ファジィ線形回帰分析における最小二乗法

信州大学工学部　正会員　奥谷　巖
信州大学工学部　正会員　高瀬　達夫
信州大学工学部　　　　　○平賀　功一郎

1. はじめに

被説明変数がファジィ数で、係数または被説明変数のどちらか一方がファジィ数の場合の線形回帰モデルは、Tanaka らによって開発された線形計画法あるいは 2 次計画法による方法で解かれるのが一般的である。しかしながらこのモデルでは、その構造上ファジィ数の幅を 0 としたケースでは問題を解くことができないため、通常の線形回帰モデルを抱合するという意味での汎用性は備えられていない。

本研究では 2 つのファジィ数間の距離を、中心と幅の 2 次元平面上でのユークリッド距離として定義し、被説明変数と説明変数のファジィ線形結合との距離の二乗和を最小にするという基準により係数を決定する新たな方法を提案し、その有効性の検討を行う。

2. ファジィ数間の距離

本研究で扱う左右対称な三角形ファジィ数Aを、簡単のために$A=(\alpha,c)_L$とて表す。ここにAのメンバシップ関数$\mu_A(x)$は

$$\mu_A(x)=L((x-\alpha)/c)$$

ただし、$L(x)=max(0,1-|x|)$である。

いま二つのファジィ数が$A_1=(\alpha_1,c_1)_L$と$A_2=(\alpha_2,c_2)_L$と与えられた場合、現在のファジィ数間の距離の定義によれば、二つの三角形ファジィ数A_1,A_2の距離は、左側と右側それぞれの辺相互の隔たりを面積で表示し加えたものとして表される。しかし、このような定義の下では、二つのファジィ数の台集合が分離している場合や、一部が重なっている場合、ファジイ数の幅は距離に全く関与せず、一方の台集合に含まれるようになると急に幅の差が距離の中に現れ始めるという、常識的な感覚とは若干かけ離れた状況が出てくる。

こうしたことに鑑みて、本研究では、A_1とA_2の距離を単純に中心と幅で作られる 2 次元平面上のユークリッド距離として次のように定義する。すなわち、いま距離をdとしたとき

$$d=\sqrt{(\alpha_1-\alpha_2)^2+(c_1-c_2)^2} \tag{1}$$

のように与える。これは当然の事ながら、距離空間の条件を満足する。

また、PHIL DIAMOND の方法を用いると、$X_1=(\alpha_1,c_1,c_2)_T$と$X_2=(\alpha_2,d_1,d_2,)_T$とすると二つのファジィ数間の距離が中心と幅で作られる 2 次元平面上の距離として以下のように定義される。

$$\begin{aligned} d(\alpha_1,\alpha_2)^2 &= [\alpha_1-\alpha_2-(c_1-d_1)]^2 \\ &+[\alpha_1-\alpha_2+(c_2-d_2)]^2+(\alpha_1-\alpha_2)^2 \end{aligned} \tag{2}$$

しかしここで$c_1=c_2=c,d_1=d_2=d$のように左右対称な三角形ファジィ数とした場合 (2) 式は

$d(\alpha_1,\alpha_2)^2=3(\alpha_1-\alpha_2)^2+2(c-d)^2$ となり (1) 式と同様の結果をもたらすことになり (1) 式の方が単純でより扱いやすいものとなっていることがわかる。

3. ファジィ最小二乗モデル

合計N組のデータがあるものとしたとき j 番目の被説明変数のファジィデータを $Y_j=(y_j,e_j)_L$、n 個の変数からなる通常の説明変数のデータを$x_j=(x_{j1},x_{j2},\cdots,x_{jn})$とする。このとき$Y_j$となる推定値$\hat{Y}_j$が

$$\hat{Y}_j=A_0+A_1x_{j1}+\cdots+A_nx_{jn} \tag{3}$$

のような可能性線形システムで与えられると考える。ここに、A_lはファジィ数（α_l, c_l）$_L$を表すものとす

るが、このA_Iをより好ましい推定値Y_jが得られるように決定する問題を設定する。ここで提案する最小二乗モデルは、式（1）のような二つのファジイ数間の距離の定義の下で、ファジィデータY_jとその推定値Y_jの距離の二乗和を最小にするようにファジイ係数A_Iを与えようとするもので具体的には次のように簡単に数式的展開がなされる。

