Application de la Quantification Optimale à la méthode SIR

Romain Azaïs, Anne Gégout-Petit, Jérôme Saracco

Application de la Quantification Optimale à la méthode SIR

Romain Azaïs, Anne Gégout-Petit & Jérôme Saracco

INRIA Equipe-Projet CQFD
Institut de Mathématiques de Bordeaux UMR CNRS 5251
Université de Bordeaux

Abstract

We consider a semiparametric regression model involving a d-dimensional quantitative explanatory variable X and including a dimension reduction of X via an index $\beta' X$. In this model, the main goal is to estimate the euclidean parameter β and to predict the real response variable Y conditionally to X. Our approach is based on Sliced Inverse Regression (SIR) method and optimal quantization in $\mathbf{L}^p$-norm. We show the convergence of the proposed estimators of β and of the conditional distribution. Simulation studies show the good numerical behavior of the proposed estimators for finite sample size.

Keywords Optimal quantization, Semiparametric regression model, Sliced Inverse Regression (SIR), Dimension reduction

Résumé

On considère un modèle de régression semi-paramétrique dans lequel la variable explicative X est de dimension d et intervient seulement via sa projection $\beta' X$. Pour ce modèle, le problème majeur est l'estimation du paramètre vectoriel β et la prédiction de la variable réponse réelle Y conditionnellement à X. Notre approche est fondée sur la méthode SIR (pour Sliced Inverse Regression) et la quantification optimale en norme $\mathbf{L}^p$. Nous démontrons la convergence des estimateurs proposés de β et de la loi conditionnelle. Des simulations montrent le bon comportement numérique des estimateurs.

Mots-clés Quantification optimale, Modèle de régression semi-paramétrique, Sliced Inverse Regression (SIR), Réduction de dimension

1 Introduction

On s'intéresse au modèle de régression semi-paramétrique

$$Y = f(\beta' X, \epsilon) \tag{1}$$

où la variable d'intérêt Y est liée via une fonction inconnue f à la variable aléatoire X d-dimensionnelle seulement par sa projection $\beta' X$. La variable aléatoire ϵ représente une erreur indépendante de X.

Il existe de nombreuses méthodes pour identifier et estimer l'espace engendré par le vecteur β : SIR (pour Sliced Inverse Regression, méthode introduite par Duan et Li, 1991), principal hessians directions (voir par exemple Cook, 1998) ou SAVE (pour Sliced Average Variance Estimation, voir Cook, 2000). On se focalise ici sur l'approche SIR, devenue la méthode standard dans ce domaine. Peu de résultats combinent l'estimation de la direction de β et celle de la fonction de lien f dans la littérature (voir par exemple Gannoun et al., 2004).

L'idée principale de SIR est de diviser l'espace dans lequel Y prend ses valeurs en tranches et de considérer la matrice de covariance de la moyenne conditionnelle de X dans chacune de ces tranches. Posons $\Sigma = \mathbf{V}\text{ar}(X)$ et supposons que l'espace d'arrivée de Y est partitionné en H tranches S_h. $\hat{Y} = \{h : Y \in S_h\}$ est une version discrète de la variable continue Y. Sous deux hypothèses (la première portant sur la loi de X et la seconde nous assurant que le cas considéré n'est pas pathologique), on peut montrer que β est un vecteur propre principal de la matrice $\Sigma^{-1}\hat{\Gamma}$ où

$$\hat{\Gamma} = \mathbf{V}\text{ar}(\mathbf{E}[X|\hat{Y}]) = \sum_{h=1}^{H} \mathbf{P}(Y \in S_h)\big(\mathbf{E}[X|Y \in S_h] - \mathbf{E}[X]\big)\big(\mathbf{E}[X|Y \in S_h] - \mathbf{E}[X]\big)'.$$

