Examen du 7 Avril 2021

Une attention particulière sera portée à la clarté de la rédaction, ainsi qu'à la justification précise de chaque résultat obtenu, en faisant appel aux résultats du cours (documents autorisés) sans les redémontrer. Les questions marquées du signe $*$ sont moins guidées, et on pourra donner des réponses non détaillées mais clairement argumentées. On pourra admettre la validité des questions non-traitées pour aborder les suivantes.

Problème 1 : Budget d'échantillonage et estimations optimales

Soit $D \subset \mathbb{R}^d$ un domaine borné et $\mu = dx$ la mesure de Lebesgue. On considère $V = L^2(D) = L^2(D, \mu)$ muni de sa norme usuelle $\|v\| = \|v\|_{L^2}$. Soit $V_n \subset V$ un espace de dimension n et $\{L_1, \dots, L_n\}$ une base L^2-orthonormée de V_n. Pour un tirage de m points $x^1, \dots, x^m \in D$ indépendants suivant une loi de probabilité σ, c'est à dire de l'échantillon $S = \{x^1, \dots, x^m\}$ suivant la loi produit $\otimes_{i=1}^m \sigma$, et pour une fonction de poids w donnée, on note

$$\mathbf{G} = \Big(\frac{1}{m}\sum_{i=1}^m w(x^i) L_j(x^i) L_k(x^i)\Big)_{j,k=1,\dots,n}$$

et $E_{1/2}$ l'évènement $\{\|\mathbf{G} - \mathbf{I}\| \le \frac{1}{2}\}$, où $\|\mathbf{M}\|$ désigne la norme spectrale d'une matrice $\mathbf{M}$.

Partie 1. Résultats de base

1. Montrer par un résultat de cours que pour un certain choix de mesure de probabilité σ et de fonction de poids w que l'on précisera, on a

$$m \ge m(n) := \lceil 10 n \ln(4n) \rceil \implies \Pr(E_{1/2}^c) \le \frac{1}{2},$$

où $E_{1/2}^c$ est l'évènement complémentaire $\{\|\mathbf{G} - \mathbf{I}\| > \frac{1}{2}\}$. On utilise dans la suite cette mesure de probabilité σ, cette fonction de poids w et cette valeur $m = m(n)$.

2. Soit u une fonction définie sur D que l'on peut mesurer exactement en des points x^i. On pose $y^i = u(x^i)$ et on note u_n l'estimateur des moindres carrés à poids

$$u_n = \operatorname{argmin}\Big\{\frac{1}{m}\sum_{i=1}^m w(x^i)|y^i - v(x^i)|^2 \ : \ v \in V_n\Big\},$$

lorsque les x^i sont tirés indépendamment suivant la loi σ avec $m = m(n)$. On pose $\tilde{u}_n = u_n$ lorsque $E_{1/2}$ est vérifié et $\tilde{u}_n = 0$ lorsque ce n'est pas le cas. En utilisant les résultats du cours donner une estimation de $\mathbb{E}(\|u - \tilde{u}_n\|^2)$ faisant apparaître l'erreur de meilleure approximation $e_n(u) = \min_{v \in V_n} \|u - v\|$ et un terme additionnel.

Partie 2. Estimateur conditionné

On tire à présente un échantillon $S^* = \{x^1, \dots, x^m\}$ avec $m = m(n)$ suivant la loi produit $\otimes_{i=1}^m \sigma$ *conditionnée* à la réalisation de l'évènement $E_{1/2}$. Pour cela, on tire successivement et indépendamment des échantillons $S_k = \{x_k^1, \dots, x_k^m\}$ suivant $\otimes_{i=1}^m \sigma$ pour $k = 1, 2, \dots$, que l'on rejette tant que $E_{1/2}$ n'est pas vérifié, et on retient $S^* := S_{k^*}$ pour k^* la première valeur telle que $E_{1/2}$ est vérifié. Cette valeur de k^* est donc aléatoire.

3. Montrer que

$$\Pr(k^* > k) \leq 2^{-k},$$

et en déduire que $\mathbb{E}(k^*) \leq 2$.

4. On note u_n^* l'estimateur u_n utilisant les évaluations de u sur l'échantillon $S^* = S_{k^*}$, c'est à dire l'estimateur u_n conditionné à l'évènement $E_{1/2}$. On rappelle que l'espérance d'une variable aléatoire z conditionnée à un évenement E vérifie

$$\Pr(E)\mathbb{E}(z|E) = \mathbb{E}(z\chi_E).$$

Montrer une estimation de la forme

$$\mathbb{E}(\|u - u_n^*\|^2) \leq 2(1 + \delta(n))e_n(u)^2,$$

où $e_n(u) = \min_{v \in V_n} \|u - v\|$ est l'erreur de meilleure approximation et $\delta(n) \to 0$ quand $n \to \infty$. En quoi cette estimation est meilleure que celle de $\mathbb{E}(\|u - \tilde{u}_n\|^2)$ obtenue dans la question 2 ?

