

ファジィ回帰分析における新たな最小化基準の研究

創価大学大学院工学研究科
情報システム学専攻修士課程
田島博之 (Hiroyuki Tajima)

1. はじめに

回帰分析における観測誤差についての研究は天文学者らによって18世紀初頭より始められた。近年、回帰分析における「誤差」というものが、単なる観測誤差というものではなく、人間にとって十分意味のある要因から生じた結果とする考え方が「ファジィ理論」とともに現われ、ファジィ回帰分析という分野で研究されるようになった。現在、ファジィ回帰分析というと田中らによる定式化が一般的である。これは独立変数、従属変数がともに実数データである場合と、独立変数が実数、従属変数がファジィ数で与えられる2つの場合がある。

本研究では独立変数、従属変数がともに1つの実数データとして与えられたファジィ単回帰モデルにおいて、新たな最小化基準を用いた定式化を提案する。

2. 従来型のファジィ回帰モデル

田中らによる定式化は、区間回帰モデルの類似として提案されている。1) 定式化に際して取り扱うファジィ数には、対称三角型ファジィ数を用いる。以下にいくつかの定義を記す。

対称三角型ファジィ数 \tilde{A} はその中心 a と中心からの広がりを e としたときこれを $\tilde{A} = (a, e)_L$ と表記する。

ファジィ数 \tilde{A} のメンバシップ関数の定義

$$\mu_{\tilde{A}}(x) = \max(0, 1 - |a - x|/e)$$

ファジィ数 \tilde{A} の α レベル集合 (α -cut) の定義

$$[\tilde{A}]_{\alpha} = \{x \mid \mu_{\tilde{A}}(x) \geq \alpha\} \quad \text{for } 0 < \alpha \leq 1$$

ファジィ数に対する演算は拡張原理により定義される。ファジィ数の和と実数倍は以下の通り。

$$\begin{aligned} \tilde{A} + \tilde{B} &= (a, e)_L + (b, f)_L = (a + b, e + f)_L \\ k \cdot \tilde{A} &= k \cdot (a, e)_L = (k \cdot a, |k| \cdot e)_L \end{aligned}$$

ファジィ線形回帰モデルは次のように定める。

$$\begin{aligned}\tilde{Y}(\mathbf{x}_p) &= \tilde{A}_0 + \tilde{A}_1 x_{p1} + \tilde{A}_2 x_{p2} + \dots + \tilde{A}_n x_{pn} \\ &= (a_0, e_0)_L + (a_1, e_1)_L x_{p1} + (a_2, e_2)_L x_{p2} + \dots + (a_n, e_n)_L x_{pn} \\ &= (a(\mathbf{x}_p), e(\mathbf{x}_p))_L\end{aligned}$$

$$\begin{aligned}a(\mathbf{x}_p) &= a_0 + a_1 \cdot x_{p1} + a_2 \cdot x_{p2} + \dots + a_n \cdot x_{pn} \\ e(\mathbf{x}_p) &= e_0 + e_1 \cdot |x_{p1}| + e_2 \cdot |x_{p2}| + \dots + e_n \cdot |x_{pn}|\end{aligned}$$

定式化

実数の入出力データ (\mathbf{x}_p, y_p) , $p=1, 2, \dots, m$ が与えられた場合での基本的なファジィ回帰分析の問題は、与えられたすべてのデータ点を h レベル集合内に包含するようなあいまいさ最小のファジィ回帰モデルを求める問題である。すなわち

$$y_p \in [\tilde{Y}(\mathbf{x}_p)]_h, \quad p=1, 2, \dots, m$$

という制約のもとで、ファジィ線形回帰モデルによる推定ファジィ数 $\tilde{Y}(\mathbf{x}_p)$ の広がり
の最小化である。ただし、 $0 < h \leq 1$ であり、上式右辺のレベル集合は次のような閉区間となる。

$$\begin{aligned}[\tilde{Y}(\mathbf{x}_p)]_h &= [(a(\mathbf{x}_p), e(\mathbf{x}_p))_L]_h \\ &= [a(\mathbf{x}_p) - (1-h)e(\mathbf{x}_p), a(\mathbf{x}_p) + (1-h)e(\mathbf{x}_p)]\end{aligned}$$

