

ベキ重み付き平滑化スプラインによる 分散均一性の診断

大阪大学 大学院基礎工学研究科 坂本 亘

要 旨 ノンパラメトリック回帰モデルでは多くの場合に分散均一性などの要件が暗黙裡に仮定されるため、回帰関数の推定に悪い影響が生じるおそれがある。分散均一性の診断を未知の非線形回帰構造の推定と同時に行うことを目的として、ベキ重み付き平滑化スプライン(PWSS)モデルが提案される。加法回帰モデルの応答にベキ変換を施し、変換後に分散が一定になることが仮定される。罰則付き最尤法により推定関数として平滑化スプラインが得られ、重み可変型の後退あてはめアルゴリズムが構築される。ベキ変換パラメータと関数の滑らかさを制御する平滑化パラメータは平滑化スプラインに対する Bayes 流接近法に基づき最大周辺尤度法によって推定される。平滑化スプラインが線形混合モデルの最良線形不偏予測量であるという性質を利用して、比較的計算が容易な周辺対数尤度の形が導出される。文献事例のデータ集合に対する検討およびシミュレーション実験により、PWSS モデルによって推定されたベキ変換が非線形の構造を考慮に入れながら分散を均一化しうることが示される。

1. 序

ノンパラメトリック回帰は「探索的データ解析」の代表的な道具の一つである。未知の回帰構造を柔軟で包括的な関数空間の中で探索したり、逆に仮定したパラメトリック・モデルの妥当性を診断したりすることで、パラメトリック方式との相補的な役割を果たす。中でも平滑化スプラインはノンパラメトリック回帰の最も有用な手法の一つとして知られている。その理由として、罰則付き接近法の最適解として得られ拡張性に富むこと、Bayes 流接近法により線形混合モデルの最良線形不偏予測子(BLUP)として表現されること、計算を実行するためのプログラムやソフトウェアが豊富に開発されていること、などを挙げることができる。

ところで、連続値応答が数個の説明変数とともに観測されている場合に、回帰構造の探索を行うには、Hastie and Tibshirani (1986, 1990) などで詳細に論じられ、S-PLUS などの統計ソフトウェアにも採用されている、加法回帰モデルの適用を検討することができる。ところが、加法回帰モデルでは (i) 加法性、(ii) 分散均一性、(iii) 正規性(最小二乗推定を行うことにより暗黙裡に仮定される)、(iv) 独立性、という四つの要件が仮定される。これらの要件を同時にすべて満たすことは困難であるにもかかわらず、加法回帰モデルの「堅い」仮定のために回帰関数の推定に悪い影響が生じるおそれがある。そこで、仮定の一部を緩めることにより、これらの要件の妥当性を診

断することが要請される．相関のある場合については多くの論文で取り扱われている (Diggle and Hutchinson (1989), Altman (1990), Wang (1999), 坂本 (2002)) が, (i)~(iii) の三つの要件の妥当性を診断するには応答の変換を行うことが考えられる．

応答の変換方式としてはノンパラメトリック, パラメトリックの両方向からの研究が進んでいる．ノンパラメトリック変換は, 応答の滑らかな変換関数を平滑化などの方法で推定するもので, ACE(交替条件付き期待値)(Breiman and Friedman (1985)), AVAS(加法化および分散安定化)(Tibshirani (1988)), ノンパラメトリック両辺変換(Wang and Ruppert (1995))などが提案されている．しかしながら, 実際の適用では変換関数は単調であることが望ましい上に, 測定された応答が濃度や質量といった意味をもつことが多いため, パラメトリック変換の方が解析結果の解釈を行う上で有益な示唆を行う場合がある．パラメトリック変換の代表として, Box and Cox (1964) によって提案されたベキ変換

$$y^{(\phi)} = \begin{cases} (y^\phi - 1)/\phi, & \phi \neq 0, \\ \log y, & \phi = 0 \end{cases} \quad (1)$$

(ϕ はベキ・パラメータ)の利用が Hastie and Tibshirani (1990) や Linton *et al.* (1997) の中で検討されている．これらは応答の変換を行った後に (i)~(iii) の要件を同時に満たすことを仮定しているが, これらの仮定に関する妥当性を診断するという目的をより明確にするためには, 変換前の仮定と変換後に満たすべき要件とを分離することが必要である (Goto (1995)).

本稿では (ii) 分散均一性の診断に着目する．Box and Hill (1974) では線形モデルの適用の中で元の応答に (i) 加法性(線形性)と (iii) 正規性を仮定し, ベキ変換によって (ii) 分散均一性を満たすことを要請しており, この方式は Goto (1995) によってベキ重み付き変換(PWT)と呼ばれている．本稿ではこの着想をノンパラメトリック加法回帰モデルに取り入れたベキ重み付き平滑化スプライン(PWSS: power-weighted smoothing splines)モデルを提案する．本モデルは加法回帰モデルの応答にベキ変換を施し, 変換後に分散が一定になることを仮定するものであり, 分散均一性の診断と同時に未知の非線形回帰構造の推定を行うことを目的とする．ベキ・パラメータと関数の滑らかさを制御する平滑化パラメータの推定は, 分散を均一にする応答の変換, および応答と説明変数の間に潜在する構造の両方を推定するという点で重要であり, 本稿では Bayes 流接近法に基づく最大周辺尤度法の利用を提案する．

本稿では, 2 節で加法回帰モデルとその推定方式について概説した後, 3 節で PWSS モデルを提案し, その推定方式について論じる．さらに 4 節で文献事例への適用および簡単なシミュレーション実験の結果について述べる．

2. 加法平滑化スプライン

2.1. 加法回帰モデル

応答 y_i ($i = 1, \dots, n$) の各々が説明変数 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ および $\mathbf{t}_i = (t_{i1}, \dots, t_{ir})^T$ とともに独立に観測されているとする．通常の加法回帰モデルでは, 応答 y_i は期待値が

$$\mu_i \equiv E[y_i | \mathbf{x}_i, \mathbf{t}_i] = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}_1 + f_1(t_{i1}) + \dots + f_r(t_{ir}) \quad (2)$$

によって表され, 一定の分散 σ^2 をもつと仮定される．ここに, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T)^T$ は回帰パラメータ

のベクトル, f_1, \dots, f_r は平滑関数であり, これらを推定する. なお, 最適解の一意性を保証するため, $\sum_{i=1}^n f_j(t_{ij}) = 0$ の条件を課す.

推定されるパラメータおよび平滑関数の計算方式として, 後退あてはめアルゴリズム (Hastie and Tibshirani (1986)) がよく知られており, 以下の更新方程式の収束解を推定値とする.

$$\hat{\boldsymbol{\beta}}^{\text{new}} = (X_L^T X_L)^{-1} X_L^T (\mathbf{y} - \hat{\mathbf{f}}_1 - \dots - \hat{\mathbf{f}}_r), \quad (3a)$$

$$\hat{\mathbf{f}}_j^{\text{new}} = S_j (\mathbf{y} - X_L \hat{\boldsymbol{\beta}} - \sum_{k \neq j} \hat{\mathbf{f}}_k) \quad (j = 1, \dots, r), \quad (3b)$$

ここに, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{f}_j = (f_j(t_{1j}), \dots, f_j(t_{nj}))^T$, $X_L = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n]^T$, $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i^T)^T$ ($i = 1, \dots, n$) であり, 記号 $\hat{\cdot}$ は推定値を表す. また, S_j ($j = 1, \dots, r$) は第 j 説明変数の値 (t_{1j}, \dots, t_{nj}) に対する平滑作用素で, 核平滑化, 局所平滑化, スプライン平滑化など, 種々の平滑化手法に対応したものが用いられる.

スプライン平滑化を用いることで得られる推定値は罰則付き接近法の最適解であることが知られている. 加法回帰モデル (2) に対して, 通常の残差平方和に罰則項を加えたものを最小化することで推定値 (平滑化スプライン) が得られる. これは, 元の応答 y_i が平均 μ_i , 分散 σ^2 の正規分布に従うことを仮定すれば, 罰則付き対数尤度

$$l_P(\boldsymbol{\beta}, f_1, \dots, f_r) = -\frac{1}{2\sigma^2} \sum_{i=1}^m (y_i - \mu_i)^2 - \frac{1}{2} \sum_{j=1}^r \lambda_j J(f_j) \quad (4)$$

を (所与の σ^2, λ_j に対して) 最大にする $\boldsymbol{\beta}, f_1, \dots, f_r$ を求めることと同等である. ここに, $\lambda_j (> 0)$ は平滑化パラメータと呼ばれ, 推定関数の滑らかさを制御する. $J(f_j)$ は関数 f_j の乱雑度を表す罰則である. 罰則として f_j の m 次導関数の 2 乗積分 $J(f_j) = \int \{f_j^{(m)}(t_j)\}^2 dt_j$ を用いるとき, 最大罰則付き尤度推定値 (MPLE) は $(2m - 1)$ 次自然スプラインとなることが知られている (Green and Silverman (1994)). S_j を $\lambda_j^* = \sigma^2 \lambda_j$ の値に依存する (中心化された) 平滑化スプライン作用素として, MPLE の値を更新方程式 (3) によって求めることができる.

2.2. 平滑化スプラインに対する Bayes 流接近法

関数 f_j の MPLE, すなわち $(2m - 1)$ 次平滑化スプラインについては, Bayes 流接近法による表現が知られている. 関数 f_j に対する事前分布を

$$\mathbf{f}_j = X_{S_j} \boldsymbol{\delta}_j + Z_{S_j} \mathbf{a}_j \quad (5)$$

と表現するとき, f_j の事後平均および事後モードは $(2m - 1)$ 次平滑化スプラインになる (Green (1987)). ここに X_{S_j} は $q_j \times (m - 1)$ 行列, Z_{S_j} は $q_j \times (q_j - m)$ 行列, $\boldsymbol{\delta}_j$ は $(m - 1)$ 次元固定パラメータ・ベクトル, そして \mathbf{a}_j は $(q_j - m)$ 次元ランダム・パラメータ・ベクトルで各要素は独立に平均 0, 分散 λ_j^{-1} の正規分布に従う. 制約 $\sum_j f_j(t_{ij}) = 0$ のため, X_{S_j} は定数項部分に対応する列を含めないことに注意されたい (X_{S_j}, Z_{S_j} の具体的な表現については補遺 A.1 を参照).