5*. La construction de l'estimateur u_n^* nécessite l'évaluation de u aux m points de l'échantillon S^*, mais aussi la construction de tous les échantillons S_k et l'examen des matrices $\mathbf{G}$ correspondantes pour $k = 1, \dots, k^*$ ce qui n'est pas le cas pour l'estimateur $\tilde{u}_n$. Dans quels type d'applications peut-on affirmer qu'il est néammoins intéressant d'utiliser l'estimateur u_n^* plutôt que $\tilde{u}_n$?

Partie 3. Budget d'échantillonage quasi-optimal

On se propose de réduire la taille de l'échantillon S^* à la valeur quasi-optimale $\overline{m} = 2n$ en préservant certaines propriétés de la méthode des moindres carrés quitte à modifier les poids. On rappelle que si $\mathbf{a} = (a_1, \dots, a_n)^T$ est un vecteur colonne de $\mathbb{R}^n$, on peut lui associer une matrice

$$\mathbf{a}\mathbf{a}^T = (a_j a_k)_{j,k=1,\dots,n},$$

qui est symmétrique positive et de rang 1. On va utiliser le résultat suivant dû à Batson, Spielman et Srivastava : soit $m \geq n$ et soit $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ un ensemble de m vecteurs de $\mathbb{R}^n$ tels que

$$\sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^T = \mathbf{I}.$$

Alors pour tout entier $k > 1$ il existe des poids $s_i \geq 0$ tels que

$$\#\{i \,:\, s_i \neq 0\} \leq kn,$$

et

$$\mathbf{I} \leq \sum_{i=1}^m s_i \mathbf{a}_i \mathbf{a}_i^T \leq \Big(\frac{\sqrt{k}+1}{\sqrt{k}-1}\Big)^2 \mathbf{I},$$

au sens de l'ordre des matrices symétriques : $\mathbf{M}_1 \leq \mathbf{M}_2$ si et seulement si $\langle \mathbf{M}_1 \mathbf{x}, \mathbf{x}\rangle \leq \langle \mathbf{M}_2 \mathbf{x}, \mathbf{x}\rangle$ pour tout $\mathbf{x} \in \mathbb{R}^n$, c'est à dire $\mathbf{M}_2 - \mathbf{M}_1$ est symétrique positive.

6*. Montrer que ce résultat implique la variante suivante : si $(\mathbf{a}_i)_{i=1,\dots,m}$ est une suite de vecteurs de $\mathbb{R}^n$ tels que

$$\frac{1}{2}\mathbf{I} \leq \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^T \leq \frac{3}{2}\mathbf{I},$$

il existe des poids $s_i \geq 0$ tels que

$$\#\{i \,:\, s_i \neq 0\} \leq 2n$$

et

$$\frac{1}{2}\mathbf{I} \leq \sum_{i=1}^m s_i \mathbf{a}_i \mathbf{a}_i^T \leq \frac{3}{2}\Big(\frac{\sqrt{2}+1}{\sqrt{2}-1}\Big)^2 \mathbf{I}.$$

Indication : considérer les vecteurs $\tilde{\mathbf{a}}_i = \mathbf{M}^{-1/2}\mathbf{a}_i$ où $\mathbf{M} = \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^T$.

7. En notant qu'il existe une certaine valeur $\delta \in]0,1[$ telle que

$$\frac{1+\delta}{1-\delta} = 3\Big(\frac{\sqrt{2}+1}{\sqrt{2}-1}\Big)^2,$$

montrer qu'avec des poids renormalisés de la forme $t_i = c s_i$ pour un certain $c > 0$, on a

$$(1-\delta)\mathbf{I} \leq \sum_{i=1}^m t_i \mathbf{a}_i \mathbf{a}_i^T \leq (1+\delta)\mathbf{I}.$$

8. On considère l'échantillon $S^* = \{x^1, \dots, x^m\}$ construit dans la question 3, pour lequel l'évènement $E_{1/2}$ est réalisé et on définit les vecteurs

$$\mathbf{a}_i = \Big(\frac{w(x^i)}{m}\Big)^{1/2} (L_1(x^i), \dots, L_n(x^i))^T \in \mathbb{R}^n.$$

Montrer qu'on a $\frac{1}{2}\mathbf{I} \leq \sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^T \leq \frac{3}{2}\mathbf{I}$.

