Séminaire Mathématique de Béjaïa (LaMOS)
Volume 11, 2012, pp. 69 - 72

16

Application des fonctions splines

Sonia AMROUN et Smail ADJABI

Laboratoire de Modélisation et d'Optimisation des Systèmes (LAMOS)
Université de Béjaïa, Béjaïa 06000, Algérie
Tél. (213) 34 21 51 88

Résumé La régression non paramétrique est un outil statistique permettant de décrire une relation entre une variable dépendante et une variable explicative, sans spécifier la forme de cette relation. L'objectif de ce travail est de comparer deux méthodes non paramétrique, la méthode du noyau et la méthode des fonctions splines, pour estimer la courbe de régression de la moyenne. Nous avons donné l'estimateur spline de lissage et fait la comparaison par simulation sur deux modèles cibles de régression et sur un jeu de données réels. Les résultats numériques et graphiques montrent que la méthode des splines est meilleure que la méthode du noyau. Cependant quand la taille de l'échantillon observé est suffisamment grande les deux méthodes sont équivalentes.

Mots cléfs : Estimation, courbe de régression de la moyenne, noyau, fonction spline, paramètre de lissage, matrice de lissage.

16.1 Introduction

Le terme "fonctions splines" a été introduit par Schoenberg $(1946a,\ b)$, bien que leur origine soit due aux travaux de Whittaker (1923) sur les méthodes de graduation de données. Ces recherches ont poussé Schoenberg (1964) et Reinsch (1967) à dériver les splines de lissage classiques comme classe d'estimateurs. Par ailleurs, Ahlberg, Nilson et Walsh $(1964,\ 1967)$ mentionnent les splines dans un contexte d'interpolation de données sans bruit. Ce n'est donc qu'au début des années 1960 que la théorie des splines s'est développée. Les splines étaient alors considérées comme un outil d'analyse numérique pour modéliser des fonctions. les fonctions splines se sont énormément développées et sont devenues importantes dans différentes branches des mathématiques telles que :

- la théorie de l'approximation.
- L'analyse numérique.
- Les statistiques.

16.2 Spline cubique naturelle

Définition 16.1 *Soient $x_1 < x_2 < \cdots < x_n$ n points d'un intervalle $[a, b]$, une fonction f définie sur $[a, b]$ est une spline cubique si les deux conditions suivantes sont satisfaites :*

1. *Sur chaque intervalle $(a, x_1),\ (x_1, x_2), \cdots, (x_n, b)$, f est un polynôme cubique.*
2. *La fonction f est deux fois continûment differentiable sur $[a, b]$, et donc f et ses dérivées d'ordre 1 et 2 sont continues aux points x_i.*

16.3 Les splines de lissage

Définition 16.2 *Les splines de lissage déterminent la valeur de l'estimateur en minimisant un critère bien précis défini par :*

$$S(m) = \sum_{i=1}^{n} (y_i - m(x_i))^2 \ + \ \lambda \int_a^b (m^{(r)}(x))^2 dx, \tag{16.1}$$

Théorème 16.1 *Supposons que $n \geq 0$ et que le paramètre de lissage λ est positif, alors $\hat{m}_\lambda$ est une spline cubique naturelle telle que :*

$$M = (I + \lambda K)^{-1} Y = A_\lambda Y,$$

et pour toute m dans $W_2^2[a,b]$, on a

$$S(\hat{m}_\lambda) \leq S(m).$$

16.4 Résultats numériques

L'estiamteur spline de lissage est comparé à celui du noyau.On
considérons le modèle de régression $y = m(x) + z$. On prend les modèles de fonctions suivantes :

1- Le modèle 1 **avec** $m_1(x) = x + 0.5\exp(-50(x-0.5)^2)$.
2- Le modèle 2 **avec** $m_2(x) = 4.26(\exp(-3.25x) - 4\exp(-6.5x) + 3\exp(-9.75x))$.
$m_3(x) = \sqrt{x(1-x)}\sin(\frac{2.1\pi}{x+0.5})$.

