

B-スプラインによる非線形回帰モデルと 情報量規準

九州大学* 井元清哉**・小西貞則

(受付 1999 年 4 月 9 日; 改訂 1999 年 7 月 1 日)

要 旨

B-スプラインを用いた非線形回帰モデルに基づいてデータから情報抽出を行う方法について検討する。確率分布によって表現された B-スプライン非線形回帰モデルのパラメータを罰則付き対数尤度関数に基づいて推定するとき、平滑化パラメータと節点の個数の選択がモデル推定において重要となる。従来は、その選択規準として、交差検証法、一般化交差検証法、情報量規準 AIC の自由パラメータ数を調節する方法や赤池のベイズ型情報量規準 ABIC が用いられてきた。本稿では、情報量の観点から構成したモデル評価規準に基づいて平滑化パラメータの選択を行い、B-スプライン非線形回帰モデルを推定する方法について述べ、数値実験を通して従来の手法と比較し、その特徴と有効性を検証する。

キーワード：B-スプライン、非線形回帰、モデル評価、平滑化パラメータ、情報量規準。

1. はじめに

説明変数 X と目的変数 Y に関して観測されたデータから、非線形構造を捉えるために様々なモデルが提案されてきた (例えば, Ramsay (1988); Green and Silverman (1994); Simonoff (1996); Kitagawa and Gersch (1996) など)。非線形回帰においては、データの平均構造を記述するために移動平均、スプライン、核関数などが用いられるが、ここでは等間隔節点をもつ B-スプライン (de Boor (1978); Eilers and Marx (1996)) を用いた方法を取り上げる。B-スプライン非線形回帰モデルのパラメータ推定は、対数尤度に曲線の局所変動の程度を考慮に入れたいわゆる罰則付き対数尤度関数に基づいて行なわれる。その際、平滑化パラメータと節点の個数の選択がモデル推定の本質的な問題となり、その選択規準として様々な方法が提案されてきた。

本稿では、B-スプライン非線形回帰モデルの平滑化パラメータと節点の個数の選択規準として交差検証法 CV、一般化交差検証法 GCV (Craven and Wahba (1979); Green and Silverman (1994)), 赤池のベイズ型情報量規準 ABIC (Akaike (1980a, 1980b)), Eilers and Marx (1996) と Hurvich et al. (1998) の赤池情報量規準 AIC (Akaike (1973, 1974)) の自由パラメータ数を調整した修正 AIC、井元・小西 (1999) のスプライン情報量規準 SPIC の 6 つの選択規準を取り上げる。

* 大学院数理学研究科：〒812-8581 福岡市東区箱崎 6-10-1.

** 日本学術振興会特別研究員。

B -スプラインを用いた曲線推定の問題に関しては、これまで様々な方法が提案されてきた。市田・吉本 (1979, p. 89), Hiragi et al. (1985) は, B -スプライン曲線の滑らかさを節点の個数とその位置によって規定し, その選択に情報量規準 AIC を用いた。また, 田辺 (1976) は, 逆問題のフレームワークで B -スプラインの節点の個数およびその位置を AIC を用いて選択する方法を提案した。これらの研究では, パラメータ推定には, 最尤法あるいは最小二乗法が用いられている。石黒・荒畑 (1982), 田辺・田中 (1983) は, ABIC に基づいて B -スプライン, 離散スプライン非線形回帰モデルの平滑化パラメータを選択する方法を提案した。また, 田辺 (1989, pp. 379-380) は, 正規分布に基づく B -スプライン非線形回帰モデルを通して, 曲線の滑らかさを調節する平滑化パラメータの選択をベイズモデルから捉え, ABIC との関係を明確に与えている。Eilers and Marx (1996), Hurvich et al. (1998) の提案した二つの情報量規準は, どちらも AIC を非線形回帰モデルの評価に用いるためにモデルに含まれる自由パラメータ数を, いわゆるハット行列のトレースを用いて修正したものである。しかし, AIC は最尤法によって推定したモデルを評価するための規準であり, 罰則付き対数尤度関数により推定されたモデルの評価に直接適用するには問題がある。また, ハット行列による自由パラメータ数の修正法に対しても検討の余地がある。これに対して, 井元・小西 (1999) は, これらの点を理論的に解決し, 罰則付き最尤法に基づいて推定したモデルの良さを測る一つの情報量規準 SPIC を提案した。本稿では, 数値実験を通して種々のモデル評価規準を比較検討し, その特徴・性質および有効性に関して考察する。

2. モデル

説明変数 $X (\in \mathbb{R})$ と目的変数 $Y (\in \mathbb{R})$ に関して大きさ n のデータ $\{(x_\alpha, y_\alpha); \alpha = 1, 2, \dots, n\}$ が観測されたとする。一般に回帰モデルは, 各 x_α におけるデータ y_α の確率的変動を表す成分と, その条件付き期待値 $E[Y_\alpha | x_\alpha] = \mu_\alpha (\alpha = 1, 2, \dots, n)$ に対して仮定する系統的成分からなる。ここでは, データは

$$(2.1) \quad y_\alpha = \mu_\alpha + \varepsilon_\alpha, \quad \varepsilon_\alpha \sim n(0, \sigma^2), \quad \alpha = 1, 2, \dots, n$$

に従って生成されたとする。ただし, $n(0, \sigma^2)$ は, 平均 0, 分散 σ^2 の正規分布を表す。

データの構造を説明変数の線形結合 $\mu_\alpha = \beta_0 + \beta_1 x_\alpha$ で表したものが, 周知の正規線形モデルであり, また, 十分滑らかな関数 $w(\cdot)$ を用いて $\mu_\alpha = w(x_\alpha)$ と仮定したものはノンパラメトリック回帰モデルと呼ばれる。実用上, 関数 $w(\cdot)$ としては, 区分的多項式で与えられる自然 3 次スプラインを用いることが多い (Green and Silverman (1994), p. 12)。曲線 $w(x)$ の推定は, 各データ x_α での値 $w(x_\alpha) = w_\alpha (\alpha = 1, \dots, n)$ を推定する問題に帰着され, $w = (w_1, \dots, w_n)'$ とパラメトライズする。しかし, 自然 3 次スプラインを用いた非線形回帰モデルは, (1) モデルのパラメータ数が標本数を上回る, (2) データ点 x_α が重複するとパラメータ推定の際に困難が生じる (井元・小西 (1999)) などの欠点を持つ。そこで, 本稿では, 滑らかな回帰関数を B -スプラインを用いて構成した非線形回帰モデルを取り上げ, 曲線推定を行う方法について検討する。

B -スプラインとは, 基底関数の線形結合で滑らかな関数を表現する手法である。図 1 には次数が 3 の B -スプライン基底関数の例を与える。各基底関数 B_j は, 節点と呼ばれる等間隔に配置された点 t_i において滑らかに連結した既知の区分的多項式で構成される。例えば, 1 番目の基底関数 B_1 は, 5 つの節点 t_1, \dots, t_5 において滑らかに連結した 4 つの 3 次多項式で構成されている。ただし, t_1 と t_5 では軸に滑らかに連結している。ここで, 3 次 B -スプライン基底関数に対して「滑らかさ」とは, 2 階微分導関数が連続であることをいう。また基底関数は, データ

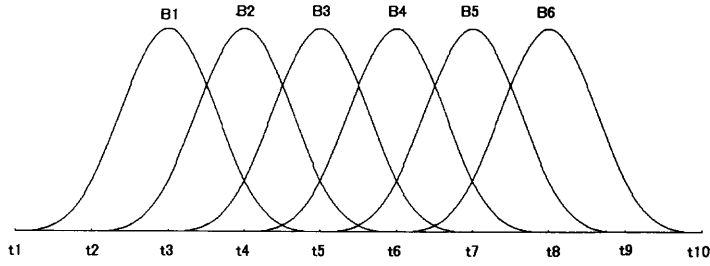


図1. 節点 $t_i (i = 1, \dots, 10)$ をもつ次数3の B-スプライン基底関数.

