

適応型モデル選択基準を用いた一般化加法モデルの推定

宮田 敏*

Estimation of Generalized Additive Models Based on Adaptive Model Selection Criterion

Satoshi Miyata*

一般化加法モデル (GAM, Generalized Additive Model) は, 一般化線型モデルの非線型モデルへの自然な拡張として, 近年盛んに用いられるようになっている。従来用いられた GAM では, 共変量 x_i を非線型変換する際, 局所重み付き多項式回帰や平滑化スプラインなど「滑らかな」変換が用いられてきた。しかし現実のデータにおいては微分不能な点や不連続点など「滑らかでない」関数関係も多く見受けられる。本研究では GAM の非線型変換に多重節点を持った回帰スプラインを導入し, 不均一な滑らかさをもつ関数関係を表現できるように GAM を拡張することを提案する。GAM の非線型予測子 $\eta = \beta_0 + \sum_{j=1}^p f_j(x_j)$ は, 適応型モデル選択基準 (AMSC, Adaptive Model Selection Criterion) を用いて最適化される。AMSC は Shen *et al.* (2004) により, 指数型分布族に対するモデル選択基準として導入され, Miyata and Shen (2003, 2005) により可変節点スプラインの節点最適化に応用された。一般にモデル選択基準の最適化は複雑な非線型問題となるが, 本研究では遺伝的アルゴリズムを用いた大域的な最適化を提案する。本発表で提案された方法は, シミュレーションによる数値実験により検証され, 実際のデータに対して応用される。

Recently, the generalized additive models (GAM) have been frequently used as natural enhancing to non-linear model of the generalized linear models. Conventional GAM models adopted “smooth” functions, such as loess or smoothing spline, to apply non-linear transformation to covariates. In real datasets, however, inhomogeneous smooth structures, such as discontinuity or indifferentiability, are not rare. In this paper, we propose to adopt adaptive free-knot splines (AFKS) with multiple knots to extend GAM so that the model is capable of capturing complex structure of the data. The non-linear predictor of the GAM, $\eta = \beta_0 + \sum_{j=1}^p f_j(x_j)$, will be optimized by the adaptive model selection criterion (AMSC). AMSC has been introduced as a model selection criterion for the exponential family distributions by Shen *et al.* (2004), which was applied to optimize the knot sequence of AFKS by Miyata and Shen (2003, 2005). In general, the optimizations of the model selection criteria are complex problem. In this research, we utilize the evolutionary algorithm to achieve the global optimum. The proposed methodology will be examined by the simulation study and applied to the real data set.

キーワード: 一般化加法モデル, 可変節点スプライン, 適応型モデル選択基準, 遺伝的アルゴリズム, 不均一な滑らかさ

* (財) 癌研究会ゲノムセンター情報解析グループ, 〒 135-8550 東京都江東区有明 3-10-6 (E-mail: satoshi.miyata@jfc.or.jp).

1. はじめに

近年の計算機技術の発達に伴い、非線型な構造を持つ大規模データの集積と解析が可能となり、広範な統計解析ツールの中から適切なモデルを選択し、データに適合した柔軟な解析を行う必要性が増している。一般化加法モデル (Generalized Additive Models, GAM) は一般化線型モデルの非線型モデルへの自然な拡張として、各種ブースティング、カーネルマシン、サポートベクターマシンと並び、有効な学習機械として広く利用されている (GAM の一般的な解説については、Hastie and Tibshirani (1990), Wood (2006), 辻谷-外山 (2007) 等を参照)。本稿では、従来用いられてきた GAM を拡張し、不連続点や微分不能な点など「滑らかではない」非線型構造も含むデータに対して適合的なモデルを提案する。

いま応答変数 y_i と p -次元共変量 $x_i = (x_{i1}, \dots, x_{ip})'$ の n 組のデータ $\{(y_i, x_i)\}_{i=1}^n$ が観察されたとする。ただし、応答変数 y_i は以下の確率分布

$$p(y_i | \mu_i) = \exp(\phi(\mu_i)y_i + \alpha(\mu_i) + m(y_i)) = \exp(\theta_i y_i + \alpha(\mu_i) + m(y_i)) \quad (1.1)$$

を持つ指数型分布族に従うとする。ただし、 $\mu_i = E(Y_i)$, $\sigma_i^2(\mu_i) = \text{var}(Y_i) = 1/\phi'(\mu_i)$ であり、 ϕ, α, m は散布度パラメタ (dispersion parameter) ψ に依存する場合もある。

一般化加法モデルにおいて、 $\mu_i = E(Y_i | x_i)$ は非線型関数 f_1, \dots, f_p を用いて共変量 x_i

$$g(\mu_i) = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) \quad (1.2)$$

によって連結される (Hastie and Tibshirani (1990))。ただし $g(\cdot)$ は連結関数 (link function) であり、二階連続微分可能な狭義の単調関数である。従来用いられた GAM では、共変量 x_i を非線型変換する際、局所重み付き多項式回帰 (locally weighted polynomial regression, loess) や平滑化スプライン (smoothing spline) など「滑らかな」変換が用いられてきた。しかし現実のデータにおいては、データの「滑らかさ」が局所的に変わる場合も多く、そのような場合 loess や平滑化スプラインにより共変量の平滑な関数を求めることは、後述するように深刻な問題を引き起こす場合がある。本研究では非線型変換 $f_j, j = 1, \dots, p$ に多重節点を持った可変節点スプライン (Adaptive free-knot spline, AFKS) を導入し、不均一な滑らかさを持つ関数関係を表現できるように GAM を拡張する。

GAM のモデル推定においては、非線型関数によるデータの平滑化の程度を調節する平滑化パラメタの選択が重要な問題となる。データに対し最適なパラメタを選択するためには、適切なモデル選択基準とその最適化法が必要である。例えば、従来用いられた平滑化スプラインによる平滑化では、ペナルティ付き残差平方和 (Penalized Residual Sum of Squares, PRSS) の最小化が行われ、平滑化パラメタ λ は Cross Validation (CV), Generalized Cross Validation (GCV) などの最小化によって最適化される。しかし非線型回帰モデルの推定

において、CV または GCV はしばしば不安定であり (Efron (1983)), 過小推定 (under estimate) をもたらす傾向があることが知られている (たとえば Simonoff (1996) を参照). 本稿では、適応型モデル選択基準 (AMSC, Adaptive Model Selection Criterion) を用いたパラメタの最適化を提案する. AMSC は Shen *et al.* (2004) により、指数型分布族に対するモデル選択基準として導入され、Miyata and Shen (2003, 2005) により AFKS の節点最適化に応用された (Ye (1998), Shen and Ye (2002) も参照).

一般に、モデル選択基準をパラメタについて最適化する問題は複雑な非線形問題となる. 最適化問題の目的関数となるモデル選択基準は一般に多くの極小値を持ち、パラメタについて不連続もしくは微分不能である. 本稿では確率的最適化法の一つである遺伝的アルゴリズム (Evolutionary Algorithm, EA) を用い、パラメタの最適化を行う. EA による数値最適化は非常に緩やかな仮定のもとで大域的収束が保証されており、解の探索は頑健である.

本稿の構成は以下の通りである. 第2節では、一般化加法モデルを導入する. あわせて従来用いられてきたモデルの問題点を指摘し、AFKS を用いた GAM の拡張を提案する. 第3節では、Shen *et al.* (2004) にしたがって AMSC を導入し、GAM に対する AMSC を定式化する. さらに、EA による確率的最適化法を用いた AMSC の最適化を行う. 第4節では、数値実験を通じ提案された方法の有効性を検証する.

2. 一般化加法モデル

応答変数 y_i と p -次元共変量 $x_i = (x_{i1}, \dots, x_{ip})'$ の n 組のデータ $\{(y_i, x_i)\}_{i=1}^n$ が与えられた時、伝統的な線形回帰モデルは以下のように与えられる.

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i, \epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), i = 1, \dots, n. \quad (2.1)$$

本節では、以下 (2.1) を拡張して加法モデル、一般化線形モデルを導入し、最終的に一般化加法モデルを定式化する.

