統計数理研究所彙報
第30巻　第1号（1982）

ベイズ型スプライン回帰

統計数理研究所　石　黒　真木夫
〃　荒　畑　恵美子

（1982年5月　受付）

1. はじめに

スプラインを回帰分析に応用する研究は数多い（たとえば市田と吉本，[2]）．節点の決定が実用上の最大の難関と考えられ，これを解決するためにいくつかの提案がなされてきた．たとえば，赤池の AIC 最小化による方法（吉本ら，[3]）などである．

本稿の主張は，AIC と同じく赤池によって導入されたベイズ型のモデルを応用することによって，節点の決定にわずらわされることのない実用的な回帰曲線の推定法が得られるということにある．

問題を明確にするために，ある区間で定義された回帰曲線 $h(\cdot)$ を等間隔の節点を持つ所定の次数のスプラインで推定することを考えよう．この場合に，節点数が多いほど柔軟なモデルとなる一方，パラメータの推定を単純な最小二乗法によるかぎり，節点が多ければ多いほど結果は不安定になる．最小二乗法に固執するかぎり，この不安定性を制御する唯一の方法は，節点の数を必要最小限におさえることである．ところで，‘単純な最小二乗法に固執すれば’という前提は必然的なものではない．他に推定値の安定性を保証する方法があれば，節点の数は非常に多くとってよいことになる．次節では，スプラインのパラメータに事前分布を仮定することによってこの目的を達成する方法を示し，3節で既存の方法との関連を論じ4節にその実用性を示す数値例を与える．

2. ベイズ型モデル

n 対のデータ $(x_i,\ y_i)$ $i=1,\ 2,\ \cdots,\ n$ が与えられているものとする．x_i と y_i の間に

(1) $$y_i=h(x_i)+\varepsilon_i \qquad (i=1,\ 2,\ \cdots,\ n)$$

なる関係を想定して，$h(\cdot)$ を推定することが回帰分析の目的となる．ここで $\{\varepsilon_i\}$ は互に独立に平均0，分散 σ^2 の正規分布に従い，すべての x_i が区間 $(b,\ c]$ に落ちていると仮定する．$h(\cdot)$ を近似する関数として，等間隔節点 $\{t_i\}$ を持つ2次の B–スプライン

(2) $$h(x|\boldsymbol{a})=\sum_{j=1}^{m} a_j s_j(x)$$

を採用することにする．ここで

$$d=\frac{c-b}{m}$$

として，i 番目の節点を

$$t_i=b+di \qquad (i=0,\ 1,\ \cdots,\ m)$$

で与える．通常3次のスプラインを用いることが多いが，m を十分に大きくとることにすれば2次のスプラインで十分であると考えられる．もちろん以下の方法を高次に拡張することも可能であるが，その必要はないと考える．

$s_j(\cdot)$ を定義する前に準備として2つの関数 $j(x)$ と $d(j)$ を定義しておこう．$j(x)$ を $x\in(c, d]$ に対して，

$$j(x)=j \quad \text{if } x\in(t_{j-1},\ t_j] \tag{3}$$

で定義される関数とする．$(t_{j-1}, t_j]$ を j-区間と名づける．$d(j)$ は j-区間の中点，すなわち

$$d(j)=\frac{t_{j-1}+t_j}{2}=b+dj-\frac{d}{2} \tag{4}$$

とする．これらを用いて $s_j(x)$ を

$$s_j(x)=\begin{cases} g(x-d(j-1)) & \text{if } j(x)=j-1 \\ f(x-d(j)) & \text{if } j(x)=j \\ g(d(j+1)-x) & \text{if } j(x)=j+1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

で定義する．ただし，

$$\begin{aligned} f(x)&=-\frac{1}{d^3}x^2+\frac{3}{4d} & -\frac{d}{2}<x\leqq\frac{d}{2} \\ g(x)&=\frac{1}{2d^3}\left(x+\frac{d}{2}\right)^2 & -\frac{d}{2}<x\leqq\frac{d}{2} \end{aligned} \tag{6}$$

