

直接探索法を用いた赤池情報量基準の最適化

山口拓未

摘要

いくつかのパラメータを含む数理モデルを考えると、現実のデータから適切なパラメータを推定することが重要となる。その際に、選ばれたパラメータの良さを測る基準の1つとして赤池情報量基準 (Akaike's Information Criterion; AIC) がある。AIC が最小となるパラメータを選べば、モデルの簡単さと、データとの適合性のバランスの取れたモデルが選択できると考えられている。

AIC は選択されるパラメータの数と最大対数尤度で定義されている。このパラメータ数を連続関数で表すことができないため、AIC を最小とするパラメータを求める問題は混合整数非線形計画問題として定式化される。しかし、選択可能なパラメータ数が多いときにはその問題を解く計算コストが膨大になる。そこで、そのような大規模な問題ではパラメータ数をある連続関数で近似した非線形計画問題を考え、その問題を解くことによって近似解を得ることが提案されている。しかし、扱う問題によっては近似精度が悪くなり、望ましいパラメータが求まらないことがある。

本報告書では、上記のパラメータ数を近似した関数に係数を加えることによって、より近似精度の高い非線形計画問題を考える。このとき、ある固定された係数に対して近似した問題を解くことによって得られる AIC はその係数の関数とみなすことができる。しかし、その関数は数式で陽に表せないため、通常非線形計画問題に対する解法では、その関数の最小化ができない。そこで、直接探索法を用いて AIC が最小となるような係数を求めることを提案する。また、多項式モデルに関して提案手法の数値実験を行い、その有効性を検証した。

目次

1	序論	1
2	赤池情報量基準と直接探索法	3
2.1	多項式モデルの最大対数尤度	3
2.2	赤池情報量基準	4
2.3	直接探索法	5
3	赤池情報量基準の最小化	6
3.1	近似を用いた従来手法	6
3.2	直接探索法を用いた赤池情報量基準の最小化	7
4	数値実験	8
4.1	数値実験結果	9
4.2	実験結果に対する考察	14
5	結論	15

1 序論

工学や社会など様々な分野において、大量に蓄積されたデータから本質的な情報を抽出し、予測・制御を行うことは必要不可欠である。その際にはデータから構築される数理モデルが重要な役割を果たす。良いモデルが得られれば、そのモデル上で精度の高い分析ができる。したがって、現実の複雑な問題を解くためには、まずより良いモデルを構築することが必要である。本報告書では、数理モデルとして入力 $x \in \mathbb{R}$ に対し出力 $y \in \mathbb{R}$ を与える関数を考える。つまり関数 $y = f(x; \alpha_0, \alpha_1, \dots, \alpha_m)$ を考える。以下では、関数 $f(x; \alpha_0, \dots, \alpha_m)$ をモデル関数と呼ぶ。ここで、 α_j ($j = 0, \dots, m$) はモデル関数を定めるパラメータであり、 $\alpha = (\alpha_0, \dots, \alpha_m)^T$ とする。このときモデル関数は $f(x; \alpha)$ と表せる。

はじめに、良いモデルの基準を与える。まず、現実との誤差が小さいほど良いモデルであるということができる。現実の入出力のデータ (x_i, y_i) ($i = 1, \dots, n$) が与えられたとき、そのような誤差として 2 乗和

$$\sum_{i=1}^n |f(x_i; \alpha) - y_i|^2$$

がよく用いられる。この 2 乗和を最小とするようなパラメータ α_j を選べば、 $f(x; \alpha)$ は良いモデルとなることが期待できる。このように 2 乗和を最小にするモデルでは、一般にパラメータ数 m を増やせば増やすほど誤差を小さくできる。実際、モデル関数が多項式 $f(x; \alpha) = \sum_{j=0}^m \alpha_j x^j$ で与えられているとき、 m が大きければ任意の関数を表すことができる。そのため $m > n$ であれば、すべてのデータに対して $f(x_i; \alpha) = y_i$ とすることができる。しかし、データにノイズが存在するとき、実際の測定対象とは関係の無いノイズにも適合させてしまうため、用いたデータ以外には役に立たないモデル関数となってしまうことがある（過学習が起きる）。そのような観点から、より簡単なモデル関数、つまりパラメータ数をできるだけ少なくするように選ぶことも重要である。以上をまとめると、本報告書で考えるより良いモデルとは現実のデータとの誤差が小さく、パラメータ数の少ないモデル（シンプルモデル）である。

そのようなモデルを選択する基準として様々なものがあるが、本報告書では赤池情報量基準 (Akaike's Information Criterion; AIC) を考える。これは赤池 [1],[2] が 1971 年に考案し、1973 年に発表したもので、

$$\text{AIC} = -2 \ln L + 2 \text{card}(\hat{\alpha})$$

と表される。ここで、 L は与えられたデータから得られるモデルの最大尤度、 $\text{card}(\hat{\alpha})$ は $\alpha_j \neq 0$ の数、つまり自由パラメータの数を表す。ただし、 $\hat{\alpha}$ は最大尤度を与えるモデルのパラメータである。ここで、 $-2 \ln L$ がモデルとの誤差を表し、 $\text{card}(\hat{\alpha})$ がモデルのシンプルさを表していることに注意する。そのため、上記のように赤池情報量基準が最小となるモデルを選べば、多くの場合それが良いモデルと考えることができる。

本報告書では、与えられたデータから赤池情報量基準を最小とするパラメータ α を求める手法を提案する。モデル関数のパラメータ数が少なければ、モデルに使用しないパラメータの組み合わせ

