Examen Partiel – Octobre 2018

Durée 2h00 – Sans documents

Important. Suivant les réglements en vigueur,

1. les enseignants présents lors de l'épreuve ne peuvent communiquer que sur les fautes d'énoncé potentielles. Toute autre question durant la composition ne sera pas acceptée.
2. les étudiants sont tenus de se lever au moment de l'annonce de fin de la composition. En cas de refus, le responsable de l'UE sera fondé à ne pas prendre en compte la copie incriminée.
3. l'identification des copies et intercalaires doit se faire au moment de la remise de chaque copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.

Exercice 1 (6 pts). Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires *i.i.d.* uniformes sur $[0,\theta]$, où $\theta > 0$ est un paramètre inconnu. On rappelle la densité d'une variable aléatoire uniforme sur $[0,\theta]$:

$$f(x|\theta) := \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x).$$

1. *(1 pt)* Soit

$$\widehat{\theta}_n := \frac{2}{n}\sum_{i=1}^{n} X_i.$$

Justifier la convergence presque sûre de cet estimateur.

2. *(0.5 pt)* En déduire un estimateur qui converge presque sûrement vers $\mathrm{Var}_\theta(X_1)$.

3. *(1.5 pts)* Montrer que

$$\sqrt{3n}\left(1-\frac{\theta}{\widehat{\theta}_n}\right) \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1).$$

4. *(1.5 pts)* Énoncer la méthode delta. Retrouver en utilisant cette méthode le résultat précédent.

5. *(1.5 pts)* Soit $\alpha \in]0,1[$.

(a) Trouver $q > 0$ tel que

$$\mathbb{P}_\theta\left(\theta \in \left[\left(1-\frac{q}{\sqrt{3n}}\right)\hat{\theta}_n, \left(1+\frac{q}{\sqrt{3n}}\right)\hat{\theta}_n\right]\right) \xrightarrow[n\to+\infty]{} \alpha.$$

(b) On choisit $\alpha = 0.9$. Donner la commande R qui renvoie la valeur de q.

Exercice 2 (9 pts). Soit $X_1,\dots,X_n$ un échantillon de variables aléatoires *i.i.d* de loi de Cauchy $\mathcal{C}(\beta)$. On rappelle que la densité de la loi de Cauchy est :

$$f_X(x) = \frac{1}{\pi(1+(x-\beta)^2)},$$

et sa fonction caractéristique : $\phi_\beta(t) = \mathbb{E}_\beta[\exp(itX)] = \exp(\beta it - |t|)$. L'objectif est de comparer deux estimateurs de β.

1. *(1 pt)* Déterminer la limite en loi de $\frac{1}{n}\sum_{i=1}^n X_i$.

2. *(0.5 pt)* Montrer que la médiane de $\mathcal{C}(\beta)$ est β.

On propose deux estimateurs $\hat\beta_n^1$ et $\hat\beta_n^2$ de β. Si $(X_{(i)})$ est la version ordonnée de (X_i), et $p \in]0, 1/2[$, on définit :

$$\hat\beta_n^1 = X_{\lceil n/2\rceil} \quad \text{et} \quad \hat\beta_n^2 = \frac{1}{(1-2p)n}\sum_{i\in\lceil pn\rceil}^{\lceil n-pn\rceil} X_{(i)}.$$

Le premier estimateur est la médiane empirique, le second correspond à la *moyenne tronquée* (en anglais *trimmed mean*), c'est à dire la moyenne calculée sur le sous échantillon constitué des valeurs les moins extrêmes.

3. *(3.5 pts)* Soit $q_n(p)$ le quantile d'ordre p de l'échantillon $\{X_1, \dots, X_n\}$, et $q(p)$ le quantile théorique d'ordre p de la loi $\mathcal{C}(\beta)$. On rappelle que

$$q_n(p) = \inf\left\{x, F_n(x) = \frac{1}{n}\sum \mathbb{1}_{X_i<x} \geq p\right\} \quad \text{et} \quad q(p) = \inf\{x, F(x) \geq p\},$$

avec F_n la fonction de répartition empirique associée à l'échantillon $\{X_1, \dots, X_n\}$, et F la fonction de répartition associée à $\mathcal{C}(\beta)$.

