DM 8 : Estimation statistique

Exercice 1 (Loi exponentielle translatée)**.** On observe $\mathbf{x} = (x_1, \ldots, x_n)$ que l'on considère comme la réalisation du vecteur aléatoire $\mathbf{X} = (X_1, \ldots, X_n)$, où les X_i sont des variables aléatoires i.i.d. de la loi $\mathbb{P}_\theta$ de densité $f_\theta(x) = \mathrm{e}^{-(x-\theta)}\mathbb{1}_{\{x\geq\theta\}}$ de paramètre inconnu $\theta \in \mathbb{R}$.

1. Soit Y une v.a. de loi exponentielle de paramètre 1. Montrer que X_i a même loi que $Y + \theta$.
2. Calculer l'estimateur par la méthode des moments $\hat{\theta}_n^{\mathrm{MM}}$ de θ. On utilisera le moment d'ordre 1.
3. Montrer que l'estimateur $\hat{\theta}_n^{\mathrm{MM}}$ est convergent et asymptotiquement normal. Déterminer son risque quadratique moyen.
4. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ de θ, et déterminer sa loi.
5. Vérifier si l'estimateur $\hat{\theta}_n^{\mathrm{MV}}$ est convergent.
6. Déterminer son risque quadratique moyen $\mathrm{RQM}_\theta(\hat{\theta}^{\mathrm{MV}})$.
7. Étudier la convergence en loi de $n^\alpha(\hat{\theta}_n^{\mathrm{MV}} - \theta)$ lorsque n tend vers l'infini. En déduire que $\hat{\theta}_n^{\mathrm{MV}}$ n'est pas asymptotiquement normal et préciser la vitesse de convergence de l'estimateur.

Solution.

1. Notons $V = Y + \theta$. Il est clair que $V > \theta$ p.s. La fonction de répartition de $V = Y + \theta$ vaut $F_V(v) = 0$ pour $v \leq \theta$, et pour $v > \theta$
$$F_V(v) = \mathbb{P}_\theta(V \leq v) = \mathbb{P}_\theta(Y \leq v - \theta) = F_Y(v - \theta) = 1 - \mathrm{e}^{-(v-\theta)},$$
car $v - \theta > 0$. En dérivant on trouve que la densité de V est f_θ. Donc, V et X ont bien la même loi.
2. On a $\mathbb{E}_\theta[X_1] = \mathbb{E}_\theta[Y + \theta] = 1 + \theta$. Pour l'EMM, en posant $1 + \theta = \bar{x}_n$, on obtient $\hat{\theta}^{\mathrm{MM}} = \bar{X}_n - 1$.
3. Par la LFGN, on a bien $\hat{\theta}_n^{\mathrm{MM}} = \bar{X}_n - 1 \to 1 + \theta - 1 = \theta$ *p.s.*. D'où la convergence de l'EMM. D'après le TCL, on a
$$\sqrt{n}(\hat{\theta}_n^{\mathrm{MM}} - \theta) = \sqrt{n}((\bar{X}_n - 1) - \theta) = \sqrt{n}(\bar{X}_n - \mathbb{E}_\theta[X_1]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Var}_\theta(X_1)),$$
avec $\mathrm{Var}_\theta(X_1) = \mathrm{Var}_\theta(Y + \theta) = 1$. L'EMM est donc bien asymptotiquement normal avec une vitesse de convergence de $n^{-1/2}$. Quant au risque quadratique moyen, on trouve
$$\mathrm{RQM}_\theta(\hat{\theta}^{\mathrm{MV}}) = (\mathbb{E}_\theta[\hat{\theta}^{\mathrm{MM}}] - \theta)^2 + \mathrm{Var}(\hat{\theta}^{\mathrm{MM}}) = 0 + \frac{1}{n}\mathrm{Var}_\theta(X_1) = \frac{1}{n}.$$
4. Pour l'EMV on calcule la fonction de vraisemblance
$$\mathcal{L}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \mathrm{e}^{-x_i+\theta}\mathbb{1}\{x_i \geq \theta\} = \mathrm{e}^{-\sum_{i=1}^n x_i + n\theta}\mathbb{1}\{x_{(1)} \geq \theta\},$$
où $x_{(1)} = \min\{x_1, \ldots, x_n\}$. D'une part, la fonction $\theta \mapsto \mathbb{1}\{x_{(1)} \geq \theta\}$ est maximale pour tout $\theta \leq x_{(1)}$. D'autre part, la fonction $\theta \mapsto \mathrm{e}^{-\sum_{i=1}^n x_i + n\theta}$ est croissante en θ. Cela implique que le maximum est atteint en $\theta = x_{(1)}$. Donc, l'EMV est $\hat{\theta}_n^{\mathrm{MV}} = X_{(1)}$.
Pour la loi de $\hat{\theta}_n^{\mathrm{MV}}$, il est clair que $\hat{\theta}_n^{\mathrm{MV}} \geq \theta$ *p.s.*. La fonction de répartition est donnée pour $t > \theta$ par
$$\begin{aligned} F_{\hat{\theta}_n^{\mathrm{MV}}}(t) &= \mathbb{P}_\theta(X_{(1)} \leq t) = 1 - \mathbb{P}_\theta(X_{(1)} > t) = 1 - [\mathbb{P}_\theta(X_1 > t)]^n = 1 - [\mathbb{P}_\theta(Y > t + \theta)]^n \\ &= 1 - \mathrm{e}^{-n(t+\theta)}. \end{aligned}$$
Donc, la densité de $\hat{\theta}_n^{\mathrm{MV}}$ est donnée par $f_{\hat{\theta}_n^{\mathrm{MV}}}(t) = n\mathrm{e}^{n(t-\theta)}\mathbb{1}\{t > \theta\}$. Autrement dit, notons Z_n des v.a. de loi exponentielle $\mathrm{Exp}(n)$, alors $\hat{\theta}_n^{\mathrm{MV}}$ a même loi que $Z_n + \theta$ pour tout n.