さらに, 平滑化スプラインに対する線形混合モデル表現が知られており, 後述する平滑化・分散パラメータの推定方式に利用される. 加法回帰モデル (2) の場合, $\boldsymbol{\delta} = (\boldsymbol{\delta}_1^T, \dots, \boldsymbol{\delta}_r^T)^T$, $\mathbf{a} = (\mathbf{a}_1^T, \dots, \mathbf{a}_r^T)^T$ と書くとき, MPLE ($\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\mathbf{a}}$) は線形混合モデル

$$\mathbf{y} = X_L \boldsymbol{\beta} + X_S \boldsymbol{\delta} + Z_S \mathbf{a} + \boldsymbol{\epsilon}, \quad \mathbf{a} \sim N(\mathbf{0}, \Lambda^{-1}), \quad \boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (6)$$

の最良線形不偏予測量 (BLUP) であることが知られている (Robinson (1991), Speed (1991), Zhang *et al.* (1998)). ここに $X_S = [X_{S1}, \dots, X_{Sr}]$, $Z_S = [Z_{S1}, \dots, Z_{Sr}]$ であり, Λ は行列 $\lambda_1 I, \dots, \lambda_r I$ (I は単位行列) を対角ブロックにもつ対角行列である.

2.3. 平滑化パラメータの選定

ノンパラメトリック回帰, 中でも平滑化スプラインに関する文献の中で標準的に用いられている平滑化パラメータの選定方式は, Craven and Wahba (1979) の一般化交差確認法に代表される交差確認法や, 情報量基準の最適化 (Hastie and Tibshirani (1990), Eilers and Marx (1996)) など, 「応答の最適予測」を目標としている. ところが, 序章で述べたように, ノンパラメトリック回帰の本来の目的は「回帰構造の探索」であり, 応答の予測値に基づく選定方式はその目的に合致しないと考えられる (坂本 (2002)). さらに, 予測基準では, 推定しようとする構造が複雑であるほど, それらの情報を取り入れることが困難であり, 過小平滑化につながるものが指摘されている (Simonoff (1996), Diggle and Hutchinson (1989), Altman (1990), Wang (1998)).

平滑化パラメータを選定する別の方法は 2.2 節で述べた平滑化スプラインに対する Bayes 接近法に基づく. 最大周辺尤度法 (MML) は線形混合モデル (6) に対して \mathbf{y} の周辺密度

$$p(\mathbf{y}) = \iiint p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a}) p(\boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\mathbf{a}) d\boldsymbol{\beta} d\boldsymbol{\delta} d\mathbf{a}$$

に基づく尤度を最大にする. ここに, $p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a})$ は $\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a}$ が与えられたときの \mathbf{y} の条件付き密度である. また, $p(\boldsymbol{\beta}), p(\boldsymbol{\delta}), p(\mathbf{a})$ はそれぞれ $\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a}$ の事前確率密度であり, $\boldsymbol{\beta}, \boldsymbol{\delta}$ については変則 (improper) 一様分布 $p(\boldsymbol{\beta}) \propto 1, p(\boldsymbol{\delta}) \propto 1$ に従うとみなされる.

制限付き最尤推定法 (REML) は分散成分のパラメータの推定方式として Patterson and Thompson (1971) によって提案されたもので, 誤差対比 (すなわち期待値が 0 である対比) の線形独立集合の密度に基づく尤度 (制限付き尤度) を最大にする方法である. 加法回帰モデル (2) に対して, 応答の正規性を仮定するとき, REML は MML と同等な方法である. 周辺 (制限付き) 対数尤度の式は, 後述の PWSS モデルに対して 3.3 節で導出する周辺 (制限付き) 対数尤度の特別な形をもつ. Zhang *et al.* (1998), 坂本 (2002) も参照されたい.

3. ベキ重み付き平滑化スプライン

3.1. ベキ重み付き変換を用いた線形モデル

ベキ重み付き平滑化スプラインについて論じる前に, Box and Hill (1974) などで研究されている, 線形回帰モデルに対するベキ重み付き変換について述べる. 応答 y_i ($i = 1, \dots, n$) は期待値が説明変数 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ を用いて

$$\mu_i \equiv E[y_i|\mathbf{x}_i] = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}_1 \quad (7)$$

と表され, 一定でない分散 $\sigma^2 w_i^{-1}$ (w_i は未知の重み) をもつ正規分布 $N(\mu_i, \sigma^2 w_i^{-1})$ に独立に従うと仮定される. さらに, (1) で定義される応答の Box-Cox ベキ変換 $y_i^{(\phi)}$ が一定の分散 σ^2 をもつと仮定される. 応答の変換前後の分散の間には近似的に $\text{var}\{y_i^{(\phi)}\} \approx \text{var}(y_i)(dy^{(\phi)}/dy)_{\mu_i}^2$ という関係が成り立つ (Bartlett (1947)) ことから, 変換前の分散の重みを $w_i \approx \mu_i^{2\phi-2}$ と考えることができる.

ベキ・パラメータ ϕ を回帰パラメータ $\beta = (\beta_0, \beta_1^T)^T$ とともに推定することで、線形モデルにおける分散の不均一性が修正される。

Box and Hill (1974) では、所与の ϕ に対して重みを $w_i = \hat{\mu}_i^{2\phi-2}$ ($\hat{\mu}_i = \hat{\beta}_0 + \mathbf{x}_i^T \hat{\beta}$, $\hat{\beta}_0, \hat{\beta}_1$ は β_0, β の重み付き最小二乗推定値) によって更新し、 $\hat{\beta}_0, \hat{\beta}_1$ が収束するまで更新を繰り返すことで推定値を求めている。ベキ・パラメータ ϕ の推定には最大周辺尤度法が用いられ、 ϕ の推定値が 1 から離れるほど分散不均一性が顕著であることが示唆される。Goto (1995) などはこのモデルの応答変換をベキ重み付き変換と呼び、 (β, ϕ, σ^2) に関する通常の最尤推定を行っている。なお、Carroll and Ruppert (1987) などによって研究されている両辺変換モデル

$$y_i^{(\phi)} = \mu_i^{(\phi)} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

は応答のベキ重み付き変換と類似した結果を与えることが知られている。このことは現実のデータへの適用の際に解釈上の重要な意味を与える。

3.2. ベキ重み付き平滑化スプライン

2節と同様に、応答 y_i ($i = 1, \dots, n$) の各々が説明変数 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ および $\mathbf{t}_i = (t_{i1}, \dots, t_{ir})^T$ とともに独立に観測されているとする。応答 y_i は期待値が

$$\mu_i \equiv E[y_i | \mathbf{x}_i, \mathbf{t}_i] = \beta_0 + \mathbf{x}_i^T \beta_1 + f_1(t_{i1}) + \dots + f_r(t_{ir}) \quad (8)$$

と表され、一定でない分散 $\sigma^2 w_i^{-1}$ をもつ正規分布 $N(\mu_i, \sigma^2 w_i^{-1})$ に独立に従うと仮定される。さらに (1) で定義される応答の Box-Cox ベキ変換 $y_i^{(\phi)}$ が一定の分散 σ^2 をもつと仮定する。回帰パラメータ $\beta = (\beta_0, \beta_1^T)^T$ および平滑関数 f_1, \dots, f_r をベキ・パラメータ ϕ とともに推定する (2節と同様に、最適解の一意性を保証するため、 $\sum_{i=1}^n f_j(t_{ij}) = 0, j = 1, \dots, r$ の条件を課す)。3.1節と同様に、変換前の分散の重みを $w_i \approx \mu_i^{2\phi-2}$ と考えることができる。そこで、ベキ・パラメータ ϕ を回帰パラメータ $\beta = (\beta_0, \beta_1^T)^T$ および平滑関数 f_1, \dots, f_r とともに推定することにより、加法モデルにおける分散均一性を診断することが可能となる。

モデル (8) に対して罰則付き対数尤度

$$l_p(\beta, f_1, \dots, f_r) = -\frac{1}{2\sigma^2} \sum_{i=1}^m w_i (y_i - \mu_i)^2 - \frac{1}{2} \sum_{j=1}^r \lambda_j J(f_j) \quad (w_i = \mu_i^{2\phi-2}) \quad (9)$$

を (所与の $\phi, \sigma^2, \lambda_j$ に対して) 最大にする β, f_1, \dots, f_r を求める。ここに $\lambda_j (> 0)$ は平滑化パラメータ、 $J(f_j)$ は f_j の乱雑度罰則 $J(f_j) = \int \{f_j^{(m)}(t_j)\}^2 dt_j$ である。2.1節の加法モデルの場合と同様に、モデル (8) における f_1, \dots, f_r の最大罰則付き尤度推定値 (MPLE) は $(2m-1)$ 次自然スプラインであることが容易に示される。これらの MPLE \hat{f}_j を本稿ではベキ重み付き平滑化スプライン (PWSS) と呼ぶ。

(9) の最大化では Fisher 評点化のような反復アルゴリズムの適用も考えられるが、重み $w_i = \mu_i^{2\phi-2}$ が β と f_j ($j = 1, \dots, n$) に依存するため、 l_p の導関数の計算がやや複雑になる。そこで、ここでは Box and Hill (1974) に倣い、所与の重み w_i に対する β および f_j の推定方程式を導出し、 w_i を更新しながらこれらの推定方程式を反復的に用いることを考える。重み w_i が与えられているとき、本稿の 2.1 節や Hastie and Tibshirani (1986) と同様にして、加法モデルに対する (重み付きの) 更

新方程式

$$\hat{\boldsymbol{\beta}}^{\text{new}} = (X_L^T W X_L)^{-1} X_L^T W (\mathbf{y} - \hat{\mathbf{f}}_1 - \cdots - \hat{\mathbf{f}}_r), \quad (10a)$$

$$\hat{\mathbf{f}}_j^{\text{new}} = S_{W_j} (\mathbf{y} - X_L \hat{\boldsymbol{\beta}} - \sum_{k \neq j} \hat{\mathbf{f}}_k) \quad (j = 1, \dots, r) \quad (10b)$$

を導出することができる．ここに $W = \text{diag}(w_1, \dots, w_n)$, $w_i = \hat{\mu}_i^{2\phi-2}$ であり, S_{W_j} は(中心化された)重み付き平滑化スプライン作用素で, $\{t_{1j}, \dots, t_{nj}\}$ および $\lambda_j^* = \sigma^2 \lambda_j$ に依存する．そこで, 所与の ϕ および $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_r^*)$ に対して, $\hat{\boldsymbol{\beta}}$ および $\hat{\mathbf{f}}_j$ ($j = 1, \dots, r$) が一通り更新されるごとに重み w_i を更新し, $\hat{\boldsymbol{\beta}}$ および $\hat{\mathbf{f}}_j$ の値が収束するまで反復を繰り返す．すなわち, これは重み可変型の後退あてはめアルゴリズムと考えることができる．更新方程式 (10) の導出の詳細は補遺 A.2 を参照されたい．

3.3. ベキ・平滑化パラメータの推定

PWSS モデル (8) の目的は加法モデルにおける分散均一性を診断することであり, ベキ・パラメータ ϕ がその指標となるため, ϕ の推定が重要となる．Box and Hill (1974) からの類推により, 最大周辺尤度法(MML)を用いて ϕ を推定する方法が考えられる．他方, 2.3 節の方法と同様に, 平滑化スプラインの Bayes 流接近法に基づいて, MML により平滑化パラメータ $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)$ の推定を行う方法が考えられる．そこで, ベキ・平滑化の両方のパラメータを同時に MML によって推定する方法を提案する．

