1 Exercices

Exercice 1. Soit $\{X_k,\ k \in \mathbb{N}\}$ un échantillon de loi $\Gamma(2,\theta)$:

$$x \mapsto p(\theta,x) = \theta^2 x \mathrm{e}^{-\theta x} \mathbb{1}_{\mathbb{R}_+}(x)\ , \quad \theta \in \Theta := \mathbb{R}_+^*.$$

On rappelle que l'espérance d'une loi $\Gamma(a,b)$ est a/b et que sa variance est a/b^2. Enfin, si $X \sim \Gamma(a,b)$, $Y \sim \Gamma(a',b)$ et X,Y sont indépendantes, alors $X+Y \sim \Gamma(a+a',b)$

1. Calculer l'estimateur du maximum de vraisemblance.
2. Calculer le biais et le risque quadratique de l'estimateur du maximum de vraisemblance.
3. Déterminer la loi limite de l'estimateur du maximum de vraisemblance.

Exercice 2. Une variable aléatoire X à valeurs dans $\mathbb{R}_+^*$ suit une loi log-normale de paramètre $(\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ si la variable $Y = \log(X)$ est normale de moyenne μ et variance σ^2. Cette distribution est très couramment utilisée pour modéliser le prix des actifs financiers. Soit $\{X_k,\ k \in \mathbb{N}\}$ une suite de variables aléatoires indépendantes de loi log-normale de paramètre $(\theta,\theta) \in \Theta = \mathbb{R}_*^+$.

1. Déterminer l'estimateur du maximum de vraisemblance.
2. Déterminer la loi limite de l'estimateur du maximum de vraisemblance.
3. Construire un intervalle de confiance de probabilité de couverture asymptotique $1-\alpha$ pour $\alpha \in]0,1[$.

Exercice 3. Soit $\{X_n,\ n \in \mathbb{N}\}$ un échantillon de densité

$$x \mapsto f(x-\theta), \qquad x \in \mathbb{R}, \theta \in \Theta := \mathbb{R},$$

où f est une densité ~~continue~~ par rapport à la mesure de Lebesgue sur $\mathbb{R}$ telle que $f(x) = f(-x)$ pour tout $x \in \mathbb{R}$. Nous allons étudier les Z-estimateurs associés à

$$\psi(\theta,x) = \gamma(x-\theta)$$

où

H1 $\gamma : \mathbb{R} \to \mathbb{R}$ est une fonction impaire, croissante (au sens large) et telle que

$$\lim_{-\infty} \gamma < 0, \qquad \lim_{+\infty} \gamma > 0, \qquad \sup_{\theta \in \mathbb{R}} \int_{\mathbb{R}} |\gamma(x-\theta)| f(x) \lambda_{\mathrm{Leb}}(\mathrm{d}x) < +\infty. \tag{1}$$

On définit, pour tout $\theta \in \Theta$

$$\psi_n(\theta) := \frac{1}{n} \sum_{k=1}^{n} \psi(X_k,\theta) = \frac{1}{n} \sum_{k=1}^{n} \gamma(X_k - \theta).$$

La figure 1 montre cette fonction pour différentes valeurs de n, dans le cas où γ est égale à $\gamma_0 = \mathrm{signe}(x)$ (gauche), $\gamma_\infty(x) = x$ (centre) et $\gamma_1(x) = \mathrm{signe}(x)\,(|x| \wedge 1)$ (droite), et f est la loi

$\mathcal{N}(0,1)$; de plus, $\{X_k, k \geq 1\}$ sont i.i.d. de loi $\mathcal{N}(\theta_0, 1)$ avec $\theta_0 = 2$. En trait épais, on trace aussi la fonction

$$\begin{aligned}\theta \mapsto \Psi_{\theta_0}(\theta) &:= \mathbb{E}_{\theta_0}\left[\gamma(X_1 - \theta)\right] \\ &= \int \gamma(x-\theta) f(x-\theta_0) \lambda_{\text{Leb}}(\mathrm{d}x) = \int \gamma(x+\theta_0-\theta) f(x) \lambda_{\text{Leb}}(\mathrm{d}x)\end{aligned}$$

Noter que la condition (1) justifie l'existence de Ψ_{θ_0} pour tout θ.