9. En déduire qu'il existe un sous échantillon $\overline{S} \subset S$ de taille $2n$, que l'on notera $\{x^1, \dots, x^{2n}\}$ quitte à re-indexer les x^i, et des poids $w_i \geq 0$ tels que la matrice

$$\overline{\mathbf{G}} = \Big(\frac{1}{2n}\sum_{i=1}^{2n} w_i L_j(x^i) L_k(x^i)\Big)_{j,k=1,\dots,n},$$

vérifie

$$(1-\delta)\mathbf{I} \le \overline{\mathbf{G}} \le (1+\delta)\mathbf{I}.$$

10. On note $\overline{u}_n$ l'estimateur des moindres carrés basé sur l'échantillon $\overline{S}$ et les poids w_i, c'est-à-dire

$$\overline{u}_n = \operatorname{argmin}\Big\{\frac{1}{2n}\sum_{i=1}^{2n} w_i|y^i - v(x^i)|^2 \ : \ v \in V_n\Big\}.$$

On suppose que u et les fonctions de V_n sont contenues dans $\mathcal{C}(\overline{D})$ et que les fonctions constantes sont contenues dans V_n. Montrer qu'on a

$$\|u-\overline{u}_n\| \le Ce_n(u)_\infty, \qquad e_n(u)_\infty = \min_{v\in V_n}\|u-v\|_{L^\infty},$$

en précisant la valeur de C par rapport à δ et $|D| = \mu(D)$.

11*. Pourquoi n'est-il pas clair que l'on puisse espérer obtenir une estimation de la forme

$$\mathbb{E}(\|u-\overline{u}_n\|^2) \le Ce_n(u)^2,$$

si on essaie d'adapter les techniques de preuve du cours avec l'échantillon $\overline{S}$?

12*. Soit $\mathcal{K}$ une partie compacte de $\mathcal{C}(\overline{D})$. En utilisant le résultat établi dans la question 10, établir pour tout $n > 0$ l'inégalité

$$r_{2n}(\mathcal{K})_V \le Cd_{n-1}(\mathcal{K})_{L^\infty},$$

où

$$d_{n-1}(\mathcal{K})_{L^\infty} = \inf_{\dim(V_{n-1})=n-1}\ \max_{u\in\mathcal{K}}\ \min_{v\in V_{n-1}}\|u-v\|_{L^\infty},$$

désigne l'épaisseur de Kolmogorov de $\mathcal{K}$ dans l'espace $\mathcal{C}(\overline{D})$ muni de la norme L^∞, et où

$$r_n(\mathcal{K})_V := \inf_{x^1,\dots,x^n\in D}\ \inf_R\ \sup_{u\in\mathcal{K}}\|u - R(u(x^1),\dots,u(x^n))\|_V,$$

est le nombre de reconstruction optimale défini dans le cours, ipour la norme $\|v\|_V = \|v\|_{L^2}$, et où C est la constante de la question 10.

Problème 2 : Approximation polynomiale en grande dimension

On s'intéresse à l'approximation de fonctions u définies sur $D = [-1,1]^d$ par des polynômes. Pour $\Lambda \subset \mathbb{N}^d$ un ensemble fini de multi-indices on définit l'espace

$$\mathbb{P}_\Lambda := \operatorname{vect}\{x \mapsto x^\nu \ : \ \nu \in \Lambda\},$$

où $x^\nu = \prod_{j=1}^d x_j^{\nu_j}$ pour $x = (x_1,\dots,x_d)$ et $\nu = (\nu_1,\dots,\nu_d)$.

1. On dit que Λ est un ensemble plein si

$$\nu \ \in \Lambda \quad \text{et} \quad \tilde{\nu} \leq \nu \implies \tilde{\nu} \in \Lambda,$$

où $\tilde{\nu} \leq \nu$ signifie que $\tilde{\nu}_j \leq \nu_j$ pour tout $j = 1, \dots, d$. Exhiber par un dessin en dimension $d = 2$ un exemple d'ensemble Λ plein et un exemple d'ensemble Λ qui ne l'est pas. Que signifie être un ensemble plein en dimension $d = 1$?

2. Montrer que si Λ est un ensemble plein, une base orthonormée de $\mathbb{P}_\Lambda$ pour $L^2(D, \mu)$ où μ est la mesure de probabilité uniforme sur D est fournie par les fonctions

$$L_\nu(x) = \prod_{j=1}^{d} L_{\nu_j}(x_j), \qquad \nu \in \Lambda,$$

où les fonctions d'une variable $t \mapsto L_k(t)$ sont les polynômes de Legendre de degré k avec la normalisation

$$\int_{-1}^{1} |L_k(t)|^2 \frac{dt}{2} = 1.$$

3. Pour un ensemble plein Λ, on s'intéresse à la fonction de Christoffel de l'espace $V_n = \mathbb{P}_\Lambda$, avec $n := \dim(V_n) = \#(\Lambda)$, qui est définie par

