数理解析研究所講究録
1079 巻 1999 年 190-195

実数データを用いたファジィ回帰モデル同定法の研究

A study on methods to fix the fuzzy regression models with the real number data.

創価大学大学院工学研究科
田島 博之（D２）
Graduate school of Engineering, Soka University
Hiroyuki Tajima

１．はじめに

ファジィ回帰モデルによる分析は、従属変数に視覚、味覚、聴覚等の感覚また、意思決定のような人間の思考が絡む状況を表したデータが実数値あるいはファジィ数値として与えられた場合に有効である。これらの曖昧さを多分に含んだデータを分析するうえで、ファジィ回帰モデルはそのデータをよく表現することができる。

従来の最小二乗推定法において、データとモデルの予測値の差は測定誤差として考え、その値は正規分布に従うと仮定されていたが、このファジィ回帰モデルにおいては、モデルの表現する幅自体に重要な意味を持たせている。曖昧さには本来その要因が未知という意味と、その要因があまりにも複雑な関係からなるために特定するのが難しいという意味がある。どちらの場合においても、ファジィ回帰モデルはデータに内在的に含まれる、未知の要因を明らかにすることなくモデルに表現することができる。区間回帰モデルの応用であるファジィ回帰モデルは全てのデータを包括したモデルを同定する所に特徴があり、その点は最小二乗推定法の概念とは異なる。

そこで本研究では、従来の区間回帰モデルと最小二乗推定法の両方の概念を導入した新しい２つのタイプのファジィ回帰モデル同定法を提案する。新提案モデルは同定する推定ファジィ数の中心値を表す関数に重点をおいている。なお、本研究で着目するモデルは、実数値データが与えられた場合を考えている。

２．数学的準備

ファジィ数、ファジィ回帰モデルに対する諸定義は以下に示す。従来型モデルは、対称三角型ファジィ数を用いていたが、本提案型モデルでは、非対称型三角ファジィ数を用いる。

ファジィ数 $\tilde{A}$ はその中心 a と左（減少方向）への広がりの幅 l および右（増加方向）への広がりの幅 r を用いて $\tilde{A}=(a,l,r)$ と表記する。メンバーシップ関数は次のように定義する。

$$\mu_{\tilde{A}}(x)=\begin{cases}\max\left(0,1-\dfrac{a-x}{l}\right) & for\ x<a\\ \max\left(0,1-\dfrac{x-a}{r}\right) & for\ x\geqq a\end{cases}$$

このファジィ数 $\tilde{A}$ の α レベル集合は以下の閉区間となる。

$$[\ \tilde{A}\ ]_\alpha=[(a,l,r)]_\alpha$$
$$=[a-(1-\alpha)l,a+(1-\alpha)r]$$

ファジィ数間の演算は次に定義する。

$$\tilde{A}+\tilde{B}=(a,s,t)+(b,u,v)$$
$$=(a+b,s+u,t+v)$$

$$k\cdot\tilde{A}=k(a,s,t)=(k\cdot a,k\cdot s,k\cdot t)$$

以上よりファジィ回帰モデルを定義する。

$$\tilde{Y}(\mathbf{x})=\tilde{A}_0+\tilde{A}_1x_1+\cdots+\tilde{A}_nx_n$$
$$=(a_0,l_0,r_0)+x_1\cdot(a_1,l_1,r_1)+\cdots$$
$$\cdots+x_n\cdot(a_n,l_n,r_n)$$
$$=(a_0+\cdots+a_nx_n,l_0+\cdots+l_n|x_n|,r_0+\cdots+r_n|x_n|)$$
$$=(a(\mathbf{x}),l(\mathbf{x}),r(\mathbf{x}))$$

以上は本論で提案するモデルの定義である。なお、従来型では$l(\mathbf{x})=r(\mathbf{x})$なので、これを$e(\mathbf{x})$として考えている。

