Examen TIAD - Master 1 IM
Mardi 12 Mai 2015

Durée 2 h. Notes de cours et calculatrices autorisées.

Livres, notes de TD et codes informatiques NON autorises.

Tran Viet Chi, `chi.tran@univ-lille1.fr`, bureau 316 (bâtiment M3).

Exercice 1 (Evolution de la population américaine par région)
On dispose des populations dans les 50 états U.S., groupés par régions, entre 1920 et 1970 (données de l'U.S. Bureau of the Census, 1979, voir **Annexe 1** où les tailles de populations sont données en milliers d'individus). L'Alaska et Hawaii sont traitées en régions supplémentaires, car ces états n'ont pas appartenus aux U.S. sur l'ensemble de la période et ne sont pas physiquement connectés aux 48 autres états, ce qui entraîne des différences dans leurs évolutions. On pondère ainsi les 48 états de l'étude par une variable `w` qui vaut 1000 (puisque les chiffres sont en milliers d'individus) et les états supplémentaires par -1000. On réalise une analyse de ces données de ce tableau, considéré comme un tableau de contingence, et dont le code est en **Annexe 5**.

1. En vous aidant de l'ensemble des annexes, quels sont les individus, quelles sont les variables et quelle méthode d'analyse des données a-t-on utilisé ? Combien y a-t-il d'individus ?

2. dans les **Annexes 2**, il manque 1 valeur dans le tableau de contingence, 1 valeur dans les contributions au χ^2, 2 valeurs dans les profils lignes et 2 valeurs dans les profils colonnes. Recalculer ces valeurs.

3. Quel est le nombre de degrés de liberté qui a été effacé à l'**Annexe 3** ?

4. Déterminer le vecteur directeur de l'axe principal 1 pour le nuage des modalités lignes, $u_1 \in \mathbb{R}^6$.

5. Retrouver la coordonnée sur l'axe 1 de la région `Great Lakes`, qui est manquante dans l'**Annexe 3**.

6. Commenter le choix de conserver deux axes principaux dans l'étude.

7. Commenter les résultats de l'**Annexe 3**.

Exercice 2 (Estimateur à ondelettes)
On considère une analyse multi-résolution de $L^2(\mathbb{R})$ (muni de sa norme L^2 notée $\|.\|_2$ et du produit scalaire associé $\langle .,. \rangle$) engendrée par une ondelette père φ et une ondelette mère ψ, supposées à supports compacts dans $[0,1]$. Soient $X_1, \dots X_n$ n variables aléatoires réelles i.i.d. dont la loi commune a une densité f par rapport à la mesure de Lebesgue sur

$\mathbb{R}$. On supposera par la suite que f est dans l'ensemble $\mathcal{H}(m, L)$ des fonctions de classe $\mathcal{C}^m$ telles que $\|f^{(m)}\|_2 \leq L$ pour un $m \in \mathbb{N}^*$ et $L > 0$. On note :

$$f(x) = \sum_{k\in\mathbb{Z}} \alpha_{0k}\varphi_{0k}(x) + \sum_{j=0}^{+\infty} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x), \tag{1}$$

où $\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k)$ et $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$. On estime f par :

$$\widehat{f}(x) = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{0k}\varphi_{0k}(x) + \sum_{j=0}^{J_n} \sum_{k\in\mathbb{Z}} \widehat{\beta}_{jk}\psi_{jk}(x), \tag{2}$$

où J_n est un niveau de détail à déterminer en fonction de n et où

$$\widehat{\alpha}_{0k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{0k}(X_i), \qquad \widehat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(X_i).$$

1. Soit $x \in \mathbb{R}$. Montrer que les sommes $\sum_{k\in\mathbb{Z}} \widehat{\alpha}_{0k}\varphi_{0k}(x)$ et $\sum_{k\in\mathbb{Z}} \widehat{\beta}_{jk}\psi_{jk}(x)$ pour j fixé sont des sommes finies. Combien de termes comptent-elles au maximum ?