2.1 加法モデル

加法 (回帰) モデル (additive (regression) model) においては、(2.1) の線形部分 $\beta_j x_{ij}$ が以下のように非線形関数で置き換えられる.

$$y_i = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) + \epsilon_i, \epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), i = 1, \dots, n. \quad (2.2)$$

従来用いられた加法モデルでは、非線形関数 $f_j, j = 1, \dots, p$ は滑らかな (例えば、可測で二階微分可能な) 関数とされ、散布図平滑化法 (scatterplot smoother) を用いて以下の後退当てはめ法 (backfitting algorithm) により推定される (Hastie and Tibshirani (1990) 参照).

Algorithm 1: 後退当てはめ法

1. 初期化: $\beta_0 = n^{-1} \sum_{i=1}^n y_i$, $f_j = f_j^0 = 0$, $j = 1, \dots, p$.
2. $j = 1, \dots, p$ に対して, 残差

$$y - \sum_{k \neq j} f_k(x_k)$$

を求め, 散布図平滑化法の一つを用いて x_j に対して平滑化を行う.

3. 個々の f_j が収束するまで, 2 を繰り返す.

後退当てはめ法の 2 で用いられる散布図平滑化法には様々なものが考えられるが, 例えば, 平滑化スプライン, 自然スプライン, loess などが多いようである. 平滑化スプラインの場合は, 以下のペナルティ付き残差平方和 PRSS (Penalized Residual Sum of Squares)

$$PRSS_j = \sum_{i=1}^n (y_i - f_j(x_{ij}))^2 + \lambda \int (f_j''(x))^2 dx \quad (2.3)$$

を最小にする f_j として与えられ, これは x_{ij} を節点とする自然 3 次スプラインとなる (Wood (2006) 参照). ただし $\lambda > 0$ は与えられた定数で, f_j の曲げ弾性に対するペナルティである.

2.2 一般化線形モデル

一般化線形モデル (Generalized Linear Model, GLM) は, 線形回帰モデル (2.1) のもう一つの拡張として導入される. GLM を定式化するため, (2.1) を以下の 3 つの構成要素に分離して考える (McCullagh and Nelder (1989)).

LM1: ランダム成分 $\mathbf{Y} = (Y_1, \dots, Y_n)'$ の各要素は, 独立に期待値 $E(Y_i) = \mu_i$, $i = 1, \dots, n$, 分散 $\sigma^2 > 0$ の正規分布に従う.

LM2: 系統的成分 共変量 x_1, \dots, x_p , $x_j = (x_{1j}, \dots, x_{nj})'$ は, 以下の線形予測子 $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)'$ を与える.

$$\boldsymbol{\eta} = \beta_0 \mathbf{1}_n + \sum_{j=1}^p x_j \beta_j.$$

ただし, $\mathbf{1}_n = (1, \dots, 1)'$.

LM3: 連結関数 $E(\mathbf{Y}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ と線形予測子 $\boldsymbol{\eta}$ は, 恒等変換により連結される.

$$\boldsymbol{\mu} = \boldsymbol{\eta}.$$

このとき GLM は, LM1, LM3 を拡張して, 以下の 3 つの構成要素により定式化される.

GLM1: ランダム成分 $\mathbf{Y} = (Y_1, \dots, Y_n)'$ の各要素は, 独立に (1.1) の指数型分布族に従う.

GLM2: 系統的成分 共変量 x_1, \dots, x_p , $x_j = (x_{1j}, \dots, x_{nj})'$ は, 以下の線形予測子 $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)'$ を与える.

$$\boldsymbol{\eta} = \beta_0 \mathbf{1}_n + \sum_{j=1}^p x_j \beta_j.$$

GLM3: リンク関数

$$\mathbf{g}(\boldsymbol{\mu}) = \boldsymbol{\eta} \iff g(\mu_i) = \eta_i, i = 1, \dots, n.$$

ただし $g(\cdot)$ は連結関数 (link function) であり, 二階連続微分可能な狭義の単調関数.

簡単化のため, 上記構成要素 GLM3 において自然リンク関数 (canonical link function) を考えると, (1.1) において $\theta_i = \eta_i = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j$ であり, $\mathbf{Y} = (Y_1, \dots, Y_n)'$ の対数尤度は以下で与えられる.

$$l = \sum_{i=1}^n \log p(y_i | \mu_i) = \left\{ \sum_{i=1}^n \sum_{j=0}^p y_i x_{ij} \beta_j + \sum_{i=1}^n \alpha \left(g^{-1} \left(\sum_{j=0}^p x_{ij} \beta_j \right) \right) + \sum_{i=1}^n m(y_i) \right\}. \quad (2.4)$$

ただし, $x_{i0} = 1, i = 1, \dots, n$.

一般化線形モデルのパラメタ $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ は最尤法により推定される. $\boldsymbol{\beta}$ の最尤推定量は, 以下のスコア方程式の解として与えられる.

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta} | \mathbf{y}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^n \log p(y_i | \mu_i) = \mathbf{0} \\ \iff \sum_{i=1}^n x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) V_i^{-1} (y_i - \mu_i) &= 0, j = 0, 1, \dots, p. \end{aligned}$$

ただし, $V_i = \text{var}(Y_i)$. スコア方程式は一般に $\boldsymbol{\beta}$ の非線形方程式であり, $\boldsymbol{\beta}$ の最尤解は以下に述べる繰り返し再重み付け最小二乗法 (IRLS, iteratively reweighted least squares) により近似される. IRLS 法は, スコア方程式の解をヘッシアンの変わりにフィッシャー情報量を用いた Newton-Raphson 法によって求めることに相当している.

IRLS 法

1. 修正応答変数 (adjusted response variable)

$$z_i = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}, \quad \eta_i = g(\mu_i) = \sum_{j=0}^p x_{ij} \beta_j, \quad i = 1, \dots, n$$

を求める.

2. 重み付け最小二乗法のための重み W_1, \dots, W_n を, 以下のように求める.

$$W_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 V_i^{-1}.$$

3. 修正応答変数 z を共変量 $x = (x_1, \dots, x_p)'$ に W の重み付きで回帰する.
回帰パラメタ β を更新する. 更新される β の相対変化が小さくなり, β が収束するまでこのプロセスを繰り返す.

2.3 一般化加法モデル

一般化加法モデル (Generalized Additive Model, GAM) は, 一般化線型モデルの構成要素 GLM2: 線型予測子 $\eta = \sum_{j=0}^p x_{ij} \beta_j$ を加法的な予測子で置き換えたものである. すなわち GAM においては GLM の構成要素, GLM1, GLM3 はそのまま踏襲した上で, GLM2 を以下の要素 GAM2 で置き換える.

GAM2: 系統的成分 共変量 x_1, \dots, x_p , $x_j = (x_{1j}, \dots, x_{nj})'$ は, 以下の加法的予測子を与える.

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^p f_j(x_{ij}). \quad (2.5)$$

GLM と加法モデルを合わせて GAM を定式化することにより, 正規分布に限らず指数型分布族に属する多様な確率モデルに, データから直接非線型構造を導入することができる. これにより共変量の非線型性をデータ適合的に抽出し視覚化できる点で, GAM はきわめて有用な統計モデルとなる.

一般化加法モデルは, 以下の局所スコア法 (local scoring procedure) により推定される (Hastie and Tibshirani (1990) 参照).

Algorithm 2: 局所スコア法

1. 初期化: $\beta_0 = g(n^{-1} \sum_{i=1}^n y_i)$, $f_j = f_j^0 = 0$, $j = 1, \dots, p$.
2. f_1, \dots, f_p を用いた IRLS: $l = 0, 1, \dots$ に対して,

(a) 修正応答変数 (adjusted response variable)

$$z_i^l = \eta_i^l + (y_i - \mu_i^l) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)_l, \quad \eta_i^l = g(\mu_i^l) = \beta_0 + \sum_{j=1}^p f_j^l(x_{ij}),$$

$$\mu_i^l = g^{-1}(\eta_i^l), i = 1, \dots, n$$

を求める.

(b) 重み付け最小二乗法のための重み W_1, \dots, W_n を, 以下のように求める.

$$W_i^l = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)_l^2 (V_i^l)^{-1}.$$

(c) 後退当てはめ:

i. 初期化: $f_j = f_j^l = 0, j = 1, \dots, p$.

ii. $j = 1, \dots, p$ に対して, 残差

$$z^l - \sum_{k \neq j} f_k^l(x_k)$$

を求め, 散布図平滑化法の一つを用いて x_j に対して W_i による重み付け平滑化 f_j^{l+1} を求める.

iii. 個々の f_j^l が収束するまで, (ii) を繰り返す.