の点在する区間を等間隔に分割し、各小区間を4つの基底関数で覆うように構成する。従って、図1では、 t_4 から t_8 の範囲にデータは存在し、その区間を3等分して6個の基底関数は構成される。ここで、節点は必ずしもデータ点 x_a に一致しないことに注意されたい。B-スプラインの基底関数の個数を調整することによって、パラメータ数をデータ数よりも大幅に減少させることが可能となる。

基底関数は、節点の幅が決まると一意に決定される関数である。すなわち、各節点の幅を h とし $t_3 = 0$ としたとき、 $B_1(x)$ は t_3 に関して対称であり

$$B_1(x) = \begin{cases} \frac{1}{6h} \left\{ \left(2 - \frac{1}{h}x \right)^3 - 4 \left(1 - \frac{1}{h}x \right)^3 \right\} & (t_3 = 0 < x < t_4 = h) \\ \frac{1}{6h} \left(2 - \frac{1}{h}x \right)^3 & (t_4 = h < x < t_5 = 2h) \\ 0 & (t_5 = 2h < x) \end{cases}$$

となる。他の基底関数 B_2, B_3, \dots は、 B_1 を幅 h ずつ平行移動すると得られる。 m 個の基底関数 $B_j(x) (j = 1, 2, \dots, m)$ を用いた B-スプラインによる回帰関数の表現は次式で与えられる。

$$(2.2) \quad E[Y_a | x_a] = \mu_a = \sum_{j=1}^m a_j B_j(x_a) = \mathbf{b}_a \mathbf{a}.$$

ここで、 $\mathbf{b}_a = (B_1(x_a), \dots, B_m(x_a))' \in \mathbb{R}^m$ 、 $\mathbf{a} = (a_1, \dots, a_m)' \in \mathbb{R}^m (m < n)$ とする。これは、複数の基底関数に基づく非線形関数 $y = \sum_{j=1}^m a_j B_j(x)$ のデータへの当てはめと考えることができる。 \mathbf{b}_a は基底関数とデータから決定される既知のベクトルであるから、B-スプラインによる曲線推定問題は、 m 次元係数ベクトル \mathbf{a} を推定することに帰着される。

(2.1) 式で想定したデータの確率的変動と (2.2) 式の B-スプラインによって、B-スプライン非線形回帰モデルは

$$(2.3) \quad f(y_a | x_a; \mathbf{a}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_a - \mathbf{b}_a \mathbf{a})^2}{2\sigma^2} \right\}$$

と表せる。ある程度多くの基底関数を用いたとき、モデルのパラメータ \mathbf{a} と σ^2 を最尤法によって推定すると、モデルの柔軟性の故にデータに強度に依存したモデルが推定される。そこで、パラメータは、対数尤度に曲線の局所変動の程度を考慮に入れた、次の罰則付き対数尤度関数

$$(2.4) \quad \begin{aligned} l_n(\mathbf{a}, \sigma^2) &= \sum_{\alpha=1}^n \log f(y_\alpha | x_\alpha; \mathbf{a}, \sigma^2) - \frac{n\lambda}{2} k(\mathbf{a}) \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{\alpha=1}^n (y_\alpha - \mathbf{b}_\alpha \mathbf{a})^2 - \frac{n\lambda}{2} k(\mathbf{a}) \end{aligned}$$

の最大化に基づいて推定する。ここで、 $\lambda (> 0)$ は平滑化パラメータと呼ばれ、推定曲線の局所変動の程度を制御するハイパーパラメータである。 $k(\mathbf{a})$ は回帰曲線の変動に対する罰則項であり、 B -スプライン非線形回帰モデルにおいては次の2次形式で与える。

$$k(\mathbf{a}) = \sum_{j=k+1}^m (\Delta^k a_j)^2 = \mathbf{a}' D_k D_k \mathbf{a}.$$

ただし D_k は、 k 階差分を与える $(m-k) \times m$ 次の行列

$$(2.5) \quad D_k = \begin{pmatrix} (-1)^0_k C_0 & \cdots & (-1)^k_k C_k & 0 & \cdots & 0 \\ 0 & (-1)^0_k C_0 & \cdots & (-1)^k_k C_k & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & (-1)^0_k C_0 & \cdots & (-1)^k_k C_k \end{pmatrix}$$

である。ただし、 ${}_k C_i$ は2項係数である。

Eilers and Marx (1996) は、3次 B -スプライン基底関数を用いたときに、2階差分の罰則 $\mathbf{a}' D_2 D_2 \mathbf{a}$ に対して

$$\sum_{j=3}^m (a_{j-2} - 2a_{j-1} + a_j)^2 = \mathbf{a}' D_2 D_2 \mathbf{a} \approx \int \left\{ \frac{d^2 w(t)}{dt^2} \right\}^2 dt = \int \left\{ \sum_{j=1}^m a_j \frac{d^2 B_j(t)}{dt^2} \right\}^2 dt$$

の近似が成り立つことを示した。石黒・荒畑 (1982) では、2次 B -スプラインに対し、2階差分罰則項について、また、小川 他 (1995) では、1次平滑化スプラインにおける罰則項について研究されている。

(2.4) 式の罰則付き対数尤度関数の最大化に基づく推定値 $\hat{\mathbf{a}}$ 、 $\hat{\sigma}^2$ を (2.3) 式の確率密度関数 $f(y_\alpha | x_\alpha; \mathbf{a}, \sigma^2)$ に代入した

$$f(y_\alpha | x_\alpha; \hat{\mathbf{a}}, \hat{\sigma}^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp \left\{ -\frac{(y_\alpha - \mathbf{b}_\alpha \hat{\mathbf{a}})^2}{2\hat{\sigma}^2} \right\}$$

が一つの統計モデルである。モデルのパラメータ \mathbf{a} 、 σ^2 の推定値は、次のアルゴリズムによって求めることが出来る。

1. 初期値を σ_0^2 とし、 λ はあらかじめ与えておく。
2. $\mathbf{a}_{(t)} = (B'B + n\sigma_{(t-1)}^2 \lambda D_k D_k)^{-1} B'y$ で更新する ($t=1, 2, \dots$)。
3. $\sigma_{(t)}^2 = \frac{1}{n} \|y - B\mathbf{a}_{(t)}\|^2$ を求め $(\sigma_{(t)}^2 - \sigma_{(t-1)}^2)^2 < \delta$ を評価し、この条件を満たすまで2のプロセスを反復する。
4. 3の条件式を満たしたパラメータを最終的な推定値とする。

推定値 $\hat{\mathbf{a}}$ 、 $\hat{\sigma}^2$ は、平滑化パラメータ λ と基底関数の個数 m に依存し、従って適切な λ と m の選択を行う必要がある。次節では、これまでに提案された種々の選択規準を取り上げる。

