Séminaire Mathématique de Béjaïa (LaMOS)
Volume 11, 2012, pp. 79 - 82

19

Régression non paramétrique à noyau : Cas multidimensionnel par la méthode du noyau

Karima LAGHA

LAboratoire de Modélisation et d'Optimisation des Systèmes LAMOS
Université de Béjaïa 06000, Algérie.

Résumé Dans ce travail, nous considérons le problème d'estimation non paramétrique par la méthode du noyau de la fonction de régression dans le cas multidimensionnel. Des propriétés statistique de l'estimateur sont déterminées dans les cas général et linéaire.
Nous avons traité le problème de la convergence ponctuelle et uniforme presque complète de l'estimateur et nous avons obtenu la vitesse de convergence de l'estimateur.

Mots clés : Régression, estimation non paramétrique , convergence presque complète, propriétés statistiques.

19.1 Introduction

L'estimation à noyau de la densité est une technique d'estimation nonparamétrique très utilisée.
L'estimateur à noyau de la densité a été d'abord introduit dans le cas univarié, entre 1956 et 1960, par Parzen-Rosenblatt ensuite a été étendu au cas multivarié entre 1990 et 2000 [voir Simonoff]. Dans le cas multidimensionnel, l'estimateur de la densité fait intervenir une fonction noyau K (dont le choix est arbitraire) et une matrice de lissage H (dont le choix est important) (voir Wand et Jones 1995).
Cet estimateur s'écrit sous la forme suivante

$$\hat{f}_H(X) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_H(X - X_i), \tag{19.1}$$

où

- $X = (x_1, \ldots, x_d)^T, X_i = (x_{i1}, \ldots, x_{id})^T, i = 1, \ldots, n$ sont des vecteurs d-dimensionnels ;
- H est la matrice de lissage (matrice des fenêtres) d'ordre d symétrique et définie positive ;
- $\mathcal{K}$ est la fonction noyau qui est une densité multivariée symétrique ;
- $\mathcal{K}_H(X) = det(H)^{-1/2}\mathcal{K}(H^{-1/2}X)$.

L'espérance conditionnelle de Y sachant X, notée $r(X) = E(Y/X)$, a un intérêt important dans l'estimation car elle nous informe sur la liaison entre Y et X.
Dans le cas univarié, l'estimateur à noyau de la régression a été introduit en 1964 par Nadaraya-Watson. Des résultats sur la convergence presque complète, pour une estimation ponctuelle de la régression r et pour une estimation uniforme sur un compact ont été obtenus. Des résultats analogues de convergence en moyenne quadratique et de convergence en moyenne quadratiques intégrées ont été obtenus (voir Ferraty et Vieu 2002).
Dans le cas multidimentionnel, l'espérance conditionnelle s'écrit

$$E(Y/X) = E(Y/X_1, X_2, \ldots, X_d) = r(X)$$

où $X = (X_1, \ldots, X_d)^T$.
La version multivariée de l'estimateur de Nadaraya-Watson, est donnée par :

$$\hat{r}_{NW}(X) = \frac{\sum_{i=1}^{n} Y_i\mathcal{K}(\frac{X_i - X}{h})}{\sum_{i=1}^{n}\mathcal{K}(\frac{X_i - X}{h})} \tag{19.2}$$

où $\mathcal{K}$ est une fonction de R^d dans R et $h = h(n)$ paramètre réel strictement positif.
Des résultats analogues au cas unidimensionnel, adaptés au cas multiple, sur la convergence presque complète, pour une estimation ponctuelle et uniforme de la régression r ont été obtenus (voir Ferraty et Vieu 2002).
On considère la généralisation de l'estimateur de Nadaraya-Watson, au cas d'une matrice de lissage donnée par :

$$\hat{r}_H(X) = \frac{\sum_{i=1}^{n} Y_i \mathcal{K}_H(X_i - X)}{\sum_{i=1}^{n} \mathcal{K}_H(X_i - X)} \tag{19.3}$$

Où $X_1, X_2, \ldots, X_n$ est un échantillon issu de X de dimension d de fonction densité f et $Y_1, Y_2, \ldots, Y_n$ un n-échantillon issu de Y.

