Statistique 3
2019-2020. Exercices: Approche Asymptotique (4)

Exercice 1

Pour $P \in \Delta(\mathbb{R})$, de fonction de répartition F, on note $\mathbb{G}_P \in \mathcal{D}(\mathbb{R})$ le processus Gaussien centré de fonction de covariance $\mathsf{cov}(\mathbb{G}_P(s), \mathbb{G}_P(t)) = \min(F(s), F(t)) - F(s)F(t), \forall s, t \in \mathbb{R}$. Calculer $\mathbb{G}_P$ (et préciser l'ensemble de discontinuité de $\mathbb{G}_P$) lorsque P est:

1. La mesure de Dirac en zéro;
2. La loi de Bernoulli de paramètre p, $p \in]0, 1[$;
3. La loi discrète uniforme sur $\{-1, 0, 1\}$.

Exercice 2 Fonction d'influence

Calculer, lorsqu'elle existe, la fonction d'influence de T en P_0, dans les cas suivants.

1. $T(P)$ est la moyenne de P, pour tout $P \in \Delta_1(\mathbb{R})$ et $P_0 \in \Delta_1(\mathbb{R})$.
2. $T(P) = \inf\{x \in \mathbb{R} : F(x) \geq 1/2\}$, pour tout $P \in \Delta(\mathbb{R})$, où F est la fonction de répartition de P, et $P_0 \in \Delta(\mathbb{R})$ admet une densité strictement positive sur $\mathbb{R}$ par-rapport à la mesure de Lebesgue.
3. $T(P)$ est la variance de P, pour tout $P \in \Delta_2(\mathbb{R})$, où $\Delta_2(\mathbb{R})$ est l'ensemble des mesures de probabilité sur $\mathbb{R}$ admettant un moment d'ordre 2, et $P_0 \in \Delta_2(\mathbb{R})$.
4. $T(P)$ est le plus grand minimiseur $\theta \in \mathbb{R}$ de $\mathbb{E}[\ell_c(X - \theta)]$, où $X \sim P$ et $\ell_c(u) = u^2$ si $|u| \leq c$, $\ell_c(u) = 2c|u| - c^2$ sinon, où $c > 0$, et $P_0 \in \Delta(\mathbb{R})$ est une loi admettant une densité strictement positive sur $\mathbb{R}$ par-rapport à la mesure de Lebesgue.
5. $T(P)$ est le plus grand minimiseur $\theta \in \mathbb{R}$ de $\Phi_P(\theta) = \mathbb{E}[\phi(X, \theta)]$, où $X \sim P$ et $\phi(x, \cdot)$ est convexe sur $\mathbb{R}$ pour tout $x \in \mathbb{R}$, et $P_0 \in \Delta(\mathbb{R})$ est telle que Φ_{P_0} est fortement convexe sur $\mathbb{R}$ et deux fois dérivable en $\theta^* = T(P_0)$.

Exercice 3 Différentiabilité au sens de Hadamard

1. Soit T une application d'un espace vectoriel métrique E sur $\mathbb{R}$ et $a \in E$. Montrer que si T est différentiable au sens de Hadamard en a, alors elle est continue en a.
2. Soit $\Delta_1(\mathbb{R})$ l'ensemble des lois de probabilités sur $\mathbb{R}$ (muni de la tribu Borélienne) admettant un moment d'ordre 1. Pour $P \in \Delta_1(\mathbb{R})$, soit $T(P) = \int_{\mathbb{R}} x \, \mathrm{d}P(x)$. Montrer que T n'est différentiable nullepart au sens de Hadamard.

3. Soit E un espace vectoriel métrique et $a \in E$. Soit T_1 et T_2 deux applications de E dans $\mathbb{R}$ différentiables au sens de Hadamard en a. $T_1 + T_2$ et $T_1 T_2$ sont-elles différentiables au sens de Hadamard en a?

Exercice 4 Différentiabilité de Hadamard tangentielle

Soit E un espace vectoriel métrique et $E_0 \subseteq E$ et D un sous-espace vectoriel de E. Soit $T : E_0 \to \mathbb{R}$ et $a \in E_0$. On dit que T est différentiable au sens de Hadamard en a, tangentiellement à D, si et seulement s'il existe une application linéaire et continue $T'_a : D \to \mathbb{R}$ telle que pour tout $h \in D$, pour toutes suites $(t_n)_{n\geq 1} \subseteq]0, \infty[$, $(h_n)_{n\geq 1} \subseteq E$ telles que $a + t_n h_n \in E_0, \forall n \geq 1$ et $t_n \to 0, h_n \to h$ lorsque $n \to \infty$, on a:

$$\frac{1}{t_n}\left(T(a + t_n h_n) - T(a)\right) \xrightarrow[n\to\infty]{} T'_a(h).$$

Par exemple, T est différentiable au sens de Hadamard en a si et seulement si elle est différentiable au sens de Hadamard en a tangentiellement à tout l'espace E.