3. モデル評価規準

各点 x_α の期待値 $\mu_\alpha = E[Y_\alpha|x_\alpha]$ の予測値 \hat{y}_α は,

$$S = B(B'B + n\sigma^2\lambda D_k D_k)^{-1}B'$$

と定義される B-スプライン非線形回帰モデルのハット行列 S に対して $\hat{\mathbf{y}} = S\mathbf{y}$ と表せる。ただし、 $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)$ 、 $n \times m$ 次行列 B は (2.2) 式で定義した基底関数ベクトル $\mathbf{b}_\alpha = (B_1(x_\alpha), \dots, B_m(x_\alpha))'$ に対して $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ とし、 D_k は (2.5) 式の差分行列とする。これまでに平滑化パラメータを選択するために用いられてきた規準としては、次のものが上げられる。

(1) 交差検証法 (Cross-Validation, CV) :

いま、 $\hat{w}^{(-\alpha)}$ は n 個のデータ $\{(x_\alpha, y_\alpha); \alpha = 1, \dots, n\}$ から α 番目のデータ (x_α, y_α) を除いた残り $n-1$ 個のデータで推定した回帰曲線とする。このとき、交差検証法は、次の値を最小にするモデルを選択する。

$$(3.1) \quad CV = \frac{1}{n} \sum_{\alpha=1}^n (y_\alpha - \hat{w}^{(-\alpha)}(x_\alpha))^2.$$

Green and Silverman ((1994), p. 31, Theorem 3.1) により、CV は次式に帰着される。

$$(3.2) \quad CV = \frac{1}{n} \sum_{\alpha=1}^n \left(\frac{y_\alpha - \hat{w}(x_\alpha)}{1 - S_{\alpha\alpha}} \right)^2.$$

ただし、 $S_{\alpha\alpha}$ はハット行列の第 α 対角成分とする。

(2) 一般化交差検証法 (Generalized Cross-Validation, GCV) (Craven and Wahba (1979); Wahba (1990)) :

交差検証法 (3.2) 式において、 $S_{\alpha\alpha}$ を S の対角成分の平均である $\text{tr } S/n$ で置き換えたものが次の一般化交差検証法である。

$$(3.3) \quad GCV = \frac{1}{n} \sum_{\alpha=1}^n \left(\frac{y_\alpha - \hat{w}(x_\alpha)}{1 - \text{tr } S/n} \right)^2.$$

GCV は次のように書き直すことができる (Green and Silverman (1994), p. 37)。

$$GCV = \frac{1}{n} \sum_{\alpha=1}^n \left\{ \left(\frac{1 - S_{\alpha\alpha}}{1 - \text{tr } S/n} \right)^2 (y_\alpha - \hat{w}^{(-\alpha)}(x_\alpha))^2 \right\}.$$

つまり、GCV は (3.1) 式の CV に対して重み付きの平均を取っていると見なせる。

交差検証法と一般化交差検証法は、データへの当てはまりを考慮に入れた予測の意味での 2 乗誤差を最小にする平滑化パラメータを選択する規準である。坂本・白旗 (1996)、Sakamoto and Shirahata (1997) は、交差検証法、一般化交差検証法を用いて平滑化パラメータの選択を実行する際の計算量を改良する方法を提案した。

(3) 赤池のベイズ型情報量規準 (ABIC) (Akaike (1980a, 1980b)) :

B-スプライン非線形回帰モデル (2.3) 式のパラメータ \mathbf{a} の事前分布を $\pi(\mathbf{a}|\lambda)$ とする。ここで、 $\lambda (> 0)$ は事前分布を規定するパラメータでハイパーパラメータと呼ばれ、事前分布の設定の仕方によっては、罰則付き対数尤度関数の平滑化パラメータに相当する。データ $\{(x_\alpha, y_\alpha); \alpha = 1, \dots, n\}$ が与えられたとき、モデルのパラメータ \mathbf{a} の実現確率を表すのが、次式で定義される事後分布である。

$$(3.4) \quad \pi(\mathbf{a}|\mathbf{y}; \lambda) = \frac{\prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \pi(\mathbf{a}|\lambda)}{\int \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \pi(\mathbf{a}|\lambda) d\mathbf{a}}$$

Akaike (1980) は、事後分布の分母に相当するモデルの周辺尤度の最大化、あるいは

$$(3.5) \quad \text{ABIC} = -2 \log \left\{ \int \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \pi(\mathbf{a}|\lambda) d\mathbf{a} \right\}$$

の最小化によって σ^2 , λ を推定し、その値 $\hat{\sigma}^2$, $\hat{\lambda}$ を代入した

$$\prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \hat{\sigma}^2) \pi(\mathbf{a}|\hat{\lambda})$$

を最大にする \mathbf{a} の値を推定値として用いる方法を提唱した。ハイパーパラメータを選択する (3.5) 式は、ベイズ型情報量規準 ABIC と呼ばれ、諸分野のデータの分析を通してその有効性が広く認められている (例えば、赤池・北川 編 (1994, 1995); Bozdogan (1994); Kitagawa and Gersch (1996) 等を参照されたい)。

(2.4) 式の罰則付き対数尤度関数は

$$\log \left\{ \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \exp \left(-\frac{n\lambda}{2} \boldsymbol{\alpha}' D_k D_k \boldsymbol{\alpha} \right) \right\}$$

と書き直すことができる。さらに、 \mathbf{a} を m 次元確率ベクトルと考えると、罰則項を規格化定数で調整すると

$$(3.6) \quad \log \left\{ \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \frac{n\lambda D_k D_k |_{+}^{1/2}}{(2\pi)^{(m-k)/2}} \exp \left(-\frac{n\lambda}{2} \boldsymbol{\alpha}' D_k D_k \boldsymbol{\alpha} \right) \right\} \\ := \log \left\{ \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \pi(\mathbf{a}|\lambda) \right\}$$

と表すことができる。ただし、 $|n\lambda D_k D_k|_{+}$ は $n\lambda D_k D_k$ の非零固有値の積を表す。従って罰則付き対数尤度法は、ハイパーパラメータ $\lambda (> 0)$ をもつ improper な事前分布をもつベイズモデルに対応していることが分かる。もし、 λ , σ^2 が与えられたとすると、パラメータ \mathbf{a} に対する罰則付き対数尤度関数の最大化は、(3.4) 式の最大化、すなわち事後分布のモードを求める問題に帰着される。

(3.6) 式で定義した improper な事前分布 $\pi(\mathbf{a}|\lambda)$ は、退化した m 次元正規分布であることから B -スプライン非線形回帰モデルに対する ABIC は

$$(3.7) \quad \text{ABIC} = -2 \log \int \left\{ \prod_{a=1}^n f(y_a|x_a; \mathbf{a}, \sigma^2) \pi(\mathbf{a}|\lambda) \right\} d\mathbf{a} \\ = (n-k) \log(2\pi) + (n-k) \log \sigma^2 - (m-k) \log(n\beta) - \log \psi \\ + \log \{ \det(B' B + n\beta D_k D_k) \} + \frac{\|\mathbf{y} - B\hat{\mathbf{a}}\|^2 + n\beta \hat{\mathbf{a}}' D_k D_k \hat{\mathbf{a}}}{\sigma^2}$$

で与えられる。ただし、 $\beta = \sigma^2 \lambda$, $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)'$ であり、 ψ は $D_k D_k$ の $m-k$ 個の非零固有値の積、 \det は行列式を表す。