3．従来型モデル同定法

実数データに対する従来型のファジィ回帰モデル同定法は、与えられた全てのデータ点をそのモデルに包括する制約条件のもと、モデルの表現する曖昧さの幅を最小にする。

以下にそのモデル同定問題を記述する。

与えられるデータセット：

$(\mathbf{x}_i;y_i)=((x_{i1},x_{i2},\cdots,x_{in});y_i)$

$i=1,2,\ldots,m$

このようなデータが与えられたとき同定するモデルのαカットのhレベル集合が全てのデータを含むような制約条件を考える。

$y_i\in[\tilde{Y}(\mathbf{x}_i)]_h$, $i=1,2,\cdots,m$

hの値はモデル同定者が同定問題を設定する際に任意に決定するパラメータである。従来型モデル同定問題は、この制約条件のもとファジィ回帰モデルの曖昧さの幅を最小化基準として目的関数に用いた数理計画問題として定義される。

> 従来型問題1
>
> 目的関数：
>
> $$\min\sum_{i=1}^{m}e(\mathbf{x}_i)$$
>
> 制約条件：
>
> $$a(\mathbf{x}_i)-(1-h)e(\mathbf{x}_i)\le y_i$$
> $$a(\mathbf{x}_i)+(1-h)e(\mathbf{x}_i)\ge y_i$$
> $$i=1,2,\cdots,m$$
>
> $$e_j>0,\ j=0,1,\cdots,n$$
>
> 決定関数：
>
> $a(\mathbf{x}),e(\mathbf{x})$

上のモデルよりも最小二乗推定法に近いという理由で、単純に曖昧さを表す幅の関数を二乗にする、最小化基準を用いた目的関数の同定法も提案されている。その同定問題を以下に定義する。

> 従来型問題2
>
> 目的関数：
>
> $$\min\sum_{i=1}^{m}e(\mathbf{x}_i)^2$$
>
> 制約条件：
>
> $$a(\mathbf{x}_i)-(1-h)e(\mathbf{x}_i)\le y_i$$
> $$a(\mathbf{x}_i)+(1-h)e(\mathbf{x}_i)\ge y_i$$
> $$i=1,2,\cdots,m$$
>
> $$e_j>0,\ j=0,1,\cdots,n$$
>
> 決定関数：
>
> $a(\mathbf{x}),e(\mathbf{x})$

以上の２タイプが実数データに対する従来

型の代表的なファジィ回帰モデル同定法である。なお、従来型ではモデル同定者の意思を反映させる変数値として、hを用いたが新提案モデルではもちいないこととする。

4．新たなモデル同定法の提案

本論で提案するモデル同定法はいずれも中心関数に重点を置いた最小化目的関数を設定する。

ここで､分析すべきデータを以下に与える。

$$(\mathbf{x}_i;y_i)=((x_{i1},x_{i2},\cdots,x_{in}),y_i)$$

$$i=1,2,\cdots,m$$

次に新たな最小化基準を定義する。単回帰モデルを例に、その基本概念を説明する。同定するモデルの上辺、下辺をそれぞれ次のように定義する。

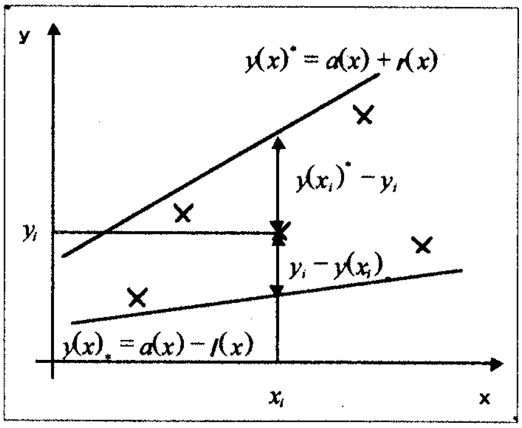
$$y(x)^*=a(x)+r(x),\ \ y(x)_*=a(x)-l(x)$$

