シミュレーションによるファジイ最小二乗法の適用安定性の検討

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

1. はじめに

被説明変数がファジイ数で、係数または被説明変数のどちらか一方がファジイ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[1)]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、その構造上ファジイ数の幅を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包含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つのファジイ数間の距離を、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イ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. ファジイ最小二乗法

2.1 ファジイ数間の距離

２つの三角形ファジイ数A_1、A_2を

$A_1=(\alpha_1,c_1)$、$A_2=(\alpha_2,c_2)$

としたとき、本研究ではこの２つのファジイ数間の距離を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。

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

この定義に従うと、$\alpha_1=\alpha_2$かつ$c_1=c_2$のとき$d=0$となり２つのファジイ数が一致しているといえる。従来のファジイ数の近さの考え方では、$\alpha_1=\alpha_2$のとき、ふたつのファジイ数の一致度が高い認識される。これに対し本手法ではdの値が小さいほうが一致度が高いとしているため、たとえ中心が一致していなくても、幅の差も含めて考慮したdの値を尊重している。

2.2 ファジイ最小二乗モデル

$Y_j=(y_j,e_j)_L$、$x_j=(x_{j1},x_{j2},\cdots,x_{jn})^T$のファジイデータ$(Y_j,x_j)$がＮ組与えられたとする。このとき$Y_j$の推定値$\hat{Y}_j$が次式で与えられると考える。

$$\hat{Y}_j=A_0+A_1x_{j1}+\cdots+A_nx_{jn} \tag{2}$$

本モデルは式(1)のような定義のもとでY_jと$\hat{Y}_j$の距離の二乗和を最小にするようにファジイ係数A_jを与えようとするもので、具体的には次のようになる。

$$\min_{\bar{\alpha},\bar{c}}\sum_j\left\{(y_j-\hat{y}_j)^2+(e_j-\hat{e}_j)^2\right\} \tag{3}$$

$$c\geq 0 \tag{4}$$

この問題は$\bar{\alpha}$と$\bar{c}$それぞれの変数に完全に分離されるから、結局

$$\min_{\bar{\alpha}}\sum_j(y_j-\hat{y}_j)^2 \tag{5}$$

および

$$\min_{\bar{c}}\sum_j(e_j-\hat{e}_j)^2 \tag{6}$$

$$c\geq 0 \tag{7}$$

という２つの問題を別々に解けばよいことになる。

本モデルの最大の特徴は、ファジイデータの幅e_jが０に近づいたとき、式中の係数A_jの幅を決定する問題が事実上なくなり、結局式(5)の重回帰分析のみを解けばよいことになり、我々の直観にもよく一致する。

なお、本研究で比較対象とする既存法は一般に用いられている可能性線形回帰分析法とし、この手法ついては寺野他[2)]に詳しいためここでは省略する。

3.シミュレーションによる適用安定性の検討

3.1 シミュレーションの方法

まず説明変数x_{ji}、誤差ε_jそれぞれについては正規乱数を、係数a_iについては三角形に従う乱数としてM1組発生させる。次にこれらの値を用いて$y_j=a_0+a_1x_{j1}+\cdots+a_nx_{jn}+\varepsilon_j$を満たす$y_j$をM1

組すべてについて行い、yの平均$\bar{y}$・標準誤差sを求める。同様のことをL回繰り返し$\bar{y}$とsの大まかな関係を規定する関数$F(y)$を求める。

再び最初の手順をM_2回実行し、求められたy_jから$s_j = F(y_j)$を計算する。そして中心をy_j、幅e_jを$e_j = \delta s_j$としたファジイ数$Y_j\,(j = 1,2,\cdots,M_2)$を作成する。ただしδは任意の定数とする。

以上のように作成したデータを用いて既存法と提案法それぞれによりファジイ係数$A_i\,(i = 1,2,\cdots,n)$を求める。そして再度最初の手順を繰り返すことによりx_jとY_jのデータを作成し、x_jとA_iより求まるY_jの推定値$\hat{Y}_j$とY_jの推定精度の比較を行う。

3.2 推定性能の比較検討

ここでは既存法と最小二乗法によるY_jの推定性能を比較検討する。任意に与えるパラメータの数値は、β＝0.1、γ=0.1、ξ=0.05、δ=3.0を採用したが、これに先立ち、予備的検討としてパラメータの値を変化させた場合の推定精度の比較を行った。その結果は講演時に示すものとするが、上記の設定値は特に最小二乗法に有利に作用していないので問題はないと言える。

表 1 は推定精度をY_jの中心と幅に分け両手法で比較したものである。まず、中心について見てみると、3変数の場合では両手法間に殆ど差がないが、7変数と 12 変数の場合では最小二乗法のほうが既存法に比べて RMSE 及び一致係数いずれの基準でも優位に立っていることがわかる。

これに対して幅の推定結果については、説明変数の数に関係なく、最小二乗法が圧倒的に既存法を凌駕している。これは、最小二乗法では$\hat{Y}_j$の幅がY_jの幅に出来るだけ近くなるようにファジイ係数A_iが定められるのに対し、既存法の最小化問題では$\hat{Y}_j$がY_jを含むようにA_iが決められてしまうため、どうしても推定幅が過大になってしまうからと思われる。

表 1 推定精度の比較

説明変数	推定変量	既存法	最小二乗法
3変数	中心	4.010	3.817
		0.959	0.961
	幅	11.783	1.397
		0.363	0.797
7変数	中心	10.456	4.365
		0.926	0.967
	幅	10.010	1.646
		0.440	0.809
12変数	中心	17.131	14.720
		0.960	0.966
	幅	23.227	4.700
		0.288	0.575

注)数値欄の上段はRMSE、下段は一致係数η

次にデータに誤差が含まれる場合の実験も行ってみた。データの転写や読み取りの際にミスが生じた場合を想定し、その影響が推定結果にどのような形で現れるかを見るためである。

ここでは以下の7つのケースを想定しシミュレーションを行った。

(ケース 1):ある 1 つのx_{ji}に誤差がある場合
(ケース 2):他の 1 つのx_{ji}に誤差がある場合
(ケース 3):上記 2 つのx_{ji}ともに誤差がある場合
(ケース 4):ある 1 つのy_jに誤差がある場合
(ケース 5):他の 1 つのy_jに誤差がある場合
(ケース 6):上記 2 つのy_jともに誤差がある場合
(ケース 7): 1 つのx_{ji}と 1 つのy_j同時に誤差がある場合

誤差の与え方は単純に$z' = z(1+\theta)$とする。ここでzは元データ、θは誤差率、z'は誤差を含んだデータである。

紙面の都合上、ここでは結果の表を割愛するが、結果から得られたことを次に示す。

ケース 1～3 では誤差がない場合に比べてあまり変化が見られなかった。それに対して、ケース 4～7 では、特に既存法のほうが誤差の影響をより強く受け、推定精度が悪化している結果が得られた。

また幅については、既存法ではそのモデル構造から予想される通りに推定精度が大きく悪化した。これに対して最小二乗法は殆ど影響を受けなかった。

4.おわりに

本研究では、ファジイ線形回帰モデルとして最小二乗法を提案し、シミュレーションを行い既存法との精度比較検討を行った。今後はさらに工夫を施した実験を行い、有効性をより確かなものとする必要がある。

参考文献

1)H.Tanaka,S.Uejima,K.Asai:Fuzzy Linear Regression Model,IEEE Trans.Systems Man Cybernet,10(4). pp.2933-2938(1980)
2)寺野、浅井、菅野:ファジイシステム入門、オーム社