罰則付き対数尤度関数とベイズ型情報量規準 ABIC との関係については、田辺・田中 (1983), Kitagawa and Gersch (1984), 田辺 (1989) を参照されたい。石黒・荒畑 (1982), 田辺・田中 (1983) は、それぞれ ABIC に基づいて B -スプライン、離散スプライン非線形回帰モデルの平滑化パラメータを選択する方法を提案した。

(4) 修正 AIC (Eilers and Marx (1996)) :

Eilers and Marx (1996) は、モデルに含まれる自由パラメータ数をハット行列の固有値の和、つまり $\text{tr } S$ を用いた

$$\text{mAIC}^* = -2 \sum_{\alpha=1}^n \log f(y_\alpha | x_\alpha; \hat{\mathbf{a}}, \hat{\sigma}_0^2) + 2 \text{tr } S$$

に基づいて B-スプライン平滑化パラメータを選択する方法を提案した。ただし、 $\hat{\sigma}_0^2$ は、 $\lambda = 0$ のときの予測値 $\hat{\mathbf{y}}_0$ を用いて $\hat{\sigma}_0^2 = \|\mathbf{y} - \hat{\mathbf{y}}_0\|^2/n = \|\mathbf{y} - B(B'B)^{-1}B'\mathbf{y}\|^2/n$ とする。しかし、この方法は、基底関数の個数をデータ数よりも十分に小さくとった場合には $B'B$ の逆行列は計算できるが、逆に基底関数を増やしていくと $B'B$ の行列式の値が極めて小さくなり、逆行列の計算が不安定となる。そこで、次節の数値実験では σ^2 を $\hat{\sigma}^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2/n = \|\mathbf{y} - B\hat{\mathbf{a}}\|^2/n$ で推定し、モデルの評価に次式を用いた。

$$\text{mAIC} = -2 \sum_{\alpha=1}^n \log f(y_\alpha | x_\alpha; \hat{\mathbf{a}}, \hat{\sigma}^2) + 2(\text{tr } S + 1).$$

ハット行列に基づく自由度の調整については Hastie and Tibshirani ((1990), p. 52) を参照されたい。

(5) Bias Corrected AIC (AICc) (Hurvich et al. (1998)) :

データを生成した真のモデルは、想定した正規線形モデル

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim n(0, \sigma^2 I) \quad (\boldsymbol{\beta} \in \mathbb{R}^p)$$

に含まれると仮定する。このとき Sugiura (1978) は、モデルの対数尤度で平均対数尤度を推定したときのバイアスは次式で与えられることを示した。

$$(3.8) \quad b = E_{F(\mathbf{y})} \left[\log f(\mathbf{y} | \hat{\boldsymbol{\theta}}) - \int f(\mathbf{z} | \hat{\boldsymbol{\theta}}) \log f(\mathbf{z} | \hat{\boldsymbol{\theta}}) d\mathbf{z} \right] = \frac{n(p+1)}{n-p-2}.$$

ただし、 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}^2)' \in \mathbb{R}^{p+1}$ は最尤推定量とする。これに対して、Hurvich et al. (1998) は、 p を $\text{tr } S$ で置き換えた次の AICc に基づいて平滑化パラメータを選択する方法を提案した。

$$\text{AICc} = -2 \sum_{\alpha=1}^n \log f(y_\alpha | x_\alpha; \hat{\mathbf{a}}, \hat{\sigma}^2) + \frac{2n(\text{tr } S + 1)}{n - \text{tr } S - 2}.$$

しかし、B-スプライン非線形回帰モデルのパラメータ \mathbf{a} 、 σ^2 は罰則付き対数尤度関数に基づいて推定されるのに対して、(3.8) 式の推定量 $\hat{\boldsymbol{\theta}}$ は最尤推定量であり、この点に理論的に問題が残ると思われる。

(6) スプライン情報量規準 (SPIC) (井元・小西 (1999)) :

罰則付き対数尤度関数に基づいて推定されたパラメータをもつ B-スプライン非線形回帰モデルを評価する情報量規準 SPIC は次式で与えられた。

$$\text{SPIC} = -2 \sum_{\alpha=1}^n \log f(y_\alpha | x_\alpha; \hat{\mathbf{a}}, \hat{\sigma}^2) + 2 \text{tr} (\hat{I}_{\lambda, m} \hat{J}_{\lambda, m}^{-1}).$$

ただし、 $\hat{I}_{\lambda, m}$ と $\hat{J}_{\lambda, m}$ は $(m+1) \times (m+1)$ 次行列で次式で与えられる。

$$\hat{I}_{\lambda, m} = \begin{pmatrix} \hat{I}_{11} & \hat{i}_{12} \\ \hat{i}'_{21} & \hat{i}_{22} \end{pmatrix}, \quad \hat{J}_{\lambda, m} = \begin{pmatrix} \hat{J}_{11} & \hat{j}_{12} \\ \hat{j}'_{21} & \hat{j}_{22} \end{pmatrix}.$$

各々の行列の要素は、

$$\begin{aligned} \mathbf{1}_n (n \times 1) &= (1, 1, \dots, 1)', & \Lambda (n \times n) &= \text{diag} [y_1 - \hat{\mu}_1, \dots, y_n - \hat{\mu}_n], \\ \mathbf{p} (n \times 1) &= (p_1, \dots, p_n)'; & p_a &= \frac{(y_a - \hat{\mu}_a)^2}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^2}, \\ \mathbf{q} (n \times 1) &= (q_1, \dots, q_n)'; & q_a &= \frac{\partial p_a}{\partial \hat{\sigma}^2} \end{aligned}$$

とおくとき,

$$\begin{aligned} \hat{I}_{11} &= B' \Lambda^2 B / \hat{\sigma}^2 - \lambda D_k' D_k \hat{\mathbf{a}} \mathbf{1}_n' \Lambda B, & \hat{I}_{12} &= B' \Lambda \mathbf{p} - \lambda \hat{\sigma}^2 \mathbf{p}' \mathbf{1}_n D_k' D_k \hat{\mathbf{a}}, \\ \hat{I}_{21} &= B' \Lambda \mathbf{p}, & \hat{I}_{22} &= \hat{\sigma}^2 \mathbf{p}' \mathbf{p}, \\ \hat{J}_{11} &= B' B + n \lambda \hat{\sigma}^2 D_k' D_k, & \hat{J}_{12} &= B' \Lambda \mathbf{1}_n / \hat{\sigma}^2, \\ \hat{J}_{21} &= \hat{J}_{12}, & \hat{J}_{22} &= -\hat{\sigma}^2 \mathbf{q}' \mathbf{1}_n \end{aligned}$$

である。このスプライン情報量規準は、Kullback-Leibler 情報量あるいは期待平均対数尤度の推定量として一般化情報量規準 (Generalized Information Criteria; Konishi and Kitagawa (1996)) の理論に基づき導出したものである。情報量規準構成の一般理論は、Konishi and Kitagawa (1996), Konishi (1999) を参照されたい。また、罰則付き対数尤度関数に基づいて推定されたモデルの評価規準については、Shibata (1989) の研究がある。

推定されたモデルの対数尤度は、 σ^2 を $\hat{\sigma}^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2/n = \|\mathbf{y} - B\hat{\mathbf{a}}\|^2/n$ と推定したとき次式で与えられる。

$$-2 \sum_{a=1}^n \log f(y_a | x_a; \hat{\mathbf{a}}, \hat{\sigma}^2) = n \log \hat{\sigma}^2 + n \log (2\pi) + n.$$