3. $\Delta(\eta^{l+1}, \eta^l)$ が収束するまで, 2 を繰り返す. ただし

$$\Delta(\eta^{l+1}, \eta^l) = \frac{\sum_{j=1}^p \|f_j^{l+1} - f_j^l\|}{\sum_{j=1}^p \|f_j^l\|}.$$

なお, 局所スコア法 2(iii) において, 散布図平滑化法に 3 次の平滑化スプラインを用いた場合, 局所スコア法は以下のペナルティ付き対数尤度最大化と等価である (Hastie and Tibshirani (1987)).

$$\sum_{i=1}^n \log p(y_i | \mu_i) + \sum_{j=1}^p \lambda_j \int \{f_j''(t)\}^2 dt. \quad (2.6)$$

2.4 一般化加法モデルの問題点とその拡張

前項で定式化した一般化加法モデルは, データから直接その非線型な構造を抽出し多様な確率分布に従う統計モデルを推定できる点で, きわめて有用である. しかし従来用いられた GAM においては, 加法的予測子 $\eta = \beta_0 + \sum_{j=1}^p f_j(x_j)$ に用いられる非線型関数 $f_j, j = 1, \dots, p$ は滑らかな関数であると仮定され, 個別に loess や平滑化スプラインなどの散布図平滑化法によって推定されると想定されてきた.

したがって, 真のモデルにおいて, f_j に不連続点や微分不能な点が存在する場合, あるいは f_j の滑らかさが局所的に変化する場合, 従来用いられた GAM による推定は, 十分機能しない可能性がある. 以下の例を考える.

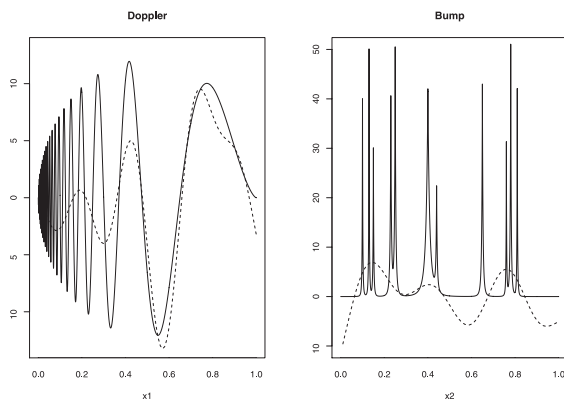


図1 真の回帰関数（左：doppler, 右：bump, 実線）と推定された曲線（点線）

2.4.1 数値例

この数値例は, $y = f + \epsilon$, $f = f_{\text{doppler}} + f_{\text{bump}}$ の加法モデルからなる. f_{doppler} , f_{bump} のテスト関数は, Donoho and Johnstone (1995) からとった.

$$y = f(x_1, x_2) + \epsilon, \epsilon \stackrel{\text{iid}}{\sim} N(0, 1), x_1, x_2 = 0.01, 0.02, \dots, 1.00.$$

$$f(x_1, x_2) = f_{\text{doppler}}(x_1) + f_{\text{bump}}(x_2),$$

$$f_{\text{doppler}}(x) = \sqrt{x(1-x)} \sin\left(2\pi \frac{1+0.05}{x+0.05}\right),$$

$$f_{\text{bump}}(x) = \sum_{j=1}^{11} h_j K\left(\frac{x-t_j}{w_j}\right), K(t) = (1 + \text{sgn}(t))/2,$$

$$(t_j) = (.1, .13, .15, .23, .25, .40, .44, .65, .76, .78, .81),$$

$$(h_j) = (4, 5, 3, 4, 5, 4.2, 2.1, 4.3, 3.1, 5.1, 4.2),$$

$$(w_j) = (.005, .005, .006, .01, .01, .03, .01, .01, .005, .008, .005).$$

上記の数値例から得られたサンプルデータに対して, 従来用いられた GAM の当てはめを行った. 解析には統計解析パッケージ R.2.9.2 (R Development Core Team (2008)) の mgcv パッケージ (Wood (2006)) を用いた. 平滑化には 3 次平滑化スプラインを用い, 平滑化パラメタは GCV の最小化により決定した.

図1 左には doppler, 右には bump を図示している. 真の関数 f_{doppler} , f_{bump} は実線で, 推定された平滑化曲線は点線でそれぞれ示している. 図左の doppler 関数においては, $x_1 \in (0.5, 1.0)$ 付近では, 推定された平滑化曲線は比較的真的関数の近くにあるが, $(0.0, 0.5)$ の範囲では推定された曲線は真的関数に追従できていない. Doppler 関数は定義域の全域で微分可能な滑らかな関数であるが, 局所的に振幅の程度が異なる. 一方 doppler 関数の推定に用いられた平滑化スプラインは, ペナルティ付き残差平方和 PRSS (2.3) の最小化に

よって得られる。すなわち平滑化スプラインの平滑化の程度は、定義域の全域で平滑化ペナルティ λ により一律に制御される。このことが、推定された平滑化曲線が真の doppler 関数の局所的な変動を捕らえられなかった理由である。一方図 1 右の bump 関数においては、推定された平滑化曲線は真の bump 関数の推定に明らかに失敗している。これは、真の bump 関数が微分不能な鋭いピークを有しているのに対して、推定に用いた平滑化スプラインが滑らかな曲線しか表現できないことに理由がある。

2.4.2 一般化加法モデルの拡張

前節で見たとおり、従来用いられてきた一般化加法モデルは、推定対象の加法モデルに不連続点や微分不能な点など不均一な滑らかさが存在する場合、有効に機能しない。この点を改善するため、本研究では非線型変換 $f_j, j = 1, \dots, p$ に多重節点を持った可変節点スプライン (Adaptive free-knot spline, AFKS) を導入し、不均一な滑らかさを持つ関数関係を表現できるように GAM を拡張することを提案する。以下では簡単化のため、 $x_{ij} \in (0, 1), i = 1, \dots, n, j = 1, \dots, p$ とする。

一般に、節点配列 (knot sequence) $t = (t_1, \dots, t_k), 0 = t_0 < t_1 \leq \dots \leq t_k < t_{k+1} = 1$ を持つ $m(m \geq 1)$ 階の可変節点スプライン $s(x)$ (ただし、節点位置 t_l において $k_l(k_l \leq m)$ の多重節点を持つとする) は、各区間 $[t_l, t_{l+1}]$ 上での m 階区分的多項式と定義される。スプライン関数は節点位置 t_l 上において $p_l = ((m-1) - k_l)$ 次連続導関数を持つ。スプライン関数の節点の総数 k は $k = \sum_{l=1}^{k_0} k_l < (n-m)$ を満たす。ただし k_0 は異なる節点の数。通常用いられるスプラインの階数は $m = 4$ であり、これはいわゆる 3 次スプライン関数に相当する。

スプライン関数は、以下の切断冪基底関数 (truncated power basis function) によって基底展開される。

$$\{x^j, j = 0, \dots, m-1; (x-t_l)_+^{(m-j)}, j = 1, \dots, k_l, l = 1, \dots, k\}$$

ただし $(\cdot)_+ = \max\{\cdot, 0\}$ 。切断冪基底関数を用いた表現は単純で直感的にも理解しやすいが、数値計算の際はしばしば不安定であり、悪条件になりやすい。実際の計算には、以下の正規化 B -スプライン基底 (normalized B -spline basis) $\{B_l(x; t), l = 1, \dots, k+m\}$ を用いて、スプライン関数を表現する。

$$B_l(x; t) = (t_l - t_{l-m})[t_{l-m}, \dots, t_l](\cdot - x)_+^{m-1}, l = 1, \dots, k+m.$$

ただし、 $[t_{l-m}, \dots, t_l]g$ は関数 g の m 階の差分商 (divided difference), $t_l = t_{\min(\max(l,0), k+1)}$; $l = 1-m, \dots, k+m$ であり、 $t_{1-m} = \dots = t_0 = 0, t_{k+1} = \dots = t_{k+m} = 1$ は外部節点 (external knots) である。 B -スプライン基底は帰納的に定義され、高速かつ安定的な計算が可能である (de Boor (2001) 参照)。

