1G-07

ベイズ型情報量基準 ABIC を用いた不等間隔データのあてはめ

高見澤淳弘† 范谷瑛† 望月大義† 藤井昭宏† 田中輝雄†
工学院大学†

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手法として，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 Bayesian Information Criterion）[1]を用いて評価し，離散点上の値で表現される関数 d-Spline を用いて，データのあてはめを行う手法が提案されている[2].

従来研究では，上記の手法を等間隔の離散点上で表現されるデータに対して適用を行ってきた[3]. しかし，実際に適用する実験データは，等間隔の離散点上に対応できるとは限らない.

そこで本研究では，ABIC を用いて一般的なデータに対してもあてはめができるようにモデル化を行った．本論文で扱う不等間隔モデルにも対応できるように，あてはめ関数 d-Spline を改良する．適用例として，田中[3]の二相問題を扱う．また，物理学実験により得られた不等間隔である実測データの二相問題へのあてはめを示す.

2. ABIC を用いたデータのあてはめ

近似関数 $\boldsymbol{f}$ は離散点 $x_j(1 \le j \le n)$上の値f_jで表す. 離散点x_jはすべてのデータを含む区間[a, b]を$1/(n-1)$等分した点である．そこで，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$\boldsymbol{y} = (y_1, \ldots, y_i, \ldots, y_N)$で表す.

データy_iをあてはめる近似関数$\boldsymbol{f}$は($\boldsymbol{f}$とy_iとの距離)+α^2×($\boldsymbol{f}$のなめらかさの強さ)の評価関数を最小とする$\boldsymbol{f}$とする.

ここで$\boldsymbol{f}$を決定するため，なめらかさとして二階差分である$f_{i-1} - 2f_i + f_{i+1}$を導入する．この近似関数$\boldsymbol{f}$を d-Spline と呼ぶ[2].

d-Spline 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．そこで，ABIC を最小にするαを使って求めた d-Spline を最適とする.

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

このとき，$Z_\alpha = [E \quad \alpha D]^T$，$\boldsymbol{b} = [\boldsymbol{y} \quad 0]^T$である. Z_αの構成要素であるEは$\boldsymbol{y}$と$\boldsymbol{f}$の対応を表す行列である．x_j上にデータy_iがあるとき，$E_{ij} = 1$とし，あとの要素は全て 0 となる$N \times n$の行列である.

また，Dは各点での 2 階差分によりなめらかさを表す$(n-2) \times n$の帯行列である.

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，区間短縮法である黄金分割法[4]を使って求める．全体のアルゴリズムを図 1 に示す.

Step 1　αの初期値を決める
Step 2　Z_αを QR 分解する
Step 3　$\boldsymbol{b} = R\boldsymbol{f}$を計算し，d-Spline を求める
Step 4　ABIC を求め，最小ならば終了
Step 5　αを更新して Step 2 にもどる

図 1　d-Spline を求める手順

3. 不等間隔データ対応へのモデルの拡張

いままで不等間隔データにあてはめを行う場合は，すべてのデータを d-Spline を表現する離散点上にくるように補正するか，離散点を非常に多く取る必要があった[5].

本研究では，図 2 のように不等間隔データ間の距離を使ってなるべく等間隔に離散点を作成するよう d-Spline を改良した.

まず任意の隣り合う 2 つのデータの最小区間$minh$を求め，その間の離散点k個を指定する．その他の隣り合う 2 つのデータの区間は，$minh$の整数倍に四捨五入で丸め，その間の離散点を先ほどの整数倍×k個で指定する.

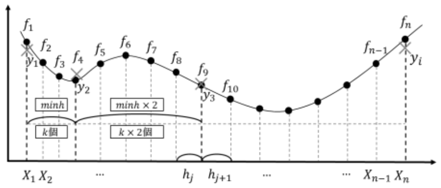

図 2　近似関数fと長さhの関係

このとき，2 章で説明した二階差分による滑らかさを表すDを以下のように変形する.