このとき、GCV, mAIC, AICc, SPIC は定数項を除いて

$$(3.9) \quad n \log \text{GCV} = n \log \hat{\sigma}^2 - 2n \log (1 - \text{tr } S/n),$$

$$(3.10) \quad \text{mAIC} = n \log \hat{\sigma}^2 + 2(\text{tr } S + 1),$$

$$(3.11) \quad \begin{aligned} \text{AICc} &= n \log \hat{\sigma}^2 + \frac{2n(\text{tr } S + 1)}{n - \text{tr } S - 2}, \\ \text{SPIC} &= n \log \hat{\sigma}^2 + 2 \text{tr} (\hat{I}_{\lambda, m} \hat{J}_{\lambda, m}^{-1}) \end{aligned}$$

となる。第1項は各規準で共通なので、実質これらの規準の差は第2項であるバイアス補正に関する項に現れることがわかる。また、

$$\begin{aligned} \frac{n(\text{tr } S + 1)}{n - \text{tr } S - 2} &= (\text{tr } S + 1) \left\{ 1 + \frac{1}{n} (\text{tr } S + 2) + \frac{1}{n^2} (\text{tr } S + 2)^2 + \dots \right\} \\ &\rightarrow \text{tr } S + 1 \quad (\text{as } n \rightarrow \infty) \end{aligned}$$

であるから、データ数が十分大きいとき AICc は漸近的に mAIC に一致する。また、(3.9) 式の GCV に関しては、

$$\begin{aligned} -2n \log (1 - \text{tr } S/n) &= -2n \{-\text{tr } S/n - (\text{tr } S/n)^2 - \dots\} \\ &\rightarrow 2 \text{tr } S \quad (\text{as } n \rightarrow \infty) \end{aligned}$$

であるから、mAIC, AICc のバイアス補正項から誤差分散の推定による補正を無視すると (3.10), (3.11) 式に含まれる $\text{tr } S + 1$ は $\text{tr } S$ となり、漸近的には GCV, mAIC, AICc は同等であることが分かる。次節では、ここで紹介した各規準について、数値実験を通してその特徴および有効性に関して考察する。

4. 数値実験

本節では、3節で紹介した6つの規準を用いてB-スプライン非線形回帰モデルの平滑化パラメータを選択する数値実験を行い、その特徴および有効性を検討する。データは

$$y_\alpha = g(x_\alpha) + \varepsilon_\alpha, \quad \alpha = 1, \dots, n$$

に従って発生させ、真の構造 $g(x)$ には次の曲線を仮定した。

1. $g(x) = e^{-3x} \sin(3\pi x)$ ($0 \leq x \leq 1$).
2. $g(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$ ($0 \leq x \leq 1$).

1の曲線は、左側で大きく変動する特徴をもち、2の曲線は、2つの極小値と全体として緩やかな正のトレンドをもつ。データ点 x_α ($\alpha = 1, \dots, n$) は、(1)一様乱数により発生、(2)等間隔に配置、の2つの設定のもとで実験を行なった。誤差項 ε_α に対しては、正規分布と混合正規分布を仮定し、その分散は次のように設定した。

$$\text{(正規分布): } \varepsilon_\alpha \sim n(0, \tau^2), \quad \tau = 0.01R_y, 0.05R_y, 0.1R_y, 0.15R_y, 0.2R_y,$$

$$\text{(混合正規分布): } \varepsilon_\alpha \sim \varepsilon n(0, (0.05R_y)^2) + (1 - \varepsilon) n(0, (0.2R_y)^2), \quad \varepsilon = 0.8.$$

ただし、 R_y は定義域 $x \in [0, 1]$ における真の曲線 $g(x)$ の値域の幅を表し、1の曲線に対しては $R_y = 0.872$ 、2の曲線に対しては $R_y = 3.738$ とした。混合正規分布は、正規分布よりも裾の重い確率分布に従う誤差を仮定している。また、データ数は、 $n = 50, 100$ として実験を行なった。図2, 3には、このモデルに従って発生したデータの1例と情報量規準 SPIC に基づく推定曲線を上げた。

情報量規準 mAIC, AICc, SPIC で用いるパラメータ $\hat{\mathbf{a}}$ 、 $\hat{\sigma}^2$ の推定値については、2章のアルゴリズムに従って求めた。そのときのハット行列は $S = B(B'B + n\hat{\sigma}^2\lambda D_k D_k)^{-1}B'$ となる。また、CV, GCV, ABIC の3つに関して \mathbf{a} の推定量は、 $\hat{\mathbf{a}} = (B'B + n\beta D_k D_k)^{-1}B'\mathbf{y}$ とし、 β を与えると推定量は求まる。このとき、ABIC に用いる分散 σ^2 の推定量は次式で与えられる。

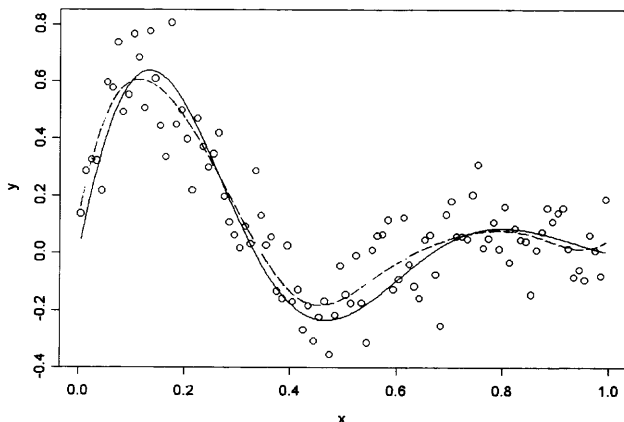


図2. $g(x) = e^{-3x} \sin(3\pi x)$, $n = 100$, $\varepsilon_\alpha \sim n(0, (0.15R_y)^2)$; 実線は真の曲線 $y = g(x)$, 破線は SPIC による推定曲線を表す。

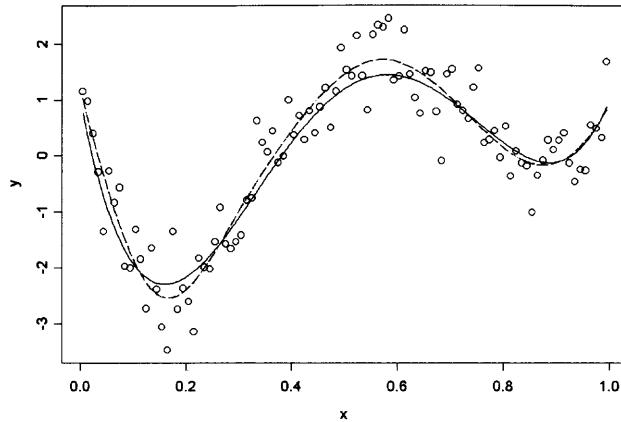


図3. $g(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$, $n = 100$, $\epsilon_a \sim n(0, (0.15R_y)^2)$; 実線は真の曲線 $y = g(x)$, 破線は SPIC による推定曲線を表す。

表1. $g(x) = e^{-3x} \sin(3\pi x)$, x : 一様乱数, $n = 100$.