関数 f_j を (5) によって表現する．重み w_i ($i = 1, \dots, n$) が与えられているとみなすとき, PWSS モデル (8) に対応する線形混合モデル

$$\mathbf{y} = X_L \boldsymbol{\beta} + X_S \boldsymbol{\delta} + Z_S \mathbf{a} + \boldsymbol{\epsilon}, \quad \mathbf{a} \sim N(\mathbf{0}, \Lambda^{-1}), \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 W^{-1}) \quad (11)$$

を考慮することができる．(11) に対する $(\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a})$ の BLUP は連立方程式

$$\begin{bmatrix} X_L^T W X_L & X_L^T W X_S & X_L^T W Z_S \\ X_S^T W X_L & X_S^T W X_S & X_S^T W Z_S \\ Z_S^T W X_L & Z_S^T W X_S & Z_S^T W Z_S + \Lambda^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} X_L^T W \mathbf{y} \\ X_S^T W \mathbf{y} \\ Z_S^T W \mathbf{y} \end{bmatrix} \quad (12)$$

の解である．ここに $\Lambda^* = \sigma^2 \Lambda$ である．BLUP $\hat{\boldsymbol{\beta}}$ および $\hat{\boldsymbol{\delta}}_j$, $\tilde{\mathbf{a}}_j$ ($j = 1, \dots, r$) を用いて, \mathbf{f}_j の MPLE は $\hat{\mathbf{f}}_j = X_{S_j} \hat{\boldsymbol{\delta}}_j + Z_{S_j} \tilde{\mathbf{a}}_j$ と表される．

線形混合モデル (11) に対して, \mathbf{y} の周辺密度

$$p(\mathbf{y}) = \iiint p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{a}) p(\boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\mathbf{a}) d\boldsymbol{\beta} d\boldsymbol{\delta} d\mathbf{a}$$

に基づく周辺尤度 $l_M(\phi, \boldsymbol{\lambda}^*, \sigma^2 | \mathbf{y}) = \log p(\mathbf{y})$ を最大化する．まず, σ^2 の MML 推定量を

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T W \{\mathbf{y} - (X_L \hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}_1 + \cdots + \hat{\mathbf{f}}_r)\}}{n - d} \quad (d = 1 + p + r(m - 1)) \quad (13)$$

と陽に求めることができる．これを代入した周辺対数尤度

$$l_M(\phi, \boldsymbol{\lambda}^*, \hat{\sigma}^2 | \mathbf{y}) = (\text{定数}) - \frac{n - d}{2} \log \hat{\sigma}^2 + \sum_{j=1}^r \frac{q_j - m}{2} \log \lambda_j^* + \frac{1}{2} \sum_{i=1}^n \log w_i - \frac{1}{2} \log |C_W| \quad (14)$$

を最大にする ϕ, λ^* をその MML 推定量とする. ここに C_W は (12) の左辺の係数行列である (正則であるとする). 通常の加法モデルでは, W が恒等行列に置き換わり, λ^* のみに対する l_M の最大点の探索を行うのに対して, PWSS では (ϕ, λ^*) に対する l_M の最大点の探索を行う. (14) の導出の詳細については補遺 A.3 を参照されたい. なお, (14) は (11) に対する制限付き対数尤度としても同一の式が得られる.

3.4. PWSS のアルゴリズム

PWSS のアルゴリズムを以下に要約する.

1. 所与の ϕ に対して

(a) 初期重みの設定

(i) ベキ重み付き変換線形回帰モデル $y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}_1 + \mathbf{t}_i^T \boldsymbol{\beta}_2 + \epsilon_i$ を考え, 初期の重みを $w_i = y_i^{2\phi-2}$ などとして, 3.1 節の方法で線形回帰推定値 $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ を求める.

(ii) 線形回帰の推定平均 $\hat{\mu}_i = \hat{\beta}_0 + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_1 + \mathbf{t}_i^T \hat{\boldsymbol{\beta}}_2$ に対して, $w_i = \hat{\mu}_i^{2\phi-2}$ を (b) での初期重みとして用いる.

(b) 所与の λ^* に対して, PWSS モデル (8) へのあてはめを 3.2 節の方法で行う.

(重み可変型の後退あてはめアルゴリズム)

(i) 重み w_i を $w_i = \hat{\mu}_i^{2\phi-2}$ によって更新する.

(ii) (10a) により $\hat{\boldsymbol{\beta}}$ を更新する.

(iii) $j = 1, \dots, n$ に対して, (10b) により \hat{f}_j を更新する.

(iv) 収束判定を行う. 収束すれば (c) へ.

(v) (i) に戻る.

(c) 3.3 節の方法で, σ^2 の MML 推定値 $\hat{\sigma}^2$ を (13) によって求め, 周辺対数尤度 $l_M(\phi, \lambda^*, \hat{\sigma}^2)$ を (14) によって計算する.

2. 格子探索法などにより $l_M(\phi, \lambda^*, \hat{\sigma}^2)$ を最大にする (ϕ, λ^*) を見つける.

PWSS のアルゴリズムを改良する余地はありうるであろう. 例えば, (b) の後退あてはめアルゴリズムの収束を改良するために, 更新の量を縮小する修正 Newton 法を適用することも可能であろう. また, (ϕ, λ^*) に対する格子探索の代わりに, これらのパラメータに関する $l_M(\phi, \lambda^*, \hat{\sigma}^2)$ の導関数 (または差分近似) を用いて, Fisher 評点化アルゴリズムを構築することも可能であろう.

BLUP $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})$ の Bayes 流事後共分散行列は $\hat{\sigma}^2 C_{W(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})}^{-1}$ によって推定される. 関数 f_j の Bayes 流信頼区間は, 逆行列 $C_{W(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})}^{-1}$ の $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})$ に対応する成分 $C_{W(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})}^{-1}$ を用いて,

$$\text{var}(\hat{f}_j) = \hat{\sigma}^2 [X_S \ Z_S] C_{W(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{a}})}^{-1} \begin{bmatrix} X_S^T \\ Z_S^T \end{bmatrix}$$

より得ることができる.

4. ベキ重み付き平滑化スプラインの適用と評価

4.1. 文献事例の検討

表 1 に列挙する 8 組のデータ集合に対して PWSS モデルを適用する. 比較対照として, ベキ重

表 1. 8 組のデータ集合.

データ/参考文献	応答 y	説明変数 x, t	標本サイズ
(1) スキーナ川の紅鮭 * Carroll and Ruppert (1987)	帰ってきた魚の数	t_1 : 年間に産卵する雌魚	$n = 28$
(2) 新聞の発行部数 ** Simonoff (1996)	日曜日の発行部数	t_1 : 平日の発行部数	$n = 19$
(3) エタノールの燃焼実験 ** S-PLUS データセット Chambers and Hastie (1992) Simonoff (1996)	NOx 濃度	t_1 : 圧縮比 t_2 : 等価比	$n = 88$
(4) ニューヨークの大気汚染 ** S-PLUS データセット Chambers and Hastie (1992)	オゾン濃度	t_1 : 日射量 t_2 : 気温 t_3 : 風速	$n = 111$
(5) 石油岩石標本の浸透率 ** Venables and Ripley (1999)	浸透率	t_1 : 横断面の面積 t_2 : 周囲の長さ t_3 : 形状 (= $t_2 / \sqrt{t_1}$)	$n = 48$
(6) 桜の木の体積 ** Bowman and Azzalini (1997) Schimek and Turlach (2000)	体積	t_1 : 直径 t_2 : 全長	$n = 31$
(7) 糖尿病データ * Hastie and Tibshirani (1990)	C-ペプチド濃度	t_1 : 年齢 t_2 : 酸性度	$n = 43$
(8) 血液の凝固時間 * McCullagh and Nelder (1989) Hastie and Tibshirani (1990)	凝固時間	t_1 : 血漿濃度 x_1 : 血漿の種類	$n = 18$

* の文献はデータが掲載されている. ** のソフトウェアや文献はサンプル・データを電子的に入手可能.

み付き変換線形回帰(PWTLR)モデル (7), および通常の加法回帰モデル (2) を同時に適用する. PWSS モデルで, $\phi = 1$ (無変換) とすれば加法回帰モデルに対応し, 平滑化パラメータを無限に大きく取れば PWTLR モデルに近づく.

PWSS モデルおよび加法回帰モデルについては, \hat{f}_j に対する等価自由度 (EDF) (Hastie and Tibshirani (1990)) を求める. ここでは, 平滑化行列 (S_j または S_{w_j}) のトレースから 1 を引いたものとして定義する. これが 1 に近いほど推定関数は直線に近く, 大きくなるほど乱雑な曲線であることを示す. また, (8) のデータ集合に対しては, 説明変数 x_1 に対応する回帰パラメータ β_1 の推定値をその標準誤差とともに算出する. 3 種類のモデルを適用した結果を表 2 に示す.

PWSS モデルの適用により, 分散を均一にするためには, (1), (5), (6) では対数変換, (2) では平方根変換, (8) では逆数変換がそれぞれ必要であり, (3), (7) では変換が不要であることが示唆される. データ集合 (1) は Carroll and Ruppert (1987) などにより詳細に解析されており, Ricker モデルや Beverton-Holt モデルといったパラメトリック非線形回帰モデルが用いられている. PWSS モデルは回帰構造に対する仮定を必要としないが, やはり対数変換により分散が均一化されることを示唆する. (4), (5), (6) については, 表 1 に示す文献では応答を最初から対数変換して加法モデルなどのあてはめを行っており, PWSS モデルの適用結果は対数変換が妥当であることを示す.

PWSS モデルと PWTLR モデルとの比較では, (3), (6), (7), (8) について $\hat{\phi}$ の値が大きく異なっている. PWSS モデルは (平滑化パラメータが十分大きいときの極限で) PWTLR モデルを含んでいるので, 平滑関数をモデルに加えることで分散を均一にする変換をより適応的に推定しようと

表 2. PWSS モデルなどの適用結果.