回帰モデル

$$Y_i = \vartheta(x_i) + \epsilon_i, \quad i = 1, \dots, n,$$

において、スプライン関数 $s(x) = \sum_{l=1}^{k+m} \alpha_l B_l(x, t)$ によって回帰関数 ϑ を推定する場合、回帰スプライン推定量 $\hat{\vartheta}_n(x, h) = \sum_{l=1}^{k+m} \hat{\alpha}_l B_l(x, t)$ は (重み付き) 最小二乗法で推定される。

$$\min_{\alpha_1, \dots, \alpha_{k+m}} \sum_{i=1}^n \left(Y_i - \sum_{l=1}^{k+m} \alpha_l B_l(x_i, t) \right)^2. \quad (2.7)$$

ただし、 $\hat{\alpha}_l$ は α_l の (重み付き) 最小二乗推定量である。スプライン関数 $s(x) = \sum_{l=1}^{k+m} \alpha_l B_l(x, t)$ のチューニングパラメタは $h = (t, k)$ であり、 h が与えられれば (3.2) を通じて $s(x)$ の形状は一意的に決定される。

本稿における GAM の拡張においては、GAM 推定のための局所スコア法 (Algorithm 2) 2(iii) の後退当てはめ (backfitting algorithm) において、上記の多重節点を持った可変節点スプライン (AFKS) を適用することで、データの非線型構造を抽出することを試みる。スプライン関数のチューニングパラメタ h の最適化には、次節で導入する適応型モデル選択基準を用いる。

3. 適応型モデル選択基準

3.1 指数型分布族に対する適応型モデル選択基準

応答変数 y_i と p -次元共変量 x_i が与えられたとき、一般化加法モデルのパフォーマンスは、局所スコア法 (Algorithm 2) 2(iii) の後退当てはめに散布図平滑加法を適用する際のチューニングパラメタによって決定される。例えば、平滑化に平滑化スプラインを用いた場合、共変量 x_i の平滑化はペナルティ付き対数尤度 (2.6) の最大化によって行われるが、そのためには平滑化パラメタ $\lambda_1, \dots, \lambda_p$ の適切な選択が必要である。

従来用いられた GAM で頻繁に用いられるモデル選択基準には CV (Cross Validation), GCV (Generalized Cross Validation), AIC などがあり、それぞれ以下のように与えられる。

$$\begin{aligned} \text{CV} &= \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i^{-i}), \\ \text{GCV} &= \frac{\frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i)}{\{1 - \text{tr}(R)/n\}^2}, \\ \text{AIC} &= \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i) + 2\text{tr}(R)\psi/n. \end{aligned}$$

ただし、 $D(y; \mu) = \sum_{i=1}^n D(y_i, \mu_i)$: 尤離度 (deviance), $\hat{\mu}_i^{-i}$: i 番目のサンプルを除いたデータで推定した第 i 応答変数の推定値, R : 局所スコア法の後退当てはめにおける重み付き加法モデルの射影行列, である。これら CV, GCV, AIC 等をチューニングパラメタに関して

最小化することで、自動的なモデル選択が行われる。しかし、GAMに限らず非線型モデルの選択において CV, GCV はしばしば不安定であり (Efron (1983), Hastie and Tibshirani (1990, 3.4.6 節)), 加法モデルにおけるモデル選択の問題はいまだ十分な検討が行われていない (Hastie and Tibshirani (1990, 6.9 節)).

本稿では、一般化加法モデルのモデル選択に対して適応型モデル選択基準 (AMSC, Adaptive Model Selection Criterion) を用いたパラメタの最適化を提案する。AMSC は Shen *et al.* (2004) により、指数型分布族に対するモデル選択基準として導入され、Miyata and Shen (2003, 2005) により可変節点スプラインの節点最適化に応用された。AMSC は Shen *et al.* (2004) に従い、以下のように定式化される。

$i = 1, \dots, n$ に対して、 $\mu_i = E(Y_i)$ に対する個別の推定量 $\hat{\mu}_i$ のパフォーマンスは、 μ_i の $\hat{\mu}_i$ に対する個別の Kullback-Leibler 損失

$$\int p(y|\mu_i) \log \frac{p(y|\mu_i)}{p(y|\hat{\mu}_i)} dy = (\phi(\mu_i)\mu_i + \alpha(\mu_i)) - (\phi(\hat{\mu}_i)\mu_i + \alpha(\hat{\mu}_i))$$

によって評価される。右辺第一項は推定量 $\hat{\mu}_i$ に依存しないので無視すると、個別の Kullback-Leibler 損失は $-\phi(\hat{\mu}_i)\mu_i - \alpha(\hat{\mu}_i) - m(y_i)$ によって定義される。これを $i = 1, \dots, n$ について合わせることで、比較 Kullback-Leibler 損失、 $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ を以下のように定義する。

$$\begin{aligned} K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) &= -\frac{1}{n} \sum_{i=1}^n (\phi(\hat{\mu}_i)\mu_i + \alpha(\hat{\mu}_i) + m(y_i)) \\ &= n^{-1} \sum_{i=1}^n ((-\log p(y_i|\hat{\mu}_i)) + \phi(\hat{\mu}_i)(y_i - \mu_i)). \end{aligned} \quad (3.1)$$

$K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ を $\hat{\mu}_i$ に関して最小化することで最適な推定量が選択されるが、 $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ は未知の $\boldsymbol{\mu}$ を含んでいるため、これを推定する必要がある。 $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ の推定量として、以下のクラスの統計量を考える。

$$-\sum_{i=1}^n \log p(y_i|\hat{\mu}_i) + \kappa. \quad (3.2)$$

(3.2) には、AIC, BIC などが含まれる。AMSC においては、Kullback-Leibler 損失 $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ の最良の推定のため、以下の基準により $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ と (3.2) のクラスの推定量の L_2 -距離 (=MSE (mean squared error)) を最小にする κ を選択する。

$$\begin{aligned} \min_{\kappa} E \left[K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) - n^{-1} \left(-\sum_{i=1}^n \log p(y_i|\hat{\mu}_i) + \kappa \right) \right]^2 \\ = \min_{\kappa} n^{-2} E \left(\sum_{i=1}^n \phi(\hat{\mu}_i)(y_i - \mu_i) - \kappa \right)^2. \end{aligned}$$

簡単な計算から、最適なペナルティ κ は、以下のように与えられる。

$$D(M) = \sum_{i=1}^n h_i(\boldsymbol{\mu}) = \sum_{i=1}^n E \phi(\hat{\mu}_i)(y_i - \mu_i) = \sum_{i=1}^n \text{Cov}(\phi(\hat{\mu}_i), y_i) \quad (3.3)$$

$E\phi(\hat{\mu}_i)(y_i - \mu_i)$ については、以下の別表現が知られている。

定理 3.1 Shen *et al.* (2004, 定理 1)

任意の μ に対して,

$$\begin{aligned} h_i(\mu) &= E\phi(\hat{\mu}_i)(y_i - \mu_i) = \frac{1}{\phi'(\mu_i)} \frac{d}{d\mu_i} E\phi(\hat{\mu}_i) \\ &= \sigma_i^2(\mu_i) \frac{d}{d\mu_i} E\phi(\hat{\mu}_i), \quad i = 1, \dots, n. \end{aligned}$$

$D(M) = \sum_{i=1}^n h_i(\mu)$ を (3.2) に代入して、最終的に適応型モデル選択基準 (Adaptive model selection criterion, AMSC) は以下のように定義される。

$$\begin{aligned} \text{AMSC} &= - \sum_{i=1}^n \log p(y_i | \hat{\mu}_i) + \sum_{i=1}^n E\phi(\hat{\mu}_i)(y_i - \mu_i) \\ &= - \sum_{i=1}^n \log p(y_i | \hat{\mu}_i) + \sum_{i=1}^n \text{Cov}(\phi(\hat{\mu}_i), y_i) \\ &= - \sum_{i=1}^n \log p(y_i | \hat{\mu}_i) + \sum_{i=1}^n \sigma_i^2(\mu_i) \frac{d}{d\mu_i} E\phi(\hat{\mu}_i). \end{aligned} \quad (3.4)$$

AMSC の第二項 $D(M)$ は、依然として未知の積分 $E\phi(\hat{\mu}_i)(y_i - \mu_i)$ を含んでいる。Shen *et al.* (2004) に従い、 $D(M)$ は以下の Monte Carlo 近似により推定される。(Shen and Ye (2002) も参照)

