Devoir 2 pour le 9 novembre

1. one-step dans les modèles exponentiels

Dans cet exercice, nous considérons un modèle exponentiel identifiable, en forme canonique. L'espace des paramètres est noté $\Theta \subset \mathbb{R}^k$. On suppose $0 \in \Theta$. $\theta^0 \in \Theta$ désigne la valeur du paramètre sous laquelle on effectue l'échantillonnage (« la vraie »). La statistique suffisante est dénotée par $T(X)$. Elle est à valeur dans $\mathbb{R}^k$. On note $\ell_n(\theta)$ la log-vraisemblance en θ pour un n-échantillon (elle n'est pas normalisée). Lorsqu'on écrit $\nabla \ell_n(\theta')$, on désigne le gradient de la log-vraisemblance par rapport au paramètre (la fonction *score*) pris en θ'. On note $\nabla^2 \ell_n(\theta')$ le Hessien de la log-vraisemblance par rapport au paramètre (l'information empirique) pris en θ'.
Dans l'énoncé $\widetilde{\theta}_n$ désigne un estimateur (pas forcément un estimateur au maximum de vraisemblance). On suppose que la suite $(\widetilde{\theta}_n)_n$ est consistante et asymptotiquement normale. On note $J(\theta^0)$ la covariance asymptotique de $\sqrt{n}(\widetilde{\theta}_n - \theta^0)$.
On définit l'estimateur à un pas (*one-step*) $\widehat{\theta}_n$ comme l'estimateur obtenu en appliquant un pas de la méthode de Newton pour approcher le maximum de vraisemblance en partant de $\widetilde{\theta}_n$:

$$\widehat{\theta}_n := \widetilde{\theta}_n - (\nabla^2 \ell_n(\widetilde{\theta}_n))^{-1} \nabla \ell_n(\widetilde{\theta}_n) .$$

On note l'estimateur au maximum de vraisemblance $\widehat{\theta}_n$. On note $I(\theta^0)$ la matrice d'information de Fisher en θ^0 ($I(\theta^0) = \nabla^2 \log Z(\theta^0)$).

(a) Quelle est la loi limite de $\frac{1}{\sqrt{n}} \nabla \ell_n(\theta^0)$?

(b) Vérifier que pour toute constante M

$$\sup_{\sqrt{n}\|\theta-\theta^0\| \leq M} \frac{1}{\sqrt{n}} \left\| \nabla \ell_n(\theta^0) - \nabla \ell_n(\theta) + \nabla^2 \ell_n(\theta^0) \left(\theta - \theta^0\right) \right\| \xrightarrow{P} 0$$

(est-il nécessaire de préciser en probabilité ?).

(c) La suite $\frac{1}{n} \nabla^2 \ell_n(\widetilde{\theta}_n)$ admet-elle une limite en probabilité ? presque sûre ? Si oui, la préciser.

(d) Montrer que

$$\sqrt{n} \left(\widehat{\theta}_n - \theta^0 \right) - I(\theta^0)^{-1} \frac{1}{\sqrt{n}} \nabla \ell_n(\theta^0) \xrightarrow{P} 0 .$$

(e) La suite $\sqrt{n} \left(\widehat{\theta}_n - \theta^0 \right)$ admet-elle une limite en loi ? Si oui, quelle est elle ?

(f) Pensez-vous qu'il soit pertinent d'itérer la descente de gradient ?

(g) Les lois Gamma définissent un modèle exponentiel. La densité de la loi Gamma de paramètre de forme $p > 0$ et de paramètre d'intensité $\lambda > 0$ est donnée par

$$\mathbb{I}_{x>0} \mathrm{e}^{-\lambda x} \frac{\lambda^p x^{p-1}}{\Gamma(p)} .$$

Proposer un estimateur de p, λ, facile à calculer via la méthode des moments. Vérifier si la suite d'estimateurs définie ainsi est consistante et asymptotiquement normale. Peut-on utiliser la méthode à un pas dans ce contexte ? Si oui, comparer les matrices de covariance de l'estimateur obtenu par la méthode des moments et de l'estimateur obtenu par la méthode à un pas.

2. Modèle de translation défini par la Loi de Cauchy

La loi de Cauchy admet comme densité sur $\mathbb{R}$

$$p(x) = \frac{1}{\pi} \frac{1}{1+x^2} \,.$$

On considère le modèle des translatées définies par

$$p_\theta(x) = p(x-\theta) \qquad \text{pour } x, \theta \in \mathbb{R} \,.$$

On note $m(\theta)$ la médiane de la loi de densité p_θ.
On note $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ les statistiques d'ordre d'un n-échantillon d'une loi p_θ.

(a) L'estimateur au maximum de vraisemblance de θ est il bien défini dans ce modèle ? Est-il unique ? Est il calculable (pour toute taille d'échantillon) ?

(b) Cacluler l'information de Fisher dans ce modèle.

(c) On estime $m(\theta)$ par la médiane empirique $\widehat{m}_{2n+1}$ dans un échantillon de taille $2n+1$. Quel est le biais de cet estimateur.

(d) La suite $(\widehat{m}_{2n+1})_n$ définit-elle une suite consistante d'estimateurs de $m(\theta)$?

(e) La suite $(\widehat{m}_{2n+1})_n$ définit-elle une suite asymptotiquement normale d'estimateurs de $m(\theta)$?

(f) Proposez un intervalle de niveau de confiance asymptotique $\alpha \in]0,1[$ pour l'estimation de $m(\theta)$.

(g) On souhaite maintenant construire un nouvel estimateur de θ, en applicant une étape de « descente de gradient » à $(\widehat{m}_{2n+1})_n$. On note $\dot{\ell}_{2n+1}(\theta')$ la dérivée par rapport à θ de la log-vraisemblance calculée sur l'échantillon de taille $2n+1$ en θ', et $\ddot{\ell}_{2n+1}(\theta')$ la dérivée seconde. L'estimateur $\widetilde{\theta}_{2n+1}$ est défini par

$$\widetilde{\theta}_{2n+1} := \widehat{m}_{2n+1} - \frac{\dot{\ell}_{2n+1}(\widehat{m}_{2n+1})}{\ddot{\ell}_{2n+1}(\widehat{m}_{2n+1})} \,.$$

Calculer la loi limite de

$$\sqrt{2n+1}\left(\widetilde{\theta}_{2n+1} - \theta\right) \,.$$

Pensez-vous qu'il soit pertinent d'itérer la descente de gradient ?

À l'aide du logiciel de votre choix, et pour les valeurs des paramètres de votre choix, effectuez une simulation numérique pour étudier le comportement des estimateurs obtenus dans l'exercice précédent, pour des valeurs de n entre 3 et 10 000. Étudiez également le comportement de la moyenne empirique et de l'écart-type empirique d'un échantillon suivant la loi de Cauchy. Commentez.