19.2 Propriétés statistique de la régression

19.2.1 Biais et variance asymptotiques

Théorème 19.1 *: [Ruppert et Wand (1994)]*

$$\begin{aligned} \mathit{bias}\left\{\hat{r}_H/X_1, \ldots, X_n\right\} \simeq \mu_2(\mathcal{K}) \frac{\nabla_r(x)^T H H^T \nabla_f(x)}{f_X(x)} + \\ \frac{1}{2}\mu_2(\mathcal{K}) tr\left\{H^T \mathcal{H}_r(x) H\right\}. \end{aligned} \tag{19.4}$$

$$\mathit{Var}\left\{\hat{r}_H/X_1, \ldots, X_n\right\} \simeq \frac{1}{ndet(H)} ||\mathcal{K}||_2^2 \frac{\sigma^2(x)}{f_X(x)} \tag{19.5}$$

dans l'intérieur du support de f_X.

19.2.2 Cas particulier : la régression linéaire

Théorème 19.2

$$\mathit{bias}\left\{\hat{r}_{1,H}/X_1, \ldots, X_n\right\} \simeq \frac{1}{2}\mu_2(\mathcal{K}) tr\left\{H^T \mathcal{H}_r(x) H\right\}. \tag{19.6}$$

$$\mathit{Var}\left\{\hat{r}_{1,H}/X_1, \ldots, X_n\right\} \simeq \frac{1}{ndet(H)} ||\mathcal{K}||_2^2 \frac{\sigma^2(x)}{f_X(x)}. \tag{19.7}$$

dans l'intérieur du support de f_X.

19.3 Convergence presque complète de l'estimateur

19.3.1 Convergence ponctuelle :

Sous les hypothèses suivantes :

$$r \text{ et } f \text{ sont 2 fois continûment différentiables autour de } x, \tag{19.8}$$

$x \in \mathbb{R}^d$ est un point fixé tel que

$$f(x) > 0. \tag{19.9}$$

$$\lim_{n\to\infty} H = O_{(d,d)}, \lim_{n\to\infty} \frac{n|H|^{1/2}}{\log n} = \infty \text{ et } \lim n^{-1}|H|^{-1/2} = 0 \tag{19.10}$$

$$\mathcal{K} \text{ est borné, intégrable et à support compact} \tag{19.11}$$

$$\int \mathcal{K}(u)\, du = 1; \int u\mathcal{K}(u)\, du = 0; \int uu^T \mathcal{K}(u)\, du = \mu_2(\mathcal{K}) I_d. \tag{19.12}$$

Où $\mu_2(\mathcal{K}) = \int {u_i}^2 \mathcal{K}(u)\, du$ est finie et indépendante de i. Pour alléger les démonstrations, on considère :

$$|Y| < M < \infty \tag{19.13}$$

où M est une constante.

19.3.2 Vitesse de convergence presque complète ponctuelle multivariée sous condition de dérivabilité.

Considérons le modèle (19.8) et supposons que les conditions (19.9),(19.10),(19.11),(19.12) et (19.13) sont satisfaites. On a

$$\hat{r}_H(x) - r(x) = O(tr(H)) + O\left(\sqrt{\frac{\log n}{n|H|^{-1/2}}}\right),\ p.co.$$

19.3.3 Elements de démonstration

$$\hat{r}_H(x) - r(x) = \frac{\hat{g}(x)}{\hat{f}(x)} \tag{19.14}$$

où

$$\hat{g}(x) = n^{-1}\sum_{i=1}^{n} Y_i \mathcal{K}_H(X_i - x) = n^{-1}\sum_{i=1}^{n} |H|^{-1/2} Y_i \mathcal{K}(H^{-1/2}(X_i - x)),$$

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n} \mathcal{K}_H(X_i - x) = n^{-1}\sum_{i=1}^{n} |H|^{-1/2} \mathcal{K}(H^{-1/2}(X_i - x)).$$

On utilise

$$\hat{r}_H(x) - r(x) = \frac{\hat{g}(x) - g(x)}{\hat{f}(x)} + (f(x) - \hat{f}(x))\frac{r(x)}{\hat{f}(x)}, rf = g. \tag{19.15}$$