Fitting Unequally Spaced Data
Using Akaike's Bayesian Type Information Criterion, ABIC
Atsuhiro Takamizawa†, Guuing Fan†, Masayoshi Mochizuki†, Akihiro Fujii† and Teruo Tanaka†
†Kogakuin University

$$D=\begin{bmatrix} h_{j+1} & -(h_j+h_{j+1}) & h_j & & & \\ & & & \ddots & & \\ & & & h_{j+1} & -(h_j+h_{j+1}) & h_j \end{bmatrix}$$

$$\frac{1}{h}\left(h_j f_{j-1}-\left(h_j+h_{j+1}\right)f_j+h_{j+1}f_{j+1}\right)$$

ここで，$h=h_j h_{j+1}(h_j+h_{j+1})$である．

Dの変更に伴い ABIC の評価式も一般化するため，以下の式を追加する．また，4 章で説明する二相問題に適用する場合も同様に対応することで一般化が可能となる．

$$ABIC(\,)=\cdots-2\sum_{j=2}^{h-1}log h_j$$

4. 二相問題あてはめへの適用

4.1 従来研究

田中[3]は，二相問題に適用するモデルを作成した．本論文では，そこに 3 章のモデル拡張を行う．二相の分岐点をrとし，rの右側と左側でなめらかさを変える．したがって，行列Dを 3 つの行列D_1(第 1 行～第$r-1$行)，D_2(第r行)，D_3(第$r+1$行～第n行)に分ける．

rを 1 からnまで変化させて ABIC を最小とするrを求め，分岐点を決める．各々のなめらかさのパラメタをα，β，γとし，Z_αと対応させると$Z_{\alpha\beta\gamma}$は次のようになる．

$$Z_{\alpha\beta\gamma}=[E \quad \alpha D_1 \quad \gamma D_2 \quad \beta D_3]^T$$

このときの ABIC の評価式を以下に示す．

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

4.2 モデルデータへのあてはめ

モデル拡張を行い，不等間隔データの二相問題へあてはめを行った結果を図 3 に示す．

適用例として区間[0,10]で以下の関数を用いた．

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

この関数から，不等間隔刻みにランダムで 32 個データ点y_iを取り，誤差として$[-1.0,1.0]$の一様乱数を加えた．×印がデータy_iである．点線が誤差を含まない真の$\boldsymbol{y}(x)$を表し，実線があてはめを行い，ABIC を用いて最適化した d-Spline である．適用することができた．

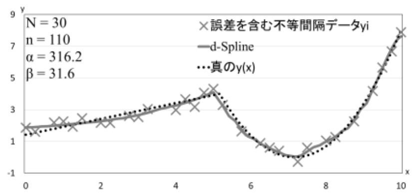

図 3　モデルデータへのあてはめ

5. 実験データへのあてはめ

5.1 使用した実験データ

マイクロ波アシスト磁気記録方式における磁界発生層の負の磁気異方性定数を変化させたときに生じる周波数についての実験データを，不等間隔である実測データの二相問題へのあてはめとして使用した[6].

5.2 結果

結果を図 4 に示す．×印がデータy_iであり，実線が ABIC を用いて最適化した d-Spline である．実験データを追従することができた．しかし，適用例と比較すると-5.0E+05J/m^3から上位はデータに追従しすぎる結果となった．

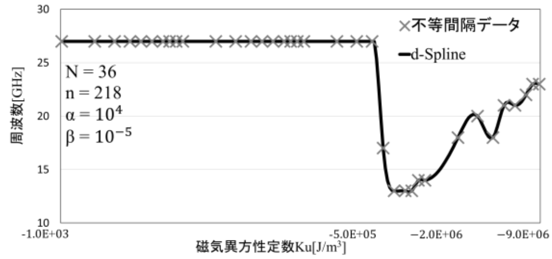

図 4　実験データへのあてはめ

6. おわりに

ABIC を用いて一般的なデータに対してもあてはめができるようにモデル化を行った．また二相問題あてはめに対して適用することができた．しかし，二相問題として扱うと，データによって挙動が異なってくる場合がある．今後の課題は，より自由度を高めるために，データごとの間隔や全体の大まかな概形から，離散点数を調整する機構などのモデル拡張が求められる．

謝辞　5 章の実験データは工学院大学情報通信工学科情報ファイル研究室 赤城文子教授にいただいた．

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] 田中亮平，村田陸，坂本真貴人，藤井昭宏，田中輝雄，ABIC を用いたデータのあてはめの二相問題への適用，情報処理学会第 76 回全国大会 pp. 203-204(2014).

[4] 戸川隼人，数値計算技法，オーム社 (1972).

[5] 川口育子，欠損データを含む二相問題に対するデータのあてはめ，工学院大学情報学部コンピュータ科学科，卒業論文 (2015).

[6] 古賀理樹，田端涼，赤城文子，吉田和悦，マイクロ波アシスト磁気記録方式における高周波磁界発生素子の負の磁気異方性が発振磁界に及ぼす影響，電子情報通信学会総合大会，pp. 23 (2015).