をすべて選び, 残りの α に対して対数尤度を最大化することによって赤池情報量基準が最小の $\hat{\alpha}$ を求めることができる. また, 混合整数非線形計画問題に定式化して, その解法を用いて効率よくしらみつぶしに調べる手法もある [4]. 混合整数非線形計画問題とは, 変数として連続変数と離散変数が存在し, 制約条件や目的関数の中に非線形なものを含む数理計画問題である. 赤池情報量基準の最小化を混合整数非線形計画問題に定式化すると, 以下のようになる.

$$\begin{aligned} \min \quad & -2l(\alpha) + 2 \sum_{j=0}^m \delta_j \\ \text{s.t.} \quad & \alpha_j - \alpha_{j,\max} \cdot \delta_j \leq 0 \\ & -\alpha_j + \alpha_{j,\min} \cdot \delta_j \leq 0 \\ & \delta_j \in \{0, 1\}, \quad j = 0, \dots, m \end{aligned}$$

ただし, m はモデル関数に含まれるパラメータの数であり, $\alpha_{j,\max}$, $\alpha_{j,\min}$ はそれぞれパラメータ α_j の上限, 下限を表す. また, l は対数尤度関数を表す. $\delta_j = 1$ のとき α_j を使用し, $\delta_j = 0$ のとき α_j を使用しないことを意味している. さらに, 自由パラメータの数 $\text{card}(\alpha)$ を離散変数 $\delta_j \in \{0, 1\}$ を用いて $\sum_{j=0}^m \delta_j$ と表している. $\delta_j = 0$ のときは制約条件より $\alpha_j = 0$ となり, $\delta_j = 1$ のときは制約条件より, $\alpha_j \neq 0$ となるので, 赤池情報量基準の最小化と等価な問題であることがわかる. 混合整数非線形計画問題は離散変数をもつため簡単に解くことはできない. さらに, その計算時間は問題の規模が大きくなると爆発的に大きくなるため, 大規模な問題には向いていない. 大規模な問題においては, $\text{card}(\alpha)$ を $\sum_{j=0}^m |\alpha_j|$ で近似した問題を解くことによって α を求める方法が提案されている. こうすることにより一般の非線形計画問題となるため, 大規模な問題でも解くことができる. しかし, 扱う問題によっては近似精度が悪くなり, 良い α が求まらない可能性がある.

本報告書では, 正のパラメータ z を用いて $\text{card}(\alpha)$ を $z \sum_{j=0}^m |\alpha_j|$ と近似することを考える. 適切な z を選べば, 従来の方法 ($z = 1$) の場合に比べて良い α が求まるはずである. しかし, 前もって適切な z を見つけることは容易ではない. そこで, 赤池情報量基準を z の関数として表すことを考える. その関数を最小化することによって, 最適な z とパラメータ α を求める. しかし, そのような関数は陽に表すことができない. そこで, 関数の形が特定されていなくても用いることができる直接探索法を用いて最適化を行うことを提案する. 直接探索法は, 微分不可能な関数の最小化に有効な手法である.

本報告書の構成は以下のとおりである. まず第 2 節では推定モデルの良さを評価するための赤池情報量基準の定義を述べ, さらに最適化を行う際に使用する直接探索法を紹介する. 第 3 節では, まず $\text{card}(\alpha)$ を凸関数で近似する手法を述べる. 次に, その手法に対してパラメータ z を導入した赤池情報量基準の最適化問題を提案する. 第 4 節では, 具体的に与えられたデータに関して数値実験を行い, その結果と考察を述べる. 最後に第 5 節で結論を述べる.

2 赤池情報量基準と直接探索法

本節では、まず多項式回帰モデルに対する最大対数尤度を導出する。次に赤池情報量基準の定義を与える。最後に本報告書で用いる直接探索法を紹介する。

2.1 多項式モデルの最大対数尤度

本報告書ではモデル関数として多項式回帰モデル

$$f(x; \alpha) = \sum_{j=0}^m \alpha_j x^j$$

を考える。そこで、この節では多項式モデルに対する最大対数尤度 $\ln L$ を導出する。以下では、 n 個の入出力データ (x_i, y_i) ($i = 1, \dots, n$) が得られているとする。さらに、実際に得られた y_i と真の値との誤差は平均 0、分散 σ^2 の正規分布 q_i に従うと仮定する。このとき、 $f(x_i; \alpha)$ を真の値とし、実際のデータ y_i との誤差を ε_i とすると、誤差が正規分布に従うことから、

$$\begin{aligned} \log q_i &= \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{\varepsilon_i^2}{2\sigma^2} \right) \right\} \\ &= \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2}{2\sigma^2} \right\} \right] \\ &= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2 \end{aligned}$$

と表せる。よって対数尤度 l は、 α と σ^2 の関数として

$$\begin{aligned} l(\alpha, \sigma^2) &= \sum_{i=1}^n \log q_i \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2 \end{aligned} \tag{2.1}$$

となる。

誤差の分散 σ^2 が既知のとき、最大対数尤度は (2.1) より、

$$\begin{aligned} \ln L &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 \\ &\quad - \min_{\alpha} \left(\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2 \right) \end{aligned} \tag{2.2}$$

と表せる。これは α の凸 2 次関数を最小化することによって得られる。

次に誤差の分散 σ^2 が未知のときを考える。ここで、 α が与えられたとき $l(\alpha; \sigma^2)$ を最大化する $\hat{\sigma}^2$ を σ^2 の代用とすることにする。 $l(\alpha; \sigma^2)$ を最大化する $\hat{\sigma}^2$ は、

$$\frac{\partial l(\alpha; \hat{\sigma}^2)}{\partial \sigma^2} = 0 \quad (2.3)$$

の解である。これは

$$-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2 = 0$$

となるから、 $\hat{\sigma}^2$ は

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2$$

である。以上より、最大対数尤度は

$$\begin{aligned} \ln L &= \max_{\alpha, \sigma^2} l(\alpha, \sigma^2) \\ &= \max_{\alpha} \left(-\frac{n}{2} \log 2\pi - \frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2} \right) \\ &= -\frac{n}{2} (\log 2\pi + 1) \\ &\quad - \frac{n}{2} \min_{\alpha} \log \left(\frac{1}{n} \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i - \dots - \alpha_m x_i^m)^2 \right) \end{aligned} \quad (2.4)$$