同定するファジィ回帰モデルは、与えられた全てのデータがそのモデルに包括されるような制約条件のもと、各データ点からモデルの上辺までの距離の二乗と、下辺までの距離の二乗を加えた値の総和を最小にする。

この考察を基に最小化基準を定義し、独立変数xが複数存在するような重回帰モデルにおいても定式化可能にした、新たなモデル同定問題を以下に提案する。

4．1　新同定法1

ここで、提案する新たなモデル同定法は2段階の手順を追う。第1段階においてモデルの上辺、および下辺の関数を求める。以下の図は単回帰モデルの例である。



与えられた全てのデータがそのモデルに包括されるような制約条件のもと、各データ点からモデルの上辺までの距離の二乗と、下辺までの距離の二乗を加えた値の総和を、最小にするような最小化基準を定義し、それを一般化したのが以下の同定問題である。

新同定問題1

First Phase

目的関数：

$$\min\sum_{i=1}^{m}\left[\{y(\mathbf{x}_i)^*-y_i\}^2+\{y(\mathbf{x}_i)_*-y_i\}^2\right]$$

制約条件式：

$$y(\mathbf{x}_i)_*\leqq y_i\quad i=1,2,\cdots,m$$,

$$y(\mathbf{x}_i)^*\geqq y_i\quad i=1,2,\cdots,m$$

決定関数：$y(\mathbf{x})^*,y(\mathbf{x})_*$

これによって決定された、モデルの上辺、下辺の関数を$\hat{y}(\mathbf{x})^*,\hat{y}(\mathbf{x})_*$とする。

次にモデルの中心関数を求めるために最小二乗法を使った **Second Phase** を示す。

Second Phase
目的関数：

$$\min \sum_{i=1}^{m} \{y_i - a(\mathbf{x}_i)\}^2$$

決定関数：
$a(\mathbf{x})$

以上が新同定法１の説明である。

４．２　新同定法２

新同定法２は、上の新同定法１とは中心関数の同定法を変化させたものである。したがって、First Phase により、モデルの上辺および下辺を決定するまでは、問題１と同じ方法をとる。

中心を表す関数 $a(\mathbf{x})$ を以下に示す新たな基準によって定める同定問題を提案する。

α_i を以下に示す関数とする。

$$\alpha_i = \begin{cases} \dfrac{y_i - \hat{y}(\mathbf{x}_i)_*}{a(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)_*} & for\ a(\mathbf{x}_i) \geqq y_i \\ \dfrac{\hat{y}(\mathbf{x}_i)^* - y_i}{\hat{y}(\mathbf{x}_i)^* - a(\mathbf{x}_i)} & for\ a(\mathbf{x}_i) < y_i \end{cases}$$