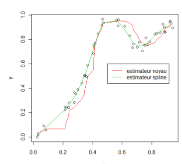

FIGURE 3.1 – Estimation de m_1, $n = 50$

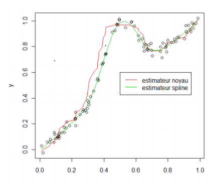

FIGURE 3.2 – Estimation de m_1, $n = 100$

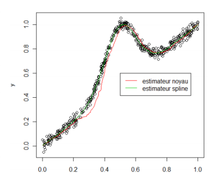

FIGURE 3.4 – Estimation de m_1, $n = 500$

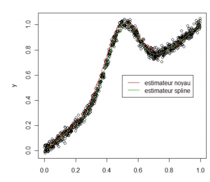

FIGURE 3.5 – Estimation de m_1, $n = 1000$

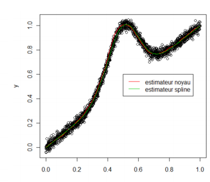

FIGURE 3.6 – Estimation de m_1, $n = 2000$

L'erreur moyenne quadratique associée au modèle m_1

n	MISE noyau	MISE spline
50	0.00639922	0.00061377
100	0.00516139	0.00059959
200	0.00376109	0.00055375
500	0.00311738	0.00053171
700	0.00126655	0.00052573
1000	0.00098935	0.00051186
1500	0.00088708	0.00050892
2000	0.00074196	0.00050615
2500	0.00066193	0.00049508
3000	0.00050997	0.00049338

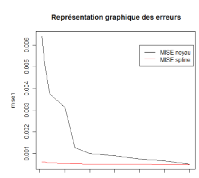

FIGURE 16.1. Estimation de m_1

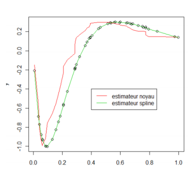

FIGURE 3.8 – Estimation de m_2, $n = 50$

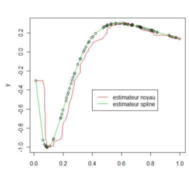

FIGURE 3.9 – Estimation de m_2, $n = 100$

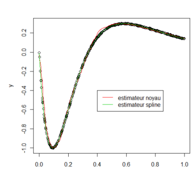

FIGURE 3.11 – Estimation de m_2, $n = 500$

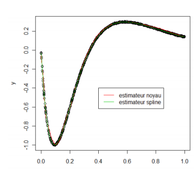

FIGURE 3.12 – Estimation de m_2, $n = 1000$

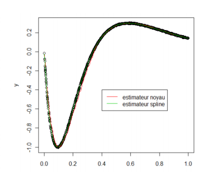

FIGURE 3.13 – Estimation de m_2, $n = 2000$

L'erreur moyenne quadratique associée au modèle m_2

n	MISE noyau	MISE spline
50	0.02555829	0.00000679
100	0.01329819	0.00000643
200	0.00139980	0.00000640
500	0.00072893	0.00000606
700	0.00058020	0.00000600
1000	0.00019613	0.00000595
1500	0.00019445	0.00000580
2000	0.00019248	0.00000575
2500	0.00016873	0.00000575
3000	0.00008621	0.00000564

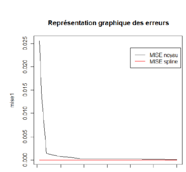

FIGURE 16.2. Estimation de m_2

3- Cas réel

x	y	x	y	x	y
0.083	525	0.25	608	0.5	665
0.75	717	1	745	1.5	803
2	859	3	940	4	1007
5	1065	6	1121	7	1183
8	1238	9	1298	10	1348
10.5	1369	11	1391	11.5	1422
12	1470	12.5	1525	13	1578
13.5	1638	14	1664	14.5	1692
15	1708	15.5	1723	16	1727
16.5	1727	17	1727	18	1729
19	1738	20	1738		

TABLE 16.1. Données numériques de croissance

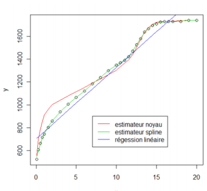

FIG. 4 – Estimation de la courbe de croissance pour la méthode du noyau, la méthode des fonctions splines et par la régression linéaire.

L'erreur moyenne quadratique associée au cas réel

MISE noyau	MISE spline	MISE régression linéaire
4427.31	31.87724	5759.474

TABLE 3.5 – Erreur moyenne quadratique associée

FIGURE 16.3. Estimation de la courbe de croissance et erreur associée par les deux méthodes.

Références

1. W-R. HEINZELMAN, A. CHANDRAKASAN, and H. BALAKRISHNAN. Energy-efficient communication protocol for wireless sensor networks. In *Proceedings of the IEEE Hawaii International Conference on System Sciences*, pages 3005–3014, Janvier 2000.
2. S. LEE and H.SIN. An energie-efficient distributed unequal clustering protocol for wirless sensor networks. *Proccedings of word Academy of Science Engineering and Technology*, 36, Decembre 2008.
3. N.KHOULALENE. *Regroupement avec Equilibrage de charge dans les Réseaux de Capteurs sans Fil*. Mémoire de magistère en informatique, Université de Béjaia, Algérie, Juin 2007.
4. M. QIN and R. ZIMMERMANN. Vca : An energy-efficient voting-based clustering algorithm for sensor networks. *Journal of Universal Computer Science*, 13(1) :87–109, Janvier 2007.