となる。

2.2 赤池情報量基準

赤池情報量基準 [1],[2],[6] とは、有効なモデル選択基準の代表的なものであり、赤池情報量基準が最小であるモデルが多くの場合良いモデルとされている。赤池情報量基準は次のように表される。

$$\text{AIC} = -2 \ln L + 2 \text{card}(\hat{\alpha}) \quad (2.5)$$

ここで、 L は最大尤度、 $\text{card}(\hat{\alpha})$ は自由パラメータの数を表す。また、 $\hat{\alpha}$ は最大尤度を与えるモデルのパラメータである。(2.5) の第 1 項は推定モデルと実際のデータとの適合度を表しており、誤差が小さいほどその値は小さくなる。第 2 項はパラメータの増加に関するペナルティと考えられる。第 1 項を小さくしようとする、つまり誤差を小さくしようするとパラメータの数が増加し、第 2 項の値が大きくなる。またその逆も同様である。赤池情報量基準を最小化するモデルは、モデルの簡単さとデータとの適合性のバランスがとれているモデルであるといえる。

ここで、 $J = \{j \mid \alpha_j = 0\}$ とする。集合 J が定まっているとき、 $\text{card}(\hat{\alpha})$ は定数となる。その定数を K とすると、赤池情報量基準は

$$\text{AIC}(J) = \min_{\alpha_j=0, j \in J} -2l(\alpha) + 2K$$

となる. 集合 J が固定されていないときは, $AIC(J)$ を最小化する問題は

$$\min_{\alpha} -2l(\alpha) + 2\text{card}(\alpha)$$

とかける. この問題を解くことが本報告書の目的である.

2.3 直接探索法

目的関数 f と実行定義域 Ω が与えられた最適化問題

$$\begin{aligned} \min \quad & f(z) \\ \text{s.t.} \quad & z \in \Omega \end{aligned} \tag{2.6}$$

を考える. 直接探索法は, 問題 (2.6) に適用できる手法である. 直接探索法はいくつかの点の候補から目的関数の値が最も小さくなる点を選択し, 更新していく手法である. 直接探索法は, 目的関数値が計算できれば用いることができ, ニュートン法や準ニュートン法 [7] とは異なり, 目的関数の勾配やヘッセ行列を利用しない. そのため, 目的関数が陽に表せない場合や, 微分情報が得られない場合に有効である. ここでは, 直接探索法の 1 つである Mesh Adaptive Direct Search(MADS)[3] を紹介する.

MADS はそれぞれの反復で, 有限個の候補点を生成して, その候補点の関数値と現在点での関数値を比較する. MADS は 2 つのステップ, ”探索ステップ” と ”収束ステップ” があり, そのステップを通して最適解を求める. なお, このアルゴリズムを説明する際に以下の記号を用いる.

$$\begin{aligned} z_0 \in \Omega &: \text{初期点} \\ \Delta_k \in \mathbb{R}_+ &: k \text{ 回目の反復でのメッシュサイズパラメータ} \\ D = [d_1, \dots, d_{n_D}] \in \mathbb{R}^{n \times n_D} &: \text{探索する方向 } d_j (j = 1, \dots, n_D) \text{ を表す行列} \end{aligned}$$

MADS では候補点はすべて次で定義されるメッシュ上から選ばれる.

定義 1 k 回目の反復におけるメッシュ M_k は次のように定義される.

$$M_k = \bigcup_{z \in S_k} \{z + \Delta_k D \mathbf{u} : \mathbf{u} \in \mathbb{N}^{n_D}\}$$

ただし, S_k は k 回目の反復の前に目的関数 f を評価した点集合を表す.

また, 評価関数として次の f_{Ω} を考える.

$$f_{\Omega}(z) = \begin{cases} f(z) & z \in \Omega \\ +\infty & \text{otherwise} \end{cases}$$

探索ステップでは, 有限個のメッシュ M_k 上の点における f_Ω の値を計算する. もし現在点よりも良い点が得られた場合は, 反復を直ちにやめるか, さらに良い点を期待して探索ステップを続ける. いずれの場合も, $f_\Omega(z_{k+1}) < f_\Omega(z_k)$ を満たす点 $z_{k+1} \in \Omega$ を次の反復の初期点とし, メッシュサイズパラメータを $\Delta_k \leq \Delta_{k+1}$ となるように更新する.

収束ステップは探索ステップで良い点が見つからなかった場合に行うステップである. もし k 回目の反復で良い点が見つからなければ, $f_\Omega(z_{k+1}) = f_\Omega(z_k)$ を満たす $z_{k+1} \in S_{k+1}$ を $k+1$ 回目の反復の初期点とし, さらにメッシュサイズパラメータを $\Delta_k > \Delta_{k+1}$ となるように更新する. 良い点が見つければ, 探索ステップと同様に更新する.

MADS のアルゴリズムは以下のように記述される.

MADS algorithm

- Step 1. 初期点の設定 : $z_0 \in \Omega, \Delta_k$ を設定する. また $k = 0$ とする.
- Step 2. 探索ステップを行い, 良い点が見つければ Step 4 に, 見つからなければ Step 3 へ.
- Step 3. 収束ステップを行う.
- Step 4. Δ_{k+1} を更新し, $k \leftarrow k + 1$ として Step 2 へ.

目的関数 f や探索方向 D が適当な仮定を満たすとき, このアルゴリズムは大域的収束することが示されている [3].