である．

以上の仮定のもとで，パラメータ $\boldsymbol{a}=(a_1,\ a_2,\ \cdots,\ a_m)^T$ および σ^2 の尤度関数は

$$L(\boldsymbol{a},\ \sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2\sigma^2}\left(y_i-\sum_{j=1}^{m}a_j s_j(x_i)\right)^2\right) \tag{7}$$

となり，各 j に対してある i が存在して $x_i\in(t_{j-1},\ t_j]$ であれば，$\boldsymbol{a}$, σ^2 の最尤推定値は (7) を最大化することによって得られる．しかしこの推定値は区間の数 m が多くなるに従って不安定となり，ある程度以上になればパラメータのあるものは不定となってしまう．

この困難を解決するために，我々は Akaike [1] によって提案されたベイズ手法を採用する．すなわち，m が十分に大きいとき，$a_1,\ a_2,\ \cdots,\ a_m$ の2次の階差

$$\Delta^2 a_j=a_j-2a_{j-1}+a_{j-2} \qquad j=1,\ 2,\ \cdots,\ m \tag{8}$$

は0の近くに分布すると期待される（2次に限る必要もないが実用上これで十分なようである．a_0 と a_{-1} に関しては後に触れる）．この期待をある程度満足しかつ尤度関数 (7) の値も大きくなるように $\boldsymbol{a}$ の推定値を決めるには，$\boldsymbol{a}$ に事前分布

$$\pi(\boldsymbol{a}|u^2)=\left(\frac{u}{\sqrt{2\pi}\,\sigma}\right)^m\exp\left\{-\frac{u^2}{2\sigma^2}\sum_{j=1}^{m}|\Delta^2 a_j|^2\right\} \tag{9}$$

を想定し，データが与えられたときの事後分布

$$\pi_{\text{post}}(\boldsymbol{a}|u^2)\propto L(\boldsymbol{a},\ \sigma^2)\,\pi(\boldsymbol{a}|u^2) \tag{10}$$

$$=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\left(\frac{u}{\sqrt{2\pi}\,\sigma}\right)^m\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n\left(y_i-\sum_{j=1}^m a_js_j(x_i)\right)^2\right.\right.$$

$$\left.\left.+u^2\sum_{j=1}^m|\Delta^2a_j|^2\right]\right\}$$

のモードを推定値とすれば良い．u^2 を与えれば (10) を最大化する $\boldsymbol{a}$ は σ^2 に無関係に決まり，

(11) $$|\boldsymbol{y}-X\boldsymbol{a}|^2+u^2|\boldsymbol{z}-D\boldsymbol{a}|^2$$

を最小化することによって得られる．ここで (11) は (10) の exp の肩の [ ] の部分であり，X は (i, j) 成分が $s_j(x_i)$ である n 行 m 列行列，$\boldsymbol{z}$ は，第1成分が $-a_{-1}+2a_0$，第2成分が $-a_0$ それ以外はすべて0の m–ベクトル，D は

$$D=\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -2 & 1 & \ddots & & & \vdots \\ 1 & -2 & 1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

で定義される m 行 m 列の行列である．

u^2 の決定と σ^2 の推定は赤池の提案によるベイズ型情報量規準

(12) $$\mathrm{ABIC}=-2\log\int L(\boldsymbol{a},\ \sigma^2)\ \pi(\boldsymbol{a},\ u^2)\,d\boldsymbol{a}$$

の最小化による．(11) を最小化する $\boldsymbol{a}$ の値を $\hat{\boldsymbol{a}}$ と書くことにすると，簡単な計算の結果 ABIC の値は

$$\mathrm{ABIC}=n\log 2\pi+\frac{|\boldsymbol{y}-X\hat{\boldsymbol{a}}|^2+u^2|\boldsymbol{z}-D\hat{\boldsymbol{a}}|^2}{\sigma^2}+n\log\sigma^2+\log\{\det(X^TX+u^2D^TD)\}$$