1. On note $\mathcal{D}(\mathbb{R})$ l'ensemble de Skorohod sur $\mathbb{R}$ et $\mathcal{D}_1(\mathbb{R})$ l'ensemble des fonctions de répartitions sur $\mathbb{R}$. Pour $F \in \mathcal{D}_1(\mathbb{R})$, soit $T(F) = \inf\{x \in \mathbb{R} : F(x) \geq 1/2\}$. Montrer que $T(F)$ est une médiane de F.
2. Soit $F \in \mathcal{D}_1(\mathbb{R})$ et supposons que F est dérivable en $m = T(F)$, avec $F'(m) > 0$.
   a) Montrer que m est l'unique médiane de F.
   b) Montrer que T est différentiable au sens de Hadamard (par-rapport à la métrique induite par la norme infinie sur $\mathcal{D}(\mathbb{R})$) en F, tangentiellement à l'ensemble des fonctions $h \in \mathcal{D}(\mathbb{R})$ qui sont continues en m.
   c) Soit $X_1, X_2, \ldots$ des variables aléatoires réelles iid de fonction de répartition F. Pour tout $n \geq 1$, soit $\hat{m}_n = T(F_n)$, où F_n est la fonction de répartition empirique de $X_1, \ldots, X_n$. A l'aide de la question précédente, démontrer que $\hat{m}_n$ est asymptotiquement normale, et calculer sa variance asymptotique (*adapter la preuve de la méthode Delta fonctionnelle vue en cours; on pourra démontrer ou admettre que le processus Gaussien $\mathbb{G}_F$ défini comme la limite en distribution, dans $\mathcal{D}(\mathbb{R})$, de $\sqrt{n}(F_n - F)$, est continu en m presque sûrement*).

Exercice 5 M-estimation convexe : une preuve alternative de la normalité asymptotique dans un cas simple

Soit $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ une fonction telle que:

- Pour tout $\theta \in \mathbb{R}$, $\phi(\cdot, \theta)$ est mesurable ;
- Pour tout $x \in \mathbb{R}$, $\phi(x, \cdot)$ est convexe et continûment dérivable ;

- $\forall x \in \mathbb{R}, \forall \theta \in \mathbb{R}, \left|\dfrac{\partial \phi}{\partial \theta}(x,\theta)\right| \leq M$, où $M > 0$ est une constante.

Soit $\bar{\theta} \in \mathbb{R}$. On note E l'ensemble des mesures signées finies ν sur $\mathbb{R}$ telles que $\phi(\cdot,\bar{\theta}) \in L^1(\nu)$, i.e., telles que $\phi(\cdot,\bar{\theta}) \in L^1(\nu^+)$ et $\phi(\cdot,\bar{\theta}) \in L^1(\nu^-)$, où ν^+ et ν^- sont les parties positive et négative, respectivement, de ν (en particulier, ν^+ et ν^- sont deux mesures (positives) finies et $\nu = \nu^+ - \nu^-$).

Pour tout $\nu \in E$, on définit la fonction $\Phi_\nu : \mathbb{R} \to \mathbb{R}$ par

$$\Phi_\nu(\theta) = \int_{\mathbb{R}} \phi(x,\theta)\,\mathrm{d}\nu(x), \quad \forall \theta \in \mathbb{R}.$$

On note à présent E_0 l'ensemble des probabilités $P \in E$ telles que Φ_P admet un unique minimiseur sur $\mathbb{R}$, qu'on note alors $T(P)$.

Pour toute la suite, soit P_0 une probabilité sur $\mathbb{R}$, telle que:

- $P \in E_0$;
- La fonction Φ_{P_0} est deux fois dérivable en $\theta_0 := T(P_0)$, avec $\Phi''_{P_0}(\theta_0) > 0$.

