Régression bayésienne avec contraintes de régularité et de forme

Christophe Abraham

Régression bayésienne avec contraintes de régularité et de forme

Christophe Abraham

Montpellier SupAgro
UMR Analyse des systèmes et biométrie
2, Place Pierre Viala, 34060 Montpellier

Mots clefs : Régression bayésienne, loi a priori impropre, contraintes de forme, régularité, mouvement brownien intégré.

Abstract
A Bayesian method for regression under shape restrictions and smoothness conditions is proposed. The support of the prior distribution is included in the set of piecewise linear fonctions. It is shown that the proposed prior can be arbitrarily close to the distribution induced by the addition of a polynomial with random coefficients and a $(m-1)$-fold integrated Brownian motion. Thanks to the piecewise linear property, it is easy to introduce shape constraints. The regression fonction is estimated by the posterior mode. The posterior mode can be computed by a simulated annealing algorithm. The shape constraint is taken into account thanks to the proposal distribution of the simulated annealing algorithm. Simulations from the posterior distribution are obtained by a Metropolis-Hastings algorithm.

Résumé
On propose une méthode bayésienne de régression sous contraintes de régularité et de forme. On construit une loi a priori qui charge les fonctions linéaires par morceaux. On montre que cette loi a priori peut être arbitrairement proche de la loi induite par l'addition d'un polynôme à coéfficients aléatoires et d'un mouvement brownien intégré $m-1$ fois. Ainsi, bien que linéaire par morceaux, la fonction de régression se comporte approximativement comme une fonction de classe C^{m-1}. La linéarité par morceaux permet d'introduire des contraintes de forme. L'estimateur choisi est le mode a posteriori. Il est calculé à partir d'un algorithme de type recuit simulé dont l'étape de proposition garantit la vérification de la contrainte de forme. Des simulations suivant la loi a posteriori sont obtenues grâce à un algorithme de type Metropolis-Hastings.

1 Introduction

Considérons le modèle usuel de régression : les observations $(t_1, y_1), \dots, (t_n, y_n)$ avec $(t_i, y_i) \in [0,1] \times \mathbb{R}$, sont supposées indépendantes et de même loi (iid) avec

$$\left\{ \begin{array}{lcl} y_i & = & f(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \\ \varepsilon_i & \sim & N(0, \sigma^2). \end{array} \right. \tag{1}$$

Le problème consiste à estimer la fonction de régression f sous des contraintes de régularité et de formes. Les contraintes de formes envisagées sont multiples : monotonie, unimodalité, convexité, succession de parties croissantes, décroissantes, convexes ou concaves, etc.

Les travaux sur la régression sous contraintes sont assez nombreux en statistique classique (pour une description des principales méthodes nonparamétriques classiques, voir Delecroix & Thomas-Hagnan, 2000). Les méthodes proposées concernent principalement la régression isotonique et se décomposent souvent en deux étapes : le lissage pour satisfaire la contrainte de régularité et l'*isotonisation* pour satisfaire la contrainte de monotonie. Par contre, en statistique bayésienne, les travaux sont plus rares et relativement récents (Cai & Dunson, 2007, Neelson & Dunson, 2004, Holmes & Heard, 2003 et Lavine & Mockus, 1996). Comme dans le cadre classique, ils concernent surtout la régression isotonique.

D'un point de vue théorique, le cadre bayésien se prête bien à la régression sous contraintes. En effet, la loi a posteriori étant absolument continue par rapport à la loi a priori, il suffit donc de choisir une loi a priori qui ne charge que les fonctions f qui vérifient les contraintes considérées. D'un point de vue pratique, le choix de la loi a priori n'est pas simple car il doit permettre le calcul de la loi a posteriori (éventuellement à partir de simulations). Sans contrainte de forme, une loi a priori commode pour effectuer une régression bayésienne nonparamétrique, consiste à supposer que f est un processus gaussien (O'Hagan, 1978). La régularité de f peut alors être contrôlée à partir de la fonction de covariance du processus. L'objet de cet article est de proposer une méthode qui, dans une certaine mesure, étend la loi a priori ci-dessus à la prise en compte de contraintes de forme.

Dans la Section 2, on propose une loi a priori qui s'adapte à des contraintes de régularité et de forme. On montre que cette loi a priori est une approximation en dimension finie d'une loi a priori proposée par Wahba (1990) induite par la somme d'un polynôme à coefficients aléatoires et d'un mouvement brownien intégré $m-1$ fois. La Section 3 est consacrée à l'inférence sous des contraintes de forme et de régularité. En particulier, on montre que l'estimateur bayésien peut être calculé à partir d'un algorithme de type recuit simulé et qu'il est possible de générer suivant la loi a posteriori à partir d'un algorithme de type Metropolis-Hastings.