3 赤池情報量基準の最小化

この節では, 赤池情報量基準の最小化問題に対する近似解法を提案する. まずその基礎となる $\text{card}(\alpha)$ の近似関数を用いた手法を紹介する. そして, その近似問題にパラメータ z を導入した問題を考えて, z に関しての最小化に直接探索法を用いる手法を提案する.

3.1 近似を用いた従来手法

2.2 節で述べたように赤池情報量基準の最小化は次のように表せる.

$$\min_{\alpha} -2l(\alpha) + 2\text{card}(\alpha) \quad (3.1)$$

ただし, $\text{card}(\alpha) = \{\alpha_j \neq 0 \text{ である数}\}$ である. ここで, (3.1) を次のように近似する.

$$\min_{\alpha} -2l(\alpha) + 2 \sum_{j=0}^m |\alpha_j| \quad (3.2)$$

$|\alpha_j|$ は $\alpha_j \neq 0$ に対する L_1 ペナルティと考えることができる. 絶対値関数の性質より, 最小解では $\alpha_j = 0$ となる j が多くなることが知られている. 一般に問題 (3.2) を解いて得られた解 $\hat{\alpha}^*$ は, $-2l(\alpha) + 2\text{card}(\alpha)$ の局所的な最小解ではない. そこで以下のようにして近似解 α^* を求めることが

提案されている.

— 近似を用いた手法 —

- Step 1. 問題 (3.2) を解き, 最小解 $\hat{\alpha}^*$ を得る.
- Step 2. 集合 $J = \{j \mid \hat{\alpha}_j^* = 0\}$ を定める.
- Step 3. $\min -l(\alpha)$, s.t. $\alpha_j = 0, j \in J$ を解き, 解 α^* を得る.
- Step 4. 得られた α^* を近似解とし, 赤池情報量基準を $-2l(\alpha^*) + 2\text{card}(\alpha^*)$ とする.

3.2 直接探索法を用いた赤池情報量基準の最小化

前節では, $\text{card}(\alpha)$ を $\sum_{j=0}^m |\alpha_j|$ で近似した. ペナルティ関数の性質より $\alpha_j = 0$ となることが多いことから, 赤池情報量基準が小さくなることが期待できる. しかし, データによっては $\sum_{j=0}^m |\alpha_j|$ は $\text{card}(\alpha)$ の良い近似とならないことがある.

そこで, 問題 (3.2) に適当なパラメータ z を加えた問題を考え, これを $P(z)$ とする. $P(z)$ の例としては, $z \in \mathbb{R}$ のときは

$$\min_{\alpha} -2l(\alpha) + 2z \sum_{j=0}^m |\alpha_j| \quad (3.3)$$

が考えられる. また, $z \in \mathbb{R}^{m+1}$ として,

$$\min_{\alpha} -2l(\alpha) + 2 \sum_{j=0}^m z_j |\alpha_j| \quad (3.4)$$

とする問題が挙げられる. 他にも z を含んだ制約条件を加えた問題も考えることができる.

前節の既存の手法では, 問題 (3.3) に対して $z = 1$ を固定していると考えられる. この z をうまく選ぶことによって, より赤池情報量基準の小さな α を求めることができると考えられる.

z によって得られる近似解 α^* は変化するため, z に応じて赤池情報量基準の値も変わると考えられる. z が与えられたとき, 3.1 節と同様にして得られる赤池情報量基準の値を $g(z)$ と表すことにする. $g(z)$ の具体的な求め方は以下のとおりである.

— $g(z)$ の求め方 —

- 入力 : z
- 出力 : $g(z)$
- Step 1. 問題 $P(z)$ を解き, 最小解 $\hat{\alpha}^*$ を得る.
- Step 2. 集合 $J = \{j \mid \hat{\alpha}_j^* = 0\}$ を定める.
- Step 3. $\min -l(\alpha)$, s.t. $\alpha_j = 0, j \in J$ を解き, 解 α^* を得る.
- Step 4. 得られた α^* を近似解とし, 赤池情報量基準 $-2l(\alpha^*) + 2\text{card}(\alpha^*)$ を計算し出力する.

関数 $g(z)$ を用いると、赤池情報量基準を最小化する問題は次のようにかかる。

$$\begin{aligned} \min \quad & g(z) \\ \text{s.t.} \quad & z \in Z \end{aligned} \tag{3.5}$$

ただし、 Z はパラメータ z が満たすべき条件集合である。問題 $P(z)$ が (3.4) で与えられているとき、問題 (3.5) は本質的に赤池情報量基準の最小化と同じになる。しかし $g(z)$ は一般的には凸関数ではない。また、関数 g は陽には数式で表せず、また微分もできない。そこで、2.3 節で紹介した直接探索法を用いて最適化を行うことを提案する。

$g(z)$ に対してそのまま直接探索法を適用してもうまく働かない可能性がある。Step 4 で求まる α^* は集合 J によって決まる。そのため、異なる z, \hat{z} を用いて $g(z)$ と $g(\hat{z})$ を計算したとしても、Step 2 で求まる J が同じであれば $g(z) = g(\hat{z})$ となってしまう。その場合、ある z の近傍では g の値は一定となる状況が考えられる。そのような z を初期状態に選んで、直接探索法を適用すると、 z のまま終了し、よい解が得られない。

この問題を解消するために、以下のような関数 $\hat{g}(z)$ を考える。

————— $\hat{g}(z)$ の求め方 —————

入力 : z

出力 : $\hat{g}(z)$

Step 1. 問題 $P(z)$ を解き、最小解 $\hat{\alpha}^*$ を得る。

Step 2. 得られた $\hat{\alpha}^*$ における $-2l(\hat{\alpha}^*) + 2\text{card}(\hat{\alpha}^*)$ を計算し出力する。