以上の関数を最小化目的関数として中心関数を求める数理計画問題を用いて新同定問題２を定式化する。

新同定問題２
First Phase（問題１と同じとする）

Second Phase
目的関数：

$$\min \ -\sum_{i=1}^{m} \alpha_i$$

制約条件式：

$$\hat{y}(\mathbf{x}_i)_* \leqq y_i \quad i = 1,2,\cdots,m \quad ,$$

$$\hat{y}(\mathbf{x}_i)^* \geqq y_i \quad i = 1,2,\cdots,m$$

決定関数：$a(\mathbf{x})$

以上が新同定法２の説明である。

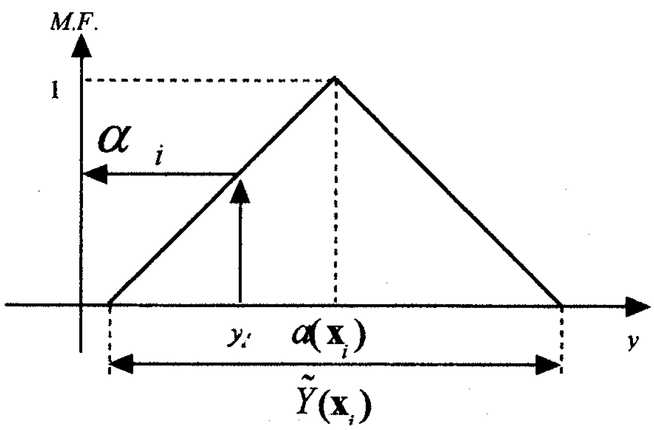
５．　評価関数の定義

本節では、モデルを比較評価するため、データとモデルの適合度基準を示す評価関数を以下に定義する。

５．１　モデル評価関数１

同定されたファジィ回帰モデルの関数を以下のように表記する。

$$\tilde{Y}(\mathbf{x}) = \left(\hat{a}(\mathbf{x}), \hat{l}(\mathbf{x}), \hat{r}(\mathbf{x})\right)$$



それぞれのデータ点について、上図で表されるような α_i を以下のように算出する。

$$\alpha_i = \begin{cases} \dfrac{y_i - \{\hat{a}(\mathbf{x}_i) - \hat{l}(\mathbf{x}_i)\}}{\hat{l}(\mathbf{x}_i)} & for\ \hat{a}(\mathbf{x}_i) \geqq y_i \\ \dfrac{\{\hat{a}(\mathbf{x}_i) + \hat{r}(\mathbf{x}_i)\} - y_i}{\hat{r}(\mathbf{x}_i)} & for\ \hat{a}(\mathbf{x}_i) < y_i \end{cases}$$

この値の平均値を評価関数とし、以下のように表記する。

$$\overline{\alpha}(\mathbf{x}, y) = \frac{1}{m}\sum_{i=1}^{m}\alpha_i$$

この値が1に近いほど、同定モデルの表現する中心関数がデータの性質を良く反映しているといえる。

5．2　モデル評価関数2

$$v(\mathbf{x}, y) = \frac{1}{m}\sum_{i=1}^{m}\left\{\overline{\alpha}(\mathbf{x}, y) - \frac{e(\mathbf{x}_i) - |y_i - a(\mathbf{x}_i)|}{e(\mathbf{x}_i)}\right\}^2$$

この関数2は$\alpha(\mathbf{x}_i)$の値が、どの程度の分散を持っているかを表現する適合度関数である。関数1の値が同じような値の場合は関数2の値が低いモデルを適合度があると判断する。

6．数値実験

次に数値実験に用いたデータと実験結果結果を示す。
なお、従来の同定法は文献[16]を参照。

データセットA

x	0.0	10.0	20.0	30.0	40.0	50.0	60.0	70.0	90.0	110	130
y	1.00	1.14	1.29	1.43	1.57	1.71	7.00	2.00	4.00	6.00	8.00

実験結果

	a0	a1	l0	l1	r0	r1	α	V
従来型1	1	5.71E-02	0	4.28E-02	0	4.28E-02	0.317	1.64E-01
従来型2	3.13	2.16E-02	2.13	7.40E-03	2.13	7.40E-03	0.333	8.60E-02
新型1	2.04	5.04E-02	1.53	2.91E-02	1.61	5.53E-02	0.536	1.16E-01
新型2	3.03	4.91E-02	2.52	2.78E-02	0.62	6.80E-02	0.57	1.26E-01

従来型のモデル同定問題を解くことによって得られるファジィ回帰モデルは、与えられたデータの多くを、そのモデルの端に追いやってしまうことがある。このことは、αレベル平均関数値α mean の値が 0.317、0.317 と低い値になっていることからも分かる。

新型では関数αの平均値が 0.536 また、0.570 といずれも高い値となっている。これは、モデル同定法の最小化基準にそれぞれデータ点の従属変数の値を取り入れているためであり、中心を表す直線がデータのある方向へ近づいているためである。このことから同定されたファジィ回帰モデルの中心値の持つ意味が、従来型よりも強いということが分かる。