2 La distribution a priori

On propose de construire f par interpolation linéaire des points $(l/j, f_{jl})$ pour $l \in \{0, \ldots, j\}$. On note $f_j = (f_{j0}, f_{j1}, \ldots, f_{jj})'$. Bien que linéaire par morceaux, on souhaite que f soit arbitrairement proche d'une fonction lisse. Ainsi, la distribution de f devra favoriser les fonctions dont les accroissements ont de faibles variations. Il convient donc de définir une version discrète des dérivées successives de f. Pour cela, on définit Δ_j la matrice de dimension $j \times (j+1)$ telle que :

$$\Delta_j = j \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

On pose $f_j^{(0)} = f_j$ et on construit par récurrence le vecteur colonne $f_j^{(m)} = \Delta_{j-m+1} f_j^{(m-1)}$. On remarque $f_j^{(m)}$ est de dimension $(j-m+1)\times 1$. En posant $f_j^{(m)} = (f_{jm}^{(m)}, \ldots, f_{jj}^{(m)})'$, $f_{jl}^{(m)}$ peut être interprété comme la dérivée à l'ordre m de f au point l/j pour $l \in \{m, \ldots, j\}$. Pour assurer la régularité de f, on impose simplement aux dérivées d'ordre m, aux points l/j pour $l \geq m$, d'être des variables aléatoires, indépendantes conditionnellement à σ^2, de loi $N(0, \sigma^2 j/\lambda)$ où λ est un réel positif. Cette condition conduit naturellement à une loi a priori impropre de f_j de densité conditionnellement à σ^2 vérifiant :

$$p(f_j|\sigma^2) \propto \exp\left\{-\frac{\lambda}{2j\sigma^2} f_j'(\Delta_j^{m\prime}\Delta_j^m) f_j\right\}, \tag{2}$$

où $\Delta_j^m = \Delta_{j-m+1}\ldots\Delta_j$. On conviendra, par la suite, qu'une variable aléatoire suit la loi impropre $\mathcal{U}$ si sa densité par rapport à la mesure de Lebesgue est constante sur $\mathbb{R}$. Si en plus de la loi impropre (2), on impose un loi inverse Gamma pour σ^2 et on obtient la loi a priori π_j suivante :

$$\begin{cases} \sigma^2 \sim IG(a,b), \\ f_{jl}^{(l)} \sim \mathcal{U}, \text{ iid, } \mathrm{l} \in \{0, 1, \ldots, \mathrm{m}-1\}, \\ f_j^{(m)} \sim N(0, \lambda^{-1} j \sigma^2 I_{j-m+1}). \end{cases} \tag{3}$$

Par exemple, si $m = 2$, f est une fonction aléatoire linéaire par morceaux définie à partir de l'ordonnée à l'origine f_{j0}, la pente à l'origine $f_{j1}^{(1)} = j(f_{j1} - f_{j0})$ et les dérivées secondes en l/j, $f_{jl}^{(2)} = j^2(f_{jl} - 2f_{j,l-1} + f_{j,l-2})$ pour $l \in \{2, \ldots, j\}$, avec $f_{j0}|\sigma^2 \sim \mathcal{U}$, $f_{j1}^{(1)}|\sigma^2 \sim \mathcal{U}$ et $f_{jl}^{(2)}|\sigma^2 \sim N(0, \lambda^{-1} j \sigma^2)$, $l \in \{2, \ldots, j\}$.

Bien que π_j soit impropre, il facile de voir à partir de (2) que la loi a posteriori est une loi normale inverse Gamma dont les paramètres sont explicitement connus (les calculs sont similaires à ceux du modèle conjugué normal inverse Gamma).

Le théorème suivant montre que la loi a priori converge lorsque le pas de discrétisation j^{-1} tend vers 0. On désigne par $[x]$ la partie entière de x, et pour tout k, l et p entiers, $M_k(l,p)$ désigne le nombre de k-uplets $(l_1, \ldots, l_k) \in \mathbb{N}^k$ tels que $p \leq l_k \leq \ldots \leq l_1 \leq l$.

Théorème 2.1 *1) Ils existent des variables aléatoires $U_0, U_1, \ldots, U_{m-1}$ et $Z_m, \ldots, Z_j$ indépendantes de lois $U_l \sim \mathcal{U}$ pour $l \leq m-1$ et $Z_l \sim N(0,1)$ pour $l \geq m$ telles que :*

$$f_j(t) = \sum_{l=0}^{m-1} \frac{M_l(l,[jt])}{j^l} U_l + \frac{\sigma}{\sqrt{\lambda}} Y_j^m(t) + O_p(j^{-1}),$$

où

$$Y_j^m(t) = j^{-m+1/2} \sum_{l=m}^{[jt]} M_{m-1}(l,[jt]) N_l.$$

2) Si $j \to \infty$, $M_l(l,[jt])/j^l \to t^l/l!$ et Y_j^m converge en loi vers un mouvement brownien intégré $m-1$ fois, noté W^m.

