4K-1

ABIC を用いたデータのあてはめの二相問題への適用

田中亮平† 村田陸† 坂本真貴人‡ 藤井昭宏† 田中輝雄†

工学院大学情報学部† 工学院大学大学院情報学専攻‡

1. はじめに

実験データには一般に誤差が含まれている．その誤差を分離し，データ本来の構造を取り出すことを考える．その手法の 1 つに，離散点上の値をパラメタとして表現した近似関数をベイズ型情報量基準 ABIC（Akaike's Bayesians information Criterion）[1]を用いて評価し，データのあてはめを行う手法がある[2]．近似関数$\boldsymbol{f}$から評価関数

$\left(\boldsymbol{f}\text{とデータとの距離}\right) + \alpha^2 \times \left(\boldsymbol{f}\text{のなめらかさの強さ}\right)$

を最小とする$\boldsymbol{f}$を最適とする．ここで，近似関数はαを小さくとるほどデータに追従し，大きくとるほどなめらかになる．最適なαを ABIC を用いて求める．

本研究ではこの ABIC を用いたデータのあてはめを二相問題に適用する．二相問題とは複数の構造で構成され，構造間の接点を分岐点と呼ぶ．その分岐点を自動で推定する手法を実現し，実際の二相問題に適用した．

2. ABIC を用いたデータのあてはめ

まず，離散点上の値をパラメタとする近似関数について述べる．

近似関数$\boldsymbol{f}$は離散点$x_j(1 \le j \le n)$上の値f_jで

$$\boldsymbol{f} = \left(f_1, \dots f_j, \dots . f_n\right)$$

と表現する．離散点x_jは区間$[\mathrm{a}, \mathrm{b}]$を$1/(n-1)$等分した点である．

また，N個のデータ$y_i(1 \le i \le N, N < n)$が得られているとき$y_i$を，$\boldsymbol{y} = (y_1, \dots y_i, \dots y_N)$と表す．データはすべて区間$[\mathrm{a}, \mathrm{b}]$上にあるとする．$f$を決定するため，なめらかさとして$f_{i-1} - 2f_i + f_{i+1}$を導入する．この近似関数を d-Spline と呼ぶ[2]．

このとき，d-Spline $\boldsymbol{f}$を評価する値である ABIC を求める式はαの関数になる．ABIC を最小にするαを使って求めた d-Spline $\boldsymbol{f}$が最適とする．

$$\mathrm{ABIC}(\alpha) = \log_e\left|\det(Z_\alpha^t Z_\alpha)\right| - 2(n-2)\log\alpha + (N-2)\log_e\|\boldsymbol{b} - Z_\alpha \boldsymbol{f}\|^2$$

$$Z_\alpha = \begin{bmatrix} E \\ \alpha D \end{bmatrix}, b = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

Eは$\boldsymbol{y}$と$\boldsymbol{f}$の対応を表す関数でx_i上にデータy_iがあるとき，$E_{ij} = 1$とし，あとの要素はすべて 0 となる$N \times n$の行列である．

Dは各点でのなめらかさを表現するため，2 階差分をとる$(n-2) \times n$の帯行列である．

$$E = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, D = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & & \\ & & & 1 & -2 & 1 \end{bmatrix}$$

ABIC を最小とするαは解析的に求めることができないため，区間縮小法である黄金分割法[3]を使って求める．

d-Spline $\boldsymbol{f}$を求めるアルゴリズムは次のようになる．

Step 1 αを定める
Step 2 Z_αを QR 分解する
Step 3 $\boldsymbol{b} = R\boldsymbol{f}$を計算し，d-Spline $\boldsymbol{f}$を求める
Step 4 ABIC を求め，最小ならば終了
Step 5 αを変更して Step 2 に戻る

3. 二相問題への適用

3.1 適用手法

データの構造がある点x_pで 2 つの相に分けられる問題を，ABIC を用いたデータのあてはめに適用させる．ここでは，分岐点x_pの両側を別の相とし，x_pの位置を推定する．

x_pに一番近い d-Spline $\boldsymbol{f}$の点をx_rとする．両側を別の相として ABIC も用いたデータのあてはめを適用するため，行列Dを 3 つの部分D_1(第 1 行～$r-1$行)，D_2(第 r 行)，D_3(第$r+1$行～$n-2$行)に分け，パラメタα, γ, βを対応させ，

$$Z_{\alpha\beta\gamma} = \begin{bmatrix} E \\ \alpha D_1 \\ \gamma D_2 \\ \beta D_3 \end{bmatrix}$$

と Z_αを置き換える．

このとき ABIC を求める式は次のようになる．

$$\mathrm{ABIC}(\alpha, \gamma, \beta, r) = \log_e\left|\det\left(Z_{\alpha\beta\gamma}^t Z_{\alpha\beta\gamma}\right)\right| + (\mathrm{N}-2)\log_e\left\|\boldsymbol{b} - Z_{\alpha\beta\gamma}\boldsymbol{f}\right\|^2 - 2(r-1)\log\alpha - 2\log\gamma - 2(n-2-r)\log\beta$$

次にx_rを推定するには，各 r でα, βを動かして ABIC が最小になる最適な r を求める．γに関して

Application of Data Fitting for Two Phase Regression Using ABIC

Ryohei Tanaka†, Riku Murata†, Makito Sakamoto‡, Akihiro Fujii† and Teruo Tanaka†
† Faculty of Infomatics, Kogakuin University
‡ Graduate School of Informatics, Kogakuin University