Algorithm 3: Monte Carlo simulation for AMSC

1. 以下の通り擬似応答変数 Y_i^* , $i = 1, \dots, n$ を生成する。

$$\tilde{Y}_i \stackrel{\text{iid}}{\sim} \begin{cases} p(y|\mu_i)|_{\mu_i=Y_i} & : \text{二項分布以外の指数型分布族に属する確率分布の場合} \\ p(y|p_i)|_{p_i=\hat{p}_i} & : \text{二項分布の場合} \end{cases}$$

$$Y_i^* = Y_i + \tau(\tilde{Y}_i - Y_i), \quad 0 \leq \tau \leq 1.$$

2. $\eta_i^{(j)} \stackrel{\text{iid}}{\sim} f(Y_i^* | Y_i)$, $i = 1, \dots, n$, $j = 1, \dots, T$.
3. $\hat{\mu}(\eta_i^{(j)})$, $i = 1, \dots, n$, $j = 1, \dots, T$ を求め,

$$\hat{b}_i = \frac{1}{T-1} \sum_{j=1}^T (\phi(\hat{\mu}(\eta_i^{(j)})) - \bar{\phi}_i)(\eta_i^{(j)} - \bar{\eta}_i)$$

とする。ただし $\bar{\eta}_i = T^{-1} \sum_{j=1}^T \eta_i^{(j)}$, $\bar{\phi}_i = \sum_{j=1}^T \phi(\hat{\mu}(\eta_i^{(j)}))$.

4. $D(M)$ を以下で近似する。

$$\hat{D}(M) = \frac{1}{\tau^2} \sum_{i=1}^n \hat{b}_i.$$

通常 $T \geq n$, $\tau \approx 0.5$ とする。

3.2 一般化線型モデル・一般化加法モデルに対する適応型モデル選択基準

本節では、前節 (3.4) で定義した適応型モデル選択基準を、一般化線型モデル、一般化加法モデルに対して適用する。まず、一般化線型モデル (GLM) の対数尤度は (2.4) で与えられるから、GLM の比較 Kullback-Leibler 損失、 $K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$ は以下のように表現される。

$$\begin{aligned} K(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) &= -n^{-1} \sum_{i=1}^n (\phi(\hat{\mu}_i) \mu_i + \alpha(\hat{\mu}_i) + m(y_i)) \\ &= -n^{-1} \sum_{i=1}^n \left(\sum_{j=0}^p \mu_i x_{ij} \hat{\beta}_j + \alpha \left(g^{-1} \left(\sum_{j=0}^p x_{ij} \hat{\beta}_j \right) \right) + m(y_i) \right) \\ &= n^{-1} \left(- \sum_{i=1}^n \log p(y_i | \hat{\boldsymbol{\beta}}) + \sum_{i=1}^n (y_i - \mu_i) \sum_{j=0}^p x_{ij} \hat{\beta}_j \right). \end{aligned}$$

したがって、一般化線型モデルに対する AMSC のペナルティ (3.3) は以下のように与えられる。

$$D(M) = \sum_{i=1}^n h_i(\boldsymbol{\mu}) = \sum_{i=1}^n E(y_i - \mu_i) \sum_{j=1}^p x_{ij} \hat{\beta}_j = \sum_{i=1}^n \text{Cov}(g^{-1}(\hat{\mu}_i), y_i). \quad (3.5)$$

特に、散布度パラメタ (dispersion parameter) が $i = 1, \dots, n$ に依存しない場合 (ロジスティック回帰, ポワソン回帰), GLM に対する AMSC のペナルティは以下の簡便な表現を得る。

定理 3.2 応答変数 y_i がベルヌーイ分布もしくはポワソン分布を持つとする。このとき、一般化線型モデルに対する適応型モデル選択基準のペナルティは、以下の形で与えられる。

$$D(M) = \sum_{i=1}^n E(y_i - \mu_i) \sum_{j=1}^p x_{ij} \hat{\beta}_j = \sum_{j=1}^p \frac{\partial}{\partial \beta_j} E(\hat{\beta}_j). \quad (3.6)$$

証明 3.2 y_i がベルヌーイ分布に従う場合について証明する。(ポワソン分布の場合も同様である。) ロジスティック回帰の場合 y_i は二項分布 $B(1, p_i)$, $0 < p_i < 1$ に従う。ただし $\mu_i = E(Y_i) = p_i$, $\sigma_i^2 = \text{var}(Y_i) = p_i(1 - p_i)$, $\phi(\mu_i) = \eta_i = \log(p_i/(1 - p_i))$,

$\alpha(\mu_i) = \log(1 - p_i) = -\log(1 + e^{\eta_i})$, $m(y_i) = 1$. このとき,

$$\begin{aligned} & \frac{\partial}{\partial \beta_j} E(\hat{\beta}_j) \\ &= \sum_{y_1} \cdots \sum_{y_n} \hat{\beta}_j \frac{\partial}{\partial \beta_j} \prod_{i=1}^n p(y_i | \mu_i) \\ &= \sum_{y_1} \cdots \sum_{y_n} \hat{\beta}_j \frac{\partial}{\partial \beta_j} \exp \left\{ \sum_{i=1}^n y_i \sum_{j=1}^p x_{ij} \beta_j + \sum_{i=1}^n \alpha \left(g^{-1} \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right) + \sum_{i=1}^n m(y_i) \right\} \\ &= \sum_{y_1} \cdots \sum_{y_n} \hat{\beta}_j \left\{ \sum_{i=1}^n y_i x_{ij} + \sum_{i=1}^n \frac{\partial}{\partial \beta_j} \alpha \left(g^{-1} \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right) \right\} \prod_{i=1}^n p(y_i | \mu_i) \\ &= \sum_{y_1} \cdots \sum_{y_n} \hat{\beta}_j \sum_{i=1}^n (y_i - \mu_i) x_{ij} \prod_{i=1}^n p(y_i | \mu_i) = E \left[\sum_{i=1}^n (y_i - \mu_i) x_{ij} \hat{\beta}_j \right]. \end{aligned}$$

ただし,

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \alpha \left(g^{-1} \left(\sum_{j=0}^p x_{ij} \beta_j \right) \right) &= \frac{\partial}{\partial \beta_j} \{-\log(1 + e^{\eta_i})\} = -\frac{e^{\eta_i}}{1 + e^{\eta_i}} \frac{\partial}{\partial \beta_j} \sum_{j=0}^p x_{ij} \beta_j \\ &= -p_i x_{ij} = -\mu_i x_{ij}. \end{aligned}$$

一般化加法モデルの加法的予測子 (2.5) が, スプライン関数 $s(x) = \sum_{l=1}^{k+m} \alpha_l B_l(x, t)$ を用いて以下のように表されるとする.

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^p s_j(x_{ij}) = \beta_0 + \sum_{j=1}^p \sum_{l=1}^{k_j+m} \theta_l^{(j)} B_l(x_{ij}, t^{(j)}). \quad (3.7)$$

ただし, k_j : s_j の内部節点 (internal knot) の個数, $t^{(j)}$: s_j の節点配列. このとき GAM の対数尤度は以下のようになり, $t^{(j)}$ が与えられれば形式的に GLM の対数尤度と同じ形になる.

$$\begin{aligned} l &= \left\{ \sum_{i=1}^n y_i \left(\beta_0 + \sum_{j=1}^p \sum_{l=1}^{k_j+m} \theta_l^{(j)} B_l(x_{ij}, t^{(j)}) \right) \right. \\ &\quad \left. + \sum_{i=1}^n \alpha \left(g^{-1} \left(\beta_0 + \sum_{j=1}^p \sum_{l=1}^{k_j+m} \theta_l^{(j)} B_l(x_{ij}, t^{(j)}) \right) \right) + \sum_{i=1}^n m(y_i) \right\}. \end{aligned}$$

よって, GAM に対する適応型モデル選択基準のペナルティは, GLM の場合と同様以下のように与えられる.

$$D(M) = \sum_{i=1}^n h_i(\boldsymbol{\mu}) = \sum_{i=1}^n E(y_i - \mu_i) \sum_{j=0}^p \sum_{l=1}^{k_j+m} \hat{\theta}_l^{(j)} B_l(x_{ij}, t^{(j)}) = \sum_{i=1}^n \text{Cov}(g^{-1}(\hat{\mu}_i), y_i). \quad (3.8)$$