On se propose d'approcher la matrice $\hat{\Gamma}$ par quantification optimale en norme $\mathbf{L}^p$: l'idée est de remplacer la variable X par sa projection optimale $\hat{X}^N$. On montre ensuite la convergence de la suite des matrices $\hat{\Gamma}_N = \mathbf{V}\text{ar}(\mathbf{E}[\hat{X}^N|\hat{Y}])$ vers $\hat{\Gamma}$ ce qui entraine la convergence de la suite des vecteurs propres principaux vers la direction principale de $\Sigma^{-1}\hat{\Gamma}$, c'est-à-dire la direction de β. La quantification optimale est souvent utilisée pour mener des calculs d'approximation d'espérances conditionnelles, voir de Saporta et al. (2010a et 2010b). Ici, nous proposons d'utiliser à nouveau cet outil pour prédire la variable Y sachant X, c'est-à-dire sachant $\beta' X$ sous le modèle (1). Nous montrons que l'erreur de prédiction tend vers 0 quand les nombres de quantifieurs tendent vers l'infini, et nous donnons une vitesse de convergence.

2 Prédiction pour un modèle de régression non-paramétrique

Avant d'aborder la prédiction dans un modèle de régression non-paramétrique, nous présentons d'abord la méthode de quantification optimale en norme $\mathbf{L}^p$.

2.1 Quantification optimale d'un vecteur X

En théorie des probabilités, la quantification optimale consiste à déterminer la meilleure approximation (en norme $\mathbf{L}^p$) d'une loi continue par une loi discrète avec un nombre fixé N de points de masse non-nulle. Cette méthode a d'abord été utilisée dans le cas uni-dimensionnel, puis multi-dimensionnel (voir Zador, 1963 et Pagès, 1998), puis pour résoudre certains problèmes d'arrêt optimal, de contrôle ou de filtrage (voir par exemple Pagès et al., 2004, Bally et al., 2005). Plus récemment, de Saporta et al. (2010a et 2010b) ont utilisé la quantification optimale afin de développer un outil numérique pour l'arrêt optimal de processus markoviens déterministes par morceaux avec une application à l'optimisation de la maintenance.

Soit X un vecteur aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et à valeurs dans $\mathbf{R}^d$. Nous supposons que la norme $\mathbf{L}^p$ de X est finie pour un certain $p \geq 1$, c'est-à-dire $\|X\|_p = \mathbf{E}[|X|^p]^{1/p}$ est finie (où $|\cdot|$ désigne la norme euclidienne sur $\mathbf{R}^d$). On cherche à approcher la loi de X par une loi discrète qui charge N points. Soit γ_N une N-grille de $\mathbf{R}^d$. On définit la projection sur cette grille ainsi :

$$\text{Proj}_{\gamma_N}(x) = \arg\min_{z \in \gamma_N} |x - z|.$$

L'erreur de quantification associée à la grille γ_N est donnée par

$$Q_N^p(\mathbf{P}_X)(\gamma_N) = \|X - \text{Proj}_{\gamma_N}(X)\|_p^p.$$

Lorsque la loi de X ne charge pas les hyperplans, on sait montrer l'existence d'une N-grille optimale minimisant $Q_N^p(\mathbf{P}_X)(\cdot)$ en annulant son gradient. Maintenant, pour tout vecteur aléatoire X dont la loi vérifie cette hypothèse, nous noterons $\hat{X}^N$ le projeté optimal de X. On peut aussi montrer que $\hat{X}^N$ vérifie la propriété de stationnarité

$$\mathbf{E}[X|\hat{X}^N] = \hat{X}^N.$$

Il existe de nombreux résultats portant sur l'asymptotique de l'erreur de quantification. Pour démontrer nos résultats de convergence (et donner des vitesses), nous avons utilisé le théorème suivant (voir Luschgy et Graph, 2000).