	CV	GCV	ABIC	mAIC	AICc	SPIC
0.01R_y						
mean	1.72×10^{-6}	9.11×10^{-7}	1.22×10^{-6}	8.07×10^{-7}	1.05×10^{-6}	8.93×10^{-7}
SD	1.51×10^{-6}	4.83×10^{-7}	2.80×10^{-7}	4.18×10^{-7}	5.77×10^{-7}	7.15×10^{-7}
MSE	9.39×10^{-6}	9.35×10^{-6}	9.31×10^{-6}	9.34×10^{-6}	9.35×10^{-6}	9.31×10^{-6}
0.05R_y						
mean	3.46×10^{-5}	4.04×10^{-5}	3.33×10^{-5}	3.58×10^{-5}	4.60×10^{-5}	2.49×10^{-5}
SD	3.47×10^{-5}	3.37×10^{-5}	8.98×10^{-6}	3.12×10^{-5}	3.63×10^{-5}	2.77×10^{-5}
MSE	1.92×10^{-4}	1.92×10^{-4}	1.83×10^{-4}	1.92×10^{-4}	1.93×10^{-4}	1.92×10^{-4}
0.1R_y						
mean	1.28×10^{-4}	1.52×10^{-4}	1.49×10^{-4}	1.37×10^{-4}	1.73×10^{-4}	1.05×10^{-4}
SD	9.72×10^{-5}	1.06×10^{-4}	3.74×10^{-5}	9.83×10^{-5}	1.14×10^{-4}	8.57×10^{-5}
MSE	7.28×10^{-4}	7.19×10^{-4}	6.89×10^{-4}	7.20×10^{-4}	7.19×10^{-4}	7.30×10^{-4}
0.15R_y						
mean	2.73×10^{-4}	2.99×10^{-4}	3.43×10^{-4}	2.78×10^{-4}	3.41×10^{-4}	2.23×10^{-4}
SD	2.23×10^{-4}	2.17×10^{-4}	1.37×10^{-4}	2.13×10^{-4}	2.24×10^{-4}	2.05×10^{-4}
MSE	1.53×10^{-3}	1.50×10^{-3}	1.42×10^{-3}	1.52×10^{-3}	1.49×10^{-3}	1.55×10^{-3}
0.2R_y						
mean	4.52×10^{-4}	5.44×10^{-4}	6.26×10^{-4}	4.82×10^{-4}	6.07×10^{-4}	3.63×10^{-4}
SD	3.11×10^{-4}	3.42×10^{-4}	2.47×10^{-4}	3.19×10^{-4}	3.64×10^{-4}	3.03×10^{-4}
MSE	2.43×10^{-3}	2.39×10^{-3}	2.30×10^{-3}	2.41×10^{-3}	2.38×10^{-3}	2.48×10^{-3}

$$\hat{\sigma}_{ABIC}^2 = \frac{1}{n-k} \{ \|y - B\hat{a}\|^2 + n\beta\hat{a}'D_kD_k\hat{a} \}.$$

この推定量は、(3.7) 式を σ^2 で偏微分することによって得られる。

ハット行列 $S = B(B'B + n\hat{\sigma}^2\lambda D_kD_k)^{-1}B'$ において負荷行列 D_kD_k の係数を n で割った $\hat{\sigma}^2\lambda$ (mAIC, AICc, SPIC) および β (CV, GCV, ABIC) のことを平滑化量と呼び、各規準がどのような平滑化量を選択し、その結果として推定した予測値と仮定した真の構造とがどの程度近いかに着目する。最終的に得られる y_a の予測値 \hat{y}_a は、平滑化量の大小で特徴づけられる。

表2. $g(x) = e^{-3x} \sin(3\pi x)$, x : 等間隔, $n = 50$.

	CV	GCV	ABIC	mAIC	AICc	SPIC
$0.01R_y$						
mean	4.76×10^{-6}	1.92×10^{-6}	2.38×10^{-6}	1.44×10^{-6}	2.64×10^{-6}	2.13×10^{-6}
SD	4.38×10^{-6}	7.07×10^{-7}	6.35×10^{-7}	5.24×10^{-7}	9.82×10^{-7}	1.89×10^{-6}
MSE	1.73×10^{-5}	1.69×10^{-5}	1.68×10^{-5}	1.69×10^{-5}	1.69×10^{-5}	1.70×10^{-5}
$0.05R_y$						
mean	5.23×10^{-5}	7.76×10^{-5}	6.72×10^{-5}	6.12×10^{-5}	1.09×10^{-4}	3.60×10^{-5}
SD	5.34×10^{-5}	5.61×10^{-5}	1.87×10^{-5}	4.93×10^{-5}	6.71×10^{-5}	3.74×10^{-5}
MSE	3.52×10^{-4}	3.50×10^{-4}	3.41×10^{-4}	3.50×10^{-4}	3.56×10^{-4}	3.54×10^{-4}
$0.1R_y$						
mean	1.94×10^{-4}	2.69×10^{-4}	2.65×10^{-4}	2.17×10^{-4}	3.60×10^{-4}	1.44×10^{-4}
SD	1.88×10^{-4}	1.99×10^{-4}	9.06×10^{-5}	1.81×10^{-4}	2.28×10^{-4}	1.56×10^{-4}
MSE	1.42×10^{-3}	1.39×10^{-3}	1.27×10^{-3}	1.40×10^{-3}	1.38×10^{-3}	1.45×10^{-3}
$0.15R_y$						
mean	3.82×10^{-4}	5.26×10^{-4}	6.28×10^{-4}	4.21×10^{-4}	7.08×10^{-4}	2.84×10^{-4}
SD	3.52×10^{-4}	3.61×10^{-4}	2.66×10^{-4}	3.39×10^{-4}	4.26×10^{-4}	2.88×10^{-4}
MSE	2.88×10^{-3}	2.82×10^{-3}	2.65×10^{-3}	2.86×10^{-3}	2.80×10^{-3}	2.94×10^{-3}
$0.2R_y$						
mean	7.28×10^{-4}	9.15×10^{-4}	1.31×10^{-3}	7.22×10^{-4}	1.27×10^{-4}	4.99×10^{-4}
SD	6.76×10^{-4}	6.36×10^{-4}	7.42×10^{-4}	5.82×10^{-4}	7.57×10^{-4}	5.52×10^{-4}
MSE	4.58×10^{-3}	4.37×10^{-3}	4.20×10^{-3}	4.48×10^{-3}	4.33×10^{-3}	4.83×10^{-3}

表3. $g(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$, x : 等間隔, $n = 100$.

