Sur une généralisation de la méthode PCO

Nicolas Marie

Sur une généralisation de la méthode PCO

Nicolas MARIE [1]

[1] *Laboratoire Modal'X, Université Paris Nanterre, Nanterre, France*
nmarie@parisnanterre.fr

Résumé. Pour le modèle de régression $Y = b(X)+\sigma(X)\varepsilon$, où la loi de X a une densité f, l'exposé portera sur une inégalité d'oracle pour un estimateur de bf, faisant intervenir un noyau au sens de Lerasle et al. (2016), sélectionné via la méthode PCO. En plus de la sélection de fenêtre pour des estimateurs à noyaux (au sens usuel) comme dans Lacour, Massart et Rivoirard (2017) ou Comte et Marie (2020), la méthode couvre la sélection de dimension pour des estimateurs par projection de f et bf dans le cas anisotrope (cf. Halconruy et Marie (2020)).

Mots-clés. Estimateurs non-paramétriques ; Estimateurs par projection ; Sélection de modèle ; Régression.

Abstract. In the regression model $Y = b(X) + \sigma(X)\varepsilon$, where X has a density f, this talk deals with an oracle inequality for an estimator of bf, involving a kernel in the sense of Lerasle et al. (2016), selected via the PCO method. In addition to the bandwidth selection for kernel-based estimators as in Lacour, Massart and Rivoirard (2017) or Comte and Marie (2020), the dimension selection for anisotropic projection estimators of f and bf is covered (see Halconruy and Marie (2020)).

Keywords. Nonparametric estimators ; Projection estimators ; Model selection ; Regression model.

Résumé détaillé

Soient $n \in \mathbb{N}^*$ variables aléatoires $(X_1, Y_1), \ldots, (X_n, Y_n)$ à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ ($d \in \mathbb{N}^*$), de même loi de probabilité absolument continue par rapport à la mesure de Lebesgue, et

$$\widehat{s}_{K,\ell}(n; x) := \frac{1}{n} \sum_{i=1}^{n} K(X_i, x)\ell(Y_i) \ ; \ x \in \mathbb{R}^d,$$

où $\ell : \mathbb{R} \to \mathbb{R}$ est une fonction borélienne et K est une application symétrique de $\mathbb{R}^d \times \mathbb{R}^d$ dans $\mathbb{R}$. Il s'agit d'un estimateur de la fonction $s : \mathbb{R}^d \to \mathbb{R}$ définie par

$$s(x) := \mathbb{E}(\ell(Y_1)|X_1 = x)f(x) \ ; \ \forall x \in \mathbb{R}^d,$$

où f est une densité de X_1. Pour $\ell = 1$, $\widehat{s}_{K,\ell}(n; .)$ est l'estimateur de f étudié dans Lerasle et al. [10], généralisant l'estimateur de Parzen-Rosenblatt et les estimateurs par

projection (cf. Parzen [12], Rosenblatt [13], Tsybakov [14], etc.), mais pour $\ell \neq 1$, il généralise des estimateurs utilisés en régression non-paramétrique. Supposons que pour tout $i \in \{1, \ldots, n\}$,

$$Y_i = b(X_i) + \sigma(X_i)\varepsilon_i \tag{1}$$

où ε_i est une variable aléatoire centrée et de variance 1, indépendante de X_i, et $b, \sigma : \mathbb{R}^d \to \mathbb{R}$ sont des fonctions boréliennes.

- Si $\ell = \mathrm{Id}_{\mathbb{R}}$, k est un noyau symétrique et

$$K(x', x) = \prod_{q=1}^{d} \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) \text{ avec } h_1, \ldots, h_d > 0 \tag{2}$$

  pour tous $x, x' \in \mathbb{R}^d$, alors $\widehat{s}_{K,\ell}(n; .)$ est le numérateur de l'estimateur de Nadaraya-Watson de la fonction de régression b (cf. Nadaraya [11] et Watson [16]). Plus précisément, $\widehat{s}_{K,\ell}(n; .)$ est un estimateur de $s = bf$ car ε_1 est indépendante de X_1 et $\mathbb{E}(\varepsilon_1) = 0$. Si $\ell \neq \mathrm{Id}_{\mathbb{R}}$, alors $\widehat{s}_{K,\ell}(n; .)$ est le numérateur de l'estimateur étudié dans Einmahl et Mason [4, 5].

