Feuille de travaux dirigés 2 : Estimation ponctuelle

Exercice 1 (Maximum de vraisemblance pour le modèle linéaire simple : rappel du td précédent)**:**
Pour tout $i = 1, \dots, n$, on considère

$$X_i = a_i \theta_1 + Z_i \tag{1}$$

où les Z_i sont des variables aléatoires indépendantes, de même loi $\mathcal{N}(0, \sigma^2)$. Les coefficients a_i sont des variables déterministes connues (et non toutes nulles). Les paramètres θ_1 et σ^2 sont inconnus. On observe $X = (X_1, \dots, X_n)^\top$. On paramètre le modèle pour X par $\theta = (\theta_1, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

1. Donner la densité p_θ de X par rapport à la mesure de Lebesgue en fonction de $a = (a_1, \dots, a_n)^\top$ et de σ^2.
2. Exprimer l'estimateur $\widehat{\theta}$ du maximum de vraisemblance pour le paramètre θ.
3. Montrer que $\widehat{\theta}_1$ (la première coordonnée de $\widehat{\theta}$) est non biaisé.

Exercice 2 (Moindres carrés dans le modèle linéaire simple)**:**
On reprend l'exemple de l'exercice ci-dessus.

1. Calculez l'estimateur de θ_1 donné par la méthode des moindres carrés.
2. Comparez à l'estimateur du maximum de vraisemblance de l'exercice précédent
3. On change la loi des bruits Z_i par une autre loi connue ; comment sont modifiés ces deux estimateurs ?

Exercice 3 (Méthode des moments)**:**
On observe les durées de vie $X_1, \dots, X_n$ de n cellules tirées au hasard avec remise (c'est-à-dire que les variables aléatoires $X_1, \dots, X_n$ sont i.i.d.) dans une population contenant une proportion $\theta \in (0, 1)$ de cellules saines et $1 - \theta$ de cellules pathogènes. On modélise la durée de vie des cellules saines par une loi exponentielle de paramètre 1, c'est-à-dire par la loi de densité $\mathrm{e}^{-x}, x \in \mathbb{R}_+$, par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$. On sait par ailleurs qu'une cellule pathogène a une durée de vie moyenne 2 fois moins grande qu'une cellule saine et suit aussi une loi exponentielle.

1. Les durées de vie $X_1, \dots, X_n$ sont donc i.i.d. de loi $(P_\theta)_{\theta \in (0,1)}$ sur $\mathbb{R}_+$. Expliquer pourquoi P_θ a pour densité
$$f_\theta(x) := \theta \mathrm{e}^{-x} + 2(1 - \theta)\mathrm{e}^{-2x}, \quad x > 0,$$
par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$.
2. Soit $p > 0$ un entier ; calculer $\mathbb{E}_\theta[X_1^p]$ et en déduire un estimateur $\widehat{\theta}_{p,n}$ par la méthode des moments.

Exercice 4 (Loi Gamma)**:**
Un opérateur mobile s'intéresse aux durées d'appel de ses clients. On observe les durées

$X_1, \ldots, X_n$ de n appels, supposées indépendantes et modélisées chacune par une loi Gamma $\mathrm{P}_\theta = \gamma(\alpha, \lambda)$, de paramètre $\theta = (\alpha, \lambda)$ avec $\alpha > 0$ et $\lambda > 0$ dont la densité est

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x) \mathbb{1}_{\mathbb{R}_+}(x) ,$$

où $\Gamma(\alpha)$ est la fonction Gamma. On s'intéresse au paramètre $\theta = (\alpha, \lambda) \in \Theta = (0, \infty)^2$.

1. Calculez l'estimateur de θ par la méthode des moments.
2. Donnez l'équation que doit vérifier l'estimateur de θ par la méthode du maximum de vraisemblance. Commentez.