となる．したがって u^2 を与えたときの σ^2 の推定値は，

(13) $$\hat{\sigma}^2=\frac{1}{n}\{|\boldsymbol{y}-X\hat{\boldsymbol{a}}|+u^2|\boldsymbol{z}-D\hat{\boldsymbol{a}}|^2\}$$

で与えられ，そのときの ABIC の値は u^2 の関数として，

(14) $$\mathrm{ABIC}_{\min}=n\log 2\pi+n+n\log\hat{\sigma}^2+\log\{\det(X^TX+u^2D^TD)\}$$

となる．u^2 を精密に決定するためには (14) を最小化する非線型最適化によらねばならないが，$\boldsymbol{a}$ の推定を目的とするかぎり，(14) を厳密に最小化する必要はない．実用上は $u^2\in\{2^P|P=-1,\ 0,\ 1,\ \cdots,\ 5\}$ 位の範囲で選択すれば十分である．

u^2 は事前分布 π のスケールを決める超パラメータ（パラメータの分布のパラメータであるので超パラメータと呼ぶ）である．これの決定をデータに依存して行うことが以上の方法の骨子である．

a_0 と a_{-1} の選択の問題を残しておいたが，(11) 式を見ると分るように，これらを π の期待値を決める超パラメータとみなすことができる．これの決定も ABIC 最小化によることがで

きる．ABIC と (a_0, a_{-1}) の関連をみると，$\mathrm{ABIC_{min}}$ の各項のうち (a_0, a_{-1}) に関係するのは $\log\hat{\sigma}^2$ の項のみである．これは (a_0, a_{-1}) を $\hat{\sigma}^2$ の値を最小化するように選択すれば良いことを示している．すなわち，(11) 式を最小化するにあたって，$\boldsymbol{z}$ を与えられたものとせずに，これに関しても最適化すれば良いわけである．つまり

$$(15)\qquad \sum_{i=1}^{n}\left\{y_i-\sum_{j=1}^{m}a_js_j(x_i)\right\}^2+u^2\sum_{j=1}^{m}(a_j-2a_{j-1}+a_{j-2})^2$$

を最小にする $(\hat{a}_{-1}, \hat{a}_0, \hat{a}_1, \cdots, \hat{a}_m)$ を求めれば良い．(13) 式の $\boldsymbol{z}$ を

$$\hat{\boldsymbol{z}}=(-\hat{a}_{-1}+2\hat{a}_0, -\hat{a}_0, 0, \cdots, 0)^T$$

で置きかえて得られる $\hat{\sigma}^2$ と (14) 式から，u^2 が与えられたときの ABIC の最小値が得られる．

我々の方法によれば，ある j-区間に落ちるデータの数が 0 になってもいっこうにさしつかえない．区間数 m を非常に大きくとれる所以である．

3. 既存の方法との関連

Wahba と Wold [4] は smoothing spline と称して，

$$(16)\qquad \frac{1}{n}\sum_{i=1}^{n}\{y_i-h(x_i|\boldsymbol{a})\}^2+\lambda\int\{h''(x|\boldsymbol{a})\}^2dx$$

を最小化する方法を提案している（次に述べるように，λ の選択と我々の u^2 の選択は実は等価な問題であるが，彼女らはこれを cross-validation によって解決する）．我々の (15) 式とこの式はよく似ている．$h(x|\boldsymbol{a})$ を2次のスプライン (2) とすると，$j(x)=k$ として (5) 式を用いて

$$\begin{aligned} h(x|\boldsymbol{a})=\sum_{j=1}^{m}a_js_j(x)&=a_{k+1}g(x-d(k))\\ &\quad+a_kf(x-d(k))\\ &\quad+a_{k-1}g(d(k)-x)\end{aligned}$$

となるから

$$(17)\qquad \{h''(x|\boldsymbol{a})\}^2=\frac{1}{d^6}(a_{k+1}-2a_k+a_{k-1})^2\qquad (j(x)=k)$$

従って

$$(18)\qquad \lambda\int\{h''(x|\boldsymbol{a})\}^2dx=\frac{\lambda}{d^5}\sum_k(a_{k+1}-2a_k+a_{k-1})^2$$