Théorème 1 *S'il existe $\delta > 0$ tel que $\|X\|_{p+\delta}$ est finie, alors il existe trois réels D_1, D_2, D_3 tels que pour tout $N \geq D_3$, on a*

$$\|X - \hat{X}^N\|_p^p \leq \frac{1}{N^{p/d}}\Big(D_1\|X\|_{p+\delta}^{p+\delta} + D_2\Big).$$

2.2 Méthode de prédiction par quantification

Considérons le modèle non-paramétrique

$$Y = \tilde{f}(U, \epsilon),$$

où $U : (\Omega, \mathcal{F}, \mathbf{P}) \to \mathbf{R}^d$ est une variable aléatoire et ϵ un terme aléatoire d'erreur indépendant de U. $\tilde{f}$ est une fonction de lien inconnue. On propose de prédire Y sachant U par une approche fondée sur la quantification optimale : soient $\hat{U}^N$ and $\hat{Y}^N$ les projetés optimaux de U et Y en norme $\mathbf{L}^p$ et $\hat{P}$ la matrice de transition de $\hat{U}^N$ vers $\hat{Y}^N$; c'est-à-dire, si γ_N et δ_N sont les N-grilles optimales de quantification pour U et Y :

$$\forall \hat{u} \in \gamma_N,\ \forall \hat{y} \in \delta_N,\ \hat{P}(\hat{u}, \hat{y}) = \mathbf{P}(\hat{Y}^N = \hat{y} | \hat{U}^N = \hat{u}).$$

On considère la variable aléatoire $\hat{Y}^c$ définie telle que $\big(\hat{U}^N, \hat{Y}^c\big)$ soit une chaîne de Markov arrêtée de transition $\hat{P}$. $\hat{Y}^c$ est notre prédicteur de Y sachant U. Dans le théorème 2, on montre que la loi discrète de $\hat{Y}^c$ sachant $\hat{U}^N = \mathrm{Proj}_{\gamma_N}(u)$ est une bonne approximation de celle de Y sachant $U = u$.

Afin d'obtenir la convergence, il est nécessaire d'introduire les hypothèses suivantes :

$(\mathcal{A}_1)$ $\exists\ p \geq 1$ t.q. $U, Y \in \mathbf{L}^p$,
$(\mathcal{A}'_1)$ $\exists\ \delta > 0$ t.q. $U, Y \in \mathbf{L}^{p+\delta}$,
$(\mathcal{A}_2)$ $\exists\ [\tilde{f}]_{Lip} > 0$ t.q. $\forall u, v \in \mathbf{R}^d$, $\|\tilde{f}(u, \epsilon) - \tilde{f}(v, \epsilon)\|_p \leq [\tilde{f}]_{Lip}|u - v|$,
$(\mathcal{A}_3)$ Les lois de U et Y ne chargent pas les hyperplans.

Théorème 2 *Soit ϕ une fonction lipschitzienne de constante de Lipschitz $[\phi]_{Lip}$, sous les hypothèses $(\mathcal{A}_{1\to3})$, on a*

$$\Big\|\mathbf{E}[\phi(Y)|U] - \mathbf{E}\big[\phi(\hat{Y}^N)|\hat{U}^N\big]\Big\|_1 \leq 2[\phi]_{Lip}[\tilde{f}]_{Lip}\|U - \hat{U}^N\|_p + [\phi]_{Lip}\|Y - \hat{Y}^N\|_p.$$

De plus, si on remplace $(\mathcal{A}_1)$ par $(\mathcal{A}'_1)$, la vitesse de convergence est donnée par

$$\Big\|\mathbf{E}[\phi(Y)|U] - \mathbf{E}\big[\phi(\hat{Y}^N)|\hat{U}^N\big]\Big\|_1 = \mathcal{O}\Big(\frac{1}{N^{1/d}}\Big).$$

3 Modèle de régression semi-paramétrique

3.1 Estimation de la direction de β

On considère ici le modèle de régression semi-paramétrique (1). Remarquons d'abord que β et f étant tous les deux inconnus, on peut seulement identifier le sous-espace engendré par β. Rappelons que $\hat{\Gamma}_N$ désigne la matrice de covariance de $\mathbf{E}[\hat{X}^N|\hat{Y}]$ où $\hat{Y} = \mathrm{Proj}_\gamma(Y)$ est la projection sur une grille (pas nécessairement optimale) γ de $\mathbf{R}$. Soit $\tilde{\beta}_N$ un vecteur propre principal de la matrice $\Sigma^{-1}\hat{\Gamma}_N$.