Ainsi, si $j \to \infty$, $f_j(t)$ tend en loi vers la fonction aléatoire suivante :

$$\sum_{l=0}^{m-1} U_l \frac{t^l}{l!} + \frac{\sigma}{\sqrt{\lambda}} W^m(t).$$

Pour une telle loi a priori, Wahba (1990) montre que l'espérance a posteriori de f est une fonction spline qui est solution du problème des moindres carrés pénalisé par une terme de la forme

$$\int_0^1 (f^{(m)}(u))^2 du.$$

3 Régression avec contraintes de formes

La fonction f étant linéaire par morceaux, il est facile de contrôler sa forme. Soit S_j l'ensemble des vecteurs f_j tels que f respecte la contrainte de forme. Par exemple, si on impose à f d'être croissante, on a :

$$S_j = \{f_j \in \mathbb{R}^{j+1} : f_{j0} \leq f_{j1} \leq \ldots \leq f_{jj}\},$$

et pour une régression unimodale, on a :

$$S_j = \cup_{l=1}^{j-1} \{f_j \in \mathbb{R}^{j+1} : f_{j0} \leq \ldots \leq f_{jl} \geq \ldots \geq f_{jj}\}.$$

Il suffit alors de choisir comme loi a priori la loi introduite dans la section précédente conditionnée par l'évènement $\{f_j \in S_j\}$. On obtient ainsi une loi a priori π_j^S dont la densité est donnée, à une constante près, par :

$$p(f_j, \sigma^2) \propto p(\sigma^2) \exp\left\{-\frac{\lambda}{2j\sigma^2} f_j'(\Delta_j^{m\prime} \Delta_j^m) f_j\right\} \mathbf{1}_{\{f_j \in S_j\}},$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de l'ensemble A. La forme est garantie par l'indicatrice $\mathbf{1}_{\{f_j \in S_j\}}$ et la régularité est contrôlée par m dans la mesure où f est proche d'un mouvement brownien intégré $m-1$ fois. Pour la loi a priori π_j^m, il est facile de voir que la densité de la loi a posteriori est proportionnelle $\mathbf{1}_{\{f_j \in S_j\}}$ multiplié par la densité de loi normale inverse gamma que l'on aurait obtenu sans considérer la restriction à S_j.

On propose d'estimer f par le mode a posteriori. En effet, contrairement à l'espérance a posteriori, le mode a posteriori vérifie nécessairement les contraintes de forme. Un algorithme de type recuit simulé est bien adapté à notre contexte pour estimer le mode de la distribution a posteriori. En effet, cet algorithme est constitué d'une étape de proposition suivi d'une étape d'acceptation. La prise en compte de la contrainte de forme sera réalisée lors de l'étape de proposition de la façon suivante. Un indice $l \in \{0, \ldots, j\}$ est choisi au hasard suivant une loi uniforme sur $\{0, \ldots, j\}$, puis la valeur de f_{jl} est changée par un tirage aléatoire uniforme sur un ensemble I_l de valeurs de f_{jl} compatibles avec la contrainte de forme. Par exemple, pour une contrainte de croissance, $I_l = [f_{j,l-1}, f_{j,l+1}]$ pour $l \in \{1, \ldots, j-1\}$, $I_0 = [f_{j1} - k, f_{j1}]$ et $I_j = [f_{j,j-1}, f_{j,j-1} + k]$ où k est un réel positif arbitraire. Une bande de confiance autour de l'estimateur pourra être obtenue à partir de simulations suivant la loi a posteriori de f. Ces simulations sont réalisées à partir d'un algorithme de type Metropolis-Hastings. En effet, cet algorithme est composé d'une étape de proposition et une étape d'acceptation. Nous proposons de prendre en compte la contrainte de forme au cours de l'étape de proposition comme pour l'algorithme de recuit simulé ci-dessus.

Bibliographie

[1] Cai, B. and Dunson, D.B. (2007) Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies, *Journal of the American Statistical Association*, **102**, 1158-1171.

[2] Delecroix, M. and Thomas-Agnan, C. (2000), Spline and Kernel Regression under Shape Restrictions, dans *Smoothing and Regression: Approaches, Computation, and Applications*, Michael G. Schimek (ed.), Jhon Wiley & Sons, New-York, 109–133.

[3] Holmes, C.C and Heard, N.A. (2003), Generalized Monotonic Regression Using Random Change Points, *Statistics in Medicine*, **22**, 623-638.

[4] Lavine, M. and Mockus, A. (1996), A nonparametric Bayes Method for Isotonic Regression, *Journal of Statistical Planning and Inference*, **46**, 235-248.

[5] Neelon, B. and Dunson, D.B. (2004), Bayesian Isotonic Regression and Trend Analysis, *Journal of Statistical Planning and Inference*, **60**, 741-759.

[6] O'Hagan, A. (1978), On curve fitting and optimal design for regression (with discussion), *Journal of Royal Statistical Society* **B, 41**, 358-367.

[7] Wahba, G. (1990), *Spline models for observational data*, CBMS-NSF series SIAM, Philadelphia.