	CV	GCV	ABIC	mAIC	AICc	SPIC
$0.01R_y$						
mean	7.28×10^{-7}	3.73×10^{-7}	8.49×10^{-7}	3.35×10^{-7}	4.23×10^{-7}	3.87×10^{-7}
SD	9.59×10^{-7}	7.62×10^{-8}	1.64×10^{-7}	7.21×10^{-8}	9.32×10^{-8}	4.65×10^{-7}
MSE	1.36×10^{-4}	1.35×10^{-4}	1.35×10^{-4}	1.35×10^{-4}	1.35×10^{-4}	1.35×10^{-4}
$0.05R_y$						
mean	1.10×10^{-5}	1.33×10^{-5}	2.89×10^{-5}	1.20×10^{-5}	1.57×10^{-5}	7.72×10^{-6}
SD	1.59×10^{-5}	1.39×10^{-5}	1.02×10^{-5}	1.24×10^{-5}	1.57×10^{-5}	8.66×10^{-6}
MSE	3.46×10^{-3}	3.50×10^{-3}	3.61×10^{-3}	3.48×10^{-3}	3.52×10^{-3}	3.44×10^{-3}
$0.1R_y$						
mean	9.20×10^{-5}	1.04×10^{-4}	1.65×10^{-4}	9.05×10^{-5}	1.28×10^{-4}	6.52×10^{-5}
SD	9.55×10^{-5}	1.07×10^{-4}	4.42×10^{-5}	9.48×10^{-5}	1.21×10^{-4}	7.84×10^{-5}
MSE	1.39×10^{-2}	1.39×10^{-2}	1.43×10^{-2}	1.39×10^{-2}	1.40×10^{-2}	1.39×10^{-2}
$0.15R_y$						
mean	2.82×10^{-4}	3.20×10^{-4}	3.79×10^{-4}	2.81×10^{-4}	3.84×10^{-4}	2.27×10^{-4}
SD	3.02×10^{-4}	2.93×10^{-4}	1.16×10^{-4}	2.76×10^{-4}	3.14×10^{-4}	2.80×10^{-4}
MSE	3.11×10^{-2}	3.06×10^{-2}	2.93×10^{-2}	3.07×10^{-2}	3.06×10^{-2}	3.12×10^{-2}
$0.2R_y$						
mean	5.43×10^{-4}	6.01×10^{-4}	7.30×10^{-4}	5.38×10^{-4}	6.99×10^{-4}	4.21×10^{-4}
SD	4.82×10^{-4}	4.73×10^{-4}	2.67×10^{-4}	4.50×10^{-4}	5.15×10^{-4}	4.51×10^{-4}
MSE	4.96×10^{-2}	4.86×10^{-2}	4.62×10^{-2}	4.88×10^{-2}	4.84×10^{-2}	5.02×10^{-2}

実験はそれぞれの設定で100回の繰り返しを行った。表1~5は、実験結果をまとめたものである。ただし、meanは選択した平滑化量の平均、SDはその標準偏差、AMSEは、真の曲線と推定されたB-スプライン曲線との予測値における平均2乗誤差を表す。なお、表中の結果は基底関数の個数を10としたときのものであるが、データの構造を捉えることができないような

表4. $g(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$, x : 一様乱数, $n = 50$.

	CV	GCV	ABIC	mAIC	AICc	SPIC
$0.01R_y$						
mean	1.05×10^{-6}	8.37×10^{-7}	1.98×10^{-6}	6.48×10^{-7}	1.12×10^{-6}	4.08×10^{-7}
SD	1.87×10^{-6}	3.89×10^{-7}	6.25×10^{-7}	2.94×10^{-7}	5.12×10^{-7}	4.65×10^{-7}
MSE	3.00×10^{-4}	2.97×10^{-4}	3.03×10^{-4}	2.96×10^{-4}	2.99×10^{-4}	2.94×10^{-4}
$0.05R_y$						
mean	5.20×10^{-5}	7.34×10^{-5}	9.36×10^{-5}	5.62×10^{-5}	1.04×10^{-4}	2.84×10^{-5}
SD	5.72×10^{-5}	7.42×10^{-5}	4.23×10^{-5}	5.85×10^{-5}	9.25×10^{-5}	3.85×10^{-5}
MSE	7.17×10^{-3}	7.28×10^{-3}	7.35×10^{-3}	7.19×10^{-3}	7.48×10^{-3}	7.11×10^{-3}
$0.1R_y$						
mean	2.43×10^{-4}	3.28×10^{-4}	3.81×10^{-4}	2.62×10^{-4}	4.59×10^{-4}	1.58×10^{-4}
SD	3.01×10^{-4}	3.50×10^{-4}	1.92×10^{-4}	2.83×10^{-4}	4.26×10^{-4}	2.08×10^{-4}
MSE	2.59×10^{-2}	2.54×10^{-2}	2.35×10^{-2}	2.57×10^{-2}	2.52×10^{-2}	2.64×10^{-2}
$0.15R_y$						
mean	6.40×10^{-4}	8.71×10^{-4}	9.51×10^{-4}	6.81×10^{-4}	1.17×10^{-3}	4.45×10^{-4}
SD	6.96×10^{-4}	7.12×10^{-4}	4.40×10^{-4}	6.51×10^{-4}	8.12×10^{-4}	5.88×10^{-4}
MSE	5.88×10^{-2}	5.71×10^{-2}	5.42×10^{-2}	5.80×10^{-2}	5.64×10^{-2}	6.13×10^{-2}
$0.2R_y$						
mean	1.27×10^{-3}	1.58×10^{-3}	1.90×10^{-3}	1.30×10^{-3}	2.06×10^{-3}	9.52×10^{-4}
SD	1.38×10^{-3}	1.32×10^{-3}	1.17×10^{-3}	1.24×10^{-3}	1.55×10^{-3}	1.20×10^{-3}
MSE	9.37×10^{-2}	9.04×10^{-2}	8.48×10^{-2}	9.24×10^{-2}	8.94×10^{-2}	9.81×10^{-2}

表5. $g(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$, 誤差: 混合正規分布.

	CV	GCV	ABIC	mAIC	AICc	SPIC
x : equal						
$n = 100$						
mean	9.52×10^{-5}	1.10×10^{-4}	1.68×10^{-4}	9.59×10^{-5}	1.34×10^{-4}	6.68×10^{-5}
SD	1.19×10^{-4}	1.25×10^{-4}	7.53×10^{-5}	1.12×10^{-4}	1.41×10^{-4}	9.35×10^{-5}
MSE	1.37×10^{-2}	1.38×10^{-2}	1.41×10^{-2}	1.38×10^{-2}	1.40×10^{-2}	1.36×10^{-2}
x : equal						
$n = 50$						
mean	1.83×10^{-4}	2.51×10^{-4}	3.26×10^{-4}	1.96×10^{-4}	3.85×10^{-5}	1.09×10^{-4}
SD	2.39×10^{-4}	2.68×10^{-4}	1.67×10^{-4}	2.27×10^{-4}	3.44×10^{-4}	1.64×10^{-4}
MSE	2.74×10^{-2}	2.74×10^{-2}	2.59×10^{-2}	2.75×10^{-2}	2.75×10^{-2}	2.81×10^{-2}
x : uniform						
$n = 100$						
mean	1.08×10^{-4}	1.43×10^{-4}	1.81×10^{-4}	1.30×10^{-4}	1.68×10^{-4}	8.86×10^{-5}
SD	1.39×10^{-4}	1.58×10^{-4}	9.00×10^{-5}	1.45×10^{-4}	1.77×10^{-4}	1.22×10^{-4}
MSE	1.33×10^{-2}	1.34×10^{-2}	1.30×10^{-2}	1.34×10^{-2}	1.34×10^{-2}	1.33×10^{-2}
x : uniform						
$n = 50$						
mean	2.40×10^{-4}	3.24×10^{-4}	3.70×10^{-4}	2.53×10^{-4}	4.45×10^{-4}	1.69×10^{-4}
SD	3.67×10^{-4}	3.15×10^{-4}	1.94×10^{-4}	2.77×10^{-4}	3.73×10^{-4}	2.65×10^{-4}
MSE	2.52×10^{-2}	2.51×10^{-2}	2.31×10^{-2}	2.53×10^{-2}	2.49×10^{-2}	2.61×10^{-2}

少ない基底関数や、逆にデータ数に近い極端に多くの基底関数をとらない限り、基底関数の個数を変えても実験の結果は本質的には変わらなかった。情報量規準 SPIC に基づく基底関数の個数選択に関しては、井元・小西 (1999) を参照されたい。また、差分行列 D_k に関しては、2 階差分行列を用いた。

