A. Dalalyan

TD7 : Estimateur du maximum de vraisemblance

Exercice 1. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. pouvant prendre les valeurs 0, 1, 2 avec probabilités $\theta/2$, $\theta/2$, $1-\theta$. Dans cet exercice, on note N_0, N_1 et N_2 le nombre de 0, de 1 et de 2 dans l'échantillon, c'est-à-dire,

$$N_a = \sum_{i=1}^{n} \mathbb{1}(X_i = a). \tag{1}$$

1. Dans quel intervalle de $\mathbb{R}$ varie θ ?
2. Proposer un estimateur $\widehat{\theta}_n$ de θ par la méthode du maximum de vraisemblance.
3. Calculer le risque quadratique de $\widehat{\theta}_n$.
4. Calculer l'information de Fisher de ce modèle. Ce dernier, est-il régulier ?

Exercice 2. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi de Cauchy $\mathcal{C}(\theta^*)$ avec $\theta^* > 0$. C'est-à-dire, chaque X_i a pour densité par rapport à la mesure de Lebesgue la fonction

$$f(x, \theta^*) = \frac{\theta^*}{\pi((\theta^*)^2 + x^2)}, \qquad \forall x \in \mathbb{R}. \tag{2}$$

1. Montrer que ce modèle est régulier.
2. Calculer la dérivée de la log-vraisemblance $\ell_n(\theta)$ et montrer qu'elle s'annule en un seul point, si l'événement $\{X_1 = \ldots = X_n = 0\}^c$ est réalisé.
3. En déduire qu'en dehors d'un événement de probabilité nulle, l'EMV existe, est unique et est asymptotiquement efficace.

Exercice 3. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi $\mathbf{P}_{\theta^*}$, où $\mathbf{P}_\theta$ est la loi sur $\mathbb{R}$ admettant pour densité

$$f(x, \theta) = \mathbb{1}_{[\theta, \theta+1]}(x), \qquad x \in \mathbb{R}.$$

1. Le modèle statistique $\{\mathbf{P}_\theta : \theta \in \mathbb{R}\}$ est-il identifiable ?
2. Le modèle statistique $\{\mathbf{P}_\theta : \theta \in \mathbb{R}\}$ est-il régulier ?
3. L'estimateur du maximum de vraisemblance du paramètre θ^* est-il unique ? Déterminer l'ensemble de points de maximum de la vraisemblance de ce modèle.
4. On considère l'estimateur $\widehat{\theta}_n = X_{(1)} = \min_{1 \le i \le n} X_i$. Sans calculer le biais, montrer que cet estimateur est biaisé.
5. Déterminer la densité $g_n(x; \theta^*)$ de la variable aléatoire $\widehat{\theta}_n$.
6. Calculer le risque de $\widehat{\theta}_n$. Proposer un estimateur $\bar{\theta}_n$ qui a exactement la même variance mais qui est sans biais. Comparer les deux risques.

7. Montrer que $\widehat{\theta}_n$ est faiblement consistant. En déduire qu'il est également fortement consistant.
8. En calculant la fonction de répartition de la variable aléatoire $Y_n = n(\widehat{\theta}_n - \theta^*)$, montrer que cette suite converge en loi vers une loi exponentielle.
9. (Facultatif) On peut également considérer l'estimateur $\widetilde{\theta}_n = \max_i X_i - 1$. Dire, sans faire des calculs, lequel des deux estimateurs $\widehat{\theta}_n$ et $\widetilde{\theta}_n$ est préférable ?