- Si $\ell = \mathrm{Id}_{\mathbb{R}}$, $\mathcal{B}_{m_q} = \{\varphi_1^{m_q}, \ldots, \varphi_{m_q}^{m_q}\}$ ($m_q \in \mathbb{N}^*$ et $q \in \{1, \ldots, d\}$) est une famille orthonormée de $\mathbb{L}^2(\mathbb{R})$ et

$$K(x', x) = \prod_{q=1}^{d} \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q)\varphi_j^{m_q}(x'_q) \tag{3}$$

  pour tous $x, x' \in \mathbb{R}^d$, alors $\widehat{s}_{K,\ell}(n; .)$ est un estimateur par projection sur $\mathcal{S} = \mathrm{span}(\mathcal{B}_{m_1} \otimes \cdots \otimes \mathcal{B}_{m_d})$ de $s = bf$.

Enfin, si $b = 0$ dans le Modèle (1), pour tout $i \in \{1, \ldots, n\}$,

$$Y_i = \sigma(X_i)\varepsilon_i \tag{4}$$

Si $\ell(x) = x^2$ pour tout $x \in \mathbb{R}$, alors $\widehat{s}_{K,\ell}(n; .)$ est un estimateur de $s = \sigma^2 f$.

Ces dernières années, plusieurs méthodes de sélection de fenêtre pour l'estimateur de Parzen-Rosenblatt ($\ell = 1$ et K défini par (2)) ont été étudiées. D'une part, la méthode de Goldenshluger-Lepski, introduite dans [6], très satisfaisante sur le plan théorique, mais pas totalement sur le plan numérique (cf. Comte and Rebafka [3]). D'autre part, dans [9], Lacour, Massart et Rivoirard ont proposé la méthode PCO (Penalized Comparison to Overfitting) et démontré une inégalité d'oracle en usant d'une inégalité de concentration pour les U-statistiques due à Houdré et Reynaud-Bouret [8]. Avec Varet, les auteurs de [9] ont établi l'efficacité de la méthode PCO sur le plan numérique dans Varet et al. [15]. Toujours dans le contexte de l'estimation de densité, la méthode PCO a été étendue à

la sélection de fenêtres pour l'estimateur récursif de Wolverton-Wagner dans Comte et Marie [1].

L'exposé portera sur l'extension suivante de la méthode PCO à la sélection de l'application symétrique K pour l'estimateur $\widehat{s}_{K,\ell}(n;.)$:

$$\widehat{K} \in \arg\min_{K\in\mathcal{K}_n}\left\{\|\widehat{s}_{K,\ell}(n;.)-\widehat{s}_{K_0,\ell}(n;.)\|_2^2+\frac{2}{n^2}\sum_{i=1}^{n}\langle K(.,X_i),K_0(.,X_i)\rangle_2\ell(Y_i)^2\right\}$$

où

$$K_0 \in \arg\max_{K\in\mathcal{K}_n}\left\{\sup_{x\in\mathbb{R}^d}|K(x,x)|\right\}$$

est un *noyau maximal* et $\mathcal{K}_n$ désigne une famille d'applications symétriques, par exemple de la forme (2) ou de la forme (3). Quelques expériences numériques seront présentées, puis il sera établi que si $\mathbb{E}(\exp(\alpha|\ell(Y_1)|))<\infty$, alors

$$\begin{aligned}\mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n;.)-s\|_2^2) \;\leqslant\; (1+\theta)\min_{K\in\mathcal{K}_n}\mathbb{E}(\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2)\\ +\frac{\mathfrak{c}}{\theta}\left(\|\mathbb{E}(\widehat{s}_{K_0,\ell}(n;.))-s\|_2^2+\frac{\log(n)^5}{n}\right).\end{aligned}$$