Le résultat du théorème découle des résultats suivants :

$$E[\hat{f}(x)] - f(x) = \frac{1}{2}\mathrm{tr}\{H\mathcal{H}_f(x)\}\mu_2(K) + o(\mathrm{tr}(H)) = O(\mathrm{tr}(H)) \tag{19.16}$$

$$E[\hat{g}(x)] - g(x) = \frac{1}{2}\mathrm{tr}\{H\mathcal{H}_g(x)\}\mu_2(K) + o(\mathrm{tr}(H)) = O(\mathrm{tr}(H)) \tag{19.17}$$

$$E[\hat{g}(x)] - \hat{g}(x) = O\left(\frac{\log(n)}{n|H|^{-1/2}}\right) \quad P.co. \tag{19.18}$$

$$E[\hat{f}(x)] - \hat{f}(x) = O\left(\frac{\log(n)}{n|H|^{-1/2}}\right) \quad P.co. \tag{19.19}$$

et

$$\exists\delta > 0, \sum_{n=1}^{\infty} P[\hat{f}(x) \leq \delta] < \infty$$

19.4 Convergence presque complète de l'estimateur

19.4.1 Convergence uniforme :

On considère, en plus, les hypothèses suivantes :

$$r \text{ et } f \text{ sont 2 fois continûment différentiables autour de } S, \tag{19.20}$$

$$\exists\theta > 0, \inf_{x\in S} f(x) > \theta \tag{19.21}$$

$$\exists\beta > 0, \exists C < \infty, \forall x \in S, \forall y \in S, |K(x) - K(y)| \leq C||x - y||^{\beta} \tag{19.22}$$

19.4.2 Vitesse de convergence presque complète uniforme multivariée sous condition de dérivabilité.

Considérons le modèle (19.20) et supposons que les conditions (19.10),(19.11),(19.12),(19.13),(19.21) et (19.22) sont satisfaites. On a

$$\sup_{x\in S}|\hat{r}_H(x) - r(x)| = O(tr(H)) + O\left(\sqrt{\frac{\log n}{n|H|^{-1/2}}}\right),\ p.co.$$

19.4.3 Elements de démonstration

On utilise le découpage suivant :

$$\sup_{x\in S}|\hat{r}_H(x)-r(x)| \leq \frac{\sup_{x\in S}|\hat{g}(x)-g(x)|}{\inf_{x\in S}|\hat{f}(x)|} + \sup_{x\in S}|f(x)-\hat{f}(x)|\frac{\sup_{x\in S}|r(x)|}{\inf_{x\in S}|\hat{f}(x)|}. \tag{19.23}$$

D'où

$$\sup_{x\in S}|E[\hat{g}(x)]-g(x)| = O(\mathrm{tr}(H))$$
$$\sup_{x\in S}|E[\hat{f}(x)]-f(x)| = O(\mathrm{tr}(H))$$
$$\sup_{x\in S}|E[\hat{g}(x)]-\hat{g}(x)| = O\left(\frac{\log(n)}{n|H|^{-1/2}}\right) \quad P.co.$$
$$\sup_{x\in S}|E[\hat{f}(x)]-\hat{f}(x)| = O\left(\frac{\log(n)}{n|H|^{-1/2}}\right) \quad P.co.$$

et

$$\exists\delta>0, \sum_{n=1}^{\infty} P[\inf_{x\in S}|\hat{f}(x)|\leq\delta] < \infty$$

Références

1. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* **29**, 832-837.
2. Parzen, E. (1962). On estiamtes models of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065-1076.
3. Simonoff, J.S. (1996). *Smoothing methods in statistics.* ISBN : 0387947167.
4. Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. *ISBN : 0412552701*, 36-39.
5. Deheuvels, P. (1977), Estimation non paramétrique de la densité par histogrammes généralisés (II) *Publications de l'institut de Statistique de l'Université de Paris*, *22*, 1-23.
6. Singh, R.S. (1976), Nonparametric Estimation of mixed partial derivatives of a multivariate density. *Journal Multivariate Annals*, *6*, 111-122.
7. Ruppert, D. et Wand, M.P. (1994), Multivariate locally weighted least squares regression. *Annals of Statistics*, *22(3)*, 1346-1370.
8. Ferraty, F. et Vieu, P. (2002/2003), Statistique Fonctionnelle : Modèles Non- paramétriques de régression. *Annals of Statistics*, *Notes de cours de DEA.*