データ	PWTLR (7)		PWSS (8)		加法モデル (2)
	応答変換 $\hat{\phi}$	回帰 $\hat{\beta}$	応答変換 $\hat{\phi}$	回帰 EDF, $\hat{\beta}$	回帰 EDF, $\hat{\beta}$
(1)	-0.040		-0.003	\hat{f}_1 : 1.42	\hat{f}_1 : 1.78
(2)	0.430		0.431	\hat{f}_1 : 1.00	\hat{f}_1 : 1.00
(3)	-0.653		1.167	\hat{f}_1 : 3.04 \hat{f}_2 : 10.10	\hat{f}_1 : 2.75 \hat{f}_2 : 10.03
(4)	—		0.188	\hat{f}_1 : 2.35 \hat{f}_2 : 4.04 \hat{f}_3 : 7.56	\hat{f}_1 : 1.77 \hat{f}_2 : 3.56 \hat{f}_3 : 3.48
(5)	—		0.076	\hat{f}_1 : 1.15 \hat{f}_2 : 3.39 \hat{f}_3 : 3.25	\hat{f}_1 : 1.92 \hat{f}_2 : 1.00 \hat{f}_3 : 1.00
(6)	-2.154		-0.107	\hat{f}_1 : 3.31 \hat{f}_2 : 1.00	\hat{f}_1 : 3.49 \hat{f}_2 : 1.00
(7)	2.266		1.054	\hat{f}_1 : 2.51 \hat{f}_2 : 2.02	\hat{f}_1 : 2.52 \hat{f}_2 : 2.01
(8)	-2.096		-5.321	\hat{f}_1 : 4.62	\hat{f}_1 : 4.91
		$\hat{\beta}_1 = -7.75$ (4.43)		$\hat{\beta}_1 = -6.93$ (0.40)	$\hat{\beta}_1 = -15.67$ (4.88)

- — は収束がうまくいかなかったために省略したものである。(4), (5) では多くの $\hat{\phi}$ に対して PWTLR モデルの収束が得られなかった。
- (4) (7) (8) に対する PWTLR, PWSS の適用では Newton 法の更新のステップサイズを 0.7 に縮小している。
- 回帰係数 $\hat{\beta}_1$ の下に表示している () 内の値は推定標準誤差である。

していることが示唆される。例として (6) に対する PWTLR モデルおよび PWSS モデルの規準化残差のプロットを図 1 に示す。PWTLR モデルでは応答の値が大きいところでの残差のばらつきがやや小さく、しかも正の方向に偏っている。他方、PWSS モデルでは残差が 0 を中心に上下にほぼ均一にばらついており、応答の値によるばらつきの変化も見られない。この結果が示唆することは、(6) の場合に PWSS モデルで変数 t_1 に対する非線形構造を考慮することで、対数変換による応答の分散均一性が満たされるという診断を正しく下すことができるということである。

PWSS モデルと加法モデルとの比較では (4), (5), (8) について推定関数の滑らかさに差が生じている。データ集合 (4) の PWSS モデルと加法モデルによる解析結果を図 2 および図 3 に示す。図 2 は規準化残差のプロット、図 3 は推定関数 \hat{f}_j ($j = 1, 2, 3$) のプロットである。(4) では $\hat{\phi} = 0.188$ のベキ変換を施すことで分散がほぼ均一になり、分散が不均一であるまま加法モデルを適用したときの結果とはやや異なる推定関数が得られることがわかる。

データ集合 (8) に対して適用されるモデルはパラメトリック回帰係数 β_1 を含むセミパラメトリック・モデルである。3 種類のモデルで全く異なる結果が得られているが、PWSS モデルに対して、すなわち分散を均一化するための変換により、 $\hat{\beta}_1$ の標準誤差の値が他の 2 種類のモデルより大幅に小さくなっていることは特筆に値する。周辺最尤推定量 $\hat{\beta}_1$ の標準誤差の値は各モデルについて推定量の標準偏差の到達しうる下限を推定しているはずであるが、このことはデータ集合 (8) が PWTLR モデルでの線形性の仮定や加法モデルでの応答の分散均一性の仮定を満たしていないことを示唆していると考えられる。

Trees Data: Standard Residuals

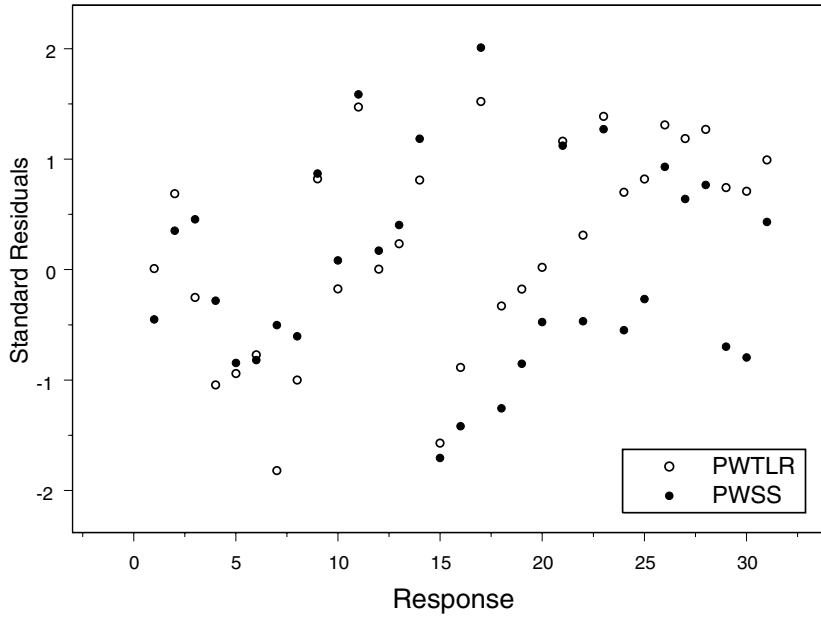


図 1. データ集合 (6) に対して **PWSS** モデルおよび **PWTLR** モデルをあてはめたときの規準化残差を (応答を横軸にとって) プロットしたものの。

Ozone Data: Standard Residuals

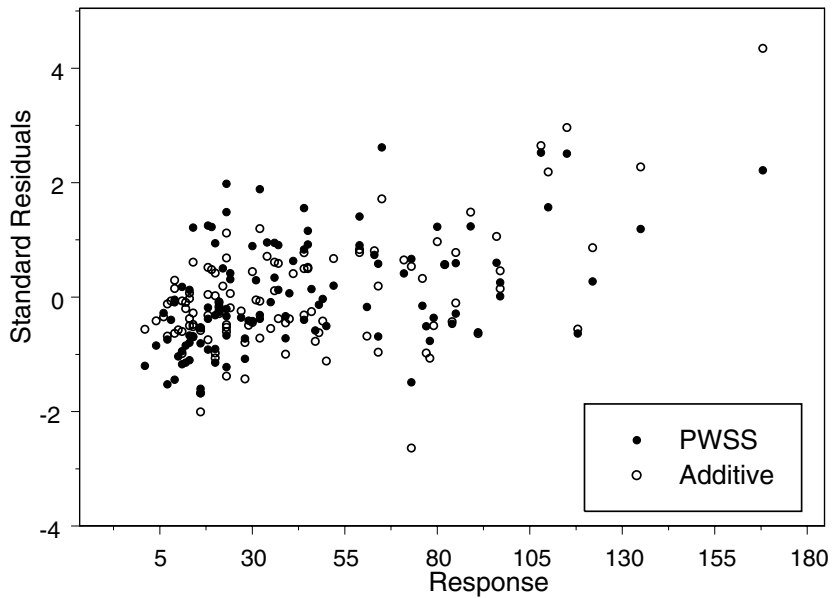


図 2. データ集合 (4) に対して **PWSS** モデルおよび加法モデルをあてはめたときの規準化残差を (応答を横軸にとって) プロットしたものの。

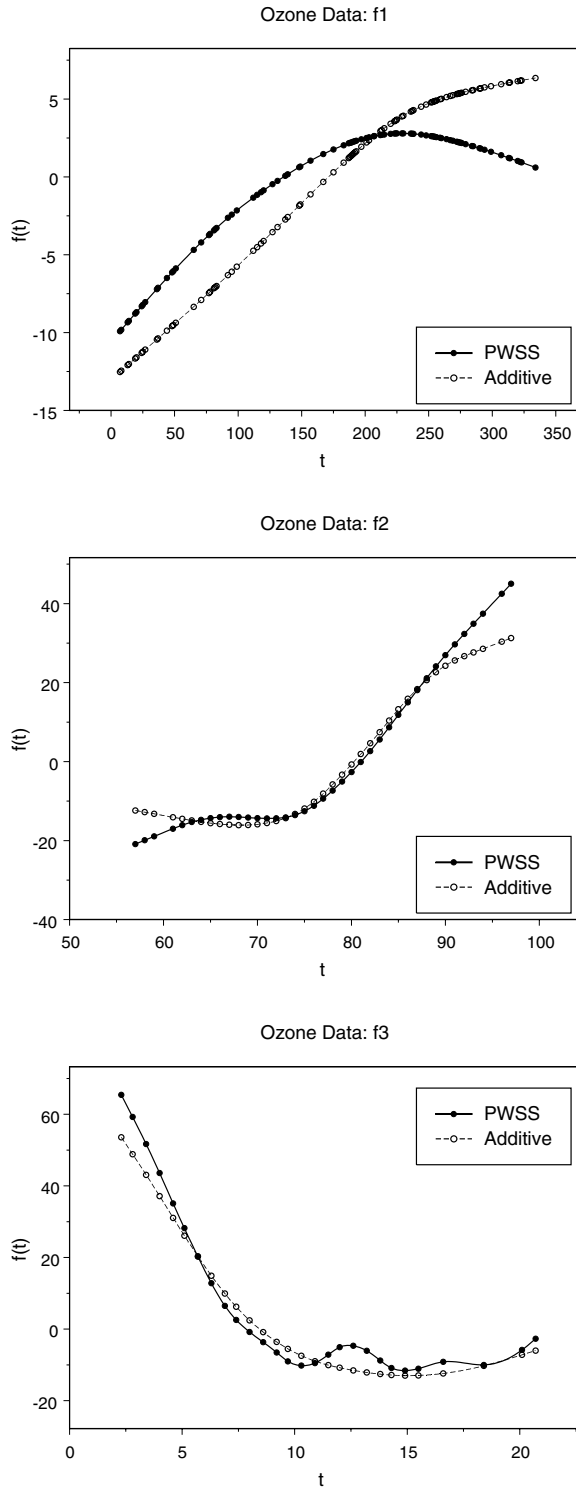


図 3. データ集合 (4) に対する PWSS モデルおよび加法モデルの推定関数 \hat{f}_j ($j = 1, 2, 3$) のプロット.

4.2. シミュレーションによる評価

PWSS モデルの性能を確かめるための簡単なシミュレーション実験について報告する。

真のモデル 応答 y_i ($i = 1, \dots, n$) の各々が説明変数 $t_i \in [0, 1]$ とともに両辺変換モデル

$$y_i^{(\phi)} = \{f(t_i)\}^{(\phi)} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \quad (\text{独立}) \quad (15)$$

に従うものとする。(3.1 節で述べたように、両辺変換モデルは応答のベキ重み付き変換に類似している。乱数の生成が容易であることから、データの生成には両辺変換モデル (15) を採用する。なお、分散均一性の診断が目標であるので右辺(説明変数側)の構造はなるべく単純にしている。)

真の関数 $f(x)$ を以下のように定める。[1] $\{f(t)\}^{(\phi)}$ が 1 次関数になるもの ($f(t)$ は $\phi = 1$ の場合を除き非線形関数)、[2] $f(t)$ 自身が 1 次関数であるもの、ただし [1],[2] とも $0 \leq t \leq 1$ に対する $\{f(t)\}^{(\phi)}$ の値域が $[0, 2]$ ($\phi > 0$)、 $[-1, 1]$ ($\phi = 0$)、 $[-2, 0]$ ($\phi < 0$) となるように単調増加関数 $f(t)$ を定める。実際に用いた ϕ の値に対する具体的な $f(t)$ の形は表 3 に示すとおりである。

表 3. 用いた真の関数.