となる．

これは Wahba と Wold の λ の決定の問題と我々の u^2 の決定の問題が等価であることを意味している．Wahba と Wold の方法と我々の違いはその決定法の違いにあるわけである．

4. 数　値　例

上述した Wahba と Wold の論文に，回帰曲線

(19) $$h(x)=4.26\,(e^{-x}-4e^{-2x}+3e^{-3x})\qquad(0\leqq x<3.25)$$

をスプラインで近似，推定する例が載っている．データとして

$$y_i=h(x_i)+\varepsilon_i$$
$$x_i=0.0325\,(i-1)\qquad(i=1,\ 2,\ \cdots,\ 100)$$

を発生し，我々の方法で推定した結果を表1に示す．

表1　ABIC の値 (m=100)

u の設定値	ABIC の値
1	45.8
2	18.0
4	− 2.0
8	−13.8
16	−15.2*
32	− 4.8

* 最小 ABIC 値

表2　ABIC の値 (m=200)

u の設定値	ABIC の値
1	100.0
2	61.6
4	30.1
8	6.3
16	−9.6
32	−16.5*
64	−11.8

* 最小 ABIC 値

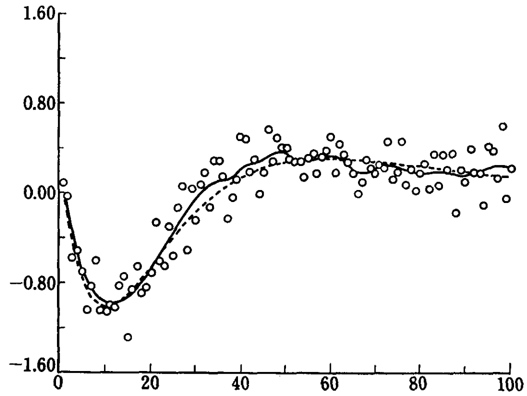

図1a　$h(\cdot)$ の推定値 (m=100, u=4)

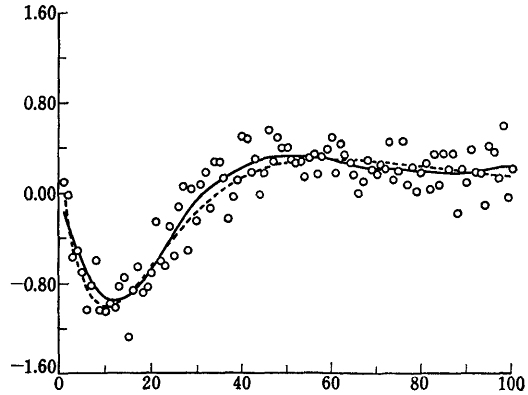

図1b　$h(\cdot)$ の推定値 (m=100, u=16)

u=4, 16, 32 とした場合の $h(\cdot)$ の推定値を図 1 a, b, c に示す. u=4 の場合の推定値が過度にデータに引かれていること, u=32 の場合はデータの特徴を追いきれなくなっていることが見てとれよう. これらの図は m=100 とした場合のものである.

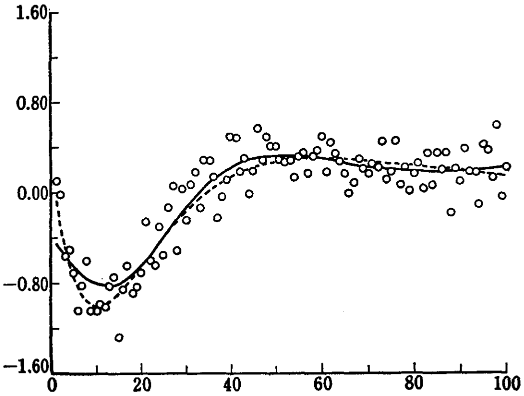

図 1 c $h(\cdot)$ の推定値 (m=100, u=32)