Pour le modèle (1), l'exposé s'achèvera sur une borne de risque pour l'estimateur quotient

$$\frac{\widehat{s}_{\widehat{K}_1,\mathrm{Id}_\mathbb{R}}(n;.)}{\widehat{s}_{\widehat{K}_2,1}(n;.)}\text{ de }b(.),$$

où

$$\widehat{K}_1 \in \arg\min_{K\in\mathcal{K}_n}\left\{\|\widehat{s}_{K,\mathrm{Id}_\mathbb{R}}(n;.)-\widehat{s}_{K_0,\mathrm{Id}_\mathbb{R}}(n;.)\|_2^2+\frac{2}{n^2}\sum_{i=1}^{n}\langle K(.,X_i),K_0(.,X_i)\rangle_2 Y_i^2\right\}$$

et

$$\widehat{K}_2 \in \arg\min_{K\in\mathcal{K}_n}\left\{\|\widehat{s}_{K,1}(n;.)-\widehat{s}_{K_0,1}(n;.)\|_2^2+\frac{2}{n^2}\sum_{i=1}^{n}\langle K(.,X_i),K_0(.,X_i)\rangle_2\right\}.$$

Ces résultats sont issus de deux travaux, l'un en collaboration avec Fabienne Comte [2], et l'autre avec Hélène Halconruy [7].

References

[1] F. Comte et N. Marie. *Bandwidth Selection for the Wolverton-Wagner Estimator.* Journal of Statistical Planning and Inference 207, 198-214, 2020.

[2] F. Comte et N. Marie. *On a Nadaraya-Watson Estimator with Two Bandwidths.* A paraître dans Electronic Journal of Statistics, 2021.

[3] F. Comte et T. Rebafka. *Nonparametric Weighted Estimators for Biased Data.* Journal of Statistical Planning and Inference 174, 104-128, 2016.

[4] U. Einmahl et D.M. Mason. *An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators.* Journal of Theoretical Probability 13, 1-37, 2000.

[5] U. Einmahl et D.M. Mason. *Uniform in Bandwidth Consistency of Kernel-Type Function Estimators.* Annals of Statistics 33, 1380-1403, 2005.

[6] A. Goldenshluger et O. Lepski. *Bandwidth Selection in Kernel Density Estimation: Oracle Inequalities and Adaptive Minimax Optimality.* The Annals of Statistics 39, 1608-1632, 2011.

[7] H. Halconruy et N. Marie. *Kernel Selection in Nonparametric Regression.* A paraître dans Mathematical Methods of Statistics, 2021.

[8] C. Houdré et P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-statistics of Order Two.* Stochastic Inequalities and Applications, vol. 56 of Progr. Proba., 55-69, Birkhauser, 2003.

[9] C. Lacour, P. Massart et V. Rivoirard. *Estimator Selection: a New Method with Applications to Kernel Density Estimation.* Sankhya A 79, 2, 298-335, 2017.

[10] M. Lerasle, N.M. Magalhaes et P. Reynaud-Bouret. *Optimal Kernel Selection for Density Estimation.* High dimensional probabilities VII: The Cargese Volume, vol. 71 of Prog. Proba., 435–460, Birkhauser, 2016.

[11] E.A. Nadaraya. *On a Regression Estimate.* (Russian) Verojatnost. i Primenen. 9, 157-159, 1964.

[12] E. Parzen. *On the Estimation of a Probability Density Function and the Mode.* The Annals of Mathematical Statistics 33, 1065-1076, 1962.

[13] M. Rosenblatt. *Remarks on some Nonparametric Estimates of a Density Function.* The Annals of Mathematical Statistics 27, 832-837, 1956.

[14] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2009.

[15] S. Varet, C. Lacour, P. Massart et V. Rivoirard. *Numerical Performance of Penalized Comparison to Overfitting for Multivariate Density Estimation.* Preprint, 2020.

[16] G.S. Watson. *Smooth Regression Analysis.* Sankhya A 26, 359-372, 1964.