ϕ	[1] 変換後に線形	[2] 変換前に線形
2	$f(t) = \sqrt{1 + 4t}$	$f(t) = 1 + (\sqrt{5} - 1)t$
1	$1 + 2t$	$1 + 2t$
0.5	$(1 + t)^2$	$1 + 3t$
0	$\exp(2t - 1)$	$\frac{1}{e} + (e - \frac{1}{e})t$
-1	$1/(3 - 2t)$	$(1 + t)/3$

なお、本稿では説明変数の値を $t_i = (2i - 1)/2n$ ($i = 1, \dots, n$) と等間隔にとり、 $\phi = -1, 0, 0.5, 1, 2$ という代表的なベキ変換を表すベキ・パラメータを用い、 $\sigma = 0.2$, $n = 25, 50, 100, 200$ としたときの結果を述べる。

データの生成 誤差 $\{\epsilon_i\}$ に相当する正規擬似乱数を IMSL サブルーチン DRNNOR を用いて生成し、これより両辺変換モデル (15) に従う $\{y_i\}$ を生成する。

モデルのあてはめ 上記で得られたデータ $\{t_i, y_i; i = 1, \dots, n\}$ に対して PWSS モデル (8) を 3.2 節の方法であてはめる。推定関数 \hat{f} は 3 次平滑化スプライン ($m = 2$) とする (計算時間の節約のため、 n が 50 を超えるときは節点を $s_k = (2k - 1)/100$ ($k = 1, \dots, 50$) の 50 個とする)。ベキ・パラメータ ϕ の推定値 $\hat{\phi}_S$ を平滑化パラメータの推定値とともに格子探索によって得る。さらに推定関数 \hat{f} の滑らかさを調べるために等価自由度 $\text{EDF} = \text{tr}(\mathcal{S}_{W_j} - 1)$ を計算する。比較対照として 3.1 節のベキ重み付き変換線形回帰 (PWTLR) モデル (7) を用い、同様に ϕ の推定値 $\hat{\phi}_L$ を求める。

モデルの性能評価 以上の操作を 1000 回繰り返す。推定値 $\hat{\phi}_L, \hat{\phi}_S$ および EDF (PWSS モデル) の経験分布を調べる。さらに $\hat{\phi}_*$ の平均平方誤差 $\text{MSE}(\hat{\phi}_*) = E[\{\hat{\phi}_* - \phi\}^2]$ ($*$ は L または S) の推定値を計算し、PWTLR モデルに対する PWSS モデルの推定効率 $\text{MSE}(\hat{\phi}_L)/\text{MSE}(\hat{\phi}_S)$ を求める。

結果と解釈 1000 個の推定値 $\hat{\phi}$ をボックスプロット (S-PLUS 6.0 を用いている) により、 $\phi = 2$ および $\phi = -1$ の場合についてそれぞれ図 4, 図 5 に示す。

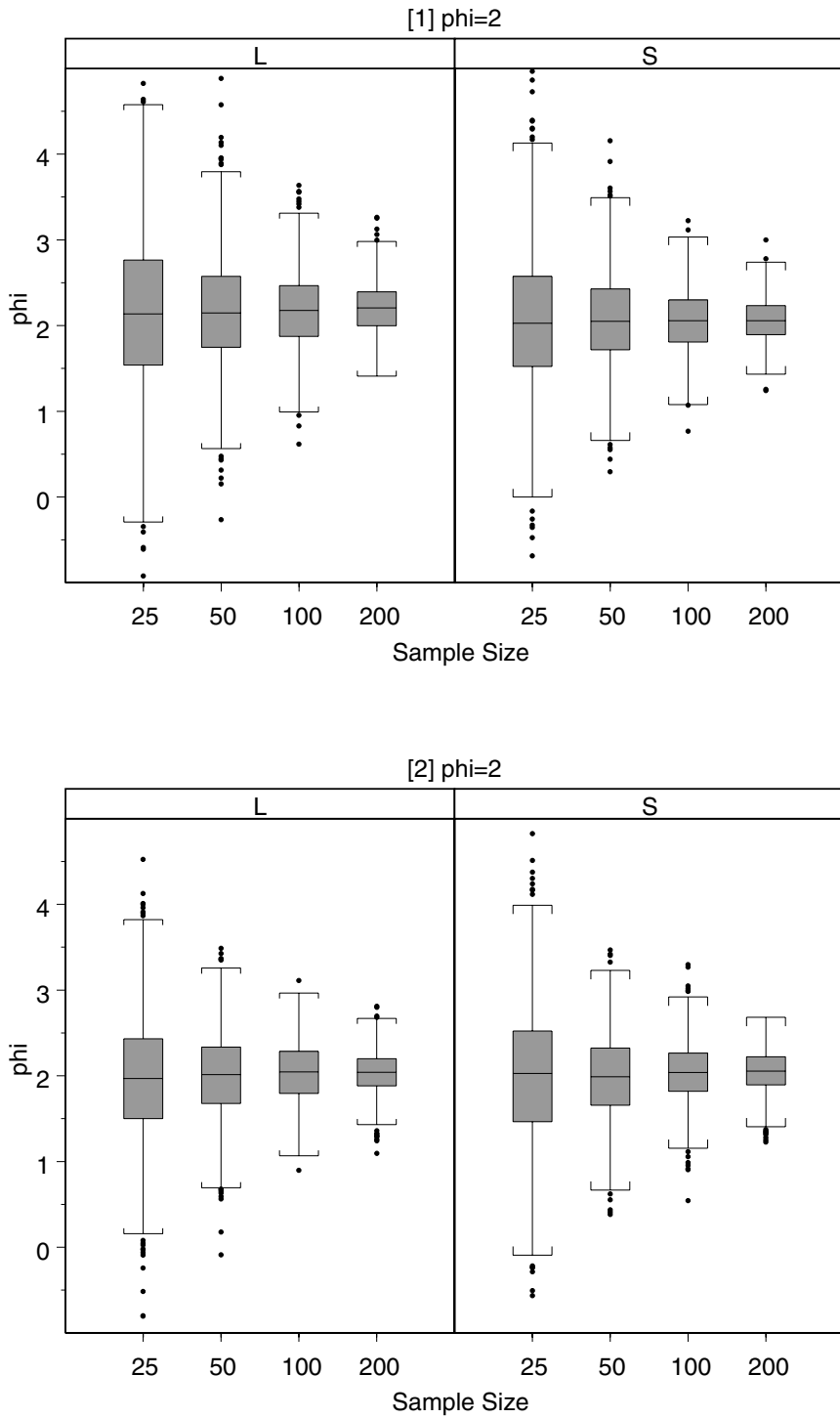


図 4. $\hat{\phi}$ のボックスプロット ($\phi=2$). L は PWTLR モデル, S は PWSS モデル.

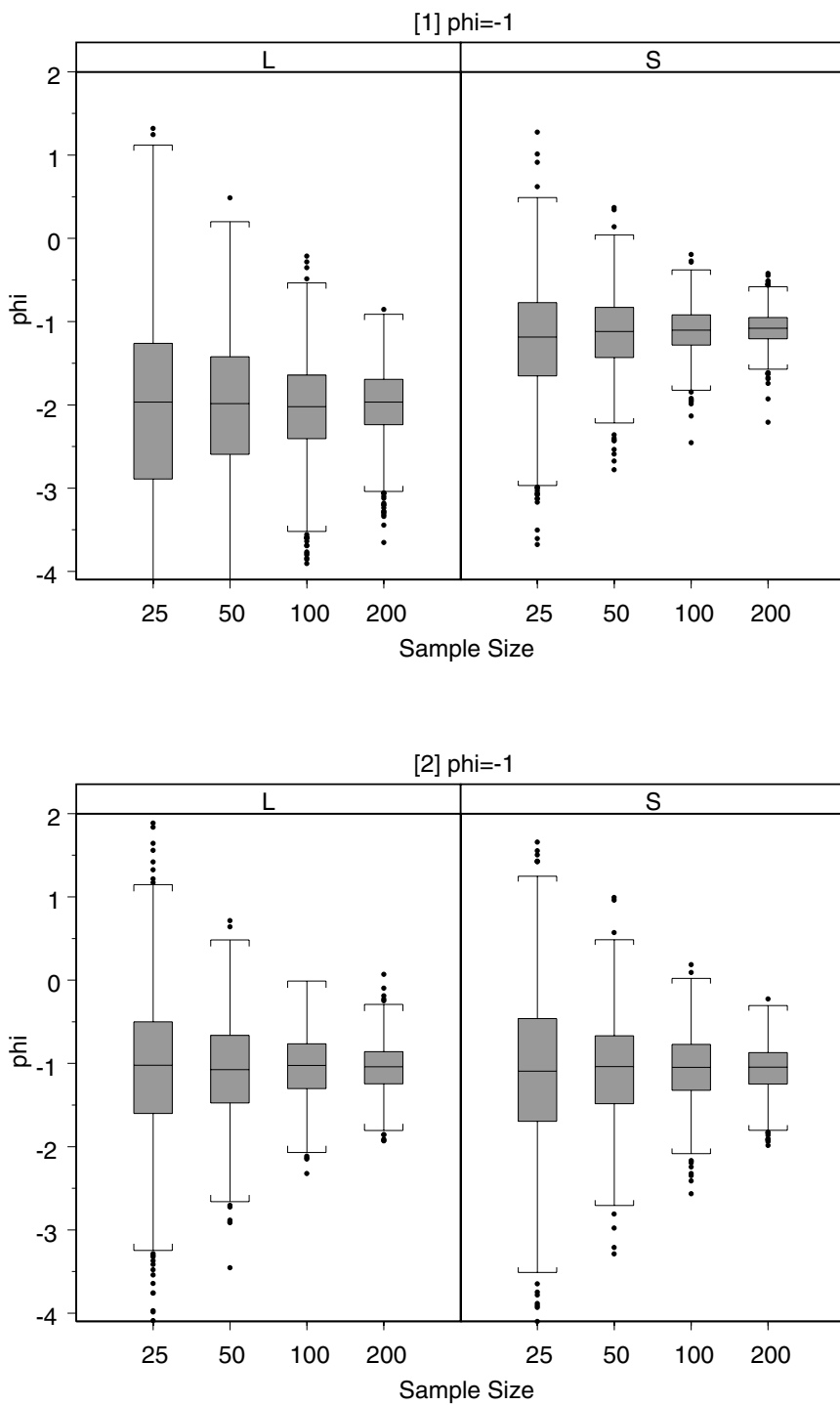


図 5. $\hat{\phi}$ のボックスプロット ($\phi = -1$). L は PWTLR モデル, S は PWSS モデル.

