat{\theta}_n$ de θ obtenu par méthode des moments. Est-il consistant? Proposer un intervalle de confiance asymptotique de niveau de confiance α .
3. Soit T_n l'estimateur du maximum de vraisemblance de θ . Montrer que pour tout réel t ,

$$\mathbb{P}_\theta^n (n(T_n - \theta) \leq t) \rightarrow e^{t/\theta} \mathbb{1}_{t \leq 0} + \mathbb{1}_{t > 0}$$

quand n tend vers l'infini. En déduire un intervalle de confiance asymptotique de niveau α .

4. Comparer les estimateurs $\hat{\theta}_n$ et T_n sur la base des longueurs moyennes des intervalles de confiance asymptotiques associés.

EXERCICE 5. — Dans une urne contenant 1000 tickets, 20 sont marqués θ et 980 sont marqués 10θ , où θ est un réel strictement positif inconnu.

1. On tire un unique ticket de valeur X . Donner le modèle. Est-il dominé par une mesure σ -finie? Donner un estimateur qui s'apparenterait à un maximum de vraisemblance $\hat{\theta}$ de θ (maximiser $P_\theta(\{X\})$). Montrer que $\mathbb{P}(\hat{\theta} = \theta) \geq 0,98$.
2. On renumérote les tickets marqués 10θ par $a_i\theta$, $1 \leq i \leq 980$, où les a_i sont des réels connus, deux-à-deux distincts, et compris dans l'intervalle $[10; 10, 1]$. Donner le nouvel estimateur du maximum de vraisemblance $\tilde{\theta}$ et montrer que $\mathbb{P}(\tilde{\theta} < 10\theta) = 0,02$. Commenter.

EXERCICE 6 (Weibull). — Soit $c > 0$ un paramètre fixé connu. On considère la loi de Weibull $P_\lambda, \lambda > 0$, de densité

$$\lambda c x^{c-1} e^{-\lambda x^c} \mathbb{1}_{]0, +\infty[}(x)$$

par rapport à la mesure de Lebesgue. On observe un n -échantillon de la loi P_λ , avec $n \geq 3$.

1. Calculer l'estimateur du maximum de vraisemblance $\hat{\lambda}_n$ de λ .
2. Calculer son risque quadratique $\mathbb{E}_\lambda^n [(\hat{\lambda}_n - \lambda)^2]$ où P_λ^n désigne la loi du n -échantillon.

EXERCICE 7. — On considère pour $\mu > 0$, $\alpha > 0$, la loi de probabilité $\mathbb{P}_{\mu, \alpha}$ dont la fonction de répartition est continue sur $\mathbb{R}$ et vaut : $\mathbb{P}_{\mu, \alpha}(X \leq x) = 1 - Cx^{-\alpha}$ si $x \geq \mu$ et 0 sinon. On se donne un n -échantillon $X_1, \dots, X_n$ i.i.d de loi $\mathbb{P}_{\mu, \alpha}$.

1. Calculer la constante C .
2. Calculer l'estimateur du maximum de vraisemblance $(\hat{\mu}_n, \hat{\alpha}_n)$ de (μ, α) .
3. Quelle est la loi limite de $n(\hat{\mu}_n - \mu)$?
4. Montrer que

$$\sqrt{n} \left(\hat{\alpha}_n^{-1} - \frac{1}{n} \sum_{i=1}^n \ln(X_i/\mu) \right) \xrightarrow{\mathbb{P}} 0,$$

puis en déduire la loi limite de $\sqrt{n}(\hat{\alpha}_n - \alpha)$.

EXERCICE 8. — * - **Un peu plus qu'identifiabilité implique consistance**

Soit $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ un modèle. On suppose que Θ est un sous-ensemble compact de $\mathbb{R}^d$, et que la métrique induite sur Θ par la distance en variation totale est équivalente à la métrique Euclidienne. Pour être plus précis, on suppose

$$H1: \forall \varepsilon > 0 \exists \delta > 0 \quad \forall u, v \in \Theta \quad \|u - v\| \leq \delta \Rightarrow d_{TV}(u, v) \leq \varepsilon,$$

$$H2: \forall \varepsilon > 0 \exists \delta > 0 \quad \forall u, v \in \Theta \quad d_{TV}(u, v) \leq \delta \Rightarrow \|u - v\| \leq \varepsilon,$$

avec $d_{TV}(u, v) = TV(P_u, P_v)$.

1. Montrer que d_{TV} définit bien une distance sur Θ . En déduire que le modèle est identifiable.
2. Soit $\varepsilon > 0$. Montrer qu'il existe $N \geq 1$ et $\theta_1, \dots, \theta_N \in \Theta$ tels que $\Theta \subset \bigcup_{i=1}^N \mathcal{B}_{TV}(\theta_i, \varepsilon)$. On notera $N(\varepsilon)$ le plus petit N nécessaire pour recouvrir Θ , et les $\theta_1, \dots, \theta_{N(\varepsilon)}$ correspondant.
3. Pour $i, j \in [1, N(\varepsilon)]$, on définit $A_{i,j} \in \mathcal{A}$ via $TV(P_{\theta_i}, P_{\theta_j}) = |P_{\theta_i}(A_{i,j}) - P_{\theta_j}(A_{i,j})|$. Soit $\theta \in \Theta$, on définit

$$\varphi : \begin{cases} [1, N(\varepsilon)] & \rightarrow & [0, 1] \\ i & \mapsto & \max_{j=1, \dots, N(\varepsilon)} |P_{\theta_i}(A_{i,j}) - P_\theta(A_{i,j})|. \end{cases}$$

Montrer qu'il existe $j_0 \in [1, N(\varepsilon)]$ tel que $\varphi(j_0) \leq \varepsilon$. Montrer que si $d_{TV}(\theta_j, \theta) > 6\varepsilon$, alors $\varphi(j) > 4\varepsilon$.

4. Soient $X_1, \dots, X_n$ i.i.d. $\sim P_\theta$. On définit $\varphi_n : i \mapsto \max_{j=1, \dots, N(\varepsilon)} |P_{\theta_i}(A_{i,j}) - P_n(A_{i,j})|$, où P_n est la mesure empirique $\frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. Montrer que

$$\mathbb{P}(\|\varphi - \varphi_n\|_\infty > \varepsilon) \leq \frac{N(\varepsilon)^2}{4n\varepsilon^2}.$$

5. On note $T_{n,\varepsilon}$ un estimateur de la forme $\theta_{\hat{i}}$, où $\hat{i} \in \arg \min_{j \in [1, N(\varepsilon)]} \varphi_n(j)$. Montrer que

$$\mathbb{P}(d_{TV}(T_{n,\varepsilon}, \theta) > 6\varepsilon) \leq \frac{N(\varepsilon)^2}{4n\varepsilon^2}.$$

6. Montrer qu'il existe $\varepsilon_n \rightarrow 0$ tel que $\frac{N(\varepsilon_n)^2}{n\varepsilon_n^2} \rightarrow 0$. Une manière de faire peut être de poser $f(\varepsilon) = (N(\varepsilon)/\varepsilon)^2$, de remarquer que f est strictement décroissante, et de définir $f^{-1}(u) = \sup\{t > 0 \mid f(t) \geq u\}$.

7. En déduire la construction d'un estimateur $\hat{\theta}_n$ tel que

$$d_{TV}(\hat{\theta}_n, \theta) \xrightarrow{\mathbb{P}} 0,$$

et conclure quant à la consistance de $\hat{\theta}_n$.