$\hat{g}(z)$ の返す値は、最大対数尤度を計算していないため、赤池情報量基準の近似値になっている。一方、 $\hat{g}(z)$ は z によって異なる値をとると考えられるため、直接探索法がうまく働くことが期待できる。そこで、直接探索法によって

$$\begin{aligned} \min \quad & \hat{g}(z) \\ \text{s.t.} \quad & z \in Z \end{aligned} \tag{3.6}$$

を解き、その解 $\bar{\alpha}$ から集合 J を決定して、 $-l(\bar{\alpha})$ を最小化して近似解を求めることを考える。

4 数値実験

本節では、前節で提案した手法を用いて赤池情報量基準の最小化を行った数値実験結果を報告する。

モデル関数の最大次数 m を 20 とした多項式モデルにおいて数値実験を行った。また、問題 $P(z)$ としては、 $z \in \mathbb{R}$ の例 (3.3) を用いた。入出力データ (x_i, y_i) は以下の多項式関数を元に生成した。

$$\text{関数 1 : } y = 0.5x^{20} - 0.3x^{10} - 0.2x^2$$

$$\text{関数 2 : } y = 5x^{20} - 3x^{10} - 2x^2$$

$$\text{関数 3 : } y = 50x^{20} - 30x^{10} - 20x^2$$

関数 2, 3 は関数 1 をそれぞれ 10 倍, 100 倍したものである. 入力データは区間 $(-1,1)$ において等間隔となるように選び, 出力は元となる関数の値に平均 0, 分散 σ^2 の正規分布に従う誤差を加えたものを用いた. 入出力データ数 n は 30, 50, 100 の 3 通りとし, 関数 1, 2, 3 の値に加える誤差の分散として以下の場合をそれぞれ考えた.

$$\text{関数 1 : } \sigma^2 = 0.000025, 0.0001, 0.0004$$

$$\text{関数 2 : } \sigma^2 = 0.0025, 0.01, 0.04$$

$$\text{関数 3 : } \sigma^2 = 0.25, 1, 4$$

また, パラメータ z の条件集合は $Z = \{z \mid z \geq 0\}$ とした.

以上のような問題において, 分散 σ^2 が既知として解く場合 (2.2) と未知として解く場合 (2.4) で実験を行った. また, 関数 g と \hat{g} それぞれにおいて最適化を行った. さらに, 比較を行うために従来手法 ($z = 1$) を用いた場合の実験も行った.

本数値実験は CPU が Intel(R)Core(TM)2Duo3.00GHz, メモリが 3.9GB の計算機上で行った. 一般に (3.3) のような絶対値を含む最小化問題は, パラメータの絶対値を別の変数に置き換えて, それに伴う制約条件を加えた問題に変換して解く. 例えば (3.3) は次のように変換する.

$$\begin{aligned} \min_{\alpha, \beta} \quad & -2l(\alpha) + 2z \sum_{j=0}^m \beta_j \\ \text{s.t.} \quad & \alpha_j \leq \beta_j \\ & -\alpha_j \leq \beta_j, \quad j = 0, \dots, m \end{aligned} \quad (4.1)$$

問題 (4.1) は MATLAB のソルバ quadprog, fmincon を用いて解いた. また, z に関する最適化問題 (3.5), (3.6) は MADS を実装した NOMADm [5] を用いて解いた. NOMADm のパラメータはメッシュサイズパラメータの縮小率を $0.7(k$ 回目の反復でよい点が見つからなかったとき, $\Delta_{k+1} = 0.7\Delta_k$ と更新する) とし, ほかのパラメータは初期設定のものを用いた. また, NOMADm を用いる際の初期点は $z_0 = 1$ とした.

4.1 数値実験結果

表 1~3 には各 n, σ^2 において生成したデータを用いて実験を行い, 関数 g を用いた手法, 関数 \hat{g} を用いた手法, 従来手法で最適化を行った際の最適パラメータ z^* とそのときの赤池情報量基準をのせている. ただし, z_1^* と AIC_1 はそれぞれ σ^2 を既知として解いた場合の最適パラメータとそのときの赤池情報量基準を表し, z_2^* と AIC_2 はそれぞれ σ^2 を未知として解いた場合の最適パラメータとそのときの赤池情報量基準を表している.

表 1 数值実験結果 (関数 1)

σ^2	n	手法	z_1^*	AIC ₁	z_2^*	AIC ₂
0.000025	30	g	2.49	-233.3794	3.4	-238.9854
		\hat{g}	0.19921	-229.2215	2.2394	-236.0495
		従来	1	-228.9586	1	-235.1838
	50	g	3	-391.0378	2	-394.2786
		\hat{g}	2.7548	-391.0378	1.5666	-394.2786
		従来	1	-387.5714	1	-392.4739
	100	g	2	-769.7314	2.7	-771.1394
		\hat{g}	1.664	-769.7314	2.0216	-769.3978
		従来	1	-769.2231	1	-769.0782
0.0001	30	g	1	-194.2529	2.1681	-205.4892
		\hat{g}	0.7282	-194.2529	2.0033	-205.4892
		従来	1	-194.2529	1	-202.2448
	50	g	1.0824	-310.0646	3	-312.1888
		\hat{g}	0.93557	-309.8182	2.8795	-312.1888
		従来	1	-309.8182	1	-308.8168
	100	g	2.147	-641.8215	2.51	-643.7954
		\hat{g}	2.0391	-641.8215	2.5047	-643.7954
		従来	1	-636.2402	1	-636.4266
0.0004	30	g	2	-144.5234	3	-146.1974
		\hat{g}	1.7002	-144.5234	2.4353	-144.7798
		従来	1	-140.9471	1	-142.7112
	50	g	2	-242.4262	2	-243.3749
		\hat{g}	0.50619	-242.7140	1.387	-243.3749
		従来	1	-239.1064	1	-238.7218
	100	g	1.51	-504.6412	2	-507.4943
		\hat{g}	1.3788	-504.6412	1.7645	-500.4129
		従来	1	-501.7035	1	-497.5927