[1] 真の $f(t)$ が非線形である場合：

PWTLR モデルでは $\hat{\phi}_L$ の中央値が真の ϕ の値に対して偏りをもつ(真値が $\phi > 1$ のときは上方に, $\phi < 1$ のときは下方に偏る). 真値が $\phi = 1$ から離れるほど偏りの程度が大きくなり, $\hat{\phi}_L$ の一致性が成り立たなくなる. これに対して, PWSS モデルでは $\hat{\phi}_S$ の中央値の偏りが小さい. ばらつきも PWTLR モデルと比べて小さく, 標本サイズの増加とともにばらつきが小さくなる. これらのことから $\hat{\phi}_S$ は一致性をもつと考えられる.

[2] 真の $f(t)$ が線形である場合：

PWTLR モデル, PWSS モデルともに中央値の偏りは小さいが, ばらつきは PWTLR モデルの方が PWSS モデルよりも小さくなる. PWTLR モデルでは [1] の場合よりもばらつきが小さいのに対して, PWSS モデルでは [1] の場合と比べてそのばらつきが同じかわずかに大きい程度である.

推定値 $\hat{\phi}$ の推定効率の対数値 $\log_{10} \text{MSE}(\hat{\phi}_L)/\text{MSE}(\hat{\phi}_S)$ を図 6 に折れ線グラフで示す. この値が 1 より大きいときは PWSS モデルの方が, 1 より小さいときは PWTLR モデルの方が, MSE の推定値が小さいことを意味する. なお, $\hat{\phi}$ の外れ値が $\text{MSE}(\hat{\phi})$ の推定に極端な影響を及ぼしている場合はその外れ値(1000 個のうち高々 3 個まで)を除いて計算している.

[1] 真の $f(t)$ が非線形である場合：

$\phi = 1$ の場合を除き, PWSS の方が PWTLR と比べて MSE が小さい. 真値が $\phi = 1$ から離れるほどその差は大きくなる.

[2] 真の $f(t)$ が線形である場合：

PWSS の方が PWTLR と比べて MSE が小さくなる時もあり, 対数推定効率は 0 の周辺でばらついている.

1000 回の反復での EDF の値をボックスプロットにより, $\phi = 2$ および $\phi = -1$ の場合について図 7 に示す.

[1] 真の $f(t)$ が非線形である場合：

EDF は 1 から離れたところに分布し, 標本サイズが大きいほど EDF の値も大きくなる. EDF が 1 より大きいときには平滑化パラメータの値があまり大きくなり, 非線形の推定関数が得られることを意味する.

[2] 真の $f(t)$ が線形である場合：

EDF の分布は 1 の周辺で密度が高くなる. EDF が 1 に近いときには平滑化パラメータの値が大きくなり, 線形の推定関数が得られることを意味する.

以上の結果から, PWTLR モデルと PWSS モデルについて以下のようなことが言える.

● ベキ・パラメータ ϕ の推定性能

変換前の回帰構造が線形である場合には PWTLR モデルの推定値 $\hat{\phi}_L$ は一致性をもち, パラメトリック推定の強みが発揮されるが, 変換前の回帰構造が非線形である場合には ($\phi = 1$ の場合を除き) $\hat{\phi}_L$ の一致性が成り立たなくなる. 他方, PWSS モデルでは真の回帰構造が線形であるか非線形であるかを問わず推定値 $\hat{\phi}_S$ の一致性が成り立つ. よって, 変換前の回帰構造が線形である場合には相対的に PWTLR の推定性能に及ばないが, 変換前の回帰構造が非線形である

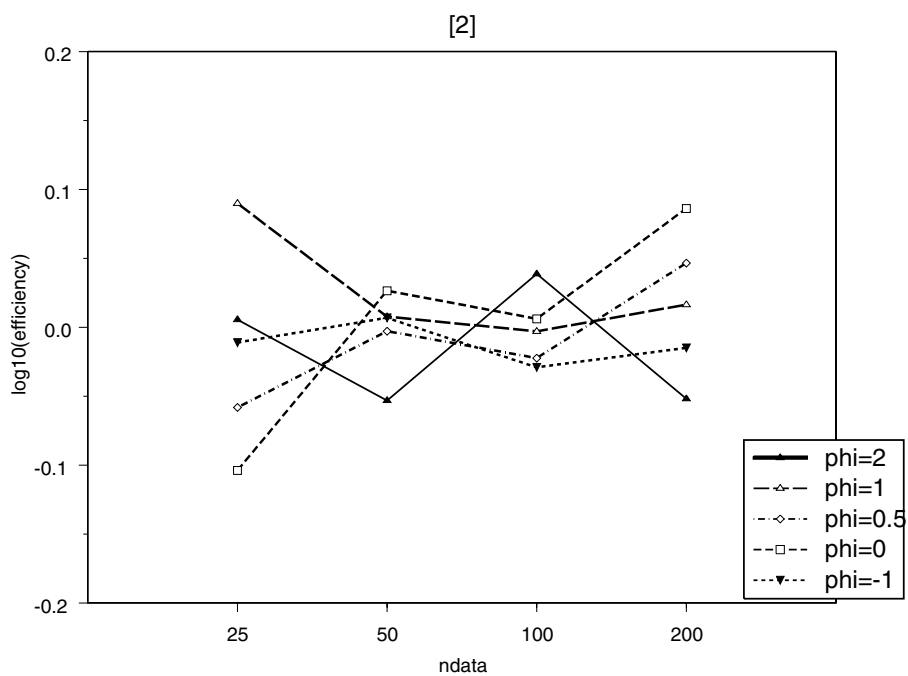
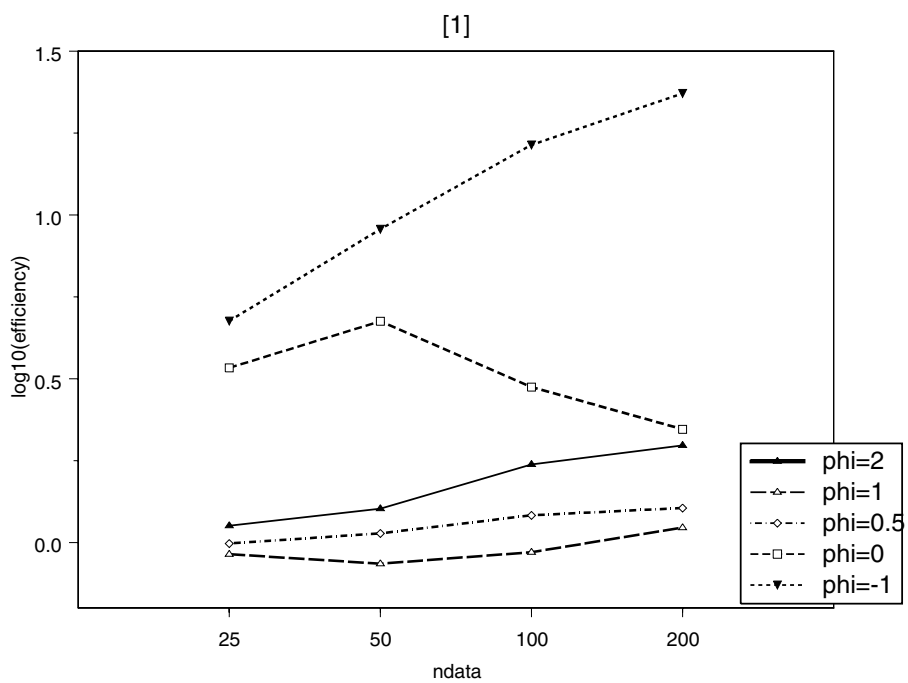


図 6. $\hat{\phi}_L$ と $\hat{\phi}_S$ の相対効率の対数 $\log_{10}\{\text{MSE}(\hat{\phi}_L)/\text{MSE}(\hat{\phi}_S)\}$.

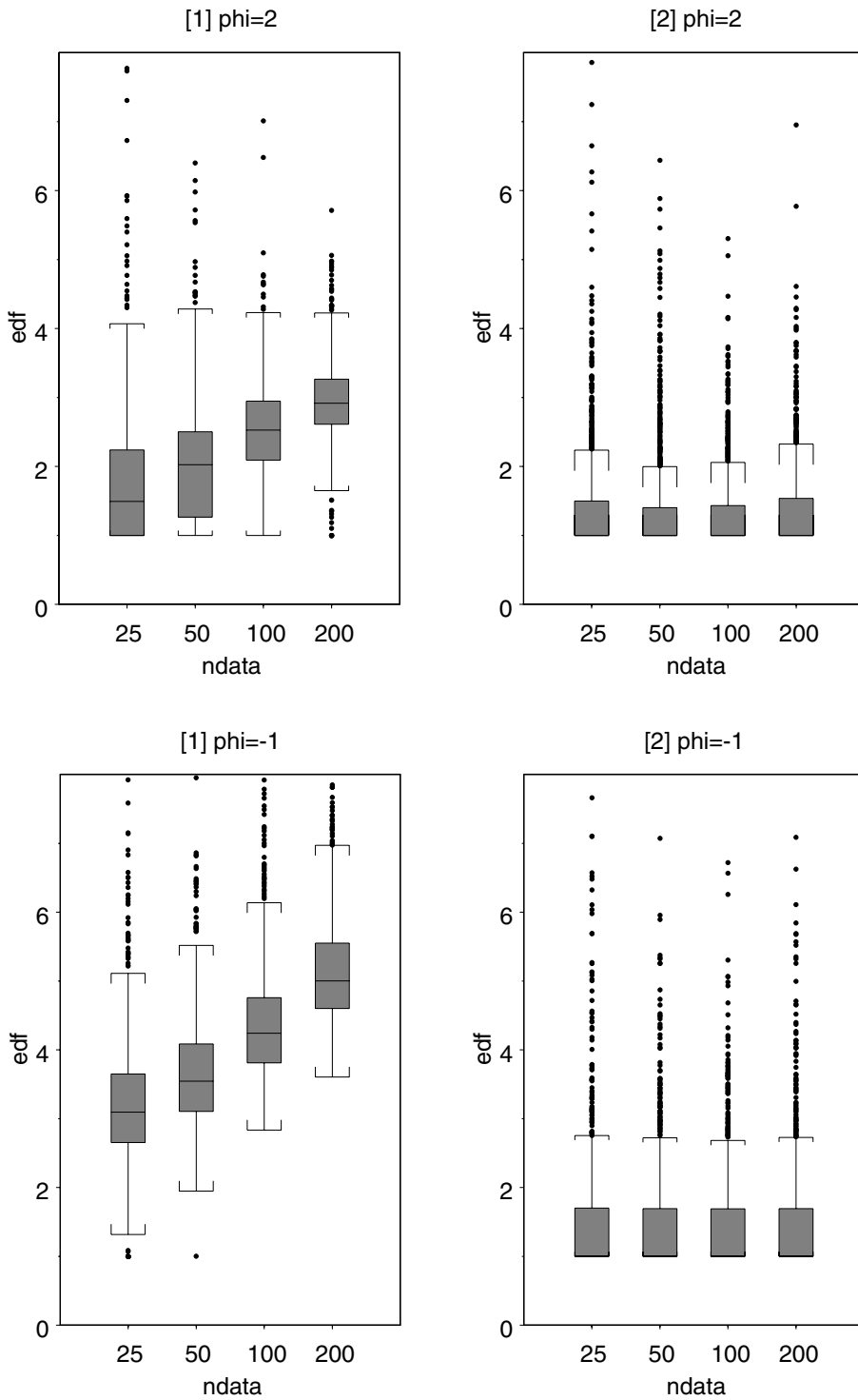


図7. PWSS モデルに対する EDF のボックスプロット. 上が $\phi=2$ の場合, 下が $\phi=-1$ の場合.