この数値実験から得られた結果は、次のように要約することができる。

- ・データに付加した誤差項の分布が正規分布の場合、その分散の大小にかかわらず ABIC の選択した平滑化パラメータに対するモデルは、真の曲線との平均 2 乗誤差 (AMSE) を小さくするという点で他の規準よりも優れている場合が多かった。

- ・ABIC は、他の規準に比べ大きな平滑化パラメータの値を選択することが多かった。また、標準偏差 (SD) が他の規準に比べて小さく、安定した平滑化パラメータの選択を行うといえる。これは、情報量規準は平均対数尤度の漸近的な推定量として導出されているのに対して、この数値実験では ABIC は周辺尤度を解析的に陽に求めることができる場合を扱っていることにも起因する。

- ・AICc は、他の 2 つの情報量規準 (mAIC, SPIC) に比べ相対的に大きな平滑化パラメータの値を選択することが多かった。これは、この 3 つの規準は 3 章の (3.10) 式から (3.12) 式で示したように、第 1 項はモデルの対数尤度に関係した共通項であり、その差は第 2 項のバイアス補正の項に現れる。 λ の値を変化させると、その補正項はほぼ同じ形状の曲線を描くが、その値は $AICc > mAIC > SPIC$ となることが多かった。つまり、AICc は mAIC, SPIC に比べ大きめの補正を与えているといえる。

- ・GCV は、CV に比べデータの分散が大きいときにより結果を得ることが多かった。分散が小さいときにはそれほど差はない。また、GCV は CV においてハット行列の各対角成分をその平均で置き換えていることから、外れ値の影響を軽減することが予想される。しかし、混合正規分布に従う誤差を付加した場合、GCV は CV に比べて、それほど顕著な改善は見られなかった。

- ・CV は、最適な平滑化パラメータとして稀に極めて小さい値を取ることがある。これは実際の解析に適用する際、注意しなくてはならない点である。

- ・SPIC は、データ数が $n = 100$ のときや、データの分散がそれほど大きくないときには、平均 2 乗誤差は他の規準と比べて小さいことが多く、標準偏差も小さいという点で優れた結果が得られた。しかし、 $n = 50$ のとき、もしくはデータの分散が大きいときには、SPIC よりも大きな平滑化パラメータを選択することの多い ABIC や AICc の方が平均 2 乗誤差を小さくする結果が得られた。

- ・混合正規分布に従う誤差に関しては、データ数が 100 の場合は SPIC が平均 2 乗誤差を小さくすることが多く標準偏差も小さかった。しかし、データ数が 50 のときは、ABIC の平均 2 乗誤差がもっとも小さかった。混合正規分布に関しては、分散の大きさ、および混合比率に関して更に検討する必要がある。

謝 辞

査読者の方々からは、大変貴重なご指摘と適切なご助言をいただきました。また、編集委員には、多数の文献をご教示いただきました。ここに記して感謝いたします。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest. (Reproduced in *Breakthroughs in Statistics*, Vol. 1 (eds. S. Kotz and N. L. Johnson), Springer New York, (1992).)
- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, AC-19, 716-723.

- Akaike, H. (1980a). On the use of predictive likelihood of a Gaussian model, *Ann. Inst. Statist. Math.*, **32**, 311-324.
- Akaike, H. (1980b). Likelihood and the Bayes procedure, *Bayesian Statistic* (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), University Press, Valencia, Spain.
- 赤池弘次, 北川源四郎 編 (1994). 『時系列解析の実際 I』, 朝倉書店, 東京.
- 赤池弘次, 北川源四郎 編 (1995). 『時系列解析の実際 II』, 朝倉書店, 東京.
- Bozdogan, H. (ed.) (1994). *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Kluwer, Dordrecht.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numer. Math.*, **31**, 377-403.
- de Boor, C. (1978). *A Practical Guide to Splines*, Springer, Berlin.
- Eilers, P. and Marx, B. (1996). Flexible smoothing with *B*-splines and penalties (with discussion), *Statistical Science*, **11**, 89-121.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman & Hall, London.
- Hiragi, Y., Urakawa, H. and Tanabe, K. (1985). Statistical procedure for deconvoluting experimental data, *Journal of Applied Physics*, **58**(1), 5-11.
- Hurvich, C. M., Simonoff, J. S. and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *J. Roy. Statist. Soc. Ser. B*, **60**, 271-293.
- 井元清哉, 小西貞則 (1999). 情報量規準に基づく *B*-スプライン非線形回帰モデルの推定, 応用統計学 (印刷中).
- 石黒真木夫, 荒畑恵美子 (1982). ベイズ型スプライン回帰, 統計研彙報, **30**(1), 29-36.
- 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 20, 教育出版, 東京.
- Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *J. Amer. Statist. Assoc.*, **79**, 378-389.
- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Springer, New York.
- Konishi, S. (1999). Statistical model evaluation and information criteria, *Multivariate Analysis, Design of Experiments and Survey Sampling* (ed. S. Ghosh), 369-399, Marcel Dekker, New York.
- Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika*, **83**, 875-890.
- 小川直子, 南 美穂子, 渋谷政昭 (1995). 一次平滑化スプラインの平滑化行列の特性, 応用統計学, **24**(1), 27-41.
- Ramsay, J. O. (1988). Monotone regression splines in action (with discussion), *Statist. Sci.*, **3**, 425-461.
- 坂本 亘, 白旗慎吾 (1996). セミパラメトリック回帰問題におけるスプライン平滑化, 計算機統計学, **9**(1), 13-35.
- Sakamoto, W. and Shirahata, S. (1997). Simple calculation of likelihood-base cross-validation score in maximum penalized likelihood estimation of regression functions, *J. Japanese Soc. Comput. Statist.*, **10**, 27-40.
- Shibata, R. (1989). Statistical aspect of model selection, *From Data to Model* (ed. J. C. Willems), 215-240, Springer, Berlin/New York.
- Simonoff, J. S. (1996). *Smoothing Method in Statistics*, Springer, New York. (竹澤邦夫, 大森 宏 訳 (1999). 『平滑化とノンパラメトリック回帰への招待』, 農林統計協会, 東京.)
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Comm. Statist. Theory Methods*, **A7**, 13-26.
- 田辺國士 (1976). 不適切問題への統計的アプローチ, 数理科学, No. 153, 1-5.
- 田辺國士 (1989). 平滑化, 『統計学辞典』(竹内 啓 編), 375-380, 東洋経済新報社, 東京.
- 田辺國士, 田中輝雄 (1983). ベイズモデルによる曲線・曲面のあてはめ, 地球, **5**(3), 179-186.
- Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

Nonlinear Regression Models Using B -spline and Information Criteria

Seiya Imoto and Sadanori Konishi

(Graduate School of Mathematics, Kyushu University)

We investigate the problem of estimating B -spline nonlinear regression models. B -spline nonlinear regression models represented by probability density function are estimated using the maximum penalized likelihood method. The essential problem of model construction is in the choice of smoothing parameter and the number of basis functions, for which several procedures such as generalized cross-validation, AIC and Akaike's Bayesian information criterion have been proposed.

The purpose of the present paper is to consider this model estimation problem from an information theoretic point of view and to give a criterion as an estimator of Kullback-Leibler information. Numerical comparisons are made to compare the properties of various types of criteria.