1. Montrer que la définition de l'espace E ne dépend pas du choix de $\bar{\theta} \in \mathbb{R}$, et que la fonction Φ_ν est donc bien définie, pour tout $\nu \in E$.
2. Montrer que pour tout $\nu \in E$, Φ_P est dérivable sur $\mathbb{R}$, avec

$$\Phi'_\nu(\theta) = \int_{\mathbb{R}} \frac{\partial \phi}{\partial \theta}(x,\theta)\,\mathrm{d}\nu(x), \quad \forall \theta \in \Theta.$$

   (On pourra écrire $\nu = \nu^+ - \nu^-$, où ν^+ et ν^- sont les parties positive et négative, respectivement, de ν)

3. Dans cette question, on cherche à démontrer que la fonction T est différentiable au sens de Hadamard en P_0. Soient $Q \in E$, $(t_n)_{n\geq 1}$ une suite de réels strictement positifs tendant vers zéro, et $(Q_n)_{n\geq 1}$ une suite d'éléments de E convergeant vers Q pour la distance de Kolmogorov, telles que pour tout $n \geq 1$, $P_0 + t_n Q_n \in E_0$. Pour tout $n \geq 1$, on note $\theta_n = T(P + t_n Q_n)$.
   a) Montrer que pour tout $n \geq 1$, θ_n satisfait:

   $$\Phi'_P(\theta_n) + t_n \Phi'_{Q_n}(\theta_n) = 0.$$

   b) Montrer que $t_n \Phi'_{Q_n}(\theta_n) \xrightarrow[n\to\infty]{} 0$.
   c) En déduire que $\theta_n - \theta_0 = O(t_n)$, lorsque $n \to \infty$.
   d) En déduire aussi que $\theta_n - \theta_0 = t_n \Phi'_{Q_n}(\theta_0) + o(t_n)$, lorsque $n \to \infty$.
   e) Montrer que l'application $\nu \in E \mapsto \Phi_\nu(\theta_0)$ est continue (par-rapport à la distance de Kolmogorov).
   f) Conclure.

4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. de loi P_0. Pour tout $n \geq 1$, on note P_n la mesure empirique associée à $X_1, \dots, X_n$, i.e.,

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\{X_i\}}.$$

   a) Montrer qu'avec probabilité 1, pour n assez grand, $P_n \in E_0$.
   b) Pour tout $n \geq 1$, on note alors $\hat{\theta}_n = T(P_n)$ si $P_n \in E_0$, $\hat{\theta}_n = 0$ sinon. En utilisant les questions précédentes, montrer que $\hat{\theta}_n$ est un estimateur consistent et asymptotiquement normal de θ_0. Calculer sa variance asymptotique et retrouver le résultat démontré en cours pour les M-estimateurs convexes.

Exercice 6 Une inégalité entre distances

Pour $P, Q \in \Delta(\mathbb{R})$, on pose $d_K(P,Q) = \sup_{x\in\mathbb{R}} |F_P(x) - F_Q(x)|$, où F_P et P_Q sont les fonctions de répartitions de P et Q respectivement, et $d_W(P,Q) = \sup\{\int_{\mathbb{R}} h(x)\,\mathrm{d}P(x) - \int_{\mathbb{R}} h(x)\,\mathrm{d}Q(x) : h \in \mathcal{H}\}$, où $\mathcal{H}$ est l'ensemble des fonctions 1-Lipschitziennes de $\mathbb{R}$ dans $\mathbb{R}$.

Soient P et Q deux probabilités sur $\mathbb{R}$. On suppose que Q admet une densité par-rapport à la mesure de Lebesgue, et que cette densité est majorée par un réel $C > 0$.

Soit $\varepsilon > 0$ quelconque. Pour $x \in \mathbb{R}$, on pose $h_x = \mathbb{1}_{\cdot\leq x}$ et $h_{x,\varepsilon}$ la fonction qui vaut 1 sur $]-\infty, x]$, 0 sur $[x+\varepsilon, \infty[$ et complétée en une fonction continue et affine sur $[x, x+\varepsilon]$. Soit $Y \sim P$ et $Z \sim Q$.

1. Montrer que pour tout $x \in \mathbb{R}$, $\varepsilon h_{x,\varepsilon}$ est 1-Lipschitzienne.
2. Montrer que $\forall x \in \mathbb{R}, \mathbb{E}[h_x(Y)] \leq \mathbb{E}[h_{x,\varepsilon}(Y)]$.
3. Montrer que $\forall x \in \mathbb{R}, \mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)] \leq C\varepsilon$.
4. Montrer que $d_K(P,Q) \leq \dfrac{1}{\varepsilon} d_W(P,Q) + C\varepsilon$.
5. En déduire que $d_K \leq \sqrt{2Cd_W}$.