$$\min_{\hat{\alpha},\hat{c}} \sum_j \left\{ (y_j - \hat{y}_j)^2 + (e_j - \hat{e}_j)^2 \right\} \quad (4)$$

$$\hat{c} \geq 0 \quad (5)$$

上記の問題はそれぞれの変数に完全に分離されるから、結局

$$\min_{\hat{\alpha},\hat{c}} \sum_j (y_j - \hat{y}_j)^2 \quad (6)$$

および

$$\min_{\hat{\alpha},\hat{c}} \sum_j (e_j - \hat{e}_j)^2 \quad \hat{c} \geq 0 \quad (7)$$

という二つの問題を別々に解けば良いことになる。このうち式（6）で表される最初の問題は、通常のデータを用いた重回帰分析に正確に対応する。後者の式（7）で表される問題は変数$\hat{c}$に非負制約がつくため、数学的技法としては 2 次計画法を使って解かねばならない。

4．比較対象とする既存モデル

4．1 線形計画法による方法

式（2）の推定値$\hat{Y}_j$がY_jなる観測値を水準ｈ以上で含むようにする方法で、次のような線形計画法を使った最小化問題として定式化される。

$$\min_{\alpha,c} \sum c^T |x_j| \quad (8)$$

$$y_j + |L^{-1}(h)| e_i \leq x_j^T \alpha + |L^{-1}(h)| c^T |x_j| \quad (9)$$

$$y_j - |L^{-1}(h)| e_i \geq x_j^T \alpha - |L^{-1}(h)| c^T |x_j| \quad (10)$$

$$c \geq 0 \quad (11)$$

他の一つは、逆に式（2）で与えられる推定値がY_jに水準ｈ以上で含まれるようにするという考え方から、式（9）、（10）の不等号の向きをかえた制約条件を作り、そのもとで式（8）で表される幅の合計を最大化するという方法である。なお最小化問題の解は常に存在するが、最大化問題の解の存在の保証はない。

4．2　2 次計画法による方法

$$Y_{*j} = A_0 + A_{*1}x_{j1} + A_{*2}x_{j2} + \cdots + A_{*n}x_{jn} \quad (10)$$

のように$Y_{*j} \subset Y_j$となる推定値Y_{*j}与えられると考える。ここで、A_{*i}はファジイ数（α_i, c_i）を表すものとする。また、$Y_j^* \supset Y_j$となる推定値Y_j^*は

$$Y_j^* = A_0 + A_1^* x_{j1} + A_2^* x_{j2} + \cdots + A_n^* x_{jn} \quad (11)$$

のように与えられると考え，A_i^*はファジイ数（α, c_i+d_i）を表すものとするが、このA_{*i}およびA_i^*をより好ましい推定値が得られるように決定する問題を設定する。この問題に対する解法は

$$\min J = \sum_{j=1}^{m} \left(d^t |x_j|^t \right)^2$$

より求めるが紙面の都合上割愛させて頂く。

5．数値実験による適用性の検討

説明変数のデータx_jと対応するファジィデータY_jを発生させ、既存解法と提案方法の適用性検討に対する数値実験を実行する。詳しい手順はここでは省略するが実験により求められた値は以下の表のようになった。

説明変数	推定変量	既存法	最小二乗法
3 変数	中心	4.010 0. 959	3.817 0.961
	幅	11.783 0.363	1.397 0.797
7 変数	中心	10.456 0.926	4.365 0.967
	幅	10.010 0.440	1.646 0.809

注）数値欄の上段は RMSE、下段は一致係数η

4.おわりに

本研究ではファジィ線形回帰分析における新たな手法を提案しその有効性を検証する。今後はシュミレーションによる様々な条件下でテストを行い、従来法に対する優位性を検証してゆくことが必要である。また、実験結果の検討は紙面の都合上省略させて頂いたが当日は発表する予定である。