7．まとめ

本稿において新たに提案したファジィ回帰モデル同定問題には、最小化目的関数に最小二乗推定法の概念を導入した。更に用いるファジィ数を従来型の対称三角型ファジィ数から、非対称三角型ファジィ数へと拡張した。

本同定法によって、同定されるモデルの中心関数は、強い意味を持ち、現実的で説得力のあるファジィ数を予測することを可能にしている。

また、データセットとモデルの適合度を明確にするために、2つのタイプの評価関数を定義した。これにより異なったタイプの同定法によって、定められたモデル間の比較を可能にした。

今後は、具体的な事例研究による汎用性の確認をするとともに、ファジィ数を左右非対称型から台形型、またLR型ファジィ数とした場合における定式化の問題などの研究が考えられる。

■

参考文献

[1] 田中,和多田,林:ファジィ線形回帰分析の三つの定式化, 計測自動制御学会論文集,Vol.22,No.10, pp.1051-1057 (1986)

[2]石渕,田中,黄:作業時間解析におけるファジィ回帰手法の比較,システム制御情報学会論文誌,Vol.3,No.3,pp.90-92 (1990)

[3]田中,石渕,黄:類似度を導入したファジィ回帰分析による職員数モデル,日本経営工学会誌Vol.41,No.2,pp.99-104 (1990)

[4]石渕:ファジィ回帰分析,日本ファジィ学会誌Vol.4 , No.1 ,pp.52-60(1992)

[5]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 pp.845-846 (1996)

[6]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 the Operations Research Society of Japan,Vol.39,No4,pp.512-524(1996)

[7]Aivars Celmins : LEAST SQUARES MODEL FITTING TO FUZZY VECTOR DATA ,Fuzzy Sets and Systems,Vol.22,pp.245-269 (1987)

[8]A.Celmins :MULTIDIMENSIONAL LEAST-SQUARES FITTING OF FUZZY MODELS , Math Modelling, Vol.9,No9,pp.669-690 (1987)

[9]Phil Diamond: Fuzzy Least SquaresInformation Sciences Vol.46 pp.141-157 (1988)

[10]坂和,矢野:ファジィ入出力データに対する多目的ファジィ回帰分析,日本ファジィ学会誌,Vol.1,No.1,pp.107-116 (1989)

[11]D.A.Savic and W.Pedrycz:Evaluation of fuzzy linear regression model,Fuzzy Sets and Systems 39,pp.51-63 (1991)

[12]Haekwan Lee, Hideo Tanaka: Dealing with Outliners by Fuzzy Regression Reflecting Central Tendency, 第 13 回ファジィシステム・シンポジウム講演論文集 pp.365-366 (1997)

[13]田島:「データの持つ特性を活かしたファジィ回帰モデルの同定法」, 第 13 回ファジィシステムシンポジウム講演論文集, pp. 377-380 (1997)

[14]田島：ある最小化基準に基づくファジィ回帰モデル, 研究集会「数理計画モデルの最適化と応用に関する研究」論文集, 8 年度文部省科学研究費総合研究(A)「数理統計学における情報抽出の理論と応用に関する研究」pp. 33-34(1996)

[15]田島：ファジィ回帰分析における新たな最小化基準の研究, 数理解析研究所講究録 891「連続と離散の最適化数理」京都大学数理解析研究所, pp. 89-94(1997)

[16]田島：中心関数を重視したファジィ回帰モデル同定法, 日本ファジィ学会誌 Vol. 10, No. 6, pp. 1135-1143(1998)

本論文に関するお問い合わせは、以下までご連絡下さい。

〒192-8577
東京都八王子市丹木町 1-236
創価大学大学院工学研究科
古川長太教授研究室所属（D2）田島博之
Email： htajima@t.soka.ac.jp