定理 3.2 と同様、ロジスティック回帰、ポワソン回帰の場合、GAM に対する AMSC のペナルティは以下の簡便な形であたえられる。

定理 3.3 応答変数 y_i がベルヌーイ分布もしくはポワソン分布を持つとする。このとき、一般化加法モデルに対する適応型モデル選択基準のペナルティは、以下の形で与えられる。

$$D(M) = \sum_{i=1}^n E(y_i - \mu_i) \sum_{j=0}^p \sum_{l=1}^{k_j+m} \hat{\theta}_l^{(j)} B_l(x_{ij}, t^{(j)}) = \sum_{j=0}^p \sum_{l=1}^{k_j+m} \frac{\partial}{\partial \beta_j} E(\hat{\theta}_l^{(j)}). \quad (3.9)$$

証明 3.3 省略

結局、非線型変換 f_j に多重節点を持った可変節点スプラインを導入した GAM の推定は、上で与えられた GAM に対する適応型モデル選択基準を、モデルのチューニングパラメタである節点配列 $t^{(j)}, j = 1, \dots, p$ について最適化することで行われる。モデル推定の詳細は、以下のように要約される。

Algorithm 4: 一般化加法モデルに対する適応型モデル選択を用いた推定

1. 固定された $\lambda > 0$ に対して、一般化加法モデルを推定する。

2. 局所スコア法

(a) 初期化: $\beta_0 = g(n^{-1} \sum_{i=1}^n y_i)$, $f_j = f_j^0 = 0$, $j = 1, \dots, p$.

(b) f_1, \dots, f_p を用いた IRLS: $l = 0, 1, \dots$ に対して、

i. 修正応答変数 (adjusted response variable)

$$z_i^l = \eta_i^l + (y_i - \mu_i^l) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)_l, \quad \eta_i^l = g(\mu_i^l) = \beta_0 + \sum_{j=1}^p f_j^l(x_{ij}),$$

$$\mu_i^l = g^{-1}(\eta_i^l),$$

を求める。 ($i = 1, \dots, n$)

ii. 重み付け最小二乗法のための重み W_1, \dots, W_n を、以下のように求める。

$$W_i^l = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)_l^2 (V_i^l)^{-1}.$$

iii. 重み付き後退当てはめ:

A. 初期化: $f_j = f_j^l = 0$, $j = 1, \dots, p$.

B. $j = 1, \dots, p$ に対して、残差 $r_i^l = z_i^l - \sum_{k \neq j} f_k^l(x_k)$, $i = 1, \dots, n$ を求め、以下を最小にする節点配列 t^l を求める。

$$\min_{t^l} \sum_{i=1}^n W_i^l \left(r_i^l - \sum_{l=1}^{k+m} \alpha_l B_l(x_i, t^l) \right)^2 + \lambda(k+m)\hat{\sigma}^2. \quad (3.10)$$

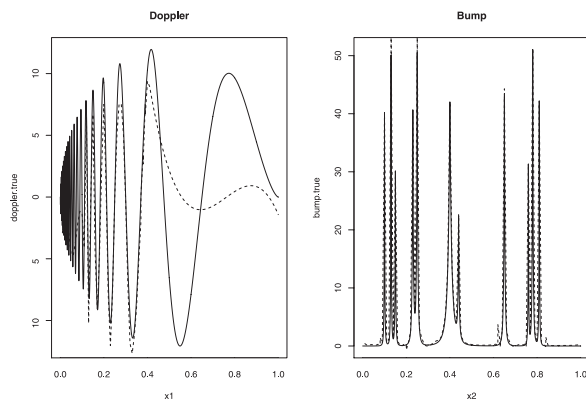


図2 真の回帰関数（左：doppler, 右：bump, 実線）と推定された曲線（点線）

C. 個々の f_j^l が収束するまで, B を繰り返す.

$\Delta(\eta^{l+1}, \eta^l)$ が収束するまで, (ii) を繰り返す. ただし

$$\Delta(\eta^{l+1}, \eta^l) = \frac{\sum_{j=1}^p \|f_j^{l+1} - f_j^l\|}{\sum_{j=1}^p \|f_j^l\|}.$$

3. Algorithm 3 により, 固定された $\lambda > 0$ に対して最適化されたモデルの AMSC を求める.

4. 1 ~ 3 を繰り返し, AMSC を最小化する $\lambda > 0$ とそれに対応する最適なモデルを探索する.

(3.10) における最小化問題は, 補遺に示す遺伝的アルゴリズムを用いて最適化する.

4. 数値例

4.1 シミュレーション

本節では 2.4.1 節において取り上げた数値例を用い, 提案した方法の有用性を数値実験により検証する. 数値実験では, 従来用いられた GAM(3 次平滑化スプラインを用い, GCV により平滑化パラメタを最適化する) と今回提案した可変節点スプラインを用いた GAM を推定し, $f = f_{\text{doppler}} + f_{\text{bump}}$, f_{doppler} , f_{bump} それぞれのパフォーマンスを平均二乗誤差 (mean squared error, MSE), $n^{-1} \sum_{i=1}^n \{y_i - \hat{f}(x_{1i}, x_{2i})\}^2$ を用いて比較した. 計算量を減らすため, Algorithm 4 における繰り返しはそれぞれ 5 回とした. 真の関数と推定された曲線の例は図 2 に示される. 100 回の数値実験による結果は, 表 1 にまとめられている.

表 1 から, 適応型モデル選択基準を用いた場合モデル全体と bump 関数のパフォーマンスは, 平滑化スプラインを用いた場合よりよい結果を示している. また, 図 2 から, doppler 関数, bump 関数の双方において, おおむね真の関数の特徴を捉えた推定が行われている.

表 1

	MSE	Doppler	Bump
GAM (smoothing spline, GCV)	167.949	36.274	192.366
GAM (AFKS, AMSC)	43.454	43.294	28.933

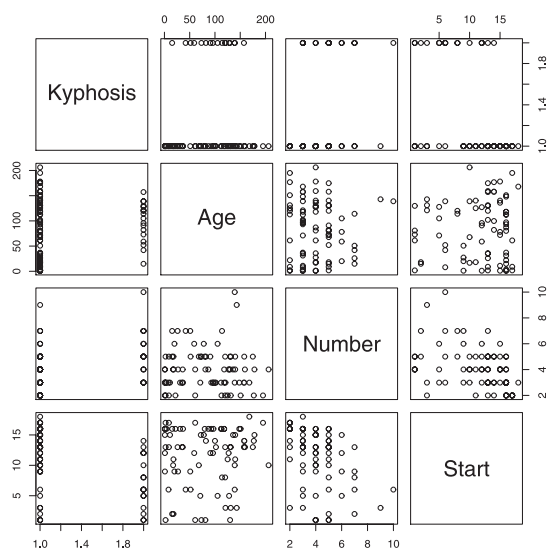


図 3 脊椎後湾症 (kyphosis) データの散布図

図 2 左, doppler 関数の $(0.5, 1.0)$ 付近の領域において, 可変節点スプラインを用いた GAM による推定曲線が真の関数を離れて平坦な形状を示すのは, GAM の後退当てはめにおいて $f_j, j = 1, \dots, p$ を逐次的に平滑化することによると思われる. すなわち, 後退当てはめにおいては, $f \approx \sum \hat{f}_j$ が達成されれば, 必ずしも $f_j \approx \hat{f}_j$ となるとは限らないため, 個々の f_j の振幅が小さくプロファイルに特徴がない場合 \hat{f}_j は全体として平坦になる傾向があるためと思われる.

4.2 脊柱後湾症データ

本節では, 提案されたモデルの実データの応用例として, 脊柱後湾症 (kyphosis) データの解析を取り上げる (Chambers *et al.* (1992)). 脊柱後湾症データは背骨の矯正手術を受けた 81 症例の子供から得られたもので, Kyphosis: 術後脊柱後湾症の有無を表わす二値変量, Age: 子供の生後月例, Number: 手術した脊椎の個数, Start: 何番目の脊椎から先を手術したかを表す数値, の四つの変量からなる. 脊柱後湾症データは統計解析パッケージ R の rpart パッケージから入手可能である. データの散布図を図 3 に示す.