表 2 数值実験結果 (関数 2)

σ^2	n	手法	z_1^*	AIC ₁	z_2^*	AIC ₂
0.0025	30	g	0.3	-84.6586	1	-85.3406
		\hat{g}	1.5626	-84.6385	0.3094	-85.3657
		従来	1	-84.6385	1	-85.3406
	50	g	3	-163.3539	1.3	-168.2344
		\hat{g}	2.3731	-163.3539	0.030413	-174.9706
		従来	1	-160.7512	1	-167.6856
	100	g	1	-321.0690	2.3	-323.3254
		\hat{g}	0.89455	-321.0690	1.1419	-323.8341
		従来	1	-321.0690	1	-321.9862
0.01	30	g	0.47493	-55.2341	1	-62.3480
		\hat{g}	1.6312	-51.8161	0.56127	-60.5346
		従来	1	-52.9741	1	-62.3480
	50	g	1.49	-81.3215	0.3	-80.0688
		\hat{g}	1.6091	-78.3269	0.85466	-79.9919
		従来	1	-79.3625	1	-79.9919
	100	g	1.7	-180.5713	2.49	-183.3367
		\hat{g}	0.15846	-180.9441	0.82243	-183.1353
		従来	1	-180.5377	1	-182.8045
0.04	30	g	2	-1.1796	1	-2.2649
		\hat{g}	0.11943	-2.8711	0.16172	-2.1888
		従来	1	0.6378	1	-2.2649
	50	g	1.3	-16.8048	1.7	-16.9391
		\hat{g}	0.91956	-14.9125	1.0825	-15.6042
		従来	1	-14.9125	1	-13.8492
	100	g	0.657	-42.2676	0.3	-44.2110
		\hat{g}	0.57323	-42.2676	0.24567	-44.2110
		従来	1	-42.1231	1	-44.1497

表 3 数值実験結果 (関数 3)

σ^2	n	手法	z_1^*	AIC ₁	z_2^*	AIC ₂
0.25	30	g	0.5401	50.5637	2	51.4052
		\hat{g}	0.0081649	51.1629	0.048049	53.8312
		従来	1	53.6242	1	185.7027
	50	g	1	73.4718	0.79	71.3786
		\hat{g}	0.038359	73.0852	0.053311	70.8520
		従来	1	73.4718	1	81.8108
	100	g	2	142.5101	0.7	136.7827
		\hat{g}	0.17065	141.9819	0.043968	139.0828
		従来	1	152.2637	1	152.2209
1	30	g	2	95.5332	2	97.2765
		\hat{g}	0.083631	95.9062	0.15151	97.6979
		従来	1	97.2448	1	187.5076
	50	g	2	160.1585	0.49	148.4161
		\hat{g}	0.045316	148.7233	0.36321	152.7942
		従来	1	160.8756	1	159.5425
	100	g	0.24958	269.6491	0.35042	264.7446
		\hat{g}	0.11206	270.5413	0.15763	266.1670
		従来	1	272.0571	1	268.0406
4	30	g	1	131.6184	2.5879	138.8906
		\hat{g}	0.11241	130.7128	0.019662	131.4304
		従来	1	131.6184	1	192.6754
	50	g	1	210.9593	0.086914	199.8945
		\hat{g}	0.039719	205.0300	0.062364	200.6144
		従来	1	210.9593	1	224.7705
	100	g	2	425.1513	1.3976	425.0967
		\hat{g}	0.17285	428.1609	0.18651	427.9567
		従来	1	426.8050	1	444.7895

また, 得られたパラメータ α が表すモデル関数の一例として, 関数 2 において $n = 30$, $\sigma^2 = 0.04$, σ^2 を既知として解いたときの結果を図 1, 2 に示す. ただし, グラフにおいて, 破線は入出力データ生成のために用意した真の関数, 点はその関数から生成した誤差を含む入出力データ, 実線は得られたパラメータ α が表すモデル関数である.

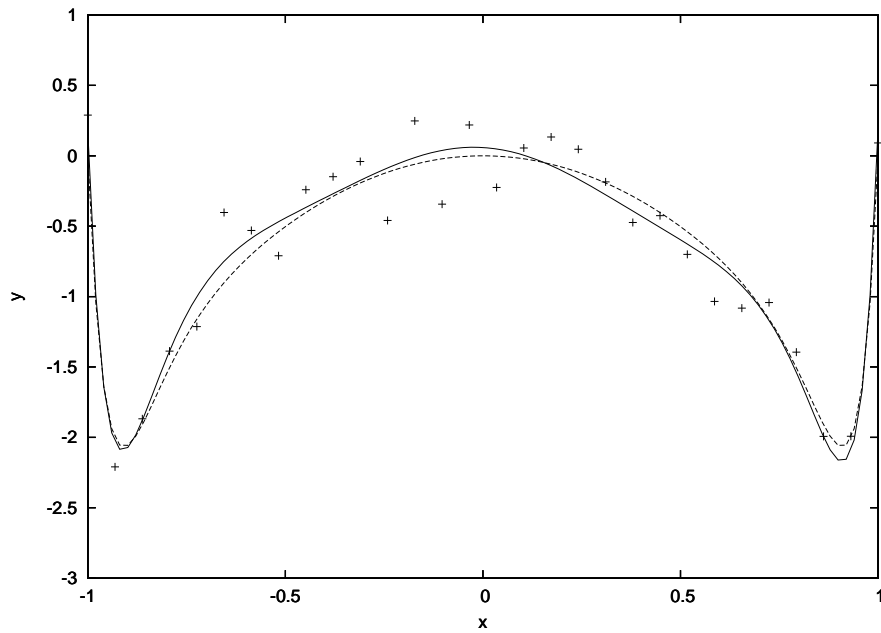


図 1 関数 2 におけるグラフ ($n = 30$, $\sigma^2 = 0.04$, σ^2 既知)