広がり
の幅の最小化基準は2種類存在しそれぞれにおいてファジィ回帰モデル同定問題が提案されている。以下にその同定問題を記す。

実数データに対するファジィ回帰モデル同定問題 1

目的関数 $\text{Min} \sum_{p=1}^m e(\mathbf{x}_p)$

制約条件

$$a(\mathbf{x}_p) - (1-h)e(\mathbf{x}_p) \leq y_p, \quad p=1, 2, \dots, m$$

$$a(\mathbf{x}_p) + (1-h)e(\mathbf{x}_p) \geq y_p, \quad p=1, 2, \dots, m$$

$$e_i \geq 0, \quad i=1, 2, \dots, n$$

決定変数 $a_i, e_i, i=1, 2, \dots, n$

実数データに対するファジィ回帰モデル同定問題 2)

$$\text{目的関数 } \text{Min} \sum_{p=1}^m \{e(\mathbf{x}_p)\}^2$$

制約条件

$$a(\mathbf{x}_p) - (1-h)e(\mathbf{x}_p) \leq y_p, p=1,2,\dots,m$$

$$a(\mathbf{x}_p) + (1-h)e(\mathbf{x}_p) \geq y_p, p=1,2,\dots,m$$

$$e_i \geq 0, i=1,2,\dots,n$$

$$\text{決定変数 } a_i, e_i, i=1,2,\dots,n$$

問題 1 と問題 2 の違いは、広がり幅の単純和か 2 乗和かの違いである。

3. 従来型の問題点

田中らの定式化においては、すべてのデータ点を包括するようなモデルを決定するために意思決定者に h という値を決めさせている。これによって決定されたモデルによって表現されるファジィ数は $\alpha = h$ における α カット集合にそのすべてのデータ点を含むようになっている。これはモデル決定者によって決定される h の値の選び方が決定者によって異なり、意思決定者の経験に任されてしまうということになる。これは、意思決定者によってモデルを自由に変化させられるという点でメリットであるのだが、その反面通常の最小 2 乗法による回帰モデルが普遍的に決定されるのと比較したとき、モデルを決定する際の一般性に欠けるという見方もできる。モデルが表現することのできるあいまいさとモデルを決定する際のあいまいさは別の問題であるので本研究ではモデルを決定する際にはできるだけ客観的にモデルを決定できるような同定法とする。

2 つめとして、決定されるファジィ回帰モデルの $e(\mathbf{x}_p)$ は必ず単調増加してしまうという点である。問題によっては、あいまいさが徐々に減る単調減少である場合も考えられるが田中らの定式化ではこの点には触れていない。本研究では Problem 2 においてあいまいさが \mathbf{x} の平均値を最小にして増減するモデルとなっている。この問題は、ファジィ数の演算としてモデルを定義したところにその原因があるのだが、本研究ではモデルの推定するファジィ数の中心値と幅をそれぞれ独立の関数として扱うことによってこの問題を回避しようと考えた。

3 つめとして、データに外れ値があるときにはその影響をまともに受けてしまうという欠点である。これに対しては従来からデータ点の組み合わせ最適化問題 3) が考えられている。最近では超楕円関数を利用したファジィロバスト回帰分析 4) が提案されているが、本研究では制約条件を用いないことにより外れ値の影響を軽減する方法をとった。

4. ファジィ単回帰モデルの提案

本研究で提案するファジィ単回帰モデルは従来の回帰モデルと異なりモデルを決定する際の演算を推定するファジィ数の中心を表す関数と幅を表す関数をそれぞれ独立の実数演算により決定する。

$$\tilde{Y} = (y_c, y_w)_L = (a(x), |e(x)|)_L$$

以下に新たなファジィ回帰モデル同定問題を2種類あげる。

与えられる実数データ： (x_p, y_p) , $p=1, 2, \dots, m$

Problem1

中心を表す関数を従来の最小2乗法により求め、幅を表す関数の決定には a_0, a_1 の標準偏差を用いる。

(フェーズ1)

$$\text{目的関数} \quad \min \sum_{p=1}^m \{y_p - a(x_p)\}^2$$

決定係数 a_0, a_1 ここで決定した a_0, a_1 の関数を $a(x) \rightarrow \hat{a}(x)$ と定める

(フェーズ2)

