Devoir 2 STT6700 hiver 2024
à remettre pour le 20 février 2024

\# 1 Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes de densité $f(\cdot|\theta)$, $\theta \in [-1, 1]$ avec

$$f(x \mid \theta) = (1 + \theta x)/2, \quad x \in [-1, 1].$$

On pose

$$Y_i = \mathbb{1}_{[0,1]}(X_i), \quad i = 1, ..., n.$$

1. Trouvez l'estimateur de $\theta$ par la méthode des moments lorsqu'on observe $X_1, \ldots, X_n$.
2. Trouvez l'estimateur de $\theta$ par la méthode des moments lorsqu'on observe $Y_1, \ldots, Y_n$.
3. Calculez les écarts quadratique moyens pour chacun des estimateurs.
4. Proposez une méthode élémentaire pour améliorer les écarts quadratique moyens. ( Indication: comparer les espaces image des estimateurs avec l'espace paramétrique. )

\# 2 Soit $X_1, X_2$ des variables aléatoires indépendandes de loi de Cauchy$(\theta, 1)$. La densité est donnée par

$$f(x \mid \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

Trouvez l'estimateur de $\theta$ par la méthode du maximum de vraisemblance.
( Indication: reparamétrisez en $\eta = \theta - m$, $m = (x_1 + x_2)/2$).

\# 3 Soit $f(\cdot \mid \theta)$ une densité de la forme

$$f(x \mid \theta) = g(\mid x - \theta \mid)$$

avec $g$ une fonction continue et décroissante. Soit $x_1, \ldots, x_n$ les observations et $x_{(1)}, \ldots, x_{(n)}$ les statistiques d'ordre. Montrez que l'estimateur de $\theta$ par la méthode du maximum de vraisemblance existe et qu'une des versions se trouve dans l'intervalle $[x_{(1)}, x_{(n)}]$.

## Solutionnaire

# 1.1 On a

$$\mathbb{E}_\theta[X_1] = \theta/3, \quad \hat{\theta}_1 = 3\bar{x}_n.$$

# 1.2 On a

$$Y_1 \sim \text{Bernoulli}\left(\frac{1}{2}(1+\frac{\theta}{2})\right), \quad \mathbb{E}_\theta[Y_1] = \frac{1}{2}(1+\frac{\theta}{2}), \quad \hat{\theta}_2 = 4\bar{y}_n - 2.$$

# 1.3 Les deux estimateurs sont sans biais. On a

$$\begin{aligned}
\text{EQM}_\theta(\hat{\theta}_1) &= \text{Var}_\theta(3\bar{X}_n) \\
&= 9\text{Var}_\theta(X_1)/n \\
&= 9\mathbb{E}_\theta[X_1^2 - (\theta/3)^2]/n \\
&= 9\left[(1/3) - (\theta/3)^2\right]/n \\
&= 3\left[1 - \theta^2/3\right]/n
\end{aligned}$$

et

$$\begin{aligned}
\text{EQM}_\theta(\hat{\theta}_2) &= \text{Var}_\theta(4\bar{Y}_n) \\
&= 16\text{Var}_\theta(Y_1)/n \\
&= 4\left[(1+\theta/2)(1-\theta/2)\right]/n \\
&= 4\left[1 - \theta^2/4\right]/n
\end{aligned}$$

# 1.4 On pose

$$\tilde{\theta}_i = \begin{cases} 1 & \text{si} \quad \hat{\theta}_i > 1 \\ \hat{\theta}_i & \text{si} \quad |\,\hat{\theta}_i\,| \le 1 \\ -1 & \text{si} \quad \hat{\theta}_i < -1 \end{cases}, \quad i = 1, 2.$$

# 2 La vraisemblance est donnée par

$$L(\theta \mid x_1, x_2) = \left(\frac{1}{\pi}\right)^2 \left\{(1+(x_1-\theta)^2)(1+(x_2-\theta)^2\right\}^{-1}$$

On pose

$$\begin{aligned}
m &= (x_1 + x_1)/2 \\
t &= |\, x_2 - x_1 \,|/2 \\
\eta &= \theta - m
\end{aligned}$$

On obtient

$$\begin{aligned}
L(\theta \mid x_1, x_2) &= \left(\frac{1}{\pi}\right)^2 \left\{(1+(x_{(1)}-\theta)^2)(1+(x_{(2)}-\theta)^2\right\}^{-1} \\
&= \left(\frac{1}{\pi}\right)^2 \left\{(1+(\eta+t)^2)(1+(\eta-t)^2\right\}^{-1} \\
&= \left(\frac{1}{\pi}\right)^2 \left\{(1+[(\eta+t)^2+(\eta-t)^2]+(\eta^2-t^2)^2\right\}^{-1} \\
&= \left(\frac{1}{\pi}\right)^2 \left\{(\eta^4-2[t^2-1]\eta^2+(1+t^2)^2\right\}^{-1} \\
&= \left(\frac{1}{\pi}\right)^2 / g(\eta^2 \mid t^2)
\end{aligned}$$

On veut minimiser la fonction $g(\cdot \mid t^2)$ sur $[0, \infty)$. La fonction $g(\cdot \mid t^2)$ est une parabole.

Si $t \leq 1$ alors $g(\cdot \mid t^2)$ est strictement croissante sur $[0, \infty]$ et $\hat{\eta} = 0$ donc $\hat{\theta} = m$.

Si $t > 1$ alors $g(\cdot \mid t^2)$ atteint son minimum en $\eta^2 = t^2 - 1$ ce qui donne deux solutions pour $\hat{\theta}$. Ainsi, pour récapituler, l'estimateur par la méthode du maximum de vraisemblance n'a pas une solution unique lorsque $t > 1$, ( on peut choisir la solution que l'on préfère ) et

$$\hat{\theta} = \left\{ \begin{array}{cc} m & \text{si} \quad t \leq 1 \\ m - \sqrt{t^2-1} \text{ ou } m + \sqrt{t^2-1} & \text{si} \quad t > 1 \end{array} \right. .$$

\# 3 La vraisemblance est donnée par

$$L(\theta \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} g(\mid x_{(i)} - \theta \mid)$$

On a

$$\begin{aligned}
t > x_{(n)} &\Rightarrow g(\mid x_{(i)} - t \mid) \leq g(\mid x_{(i)} - x_{(n)} \mid), \quad i = 1, \ldots, n \\
&\Rightarrow L(t \mid x_1, \ldots, x_n) \leq L(x_{(n)} \mid x_1, \ldots, x_n) \\
t < x_{(1)} &\Rightarrow L(t \mid x_1, \ldots, x_n) \leq L(x_{(1)} \mid x_1, \ldots, x_n)
\end{aligned}$$

Ainsi, si la vraisemblance atteint son maximum alors il existe $\theta \in [x_{(1)}, x_{(n)}]$ tel que le maximum est atteint en $\theta$. Maintenant, comme la vraisemblance est une fonction continue et $[x_{(1)}, x_{(n)}]$ est un ensemble compact la fonction atteint son maximum sur un point $\theta \in [x_{(1)}, x_{(n)}]$.