本解析では脊柱後湾症の有無に対するロジスティック判別に一般化加法モデルを応用す

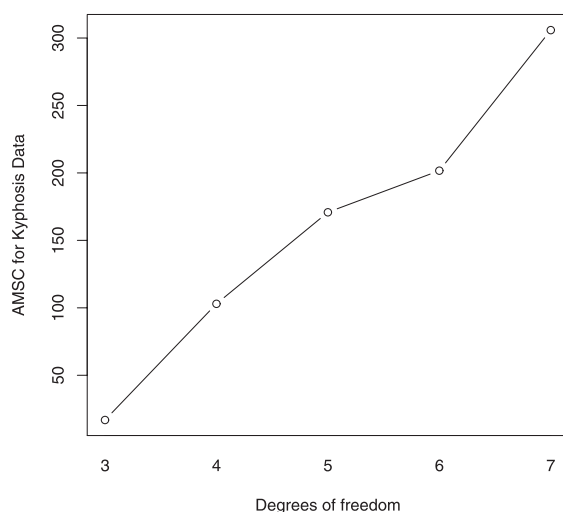


図4 脊柱後湾症データに対する GAM における AMSC

表 2

	無	有
FALSE	61	8
TRUE	3	9

るが³, 計算を簡単にするため Age のみに非線形変換を施した以下のモデルを考える.

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + s(x_{i1}) + \beta_2 x_{i2} + \beta_3 x_{i3}, \quad i = 1, \dots, 81.$$

ただし, π_i : 脊柱後湾症有の確率, x_{i1} : Age, x_{i2} : Number, x_{i3} : Start. 図3の散布図から明らかな通り, Age と Kyphosis の間には微分不能な点のような大きな構造変化は認められない. そのため Age に対する変換には多重節点を持たない等間隔節点のスプライン変換を想定し, 自由度を変えることで平滑化の程度を調節した. 提案したモデルに対する比較の対象としては, 従来用いられた GAM(3次平滑化スプラインを用い, GCV により平滑化パラメタを最適化する) による Age の平滑化を採用した. スプライン変換 $s(x_{i1})$ における自由度を $df = 3, \dots, 7$ に変化させ, それぞれ Algorithm 4 により AMSC を求めた結果を図4に示す. 計算の結果, 自由度3の場合に最も小さい AMSC を持つ最適なモデルが得られた.

GAM を用いたロジスティック判別の結果は, 表2にまとめられている. 従来法によるものも今回提案した方法によるものも同じで, 誤判別率は $(3 + 8)/81 = 0.1358$ であった.

子供の月齢に対するスプライン平滑化の予測値は図5に示される. 図5においては提案

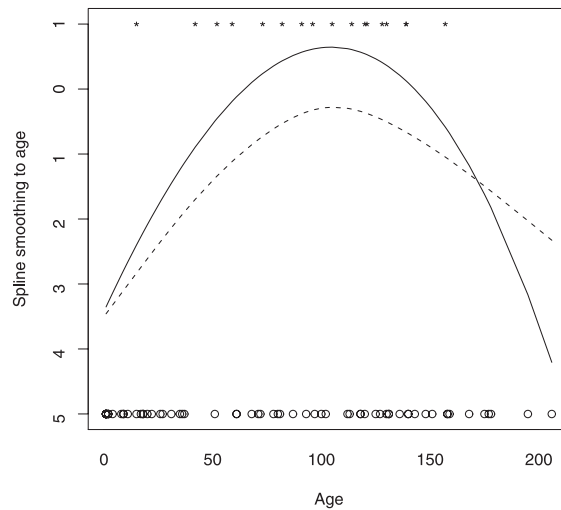


図5 脊柱後湾症データに対する GAM における月齢に対する平滑化，実線：提案した方法による平滑化，点線：従来用いられた 3 次平滑化スプライン，○：脊柱後湾症がない子供の月齢，*：脊柱後湾症がある子供の月齢

した方法による平滑化が実線で，従来用いられた 3 次平滑化スプラインが点線で示される．また脊柱後湾症がない子供の月齢を ○ で，脊柱後湾症がある子供の月齢を * で示した．

判別の結果は今回提案した方法も従来法も同じであったが，図 5 から明らかな通り脊柱後湾症が存在しない月齢 160 以上の領域では提案された方法による平滑化がより小さな値を予測しているのに対し，脊柱後湾症の存在する月齢 160 未満の領域ではより大きな値を予測しており，AMSC により平滑化パラメタの選択を行ったほうがデータの実態に合ったモデル選択がなされたことを示している．

5. まとめ

本稿では，一般化加法モデルに対する適応型モデル選択基準の導入を行った．また，データの平滑化に可変節点スプラインを用いることで，不連続点など滑らかではないデータ構造を表現できるように一般化加法モデルを拡張した．その結果，データの「滑らかさ」が局所的に変わる場合には提案された手法が従来のものよりよいパフォーマンスを示すことを，数値実験により明らかにした．

一方，今回提案した方法は計算量が莫大なものになりがちであり，より高速の計算手法の開発が必要である．また，前節においては f_j が平坦な領域において，個々の平滑化曲線の推定が困難になる現象が観察された．これは GAM において未だ解明されていない問題であり，今後の後退当てはめのような逐次的推定を必要としない方法など，新たな解析手

法の検討が必要である。

謝辞

査読者の方の丁寧かつ詳細なご指摘により、本論文の内容を大きく改善することができた。また、一橋大学経済学研究科（2010年4月より統計数理研究所）の元山齊先生には、完成前の原稿に目を通していただき貴重なコメントをいただいた。ここに記して厚く御礼申し上げる。

A. 補遺

A.1 遺伝的アルゴリズム

本稿における適応型モデル選択基準の最適化は、一般に複雑な非線形問題となる。この問題に大域的最適解を与えるため、本稿で確率的最適化法、とくに 遺伝的アルゴリズム (Evolutionary Algorithm, EA) と呼ばれる方法を用いる。本節では、Miyata and Shen (2003, 2005) に従い、EA の概略を示す。 $f: S \mapsto R$ を目的関数とする。ただし S は探索空間であり、 f は例えば拡張された GAM に対する後退当てはめ法における罰則付き残差二乗和 (3.10) である。最適化問題、

$$\min_{x \in S} f(x)$$

を考える。 \mathcal{P} と呼ばれる ξ 個の (たとえば $\xi = 10$) $x \in S$ (個体, individual) からなる集団 (population) $\mathcal{P}^{(t)} = \{x^{(t,1)}, \dots, x^{(t,\xi)}\}$ を考える。集団に含まれる個々の個体 $x^{(t,l)}$ に突然変異と呼ばれる確率的変動を加え、新たな個体 (offspring) $x^{(t,\xi+1)}$ を生成する。新たな個体の元となった $x^{(t,l)}$ を $x^{(t,\xi+1)}$ の親 (parent) と呼ぶ。

突然変異: 親となる個体に確率変数 Z を加えて新たな個体を生成する。

$$x \xrightarrow{\text{突然変異}} x + Z.$$

親の世代 (generation) に属する個体と、新たな個体からなる個体の目的関数値を評価し、値の小さい順に ξ 個の個体を選択し次の世代とする。各世代の最適値を $f_{(t)}^*$ とする。

$$\begin{aligned} f_{(t+1)}^* &= \min \left\{ \underbrace{f(x^{(t,1)}), \dots, f(x^{(t,\xi)})}_{\text{parents}}, \underbrace{f(x^{(t,\xi+1)}), \dots, f(x^{(t,\xi+\xi)})}_{\text{offspring}} \right\} \\ &= \min \{ f(x^{(t+1,1)}), \dots, f(x^{(t+1,\xi)}) \}. \end{aligned}$$

このとき、以下が成立する。

定理 A.1.1 (Miyata and Shen (2003, 2005)) $Z^{(t,l)}$ を $x^{(t,l)}$ に対する突然変異とする. 任意の $\epsilon > 0$ に対して $\delta > 0$ が存在し, 以下が成立すると仮定する.

$$P(X^{(t,\xi+l)} \in A_\epsilon) = P(x^{(t,l)} + Z^{(t,l)} \in A_\epsilon) \geq \delta, \quad (\text{A.1.1})$$

ただし, $t \geq 0$, $x^{(t,l)} \in S$, $l = 1, \dots, \xi$, $A_\epsilon = \{x \in S : f(x) < \min_{x \in S} f(x) + \epsilon\}$. このとき以下が成立する.

$$f_{(t)}^* \rightarrow f^* = \min_{x \in S} f(x) \text{ almost surely as } t \rightarrow \infty.$$

証明は, Rudolph (1997), Miyata and Shen (2003) を参照. EA の大域的収束は定理 A.1.1 により保証されるが, 収束のパフォーマンスは, EA が適用される最適化問題と確率の変動 Z の突然変異分布の設計に依存している. 具体的な問題における実装は次節以降で議論する.