はα,βよりも相対的に小さな値をとる．

このとき分岐点として選ばれるx_rは厳密にx_pと一致するとは限らない．nを大きくとることでx_pの推定がより正確になると考えられる．

3.2 適用例

例として区間[0,10]で分岐点がx =5となる関数

$$y(x)=\begin{cases}0.5x+1.5 & (x<5)\\ (x-7)^2 & (5\le x)\end{cases}$$

を用いた．0.25刻みで41個のデータ点y_i，誤差として$[-1.0,1.0]$の一様乱数を加えてデータを作成した．d-Spline $\boldsymbol{f}$は121個の点で表現する．

全体を一相問題として適用した結果を図1に，二相問題として適用した結果を図2に示す．点線が誤差を含まない真の$y(x)$を表し，実線がABICを用いて最適化したd-Spline $\boldsymbol{f}$である．

一相問題として適用した図1は，二次関数が含まれるため，全体がなめらかだと推定され，本来直線の部分も曲線に推定されてしまっている．二相問題として適用させた図2では，一次関数部分を直線に，二次関数部分を曲線に近似し，分岐点も推定できている．

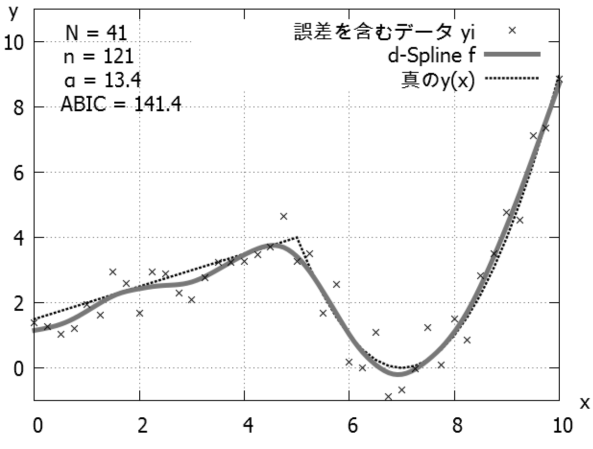

図1 一相問題として適用

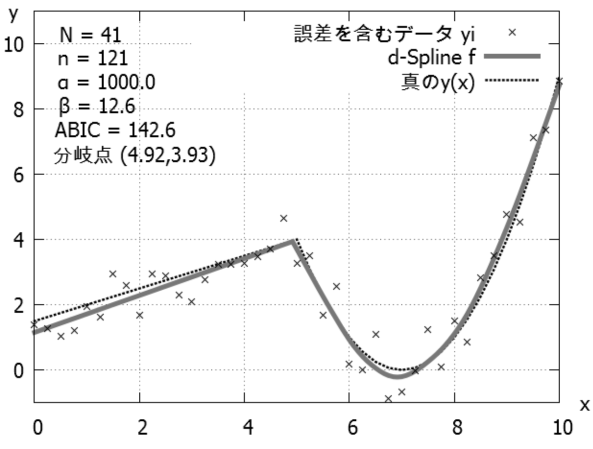

図2 二相問題として適用

4. 実験データへの適用

4.1 使用した実験データ

アクメル・クロメル熱電対で鉛の冷却過程（凝固点を含む）の熱起電力[mV]を測定したものである．熱起電力は温度に変換可能で，鉛の凝固点の熱起電力は13.35[mV]である．

4.2 適用結果

凝固点において温度が一定となるため，二相問題として適用することができる．二相問題として適用した結果を図3に示す．推定した分岐点の起電力と鉛の凝固点の起電力は測定精度の範囲内で一致していると言える．

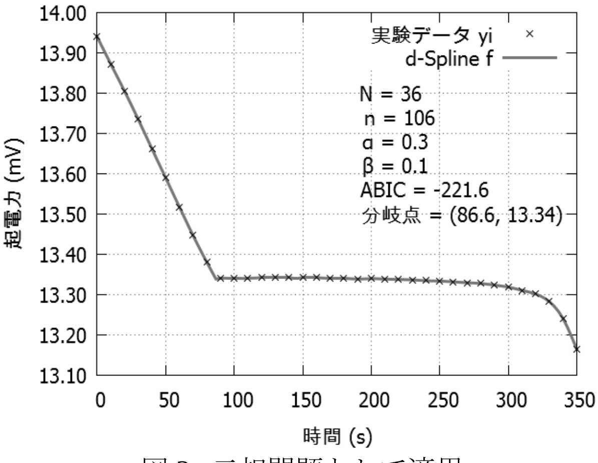

図3 二相問題として適用

5. おわりに

二相問題に対して，ABICを用いたデータのあてはめを適用することができた．分岐点を自動で推定することもできた．また，鉛の凝固点の熱起電力を推定し，期待される値に近い値を出し，実際の問題にも適用できることを示した．

二相問題の場合，最適なd-Spline $\boldsymbol{f}$の探索には，すべてのx_jにおいて，ABICが最小になるα,βを探索する．そのため，探索には多くの時間がかかる．今後の課題として，効率のよい探索方法を見つける必要がある．

謝辞

4章の実験データは，工学院大学基礎・教養教育部門（自然）の渡部隆史教授に頂きました．

参考文献

[1] Akaike, H., Likelihood and Bayes procedure, In Bayesian Statistics, J.M.Bernardo, M.H.DeGroot, D.V.Lindley and A.F.M.Smith, eds, University Press, Valencia, Spain, pp.143-166 1980.
[2] 田中輝雄，田辺國士，ベイズの方法によるデータのあてはめ，数値計算のアルゴリズムの研究 数解研講究録483，no.5，pp.86-111 1983.
[3] 戸川隼人，数値計算技法，オーム社1972.