Le théorème 3 indique que, pour N grand, la direction de $\tilde{\beta}_N$ approche bien celle de β. En fait, ce résultat est un corollaire du théorème 4 : il existe une suite (β_N) de vecteurs propres principaux de $\Sigma^{-1}\hat{\Gamma}_N$ qui tend vers β quand le nombre de quantifieurs N de la loi de X tend vers l'infini.

On a besoin des hypothèses suivantes (classiques dans l'approche SIR) :

$$\begin{array}{ll}(\mathcal{A}_5) & \exists \hat{y} \in \gamma,\ \mathbf{E}[(X-\mu)'\beta|\hat{Y}=\hat{y}] \neq 0, \\ (\mathcal{A}_6) & X \text{ a une loi à symétrie elliptique.}\end{array}$$

Et également de celles-ci (pour permettre la quantification de X) :

$$\begin{array}{ll}(\mathcal{A}_7) & \exists\ p \geq 1 \text{ t.q. } X \in \mathbf{L}^p \cap \mathbf{L}^q \text{ avec } \frac{1}{p}+\frac{1}{q}=1, \\ (\mathcal{A}_8) & \text{La loi de } X \text{ ne charge pas les hyperplans,} \\ (\mathcal{A}_9) & \exists\ \delta > 0 \text{ t.q. } X \in \mathbf{L}^{p+\delta}.\end{array}$$

Théorème 3 *Sous $(\mathcal{A}_{5\to 8})$, pour toute suite $(\tilde{\beta}_N)$ de vecteurs propres principaux de la suite $(\Sigma^{-1}\hat{\Gamma}_N)$, on a*

$$\cos^2(\tilde{\beta}_N,\beta) \to 1 \quad \textit{quand } N \to \infty$$

où $\cos^2(\tilde{\beta}_N,\beta) = \dfrac{(\tilde{\beta}'_N\beta)^2}{(\tilde{\beta}'_N\tilde{\beta}_N)\times(\beta'\beta)}$.

Pour le résultat qui suit, rappelons que pour $x \in \mathbf{R}^d$, $|x|$ désigne sa norme euclidienne.

Théorème 4 *Sous $(\mathcal{A}_{5\to 8})$, il existe une suite (β_N) telle que pour tout N, β_N est vecteur propre principal de $\Sigma^{-1}\hat{\Gamma}_N$ et telle qu'il existe $C_1, C_2 > 0$ tels que*

$$\forall N \geq C_2, \quad |\beta_N - \beta| \leq \frac{2d}{C_1}\|\Sigma^{-1}\|_\infty \|X - \hat{X}^N\|_p \|X\|_q.$$

Si de plus $(\mathcal{A}_9)$ est vérifiée, on contrôle la vitesse de convergence :

$$|\beta_N - \beta| = \mathcal{O}\Big(\frac{1}{N^{1/d}}\Big).$$

3.2 Prédiction pour le modèle semi-paramétrique

Maintenant, on utilise conjointement la méthode d'estimation de β vue en 3.1 et la méthode de prédiction par quantification pour traiter le problème de la prédiction de Y dans le cadre semi-paramétrique du modèle (1). La clé est la suivante : pour N grand, si β_N est proche de β, le modèle

$$Y = f(\beta'_N X, \epsilon)$$

peut naturellement être une bonne approximation du modèle initial. Nous avons besoin des hypothèses suivantes pour obtenir le résultat asymptotique :

$$\begin{array}{ll}(\mathcal{A}_{10}) & \exists\ [f]_{Lip} > 0 \text{ t.q.. } \forall u, v \in \mathbf{R},\ \|f(u,\epsilon) - f(v,\epsilon)\|_p \leq [f]_{Lip}|u-v|, \\ (\mathcal{A}_{11}) & Y \in \mathbf{L}^{p+\delta}, \\ (\mathcal{A}_{12}) & \text{La loi de } Y \text{ ne charge pas les hyperplans.}\end{array}$$

Nous prédisons Y sachant X par la variable aléatoire $\hat{Y}^c$ telle que $\big(\widehat{\beta'_N X}^m, \hat{Y}^c\big)$ soit une chaîne de Markov arrêtée ayant pour loi de transition le noyau de la chaîne $\big(\widehat{\beta'_N X}^m, \hat{Y}^m\big)$ où $\widehat{\beta'_N X}^m$ et $\hat{Y}^m$ sont les approximations discrètes optimales (en norme $\mathbf{L}^p$) de $\beta'_N X$ et Y avec m quantifieurs.