場合には PWTLR よりも適応的な推定が可能である。このことは PWSS モデルが(平滑化パラメータが十分大きいときの極限で)PWTLR モデルを含んでいることを考慮すれば自然な結果である。

- 非線形構造の抽出性能

EDF の分布についての考察から、PWSS モデルでは変換前の回帰構造が線形であれば線形関数を、非線形であれば非線形関数を正しく抽出することができ、標本サイズが大きいほどその可能性が高いことがわかる。

5. 結 論

本稿では、加法モデルにおける仮定の一つである分散均一性に着目し、その妥当性を診断するためのベキ重み付き平滑化スプライン(PWSS)を提案した。推定関数を求めるために、回帰方程式や平滑作用素の中に現れる重みを変更しながら後退あてはめアルゴリズムの更新を行う方法を述べた。推定される応答変換や説明変数間の構造に影響を与えるベキ・パラメータや平滑化パラメータの選定には最大周辺尤度法の適用が有用であり、比較的計算が容易な周辺対数尤度の形を導出した。文献事例のデータ集合に対する検討およびシミュレーション実験を通じて、PWSS モデルによって推定されたベキ変換が、非線形構造を考慮に入れながら適応的に分散を均一化しうることを例示した。なお、本稿では、1次元スプラインの和で表される加法モデルのみを取り扱ったが、2次元スプラインを用いるなど、交互作用を伴うモデルにも自然な形での拡張が可能であると考える。

今後の課題を以下に挙げる。まず、PWSS のあてはめを高速に行うために、推定アルゴリズムの改良、とくにベキ・平滑化パラメータを推定するための Fisher 評点化アルゴリズムの開発に取り組むべきであろう。さらに、PWSS モデルを経時対応データに適用するための拡張について検討したい。最後に、本稿では分散均一性の仮定に着目したが、加法モデルにおけるその他の仮定、すなわち加法化および正規化を目標とする変換を取り入れた平滑化スプライン・モデルを開発することが考えられる。

謝 辞 本研究への有益なコメントをくださった、大阪大学大学院基礎工学研究科の白旗慎吾教授、医学統計研究会の後藤昌司先生(元大阪大学教授)に厚くお礼を申し上げたい。また、「応用統計学」特集号への投稿の機会を設けてくださった同誌編集委員会の皆様、丁寧に本稿を査読して下さり有益な助言をくださった査読者の方々にも感謝したい。なお、本研究は文部科学省科学研究費補助金・若手研究(B)「ノンパラメトリック回帰による非線形構造の探索とその実装」(平成14年度～、課題番号14780171)の援助を受けている。

参 考 文 献

- Altman, N.S. (1990): Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.* **85**, 749–759.
 Bartlett, M.S. (1947): The use of transformations. *Biometrics* **3**, 39–52.
 Bowman, A.W. and Azzalini, A. (1997): *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, Oxford.
 Breiman, L. and Friedman, J.H. (1985): Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80**, 580–619.

- Box, G.E.P. and Cox, D.R. (1964): An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211–243.
- Box, G.E.P. and Hill, W.J. (1974): Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16**, 385–389.
- Carroll, R.J. and Ruppert, D. (1987): *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- Chambers, J.M. and Hastie, T.J. (1992): *Statistical Models in S*. Pacific Grove, California. [訳本：柴田里程 (1994): S と統計モデル. 共立出版, 東京.]
- Craven, P. and Wahba, G. (1979): Smoothing noisy data with spline functions. *Numer. Math.* **31**, 377–403.
- Diggle, P.J. and Hutchinson, M.F. (1989): On spline smoothing with autocorrelated errors. *Austral. J. Statist.* **31**, 166–182.
- Eilers, P.H.C. and Marx, B.D. (1996): Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.* **11**, 89–121.
- Eubank, R.L. (1999): *Nonparametric Regression and Spline Smoothing* (2nd Edition). Marcel Dekker, New York.
- Goto, M. (1995): Double power transformations and their performances. *Proceedings of the International Conference on Statistical Methods and Statistical Computing for Quality and Productivity Improvement (Seoul, Korea), Vol. I*, 386–397.
- Green, P.J. (1987): Penalized likelihood for general semi-parametric regression models. *Int. Statist. Rev.* **55**, 245–259.
- Green, P.J. and Silverman, B.W. (1994): *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London.
- Hastie, T. and Tibshirani, R. (1986): Generalized additive models (with discussion). *Statist. Sci.* **1**, 297–318.
- Hastie, T. and Tibshirani, R. (1990): *Generalized Additive Models*. Chapman and Hall, London.
- Linton, O.B., Chen, R., Wang, N. and Härdle, W. (1997): An analysis of transformations for additive nonparametric regression. *J. Amer. Statist. Assoc.* **92**, 1512–1521.
- McCullagh, P. and Nelder, J.A. (1989): *Generalized Linear Models* (2nd edition). Chapman and Hall, London.
- Patterson, H.D. and Thompson, R. (1971): Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.
- Robinson, G.K. (1991): BLUP is a good thing: the estimation of random effects (with discussion). *Statist. Sci.* **6**, 15–51.
- 坂本 亘 (2002): 制限付き最尤推定法による平滑化パラメータの選定：効率的な計算方式とその適用. 計算機統計学 **15**, 19–45.
- Schimek, M.G. and Turlach, B.A. (2000): Additive and generalized additive models. *Smoothing and Regression: Approaches, Computation, and Application* (Schimek, M.G., ed.), Chapter 10 (pp.277–327). John Wiley and Sons, New York.
- Simonoff, J.S. (1996): *Smoothing Methods in Statistics*. Springer-Verlag, New York. [訳本：竹澤邦夫・大森 宏 (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.]
- Speed, T. (1991): Comment on Robinson (1991), *Statist. Sci.* **6**, 42–44.
- Tibshirani, R. (1988): Estimating optimal transformations for regression via additivity and variance stabilization. *J. Amer. Statist. Assoc.* **83**, 394–405.
- Venables, W.N. and Ripley, B.D. (1999): *Modern Applied Statistics with S-Plus* (3rd Edition). Springer-Verlag, New York. [訳本：伊藤幹夫・大津泰介・戸瀬信之・中東雅樹 (2001): S-PLUS による統計解析. シュプリンガー・フェアラーク東京.]
- Wang, Y. (1998): Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.* **93**, 341–348.
- Wang, N. and Ruppert, D. (1995): Nonparametric estimation of the transformation in the transform-both-sides regression model. *J. Amer. Statist. Assoc.* **90**, 522–534.
- Zhang, Z., Lin, X., Raz, J. and Sowers, M. (1998): Semiparametric stochastic mixed models for longitudinal data. *J. Amer. Statist. Assoc.* **93**, 710–719.

補遺：PWSS のアルゴリズムの詳細

A.1. 平滑化スプラインに対する線形混合モデル表現

平滑化スプラインに対する Bayes 流事前分布の表現 (5) についての詳細を補足する。これは周辺対数尤度 (14) の中の行列式 $|C_W|$ を計算する際に必要となる。

関数 f_j ($j = 1, \dots, r$) は q_j 個の基底関数 $\varphi_{jk}(t_j)$ の線形結合 $f_j(t_j) = \sum_{k=1}^{q_j} \xi_{jk} \varphi_{jk}(t_j)$ によって表されるとする。行列 $B_j = \{\varphi_{jk}(t_{ij})\}_{i=1, \dots, n; k=1, \dots, q_j}$ を定義すると, $\mathbf{f}_j = B_j \boldsymbol{\xi}_j$ と書ける。さらに, f_j の乱雑度罰則が, $\boldsymbol{\xi}_j = (\xi_{j1}, \dots, \xi_{jq_j})$ の 2 次形式 $J(f_j) = \boldsymbol{\xi}_j^T K_j \boldsymbol{\xi}_j$ (K_j は (t_{1j}, \dots, t_{n_j}) に依存する非負値定符号対称行列) によって表されるとする。 $J(f_j) = \int \{f_j^{(m)}(t_j)\}^2 dt_j$ に対して, K_j の (k, l) 成分

は $\int \varphi_{jk}^{(m)}(t_j) \varphi_{jl}^{(m)}(t_j) dt_j$ となる。(Eubank (1999) を参照されたい。なお、3 次平滑化スプラインの場合には、基底 B_j の代わりに生起行列を用い、罰則行列 K を Green and Silverman (1994) に定義されているように構築することによっても、以下と同様に論じることが可能である。)

係数 ξ_j に対して、その事前分布を

$$\xi_j = T_j \delta_j + F_j (F_j^T F_j)^{-1} \mathbf{a}_j$$

と書くことができる。ここに、 F_j は罰則行列 K_j の最大階数分解 $K_j = F_j F_j^T$ によって得られる $q_j \times (q_j - m)$ 行列、 T_j は $q_j \times (m - 1)$ 行列で $T_j^T F_j = O$ (零行列) を満たし各列が $(1, \dots, 1)^T$ (q_j ベクトル) および F_j の各列と互いに 1 次独立になるもの、 δ_j は固定パラメータ・ベクトル、 \mathbf{a}_j は平均 $\mathbf{0}$ 、分散 $\lambda_j^{-1} \mathbf{I}$ の確率変数ベクトルである。 $X_{S_j} = B_j T_j$ 、 $Z_{S_j} = B_j F_j (F_j^T F_j)^{-1}$ とおくと (5) の表記を得る。

$X_{S_j} = B_j T_j$ の項は(定数を除いて)罰則 $J(f_j)$ に対する零空間を構成し、 T_j のとり方は上記の条件の下で任意であるから、 X_{S_j} の各行を例えば $(\psi_1(t_{ij}), \dots, \psi_{m-1}(t_{ij}))$ などとすることができる。ここに $\psi_k(t)$ は k 次の多項式である。他方、 \mathbf{a}_j は f_j の乱雑度に関係し、 $J(f_j) = \xi_j^T K \xi_j = \mathbf{a}_j^T \mathbf{a}_j$ と書くことができる。

