Examen final d'OS02 : théorie de l'estimation - 2 heures

Attention : les seuls documents autorisés sont les polycopiés de cours distribués en OS02 et un formulaire. Le barème indiqué est approximatif.

Sujet 1 ($\simeq 4$ points) . Soit $\xi_1, \xi_2, \dots, \xi_n$ une suite de variables aléatoires discrètes indépendantes issues d'une loi mélange paramétrée $P_\theta = \{p_j = \mathbb{P}_\theta(\xi = x_j),\ \ j = 1, \dots, m\}$, où θ est le paramètre de mélange :

$$P_\theta = \theta Q + (1-\theta)R \ \ \Leftrightarrow\ \ p_j = \theta q_j + (1-\theta) r_j,\ \ j = 1, \dots, m,\ \ 0 < \theta < 1,$$

où $Q = \{q_j\}$ et $R = \{r_j\}$, $j = 1, \dots, m$, sont des lois avec les espérances a et b, respectivement, **connues**. Les lois P_θ, Q et R sont définies sur le même support $\mathcal{X} = \{x_j, j = 1, \dots, m\}$.

1. Trouver un estimateur $\widehat{\theta}$ du paramètre θ par la méthode de moments (MM). On peut choisir la fonction $g(x) = x$.
2. Est-il $\widehat{\theta}$ un estimateur biaisé de θ ? Si la réponse est «oui», alors calculer le biais $\theta \to b(\theta)$ de l'estimateur MM $\widehat{\theta}$.

Sujet 2 ($\simeq 6$ points) . Soit $\xi_1, \xi_2, \dots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une distribution uniforme $U(-3\theta, \theta)$ sur $[-3\theta; \theta]$, où $\theta > 0$ est un paramètre inconnu. La densité de probabilité $p(x)$ d'une loi uniforme $U(a, b)$ est définie par :

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad x \in [a; b] \\ 0 & \text{si} \quad x \notin [a; b] \end{cases},\ a < b.$$

On cherche à estimer le paramètre θ à base de mesures $\xi_1, \xi_2, \dots, \xi_n$ à l'aide de l'estimateur (ad hoc) [1] suivant :

$$\widehat{\theta}_n = 4\xi_{(n)} + \xi_{(1)}$$

où $\xi_{(1)} = \min\{\xi_1, \xi_2, \dots, \xi_n\}$ et $\xi_{(n)} = \max\{\xi_1, \xi_2, \dots, \xi_n\}$.

1. Cet estimateur $\widehat{\theta}_n$ est-il biaisé ? Est-il asymptotiquement biaisé ?
2. Cet estimateur $\widehat{\theta}_n$ est-il convergent (en probabilité) ?

[1] **Ad hoc** est une locution latine qui signifie qui va vers ce vers quoi il doit aller, c'est-à-dire **formé dans un but précis**.

Sujet 3 ($\simeq$ 6 points) . Soit $\xi_1, \xi_2, \dots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une distribution paramètrée dont la densité est définie par :

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{si} \quad x \in [0;1] \\ 0 & \text{si} \quad x \notin [0;1] \end{cases}$$

où $\theta > 0$ est un paramètre inconnu. On cherche à estimer le paramètre $\tau = \dfrac{1}{\theta}$ à base de mesures $\xi_1, \xi_2, \dots, \xi_n$.

1. Déterminer l'estimateur $\widehat{\tau}_n$ de τ par la méthode du maximum de vraisemblance (MV).
2. Cet estimateur $\widehat{\tau}_n$ est-il biaisé ? Si la réponse est «oui», alors calculer le biais $\tau \to b(\tau)$ de l'estimateur MV $\widehat{\tau}$. Dans le cas contraire $b(\tau) = 0$.
3. Calculer la moyenne quadratique $\mathbb{E}_\tau[(\hat{\tau} - \tau)^2]$ de l'estimateur MV $\widehat{\tau}$.
4. Vérifier pour $f_\tau(x)$ les conditions de régularité r1 et r2 du Théorème de Rao-Cramer.
5. Si les conditions sont satisfaites, calculer la borne de Rao-Cramer dans la classe d'estimateurs $\mathcal{K}_{b(\tau)}$ et répondre à la question suivante : l'estimateur MV $\widehat{\tau}$ est-il efficace dans cette classe ? Utiliser le biais $\tau \to b(\tau)$ de l'estimateur MV $\widehat{\tau}$ calculé précédemment.

Sujet 4 ($\simeq$ 4 points) . Soit $\xi_1, \xi_2, \dots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une distribution normale $\mathcal{N}(\alpha, 1)$, où α est un paramètre inconnu et **aléatoire** admettant la distribution a *priori* normale $\mathcal{N}(0, \sigma^2)$, où σ^2 est une constante connue.

1. Montrer que la loi a *posteriori* de α sachant $\xi_1, \xi_2, \dots, \xi_n$ correspond à la loi normale $\mathcal{N}\left(\dfrac{\overline{\xi} n\sigma^2}{1+n\sigma^2}, \dfrac{\sigma^2}{1+n\sigma^2}\right)$, où $\overline{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$.
2. Déduire du résultat précédent l'estimateur bayésien $\widehat{\alpha}_n$ de α.

Indication : la densité a *posteriori* du paramètre α est calculée à l'aide de la formule suivante (voir les polycopiés de cours) :

$$q(a \mid x_1, \dots, x_n) = \frac{f_a(x_1, \dots, x_n) q(a)}{\int_{-\infty}^{\infty} f_s(x_1, \dots, x_n) q(s) ds}$$