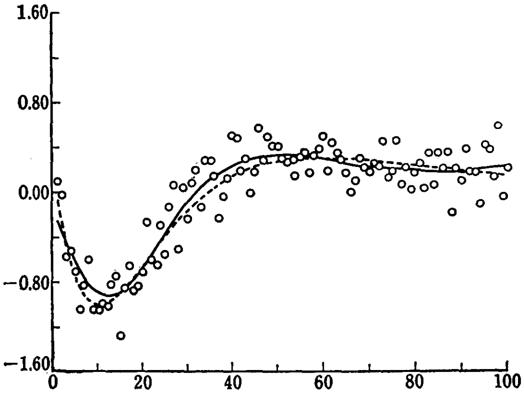

図 2 $h(\cdot)$ の推定値 (m=200, u=32)

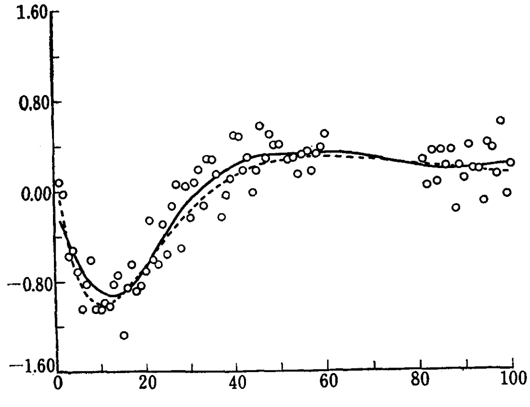

図 3 $h(\cdot)$ の推定値 (データが欠けた場合)

$m=200$ とした場合の結果を表2に示す．このとき，最小 ABIC を与えたのは $u=32$ の場合であった．このときの $h(\cdot)$ の推定値を図2に示す．($m=100$, $u=16$) として得られた図1b と図2がほぼ一致した結果を与え，$h(\cdot)$ の推定に m の選択がほとんど影響しないことが明らかであろう．

最後にいくつかのデータを省いた場合の例を図3に掲げる．内挿が成功している例である．この図は $m=100$ として得られている．

5. まとめ

ベイズ型のモデルを導入することによって，スプライン回帰が安定化できることを示した．Wahba と Wold の方法と共通点が多いが，cross-validation の煩雑な計算がなしに済むなどはるかに経済的な推定法といえる．またこの方法は赤池によって提唱されたベイズ型モデルの典型的な一つの応用例であり，理論的な意味づけが明らかな点がもう一つの利点である．

謝辞
日頃御指導を戴く統計数理研究所の赤池弘次氏および適切なコメントをくださったレフェリーに感謝します．

参考文献

[1] Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), Univ. Press, Valencia, Spain.

[2] 市田浩三，吉本富士市 (1979). スプライン関数とその応用，新しい応用の数学シリーズ 20，教育出版，東京.

[3] 吉本富士市，市田浩三，清野　武 (1976). 区分的3次関数を用いたデータ平滑化——節点の決定について——，情報処理，**17**，200-206.

[4] Wahba, G. and Wold, S. (1975). A completely automatic French curve — Fitting spline functions by cross validation—, *Commun. Statist.*, **4**, 1–17.

A Bayesian Spline Regression

Makio Ishiguro and Emiko Arahata
(The Institute of Statistical Mathematics)

A way to resolve the difficulty with the spline regression, the choice of knots, is proposed. It is obtained as a simple application of the Bayesian procedure developed by Akaike (1980, Likelihood and the Bayes procedure, *Bayesian Statistics*, (eds. J.M. Bernardo et al.) Univ. Press, Valencia, Spain).

The basic idea is to control the instability of the spline regression curve, which usually increases as the number of knots increases comparing with that of data, by assuming a prior distribution of the parameters of spline curves. The prior distribution is chosen so that the smoothness of the estimated regression curve is guaranteed.

In our procedure, the prior distribution contains several adjustable parameters (hyper parameters) which are, determined by minimizing ABIC (Akaike's Bayesian Information Criterion), which is defined as the minus twice of the logarithm of marginal likelihood.

A relationship of our procedure to the curve fitting procedure by Wahba and Wold (1975, A Completely automatic French curve, *Commun. Statist.*, **4**) is discussed. A numerical example which demonstrates the effectiveness of our approach is also included.