A.2. PWSS の更新方程式

(9) の罰則付き対数尤度は、補遺 A.1 で述べた表記を用いて、

$$l_p = -\frac{1}{2\sigma^2} (\mathbf{y} - \boldsymbol{\mu})^T W (\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2} \sum_{j=1}^r \lambda_j \xi_j^T K_j \xi_j \quad (\text{A.1})$$

と表すことができる。ここに $W = \text{diag}(w_1, \dots, w_n)$ 、 $\boldsymbol{\mu} = X_L \boldsymbol{\beta} + \mathbf{f}_1 + \dots + \mathbf{f}_r$ である。MPLE $\hat{\boldsymbol{\beta}}$ の更新方程式 (10a) を導出するためには、(A.1) に MPLE $\hat{\mathbf{f}}_k$ ($k = 1, \dots, r$) を代入することで、重み付き平方和 $(\mathbf{r} - X_L \boldsymbol{\beta})^T W (\mathbf{r} - X_L \boldsymbol{\beta})$ の最小化に帰着させればよい。ここに $\mathbf{r} = \mathbf{y} - \sum_{k=1}^r \hat{\mathbf{f}}_k$ である。他方、MPLE $\hat{\mathbf{f}}_j$ の更新方程式 (10b) を導出するためには、(A.1) に $\hat{\boldsymbol{\beta}}$ および $\hat{\mathbf{f}}_k$ ($k \neq j$) を代入し、 $\lambda_j^* = \sigma^2 \lambda_j$ とおくことにより、罰則付き平方和

$$(\mathbf{r}_j - B_j \xi_j)^T W (\mathbf{r}_j - B_j \xi_j) + \lambda_j^* \xi_j^T K_j \xi_j \quad (\text{A.2})$$

の最小化に帰着させればよい。ここに $\mathbf{r}_j = \mathbf{y} - X_L \hat{\boldsymbol{\beta}} - \sum_{k \neq j} \hat{\mathbf{f}}_k$ である。(A.2) を ξ_j に関して微分して 0 とおくことにより、 $\hat{\xi}_j = (B_j^T W B_j + \lambda_j^* K_j)^{-1} B_j^T W \mathbf{r}_j$ が得られることから、 $S_{W_j} = B_j (B_j^T W B_j + \lambda_j^* K_j)^{-1} B_j^T W$ とおくと $\hat{\mathbf{f}}_j = S_{W_j} \mathbf{r}_j$ を得る。なお、一意性の条件 $\sum_{i=1}^n \hat{f}_j(t_{ij}) = 0$ を満たすために、 $\bar{f}_j = n^{-1} \sum_{i=1}^n \hat{f}_j(t_{ij})$ とおき、 $\hat{\mathbf{f}}_j \leftarrow \hat{\mathbf{f}}_j - \bar{f}_j \mathbf{1}$ 、 $\hat{\boldsymbol{\beta}}_0 \leftarrow \hat{\boldsymbol{\beta}}_0 + \bar{f}_j$ と中心化しておく。ここに $\mathbf{1}$ は 1 を要素にもつ n -ベクトルである。

A.3. 周辺対数尤度の導出

線形混合モデル (11) における \mathbf{y} の周辺密度は

$$\begin{aligned}
 p(\mathbf{y}) &\propto \iiint \frac{1}{\sigma^n |W|^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \boldsymbol{\mu})^T W (\mathbf{y} - \boldsymbol{\mu}) \right\} \\
 &\quad \times \prod_{j=1}^r \left\{ \lambda_j^{\frac{q_j-m}{2}} \exp \left(-\frac{1}{2} \lambda_j \mathbf{a}_j^T \mathbf{a}_j \right) \right\} d\boldsymbol{\beta} d\boldsymbol{\delta} d\mathbf{a} \\
 &= \sigma^{-(n+q-rm)} |W|^{-1/2} \prod_{j=1}^r \lambda_j^{\frac{q_j-m}{2}} \\
 &\quad \times \iiint \exp \left[-\frac{1}{2\sigma^2} \left\{ (\mathbf{y} - \boldsymbol{\mu})^T W (\mathbf{y} - \boldsymbol{\mu}) + \mathbf{a}^T \Lambda^* \mathbf{a} \right\} \right] d\boldsymbol{\beta} d\boldsymbol{\delta} d\mathbf{a}
 \end{aligned}$$

($q = \sum_{j=1}^r q_j$) となる. ここに $\boldsymbol{\mu} = X_L \boldsymbol{\beta} + X_S \boldsymbol{\delta} + Z_S \mathbf{a}$ である. BLUP ($\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\mathbf{a}}$) が方程式 (12) の解であることに注意すると,

$$\begin{aligned}
 (\mathbf{y} - \boldsymbol{\mu})^T W (\mathbf{y} - \boldsymbol{\mu}) + \mathbf{a}^T \Lambda^* \mathbf{a} &= \boldsymbol{\mu}^T W \boldsymbol{\mu} - 2\mathbf{y}^T W \boldsymbol{\mu} + \mathbf{y}^T W \mathbf{y} + \mathbf{a}^T \Lambda^* \mathbf{a} \\
 &= [\boldsymbol{\beta}^T, \boldsymbol{\delta}^T, \mathbf{a}^T] C_W \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{a} \end{bmatrix} - 2\mathbf{y}^T W [X_L \ X_S \ Z_S] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{a} \end{bmatrix} + \mathbf{y}^T W \mathbf{y} \\
 &= \left\{ \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{a} \end{bmatrix} - C_W^{-1} \begin{bmatrix} X_L^T \\ X_S^T \\ Z_S^T \end{bmatrix} W \mathbf{y} \right\}^T C_W \left\{ \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \mathbf{a} \end{bmatrix} - C_W^{-1} \begin{bmatrix} X_L^T \\ X_S^T \\ Z_S^T \end{bmatrix} W \mathbf{y} \right\} \\
 &\quad + \mathbf{y}^T \left\{ W - W [X_L \ X_S \ Z_S] C_W^{-1} \begin{bmatrix} X_L^T \\ X_S^T \\ Z_S^T \end{bmatrix} W \right\} \mathbf{y} \\
 &= \begin{bmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \\ \boldsymbol{\delta} - \hat{\boldsymbol{\delta}} \\ \mathbf{a} - \hat{\mathbf{a}} \end{bmatrix}^T C_W \begin{bmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \\ \boldsymbol{\delta} - \hat{\boldsymbol{\delta}} \\ \mathbf{a} - \hat{\mathbf{a}} \end{bmatrix} + \mathbf{y}^T W \{ \mathbf{y} - (X_L \hat{\boldsymbol{\beta}} + X_S \hat{\boldsymbol{\delta}} + Z_S \hat{\mathbf{a}}) \} \\
 \iiint \exp \left\{ -\frac{1}{2\sigma^2} \begin{bmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \\ \boldsymbol{\delta} - \hat{\boldsymbol{\delta}} \\ \mathbf{a} - \hat{\mathbf{a}} \end{bmatrix}^T C_W \begin{bmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \\ \boldsymbol{\delta} - \hat{\boldsymbol{\delta}} \\ \mathbf{a} - \hat{\mathbf{a}} \end{bmatrix} \right\} d\boldsymbol{\beta} d\boldsymbol{\delta} d\mathbf{a} &\propto \sigma^{d+q-rm} |C_W|^{-1/2}
 \end{aligned}$$

($d = 1 + p + r$) であるから,

$$p(\mathbf{y}) \propto \sigma^{-(n-d)} |W|^{-1/2} \prod_{j=1}^r \lambda_j^{\frac{q_j-m}{2}} |C_W|^{-1/2} \exp \left[-\frac{1}{2\sigma^2} \mathbf{y}^T W \{ \mathbf{y} - (X_L \hat{\boldsymbol{\beta}} + X_S \hat{\boldsymbol{\delta}} + Z_S \hat{\mathbf{a}}) \} \right]$$

を得る. $X_S \hat{\boldsymbol{\delta}} + Z_S \hat{\mathbf{a}} = \hat{\mathbf{f}}_1 + \cdots + \hat{\mathbf{f}}_r$ より, 周辺対数尤度は,

$$\begin{aligned}
 l_M(\phi, \lambda^*, \sigma^2 | \mathbf{y}) &= (\text{定数}) - \frac{n-d}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n \log w_i + \sum_{j=1}^r \frac{q_j - m}{2} \log \lambda_j^* - \frac{1}{2} \log |C_W| \\
 &\quad - \frac{1}{2\sigma^2} \mathbf{y}^T W \{ \mathbf{y} - (X_L \hat{\boldsymbol{\beta}} + \hat{\mathbf{f}}_1 + \cdots + \hat{\mathbf{f}}_r) \}
 \end{aligned}$$

となる. ここで, l_M の σ^2 に関する導関数から σ^2 の MML 推定量 (13) が得られ, これを l_M の式に代入して (14) を得る.

(2002 年 12 月 26 日受付 2003 年 9 月 22 日最終修正 12 月 17 日採択)

著者連絡先: 〒 560-8531 大阪府豊中市待兼山町 1-3
大阪大学 大学院基礎工学研究科 数理科学領域
E-mail: sakamoto@sigmath.es.osaka-u.ac.jp

Diagnosing Homoscedasticity with the Power-weighted Smoothing Spline

Wataru Sakamoto

Graduate School of Engineering Science, Osaka University
1–3 Machikaneyama-cho, Toyonaka, Osaka 560–8531, Japan

Abstract

In nonparametric regression models, the requirements of homoscedasticity and so on are implicitly assumed, which leads to poor estimation of regression functions. A power weighted smoothing spline (PWSS) model, whose objective is to diagnose homoscedasticity as well as to estimate unknown nonlinear regression structure, is assumed. The responses in an additive regression model are power-transformed, and then their variances after transformation are assumed to be constant. Smoothing splines are obtained as estimated functions by maximizing the penalized likelihood, and a reweighted version of a backfitting algorithm is constructed. A power-transformation parameter and smoothing parameters, which control smoothness of the functions, are estimated by maximizing the marginal likelihood, based on Bayesian approaches to smoothing splines. A form of the marginal likelihood, which yields comparatively easy computation, is derived using the property that smoothing splines are the best linear unbiased predictor of a linear mixed model. Examination of some data sets from the literature and a simulation experiment show that the power transformation estimated with the PWSS model attains homoscedasticity while taking nonlinear structure into account.

Key words: additive model, maximum marginal likelihood, maximum penalized likelihood, nonparametric regression, transformation of responses.

E-mail address: sakamoto@sigmath.es.osaka-u.ac.jp (Wataru Sakamoto)

Received December 26, 2002; Received in final form September 22, 2003; Accepted December 17, 2003.