5. Notons $A_n = \{|\hat{\theta}_n^{\mathrm{MV}} - \theta| > a_n\}$ pour des $a_n > 0$. On a
$$\mathbb{P}_\theta(A_n) = \mathbb{P}_\theta(\hat{\theta}_n^{\mathrm{MV}} > \theta + a_n) = 1 - F_{\hat{\theta}_n^{\mathrm{MV}}}(\theta + a_n) = \mathrm{e}^{-na_n}.$$
En choisissant $a_n = 1/\sqrt{n}$, on obtient $\sum_n \mathbb{P}_\theta(A_n) < \infty$, et le théorème de Borel-Cantelli implique que $\mathbb{P}_\theta(\limsup A_n) = 0$. Autrement dit, $\hat{\theta}_n^{\mathrm{MV}} \to \theta$ *p.s.*, d'où la convergence.
6. On a $\mathbb{E}_\theta[\hat{\theta}_n^{\mathrm{MV}}] = \mathbb{E}_\theta[Z_n + \theta] = \frac{1}{n} + \theta$ et $\mathrm{Var}_\theta(\hat{\theta}_n^{\mathrm{MV}}) = \mathrm{Var}_\theta(Z_n + \theta) = \frac{1}{n^2}$. Alors, le risque quadratique moyen vaut
$$\mathrm{RQM}_\theta(\hat{\theta}^{\mathrm{MV}}) = (\mathbb{E}_\theta[\hat{\theta}^{\mathrm{MV}}] - \theta)^2 + \mathrm{Var}(\hat{\theta}^{\mathrm{MV}}) = \frac{2}{n^2}.$$
7. Remarquons que $n^\alpha(\hat{\theta}_n^{\mathrm{MV}} - \theta) > 0$ p.s. On a pour $t > 0$
$$\mathbb{P}_\theta\left(n^\alpha(\hat{\theta}_n^{\mathrm{MV}} - \theta) \leq t\right) = \mathbb{P}_\theta\left(\hat{\theta}_n^{\mathrm{MV}} \leq n^{-\alpha}t + \theta\right) = 1 - \mathrm{e}^{tn^{-\alpha+1}}.$$
Pour $\alpha = 1$ on reconnait la fonction de répartition de la loi exponentielle Exp(1). Donc, on a montré que $n(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathcal{L}} \mathrm{Exp}(1)$. La vitesse de convergence est alors de $1/n$. Dans ce modèle, l'EMV converge plus rapidement que l'EMM, et l'EMV n'est pas asymptotiquement normal.