- **(a)** Montrer que $F_n(x) \xrightarrow[n\to+\infty]{p.s.} F(x)$ pour tout $x \in \mathbb{R}$.
- **(b)** Soit $\delta > 0$, montrer que pour n assez grand et p.s. $F_n(q(p) - \delta) < p$ et en déduire que $q_n(p) \geq q(p) - \delta$.
- **(c)** Conclure que $\liminf_n q_n(p) \geq q(p)$.
- **(d)** Montrer que $F_n(q(p) + \delta) \geq p$, et en déduire que $\limsup_n q_n(p) \leq q(p)$.
- **(e)** En conclure que $\hat\beta_n^1$ converge presque sûrement vers β.

4. *(1pt)* Étant donné un vecteur d'observations `x`, écrire une fonction `estimator1` qui retourne la médiane empirique, c'est à dire $\hat\beta_n^1$.

5. *(1 pt)* Parmi les fonctions R suivantes, donner, en justifiant, celle calculant une approximation du biais de $\hat\beta_n^1$.

```r
bias.A <- function(X, k) {
        alpha.b <- numeric(k)
        for (i in 1:k) {
                alpha.b[i] <- estimator1(sample(X, replace = TRUE))
        }
        return(mean(alpha.b) - estimator1(X))
}
```

```r
bias.B <- function(X, k) {
        alpha.b <- numeric(k)
        for (i in 1:k) {
                alpha.b[i] <- estimator1(sample(X,replace=TRUE))
        }
        return(estimator1(alpha.b) - estimator1(X))
}
```

```r
bias.C <- function(X, k) {
        alpha.b <- numeric(k)
        for (i in 1:k) {
                alpha.b[i] <- estimator1(sample(X, replace = FALSE))
        }
        return(mean(alpha.b) - estimator1(X))
}
```

```r
bias.D <- function(X,k) {
        alpha.b <- numeric(k)
        for (i in 1:k) {
                alpha.b[i] <- estimator1(sample(X, replace = FALSE))
        }
        return(estimator1(alpha.b) - estimator1(X))
}
```

6. *(2 pts)* La fonction calculant $\hat{\beta}_n^2$ est la suivante, où $p \in]0, 1/2[$:

```r
estimator2 <- function(x,p) {
        return(mean(x, trim = p))
}
```

Écrire une fonction R nommée `IC.a2` qui calcule un intervalle de confiance pour $\hat{\beta}_n^2$ au niveau de confiance $1 - \alpha$ en utilisant la méthode bootstrap de votre choix.

Exercice 3 (6 pts). Soit X une variable aléatoire admettant la densité de probabilité suivante :

$$f_X(x; \theta) = \frac{\theta \exp(\theta x)}{\exp(\theta^2) - 1} \mathbb{1}_{[0,\theta]}(x)$$

où $\theta > 0$. On considère un n-échantillon $X_1, \dots, X_n$ de X.

1. *(2 pts)* Montrer que la densité est bien normalisée, c'est-à-dire, intègre à 1. Donner la fonction génératrice des moments associée à $f_X(\cdot\,; \theta)$, c'est-à-dire $\mathbb{E}[\exp(tX)]$, $t \in \mathbb{R}$. En déduire le premier moment de X.

2. *(1 pt)* La famille de densités $f_X(\cdot\,; \theta)$ indicée par θ constitue-t-elle une famille exponentielle ? Si oui, donnez le paramètre naturel $\mu = \mu(\theta)$ et la représentation cannonique en $\exp\{\mu \cdot T(x) - \psi(\mu)\}$. Si non, indiquez la partie de $f_X(\cdot\,; \theta)$ qui empêche une représentation exponentielle.

3. *(1 pt)* Montrer que la densité du n-échantillon factorise en

$$g(T(x_1, \dots, x_n); \theta) \times h(x_1, \dots, x_n)$$

où $T(x_1, \dots, x_n)$ est une statistique de dimension deux.

4. *(2 pts)* Calculer la moyenne des deux composantes de $T(X_1, \dots, X_n)$. Pouvez vous en déduire un estimateur sans biais de θ ?