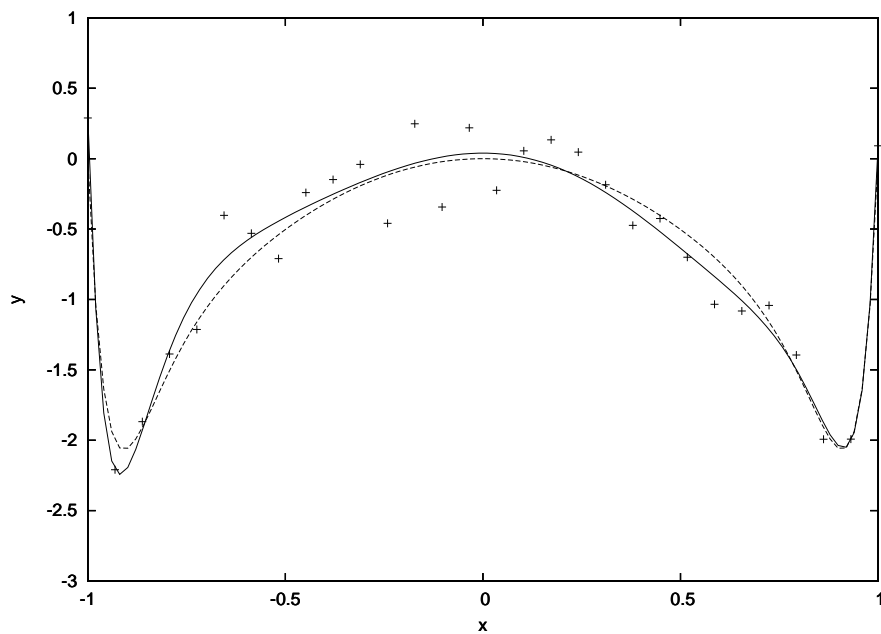


図 2 関数 2 におけるグラフ ($n = 30$, $\sigma^2 = 0.04$, σ^2 未知)

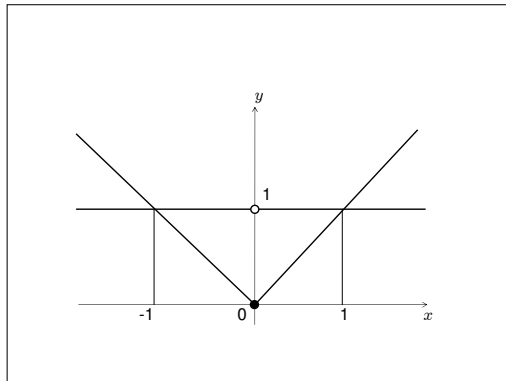


図3 $y = \text{card}(x)$, $y = |x|$ のグラフ

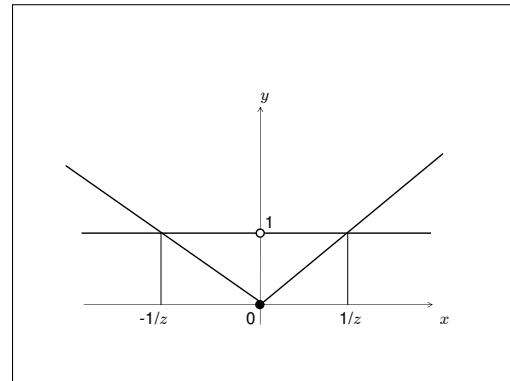


図4 $y = \text{card}(x)$, $y = z|x|$ のグラフ

4.2 実験結果に対する考察

まず、関数 g を用いた手法では、 $z = 1$ とした従来の手法と比べて赤池情報量基準が小さいまたは等しいパラメータ α が得られていることがわかる。これは、直接探索法により、 $\text{card}(\alpha)$ をより近似したパラメータ z が得られたためと考えられる。次に、関数 \hat{g} を用いた手法では、従来の手法と比べて赤池情報量基準が小さいパラメータ α も得られているが、一部では赤池情報量基準が大きなパラメータ α が得られていることがわかる。これは、直接探索法で最適化を行う際に、真の赤池情報量基準の値ではなく、赤池情報量基準の近似値を用いたためであると考えられる。

続いて、各関数ごとの最適なパラメータ z^* の値について考える。表 1~3 を見ると、最適なパラメータ z^* は関数 1 からデータを作成した場合に大きく、関数 3 からデータを作成した場合に小さくなっている傾向が読み取れる。これは、 $\text{card}(\alpha_j)$ と $z|\alpha_j|$ の近似精度によると考えられる。図 3 は $\text{card}(x)$ と $|x|$ のグラフ、図 4 は $\text{card}(x)$ と $z|x|$ のグラフを示している。図 3 から $\text{card}(x)$ と $|x|$ は $x = -1, 0, 1$ の近傍で近似精度が高いといえる。同様に図 4 から $\text{card}(x)$ と $z|x|$ は $x = -1/z, 0, 1/z$ の近傍で近似精度が高いといえる。このパラメータ z を、問題に応じて適切に選べば、 $z = 1$ と固定したときよりも近似精度の良い問題を解くことができる。これが、パラメータ z を導入した目的である。本実験でデータ作成の際に用いた関数を見ると、関数 1 の係数は小さく、関数 3 の係数は大きいものとなっている。 $\text{card}(\alpha_j)$ と $z|\alpha_j|$ の近似精度を高くするためには $z = 1/|\alpha_j|$ とすればよいと考えられるため、係数 α_j が大きいときはパラメータ z は小さくなり、 α_j が小さいときは z は大きくなると考えられる。実際に、数値実験結果ではそのような z が求まっている。