Soit K le noyau défini par : $K(x, y) = \sum_{k\in\mathbb{Z}} \varphi_{0k}(x)\varphi_{0k}(y)$. On suppose également que φ est choisi tel que pour $N > m$:

$$\forall x \in \mathbb{R}, \forall k \in \{1, \ldots, N\}, \ \int_{\mathbb{R}} (y-x)^k K(x, y)dy = 0, \quad \text{et} \ \int_{\mathbb{R}} K(x, y)dy = 1. \tag{3}$$

$$\exists F \in L^1(\mathbb{R}) \text{ positive}, \forall x, y \in \mathbb{R}, \ |K(x, y)| \leq F(y - x) \text{ et } \int_{\mathbb{R}} |x|^{N+1} F(x)dx < +\infty. \tag{4}$$

2. Calculer $\mathbb{E}(\widehat{f}(x))$, et justifier que c'est la projection, notée $P_{J_n+1}f$, de f sur $V_{J_n+1} = \mathrm{Vect}(\varphi_{J_n+1,k}, \ k \in \mathbb{Z})$.

3. Montrer que la projection $P_{J_n+1}f$ s'écrit aussi :

$$P_{J_n+1}f(x) = \int_{\mathbb{R}} 2^{J_n+1}K(2^{J_n+1}x, 2^{J_n+1}y)f(y)dy.$$

4. Lemme : (peut-être admis si vous bloquez) Montrer que pour une fonction de deux variables positive $h(x, y)$,

$$\int_{\mathbb{R}} dx\Big(\int_{\mathbb{R}} h(x, y)dy\Big)^2 \leq \Big(\int_{\mathbb{R}} \big\|h(., y)\big\|_2 dy\Big)^2.$$

Indication : on pourra montrer (en justifiant) que le membre de gauche s'écrit :

$$\int_{\mathbb{R}} dx\Big(\int_{\mathbb{R}} h(x, y)dy\Big)^2 = \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy\Big(\int_{\mathbb{R}} h(x, z)dz\, h(x, y)\Big)$$

5. Montrer que

$$P_{J_n+1}f(x) - f(x) = \int_0^1 du \int_{\mathbb{R}} dy\, 2^{J_n+1}K(2^{J_n+1}x, 2^{J_n+1}y)\frac{(1-u)^{m-1}}{(m-1)!}(y-x)^m f^{(m)}(x+u(y-x)).$$

En utilisant la quest **4.** , en déduire que :

$$\big|P_{J_n+1}f(x) - f(x)\big| \leq 2^{-(J_n+1)m} \int_0^1 du \int_{\mathbb{R}} dt\ F(t)\frac{(1-u)^{m-1}}{(m-1)!}|t|^m\big|f^{(m)}(x + 2^{-(J_n+1)}tu)\big|.$$

6. Conclure que $\|f - \mathbb{E}(\widehat{f})\|_2 \leq C2^{-(J_n+1)m}$ où C est une constante qu'on précisera.

7. Justifier pourquoi l'opérateur $f \in L^2(\mathbb{R}) \mapsto \big[x \mapsto \int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} \psi_{jk}(x)\psi_{jk}(y)f(y)dy\big]$ est l'opérateur de projection sur $V_{j+1} \setminus V_j$.

8. Montrer, en utilisant **7.** , que

$$\mathbb{E}\big(\|\widehat{f} - \mathbb{E}(\widehat{f})\|_2^2\big) = \int_{\mathbb{R}} \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} Y_i(x)\Big)^2 dx,$$

où $Y_i(x) = 2^{J_n+1}K(2^{J_n+1}x, 2^{J_n+1}X_i) - \mathbb{E}\big(2^{J_n+1}K(2^{J_n+1}x, 2^{J_n+1}X_i)\big)$.

9. Quelle est l'espérance des Y_i ? Majorer leur variance et en déduire que

$$\mathbb{E}\big(\|\widehat{f} - \mathbb{E}(\widehat{f})\|_2^2\big) \leq \frac{2^{J_n+1}}{n} \int_{\mathbb{R}} F^2(v)dv.$$

10. Déduire des questions **6.** et **9.** une majoration de $\sup_{f\in\mathcal{H}(m,L)} \mathbb{E}\big(\|\widehat{f} - f\|_2^2\big)$ en fonction de J_n et n. Quel doit être la forme de J_n en tant que fonction de n. Qu'en déduit-on ? Pouvez-vous commenter ?