$$k_n(x) = k_\Lambda(x) := \sum_{\nu \in \Lambda} |L_\nu(x)|^2.$$

On note $K_n = \|k_n\|_{L^\infty(D)}$. On admet que

$$\max_{t \in [-1,1]} |L_k(t)|^2 = |L_k(1)|^2 = 2k + 1.$$

Dans le cas de la dimension $d = 1$, montrer que l'on a exactement

$$K_n = n^2.$$

4*. En dimension $d > 1$, on veut montrer plus généralement que $K_n \leq n^2$ par récurrence sur d. Pour cela on note K la valeur maximale prise par la coordonnée ν_d sur tous les $\nu = (\nu_1, \dots, \nu_d) \in \Lambda$ et on partitionne Λ suivant

$$\Lambda = \bigcup_{k=0}^{K} \widetilde{\Lambda}_k$$

où $\widetilde{\Lambda}_k = \{\nu \in \Lambda \ : \ \nu_d = k\}$ (on pourra faire un dessin en dimension $d = 2$). En utilisant l'hypothèse de récurrence pour la dimension $d - 1$, montrer que

$$\sum_{\nu \in \widetilde{\Lambda}_k} |L_\nu(x)|^2 \leq (2k + 1)\#(\widetilde{\Lambda}_k)^2, \qquad x \in D.$$

Puis, en montrant que

$$\#(\Lambda)^2 = \sum_{k=0}^{K} \Big(\#(\widetilde{\Lambda}_k)^2 + 2 \Big(\sum_{j=0}^{k-1} \#(\widetilde{\Lambda}_j) \Big) \#(\widetilde{\Lambda}_k) \Big),$$

et que $\#(\widetilde{\Lambda}_k)$ décroit avec k, en déduire l'estimation $K_n \leq n^2$.

5. Soit $\rho = (\rho_j)_{j\geq 1}$ une suite de nombres strictement plus grands que 1. On note $\mathcal{K} = \mathcal{K}(\rho, d)$ la classe de fonctions u définies sur $D = [-1,1]^d$ et à valeurs dans $\mathbb{R}$ qui admettent un prolongement holomorphe sur un voisinage du polydisque $\{|z_j| \leq \rho_j \ : \ j = 1, \dots, d\}$ et vérifiant $|u(z)| \leq M$ sur ce polydisque. Rappeler brièvement pourquoi lorsque u appartient à cette classe les coefficients de Taylor vérifient des estimations

$$|u_\nu| \leq M\rho^{-\nu}, \quad u_\nu = \frac{1}{\nu!}\partial^\nu u(0), \qquad \rho^{-\nu} := \prod_{j=1}^{d} \rho_j^{-\nu_j}.$$

6. Soit Λ_n l'ensemble des ν correspondant aux n plus grandes valeurs de $\rho^{-\nu}$. Montrer qu'il s'agit d'un ensemble plein.

7. On suppose que la suite inverse $b = (b_j)_{j\geq 1}$ avec $b_j = \rho_j^{-1}$ appartient à $\ell^p(\mathbb{N})$ pour une valeur $0 < p < 1$. Montrer que la suite $(\rho^{-\nu})_{\nu \in \mathbb{N}^d}$ appartient à $\ell^p(\mathbb{N}^d)$ et que sa norme ℓ^p est bornée par un nombre indépendant de la dimension d.

8. En déduire que pour tout $u \in \mathcal{K}$ et pour tout $n \geq 1$, on a pour $V_n = \mathbb{P}_{\Lambda_n}$ une estimation

$$\min_{v \in V_n} \|u - v\|_{L^\infty} \leq Cn^{-s},$$

où C et s sont indépendants de la dimension d. Que cela signifie-t-il sur l'épaisseur de Kolmogorov $d_n(\mathcal{K})_{L^\infty}$? Comment peut-on interprêter cela du point de vue de la malédiction des grandes dimension ?

9. On tire m échantillons $x^1, \dots, x^m \in D$ suivant la mesure de probabilité uniforme et on considère pour $y^i = u(x^i)$ l'estimateur des moindres carrés

$$u_n = \mathrm{argmin}\Big\{ \frac{1}{m} \sum_{i=1}^{m} |y^i - v(x^i)|^2 \ : \ v \in V_n \Big\},$$

Montrer qu'en prenant $m = \lceil cn^2 \ln(n) \rceil$ pour une constante c bien choisie, et en définissant $\tilde{u}_n = u_n \chi_E$ où E est un évènement que l'on explicitera, on a pour tout $u \in \mathcal{K}$,

$$\mathbb{E}(\|u - \tilde{u}_n\|^2) \leq Cn^{-2s},$$

où $\|v\| = \|v\|_{L^2(D,\mu)}$ et où C et s sont indépendants de la dimension d.

10. Que doit-on faire pour diminuer le budget d'échantillonage et obtenir des résultats du même type ?