最後に、関数 g を用いた手法と関数 \hat{g} を用いた手法を比較する。得られた赤池情報量基準を比べると、関数 g を用いた手法が優れている場合と関数 \hat{g} を用いた手法が優れている場合とも存在する。 g が良いときは最適化の際に赤池情報量基準の値を直接用いているか、その近似値を用いているかの違いが原因であると考えられる。 \hat{g} が良いときは直接探索法でうまく最適解を見つけられたかどうかにより、違いが生じたと考えられる。その例として関数 3 において $n = 50$, $\sigma^2 = 1$, σ^2 を既知として解く場合を考える。図 5, 6 はそれぞれ関数 3 において $n = 50$, $\sigma^2 = 1$, σ^2 を既知としたときの関数 g と \hat{g} の値をプロットしたものである。図 5 より、最適解は $z^* = 0.1$ 付近に存在すると読み取れる。しかし、直接探索法で最適化すると表 3 より $z^* = 2$ となっており、あまり

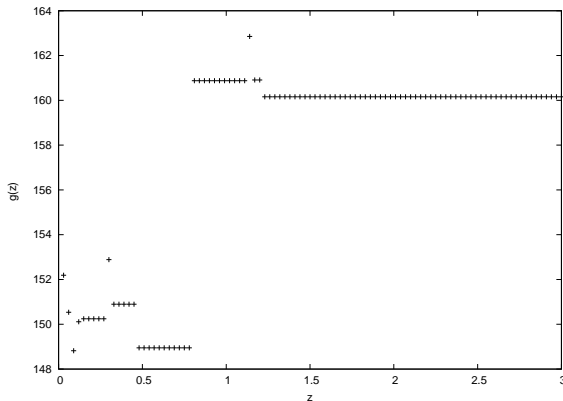


図5 関数3における $n = 50$, $\sigma^2 = 1$, σ^2 既知のときの g のグラフ

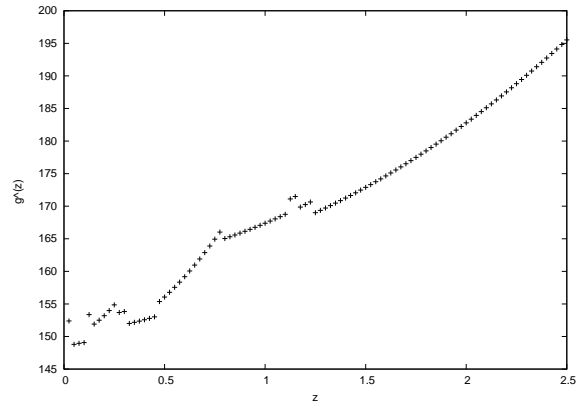


図6 関数3における $n = 50$, $\sigma^2 = 1$, σ^2 既知のときの \hat{g} のグラフ

良い解が得られていない。これは、最初のステップで初期点 $z_0 = 1$ から $z_1 = 2$ へと更新されたが、 $z = 2$ 近傍では関数 g の値が一定か $g(2)$ よりも大きいため、そのまま探索が終了したことによると思われる。図6では関数の値が z によって細かく変化するため、直接探索法がうまく働き、良い解 $z^* = 0.045316$ が得られている。ここに、関数 \hat{g} を考える利点があるといえる。

5 結論

本報告書では、赤池情報量基準を最小とするパラメータを求める手法として、選択可能なパラメータ数が多い場合でも扱うことのできる近似的な方法を考えた。その際、近似精度を良くするために係数を導入した問題を考え、直接探索法で最適化する手法を提案した。また、提案手法を実装し、数値実験を通して提案手法の有効性を検証した。その結果、従来手法よりも赤池情報量基準を小さくするパラメータを得られることが確認された。

今後の課題としては、多項式回帰モデル以外のモデル関数も考え、数値実験を通して提案手法の有効性を考察することや、パラメータ z を含んだ問題 $P(z)$ として z が多変数の場合を考え、その様々な問題に対して最適化を考えることが挙げられる。また、直接探索法は局所最適解を求めることはできるが、大域的最適解を求めることはできないと限らないため、大域的最適解を求める解法の探求も必要である。

謝辞

日頃のご教授に加え、本報告書の作成にあたり、細部にいたるまで適切なお指摘と丁寧なご指導を賜った山下信雄准教授に心より感謝の意を表します。また、日頃よりお世話になっている福島雅夫教授、林俊介助教ならびに福島研究室の皆様にも厚く御礼申し上げます。

参考文献

- [1] H. Akaike, Information theory and an extension of the maximum likelihood principle, Second international symposium on information theory, Budapest, Akademiai Kiado, pp. 267-281, 1973.
- [2] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control, Vol.19, pp. 716-723, 1974.
- [3] C. Audet and J. E. Dennis Jr., Mesh adaptive direct search algorithms for constrained optimization, SIAM J Optimization, Vol.17, pp. 188-217, 2006.
- [4] Stefan Emet, A model identification approach using MINLP techniques, Ninth WSEAS international conference on applied mathematics, Istanbul, Turkey, pp. 347-350, 2006.
- [5] NOMADm, <http://www.gerad.ca/NOMAD/Abramson/nomadm.html>.
- [6] 坂元慶行, 石黒真木夫, 北川源四郎, 情報量統計学, 共立出版, 1983.
- [7] 矢部博, 工学基礎 最適化とその応用, 数理工学社, 2006.