Théorème 5 *Sous* $(\mathcal{A}_{5\to 12})$*, pour toute fonction lipschitzienne* ϕ*, il existe trois réels* A_1, A_2, A_3*, une suite* (g_N) *admettant une limite strictement positive et deux entiers* $\overline{m}$ *and* $\overline{N}$ *tels que pour tous* $m \geq \overline{m}$ *et* $N \geq \overline{N}$*, on a*

$$\left\| \mathbf{E}\big[\phi(Y)|\beta' X\big] - \mathbf{E}\Big[\phi(\hat{Y}^m)|\widehat{\beta_N' X}^m\Big] \right\|_1 \leq \frac{A_1}{N^{1/d}} + \frac{A_2}{m} g_N + \frac{A_3}{m}.$$

4 Conclusion

La méthode que nous avons présentée permet à la fois l'estimation de la direction du paramètre β et la prédiction de la variable d'intérêt Y sachant X par une approximation de sa loi conditionnelle dans le cadre du modèle de régression semi-paramétrique (1). Pour cela, nous avons utilisé l'approche SIR standard ainsi que l'outil probabiliste de quantification optimale en norme $\mathbf{L}^p$ d'une variable aléatoire.

Nous avons également mené des études de simulation (non présentées ici) qui confirment le bon comportement des estimateurs. Pour des échantillons de grande taille ($n = 1000$), nos résultats d'estimation de β sont comparables à ceux de la méthode SIR classique. Ils sont même meilleurs quand une dépendance symétrique apparaît entre $\beta' X$ et la variable d'intérêt Y.

Pour plus de détails, les preuves et des résultats de simulation, le lecteur pourra se référer à l'article correspondant (voir Azaïs et al., 2011).

Remerciements Le Professeur François Dufour nous a donné l'idée principale de ce travail qui est l'utilisation conjointe de la quantification optimale et de la méthode SIR pour les problèmes de réduction de dimension et de prédiction dans le modèle semi-paramétrique de régression (1). Nous lui sommes très reconnaissants pour cette suggestion.

Références

[1] Romain Azaïs, Anne Gégout-Petit, and Jérôme Saracco. Optimal quantization applied to sliced inverse regression, arXiv :1101.2121v2, Submitted.

[2] V. Bally, G Pagès, and J. Printems. A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 15(1) :119–168, 2005.

[3] R. D. Cook. Principal hessian directions revisited. *JASA*, 93(441) :84–94, 1998.

[4] R. D. Cook. Save : A method for dimension reduction and graphics in regression. *Com in Stat - Theory Methods*, 29 :2109–2121, 2000.

[5] B. de Saporta, F. Dufour, C. Elegbede, and H. Zhang. Arrêt optimal pour la maintenance prédictive. In *Proceedings of* $\lambda\mu$, La Rochelle, France, 2010.

[6] B. de Saporta, F. Dufour, and K. Gonzales. Numerical method for optimal stopping of piecewise deterministic processes. *Ann. Appl. Probab.*, 20(5) :1607–1637, 2010.

[7] N. Duan and K.-C. Li. Slicing regression : a link-free regression method. *Ann. Appl. Stat.*, 19(2) :505–530, 1991.

[8] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46(1) :103–122, 2004.

[9] S. Graph and H. Luschgy. *Foundations of quantization for random vectors*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.

[10] G. Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89(1) :1–38, 1998.

[11] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pages 253–297. Birkhäuser Boston, Boston, MA, 2004.

[12] P.L Zador. *Development and evaluation of procedures for quantizing multivariate distributions*. PhD thesis, Stanford University, Stanford, CA 94305, 1963.