フェーズ1で決定した際の a_0, a_1 の標準偏差を e_0, e_1 とする。

$$\begin{aligned} \bar{x} &= \text{mean}_{p=1,2,\dots,m}(x_p) \\ \sigma^2 &= \frac{1}{m-2} \sum_{p=1}^m \{y_p - \hat{a}(x_p)\}^2 \\ e_0 &= \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{p=1}^m (x_p - \bar{x})^2}} \\ e_1 &= \sigma \sqrt{\frac{1}{\sum_{p=1}^m (x_p - \bar{x})^2}} \\ \therefore e(x) &= \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{p=1}^m (x_p - \bar{x})^2}} + \sigma \sqrt{\frac{1}{\sum_{p=1}^m (x_p - \bar{x})^2}} \times x \end{aligned}$$

Problem2

中心を表す関数を従来の最小 2 乗法により求め、幅を表す関数には $(1-k) \times 100$ [%] の信頼区間を表す区間推定の際に用いられる予測誤差の式を用いる。

(フェーズ 1)

$$\text{目的関数} \quad \min \sum_{p=1}^m \{y_p - a(x_p)\}^2$$

決定係数 a_0, a_1 ここで決定した a_0, a_1 の関数を $a(x) \rightarrow \hat{a}(x)$ と定める

(フェーズ 2)

フェーズ 1 で決定した際の誤差分散を求め $(1-k) \times 100$ [%] の信頼区間を表す区間推定時の予測誤差の値を $e(x)$ とする。

$$\bar{x} = \text{mean}_{p=1,2,\dots,m}(x_p)$$

$T_{\frac{k}{2}}$ は自由度 $m-2$ の t 分布の右片側 $\frac{k}{2}$ の確率を与える点とする。

$$\sigma^2 = \frac{1}{m-2} \sum_{p=1}^m \{y_p - \hat{a}(x_p)\}^2$$

$$e(x) = T_{\frac{k}{2}} \cdot \sigma \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{p=1}^m (x_p - \bar{x})^2}}$$

5. まとめ

以上 2 種類の定式化を行ったが、いずれも最小 2 乗法を用いることによって推定するファジィモデルの中心を表す関数を表現している。これは、田中らが提案するファジィ回帰モデルよりも推定されるファジィ数の中心の値が強い意味を持つといえる。

また、制約条件を含む計画問題の欠点である外れ値に対する対応としては、決定されたモデルから外れたデータを除外するという方法が考えられる。これは、組み合わせ最適化問題よりも非常に簡単に対応できると言える。また、注意すべき点として、 x のデータの大きさによってモデルの $e(x)$ が決定されるために起こる問題がある。これは、 x 、 y のデータデータをそれぞれ標準化することによって解決した。今後の課題としては、単回帰モデルから重回帰モデルへと拡張することが考えられる。

参考文献

1. 田中、和多田、林 「ファジィ回帰分析の3つの定式化」
計測自動制御学会論文集 Vol.22, No.10, pp.1051-1057(1986)
2. 田中、小山、李 「二次計画法による可能性回帰分析」
第12回ファジィシステムシンポジウム講演論文集 pp845-846(1996)
3. 石淵、田中、「混合0-1整数計画問題による区間回帰分析」
日本経営工学会誌 vol.40, No.5, pp312-319(1989)
4. 藪内、和多田 「超楕円関数に基づくファジィロバスト回帰分析」
Journal of Opeartion Research Society of Japan Vol.39, No.4 , December 1996
pp512-524

この論文に対するご意見ご質問は htajima@t.soka.ac.jp までお願いします。