

ファジィ回帰分析における残差分析

筑波大学大学院 システム情報工学研究科 佐藤 美佳

1. はじめに

昨今、ファジィ理論を用いて重みを推定し、それを利用した重みつき重回帰分析モデルが提案されている。これらのモデルは、従来の統計科学の分野で提案されている重みつき重回帰分析モデルと類似した構造を持ち、重みつき最小二乗法による推定が行われている。本報告では、これらのモデル間における回帰係数推定値の相違性の観点より、残差分析を行った結果について報告する。

2. ファジィ回帰モデルとその関連手法

つぎのモデルを定義する。[3]

$$\mathbf{y} = U_k X \beta_k + \mathbf{e}_k, \quad k = 1, \dots, K. \quad (1)$$

\mathbf{y} は目的変数に関するデータ、 X は p 個の説明変数に関する n 個のデータ行列、 \mathbf{e}_k は誤差、 K はクラスター数を示す。また、 β_k は回帰係数を示す。 U_k は、 X にファジィクラスタリングを適用した結果得られた個体 i のクラスター k に対する帰属度 u_{ik} , $i = 1, \dots, n$ を対角要素とし、非対角要素が 0 である $n \times n$ の対角行列を示す。(1) と関連して、ファジィc回帰モデル [2] が提案されている。これは

$$S = \sum_{k=1}^K S_k, \quad S_k = (\mathbf{y} - X \beta_k)^t (U_k)^m (\mathbf{y} - X \beta_k) \equiv \delta_k^t \delta_k, \quad \delta_k \equiv (U_k)^{\frac{m}{2}} (\mathbf{y} - X \beta_k), \quad m > 1 \quad (2)$$

を最小にする重みつき最小二乗推定量を求める。すなわち、目的変数に関するデータに存在する分類構造は説明変数に関するデータの分類構造と同一であることを仮定する。一方、(1) では、 $p \geq 2$ のとき、これらのデータ間の分類構造は必ずしも一致しないという点を考慮している。

さらに、関連する手法として、地理的重みつき回帰分析 [1] が提案されている。この方法は、(2) において重みをカーネル関数を用いて推定するものである。すなわち、重みの推定に確率構造を用いるかファジィ分類構造を用いるかの相違点がある。

また、 $\mathbf{z}_k = (z_{1k}, \dots, z_{pk})^t$ をファジィクラスター負荷量としたとき、つぎのモデルが提案されている。[4]

$$\mathbf{u}_k = X \mathbf{z}_k + \boldsymbol{\varepsilon}_k, \quad \mathbf{u}_k = (u_{1k}, \dots, u_{nk})^t, \quad k = 1, \dots, K, \quad (3)$$

このモデルを

$$\mathbf{1} = U_k^{-1} X \mathbf{z}_k + \tilde{\mathbf{e}}_k, \quad \tilde{\mathbf{e}}_k \equiv U_k^{-1} \boldsymbol{\varepsilon}_k, \quad \mathbf{1} = (1, \dots, 1)^t, \quad k = 1, \dots, K, \quad (4)$$

と変形することにより、分類構造を加味したファジィクラスター負荷量を推定するモデルが提案されている。[4] 本報告では、(1), (2), (3), (4) のモデルにおいて、誤差の正規性と均一分散の観点から残差分析を行った結果について報告する。

参考文献

- [1] Brunson, C., Fotheringham, S. & Charlton, M. (1998). Geographically Weighted Regression Modeling Spatial Non-Stationarity, *Journal of the Royal Statistical Society*, **47**, 431-443.
- [2] Hathaway, R.J. & Bezdek, J.C. (1993). Switching Regression Models and Fuzzy Clustering, *IEEE Transactions Fuzzy Systems*, **1**, 3, 195-204.
- [3] Sato-Ilic, M. and Matsuoka, T. (2002). On an Application of Fuzzy Clustering for Weighted Regression Analysis, *The 4th Conference of the Asian Regional Section of the International Association for Statistical Computing*, 55-58.
- [4] Sato-Ilic, M. (2004). On Fuzzy Clustering based Regression Models, *NAFIPS 2004 International Conference*, 216-221.