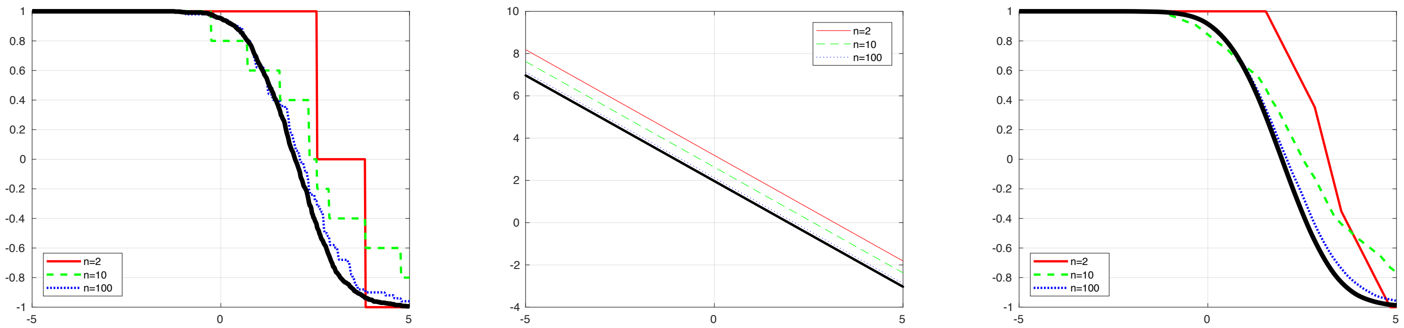

FIGURE 1 – (gauche) cas γ_0, (centre) cas γ_∞, (droite) cas γ_k avec $k = 1$. Pour chacun des cas, on trace $\theta \mapsto \psi_n(\theta)$ pour $n = 2, 10, 100$ et on trace $\theta \mapsto \Psi_{\theta_0}(\theta)$ en trait épais. On a pris $f \equiv \mathcal{N}(0,1)$ et $\theta_0 = 2$.

1. Montrer qu'il existe $\hat{\theta}_n$ tel que
$$\psi_n \geq 0 \text{ sur } \left]-\infty, \hat{\theta}_n\right[\text{ et } \psi_n \leq 0 \text{ sur } \left]\hat{\theta}_n, +\infty\right[$$
2. Montrer que si γ est continue, il existe $\hat{\theta}_n$ tel que $\psi_n(\hat{\theta}_n) = 0$.
3. Montrer que si γ est strictement croissante et continue, alors $\hat{\theta}_n$ est unique.
4. Etudier les cas où γ est une des fonctions suivantes :
$$\gamma_0(x) = \text{signe}(x), \qquad \gamma_\infty(x) = x, \qquad \gamma_k(x) = \begin{cases} x & |x| \leq k \\ k\,\text{signe}(x) & |x| > k \end{cases} \tag{2}$$

Dans la suite, on suppose que

H2 γ est continue et strictement croissante.

5. Montrer que pour tout $\theta_0 \in \mathbb{R}$, l'unique solution de "$\theta : \Psi_{\theta_0}(\theta) = 0$" est θ_0.
6. Montrer que pour tout $\varepsilon > 0$, si $\hat{\theta}_n < \theta_0 - \varepsilon$, alors $\psi_n(\theta_0 - \varepsilon) < 0$. En déduire que
$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}(\hat{\theta}_n < \theta_0 - \varepsilon) = 0\,.$$

7. Montrer que la suite d'estimateurs $\{\hat{\theta}_n,\ n \in \mathbb{N}\}$ est consistante.
8. On suppose de plus

   H3 que γ est ~~dérivable~~ C^1 et

$$\int \gamma^2(x) f(x) \lambda_{\text{Leb}}(\mathrm{d}x) < \infty, \qquad \int \gamma'(x) f(x) \lambda_{\text{Leb}}(\mathrm{d}x) > 0, \tag{3}$$

$$\forall \delta > 0, \quad \int \sup_{|y-x| \leq \delta} |\gamma'(y)| f(x) \lambda_{\text{Leb}}(\mathrm{d}x) < \infty. \tag{4}$$

   Montrer que la suite d'estimateurs $\{\hat{\theta}_n,\ n \in \mathbb{N}\}$ est asymptotiquement normale et déterminer sa distribution asymptotique. Dans la suite, on notera $V[\gamma]$ la variance limite.

Pour établir la consistance de la suite d'estimateurs, on peut remplacer le jeu de conditions H2 par "γ est croissante et f est strictement positive sur $\mathbb{R}$". Pour la normalité asymptotique, on peut juste supposer que γ est "continue, C^1 par morceaux, et vérifie (3)-(4)". On compare l'efficacité relative des estimateurs $\{\hat{\theta}_n,\ n \in \mathbb{N}\}$ obtenus avec $\gamma = \gamma_\infty$ d'une part, et $\gamma = \gamma_k$ pour $k > 0$ d'autre part (voir Eq. 2).