A.2 可変節点スプライン

本節では (3.10) での最小化問題を解くため, A.1 節で導入された遺伝的アルゴリズムを実際に適用するための方法を議論する. EA を記述するため, スプライン関数 $s(x) = \sum_{l=1}^{k+m} \alpha_l B_l(x, t)$ のチューニングパラメタである節点配列 $h = (t, k)$ を以下のように記す.

$$h^{(t,l)} = (t^{(t,l)}, k^{(t,l)}) : l\text{-th 個体 in } t\text{-th 世代, } t \in N, l = 1, \dots, \xi.$$

$h^{(t,l)} = (t^{(t,l)}, k^{(t,l)})$ は t -th 世代における突然変異によって生成される新たな個体 $h^{(t,\xi+l)} = (t^{(t,\xi+l)}, k^{(t,\xi+l)})$ の親になる. 集団 \mathcal{P} に含まれる個体の個数を $\xi = 10$ とする. $h = (t, k)$ に対する突然変異を以下のように定義する.

節点数に関する突然変異: 親の節点の個数 $k^{(t,l)}$ に対し, 切断正規分布を用いて新たな個体の節点の個数 $k^{(t,\xi+l)}$ を生成する. 切断正規分布による突然変異の範囲は $[1, n]$ とする.

$$k^{(t,\xi+l)} \sim \text{切断正規分布 } N(k^{(t,l)}, ck^{(t,l)})I_{[1,n]}.$$

ただし, $k^{(t,\xi+l)}$ は四捨五入して整数化される. パラメタ c については後述する.

中間的な個体: $j = m + 1, \dots, k^{(t,\xi+l)} - m$ に対して, 新たな個体の節点 $t_j^{(t,\xi+l)}$ を選択する. $t_j^{(t,\xi+l)}$ は親の節点配列 $\{t_{m+1}^{(t,l)}, \dots, t_{k^{(t,l)}-m}^{(t,l)}\}$ 上の一様分布により生成される. 添え字 j は t の大きさ順に並べなおされる.

節点位置に関する突然変異: $t_j^{(\xi+i,l)}$, $j = m + 1, \dots, k^{(\xi+i,l)} - m$ に対する突然変異は, 左

右に非対称な確率分布になる．まず， $j = m + 1, \dots, k^{(\xi+i,l)} - m$ に対して

$$d_j^{(\xi+i,l)} = \begin{cases} 2inc & \text{if } (t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) \leq 2inc, \\ (t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) + 2inc & 2inc < (t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) \leq 4inc, \\ 2(t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) & 4inc < (t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) \leq 6inc, \\ 4(t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}) & 6inc < (t_j^{(\xi+i,l)} - t_{j-1}^{(\xi+i,l)}). \end{cases}$$

とする．ただし， $inc = 1/n$ ．さらに $\max\{inc, \min\{s, d_j^{(\xi+i,l)}\}\}$ を，あらためて $d_j^{(\xi+i,l)}$ とする．パラメタ s については後述する．このとき， $t_j^{(\xi+i,l)}$ は， $v|Z|$ ， $Z \sim N(t_j^{(i,l)}, 1)$ と置き換えられる．ただし $1/2$ の確率で， $v = d_j^{(\xi+i,l)}$ あるいは $-d_{j+1}^{(\xi+i,l)}$ ．

多重節点: 新たな個体の節点配列 $t^{(\xi+i,l)}$ は，大小順に並べ替えられ， $|t_{j+1}^{(\xi+i,l)} - t_j^{(\xi+i,l)}| < inc$ でありかつ節点の多重度が m 以下であれば， $t_{j+1}^{(\xi+i,l)}$ は $t_j^{(\xi+i,l)}$ で置き換えられる．多重度が m を超える場合は，その節点は取り除かれる．

パラメタ c, s に対する並行焼戻し パラメタ c, s は，それぞれ節点数と節点位置に関する突然変異分布の分散を制御するパラメタである．目的関数の最適解を広い範囲から探索するため， c, s は EA における早い世代では大きな値を取る一方，世代が進行し最適化が進んだ段階では最適解周辺での微調整のために小さな値をとるように設定される．具体的には，世代 t におけるパラメタ $c = c^{(t)}$ は以下のように決定される．

1. 初期化: $c_1^{(0)} = c_2^{(1)} = c_3^{(0)} = 1.0$.
2. $c_1^{(t)} = c_1^{(t-1)} * T_0$, $c_2^{(t)} = c_2^{(t-1)} - 1/ntemp$, $zz^{(t)} = zz^{(t-1)}/T_0$, $c_3^{(t)} = 1.0 - zz^{(t)}$,
 $t = 1, 2, \dots, ntemp$.
3. $c^{(t)} \sim \text{Uniform}\{c_1^{(t)}, c_2^{(t)}, c_3^{(t)}\}$.

ただし， $ntemp (= 100)$: パラメタ減少の段階数， $T_0 (= 0.95)$: 減少率， $zz^{(0)} = T_0^{ntemp}$ ．すなわち， $c_1^{(t)}$ は指数的に， $c_2^{(t)}$ は線型に， $c_3^{(t)}$ はさらに緩慢に減少する．上記3でパラメタの値の三つの減少過程を混合する操作は「並行焼戻し (parallel tempering)」と呼ばれ，アニーリング法 (simulated annealing) の技法を参考にしたものである．同様に世代 t におけるパラメタ $s = s^{(t)}$ は， $\{s_1^{(t)}, s_2^{(t)}, s_3^{(t)}\}$ から等確率で決定される． $\{s_1^{(t)}, s_2^{(t)}, s_3^{(t)}\}$ の初期値は， x の範囲 $(x_{(n)} - x_{(1)})$ に設定され，上記と同様に更新される．

以上の突然変異により，節点配列 $h = (t, k)$ の更新を行い，前節の EA にしたがって (3.10) の最小化を行う．Miyata and Shen (2005, 定理1) により，本節で定義された EA は確率1で (3.10) の最小解に収束する．上記の突然変異における，節点数に関する突然変異，節点

位置に関する突然変異のための突然変異分布は、例えば Miyata and Shen (2005, 図 1) に与えられている。

参 考 文 献

- Chambers, J. M., Hastie, T. J. *et al.* (1992). *Statistical Models in S*, Wadsworth and Brooks/Cole, Pacific Grove, CA.
- de Boor, C. (2001). *A Practical Guide to Splines, Revised edition*, Springer.
- Donoho, D. L. and Johnstone, I. M. (1995). Adapting to Unknown Smoothness Via Wavelet Shrinkage, *J. Am. Stat. Assoc.*, **90**, 1200–1224.
- Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation, *J. Am. Statist. Assoc.*, **78**, 316–331.
- Hastie, T. and Tibshirani, R. (1987). Generalized Additive Models: Some applications, *J. Am. Stat. Assoc.*, **82**, 371–386.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman & Hall Ltd., London, New York.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall Ltd., London, New York.
- Miyata, S. and Shen, X. (2003). Adaptive Free-Knot Splines, *J. Comput. Graph. Stat.*, **12**, 197–213.
- Miyata, S. and Shen, X. (2005). Free-knot splines and adaptive knot selection, *J. Jpn. Statist. Soc.*, **35**, 303–324.
- R Development Core Team (2008). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rudolph, G. (1997). *Convergence Properties of Evolutionary Algorithms*, Verlag Dr. Kovač, Hamburg.
- Shen, X., Huang, H.-C. and Ye, J. (2004). Adaptive Model Selection and Assessment for Exponential Family Distributions, *Technometrics*, **46**, 306–317.
- Shen, X. and Ye, J. (2002). Adaptive Model Selection, *J. Am. Stat. Assoc.*, **97**, 210–221.
- Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer-Verlag Inc., Berlin, New York.
- 辻谷将明, 外山信夫 (2007). 「R による GAM 入門」『行動計量学』**34**, 111–131.
- Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC.
- Ye, J. (1998). On Measuring and Correcting the Effects of Data Mining and Model Selection, *J. Am. Stat. Assoc.*, **93**, 120–131.