9. On considère tout d'abord le cas où f est une loi $\mathcal{N}(0, \sigma^2)$. Comparer (numériquement) l'efficacité relative en fonction de k (il faut s'aider bien entendu d'un ordinateur : sur la figure 2(gauche), on représente la variance asymptotique $V[\gamma_k]$ en fonction de k - on a pris $\sigma^2 = 1$).
10. Déterminer la variance asymptotique dans le cas où

$$f(x) = (1 - \varepsilon)\phi(x) + \varepsilon \tau^{-1} \phi(\tau^{-1} x)$$

   où ϕ est la densité d'une gaussienne centrée réduite (observer que dans ce cas, $\tau^{-1}\phi(\tau^{-1}x)$ est une gaussienne centrée de variance τ^2). Sur la figure 2(droite), on représente la variance asymptotique $V[\gamma_k]$ en fonction de k - on a pris $\varepsilon = 0.01$ et $\tau = 10$.
   Ce type de modèles est utilisé en statistiques robustes pour modéliser la présence de données aberrantes : ε est la probabilité de contamination, qui est en général très petite et τ est ~~la variance~~ l'écart-type des données "contaminées" que l'on prend très grande. Commenter l'intérêt de prendre la fonction γ_k, $(k < \infty)$ plutôt que γ_∞.

Exercice 4. La loi de Pareto est couramment utilisée pour modéliser des variables à queues lourdes. Cette loi a pour densité

$$x \mapsto p_\theta(x) = \eta \alpha^\eta x^{-(\eta+1)} \mathbb{1}_{\{[\alpha, +\infty[\}}(x), \quad \theta := (\alpha, \eta) \in \mathbb{R}_*^+ \times \mathbb{R}_*^+,$$

par rapport à la mesure de Lebesgue. Soit $\{X_k,\ k \in \mathbb{N}\}$ une suite de variables aléatoires indépendantes distribuées suivant la loi de densité $\{p_\theta, \theta \in \Theta\}$. On peut vérifier que le modèle est régulier.

On suppose α connu.

1. Déterminer l'estimateur $\hat{\eta}_n$ du maximum de vraisemblance de η.

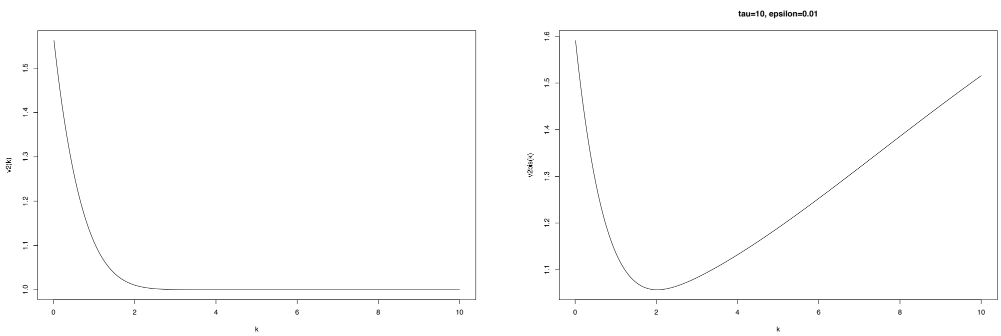

FIGURE 2 – Panneau de gauche : pas de contamination (cas $\sigma^2 = 1$) ; Panneau de droite : $\varepsilon = 0.01$ et $\tau = 10$

2. Montrer que la suite d'estimateurs $\{\hat{\eta}_n, n \in \mathbb{N}\}$ est (fortement) consistante.
3. Montrer que la suite d'estimateurs $\{\hat{\eta}_n, n \in \mathbb{N}\}$ est asymptotiquement normale.
4. Construire un intervalle de confiance de probabilité de couverture asymptotique $1-\rho$ pour le paramètre η.

On suppose maintenant que α est aussi un paramètre inconnu.

5. Déterminer l'estimateur du maximum de vraisemblance de (α, η).
6. Montrer que, pour tout α, η, nous avons
$$\mathbb{P}_\theta(\hat{\alpha}_n > y) = \left(\frac{\alpha}{y}\right)^{n\eta}, \qquad \forall y \geq \alpha,$$
où $\hat{\alpha}_n$ désigne l'estimateur du maximum de vraisemblance de α. Déterminer la moyenne et la variance de $\hat{\alpha}_n$ (on prendra n suffisamment grand).
7. Montrer que $\hat{\alpha}_n$ est un estimateur consistant de α.
8. Déterminer la loi de $n\eta \ln(\hat{\alpha}_n/\alpha)$. En déduire une fonction pivotale asymptotique pour $\hat{\alpha}_n$. Construire un intervalle de confiance de probabilité de couverture asymptotique $(1-\beta)$ du paramètre α.

Exercice 5. Soit $\{X_k,\ k \in \mathbb{N}\}$ un échantillon $\{\mathrm{N}(\theta, 1)\ :\ \theta \in \Theta = \mathbb{R}_+\}$.

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ basé sur $(X_1, \ldots, X_n)$.
2. Montrer que pour tout $\theta > 0$,
$$\sqrt{n}(\hat{\theta}_n^{\mathrm{MV}} - \theta) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0,1).$$
3. Déterminer la loi limite lorsque $\theta = 0$.