ショートノート

ファジィ数量化 I 類におけるファジィ群について†

渡辺 則生*

ファジィ数量化 I 類におけるファジィ群の役割について、線形回帰モデルを用いて考察した。最初に、ファジィ群と非線形モデルの線形近似との関係について議論した。次に、ファジィ群と線形回帰モデルの誤差分散との関係を検討する必要があることを指摘するとともに、異なるファジィ群のデータを用いたとき、結果にどの程度の差が生じるかをシミュレーションによって示した。

キーワード：ファジィ数量化 I 類、重み付き最小 2 乗法

1. ファジィ数量化 I 類と回帰モデル

量的な外的基準 y_n を、区間 $[0,1]$ 上の値をとる変数 $\mu_1(n), \dots, \mu_p(n)$ によって説明する方法としてファジィ数量化 I 類がある([2], y_n がファジィ数で与えられる場合もあるが([3]), ここでは数値の場合を考える)。ファジィ数量化 I 類においては、ファジィ群(B と表わす)が重みとして用いられる。ファジィ群 B のグレード $\mu_B(n)$ は、 n 番目の標本が、解析の対象とすべきデータの集合にどの程度属しているかを表わしていると考えられる。目的は、 y_n と $\sum_{i=1}^p a_i \mu_i(n)$ の差が小さくなるようなパラメータ $\{a_i\}$ を求めることである。 a_i を求める際、 $\mu_B(n)$ は重みという重要な役割を果たすので、 $\mu_B(n)$ を重みとして用いることの妥当性について検討する必要がある。しかし、この点についてはあまり考慮されずに応用されることが多いようである(応用例としては、たとえば[1]参照)。本稿の目的は、 $\mu_B(n)$ を重みとして用いることの妥当性について明らかにすることである。

ファジィ数量化 I 類の基本的な考え方は、 y_n と $\sum_{i=1}^p a_i \mu_i(n)$ の差をできるだけ小さくなるような a_i を求めることである。そこで $y_n = \sum_{i=1}^p a_i \mu_i$

$(n) + \varepsilon_n$ という理論モデルの存在を仮定する。 ε_n は誤差を表わす確率変数である。数量化理論は、必ずしも理論モデルを前提にした方法ではないが、方法の妥当性の理論的考察のためにはこのような仮定が適当となる。 ε_n の平均を 0 に限定しなければ、このモデルは十分一般的である。このモデルのもとで、ファジィ数量化 I 類は、 $\sum_{i=1}^p a_i \mu_i(n) \cdot \{y_n - \sum_{i=1}^p a_i \mu_i(n)\}^2$ を最小化する重み付き最小二乗法と同等となる。そこで問題となるのは、 $\mu_B(n)$ が回帰式にどのような影響を及ぼすのか、さらに、どのような $\mu_B(n)$ が妥当かという点である。本稿では、このような問題について、ふたつの観点から考察する。第一は、線形モデルをあてはめると重みの関係であり、第二は、不均一分散の線形モデルにおける、分散と重みの関係である。

理論モデルの一般形は $y_n = f_n(\mu(n)) + e_n$, $E(e_n) = 0$, $E(e_n^2) = \sigma_n^2(\mu(n))$ とおける。ただし $\mu(n) = (\mu_1(n), \dots, \mu_p(n))$ 。ここでは e_n が系列相関を持たない場合のみを扱い、 $E e_n e_m = 0 (n \neq m)$ とする。 f_n を線形モデルで近似するという立場から $f_n(\mu(n)) = \sum a_i \mu_i(n) + v_n(\mu(n))$ と表わす。 f_n が非線形ならば $v_n \neq 0$ であり、このことから、例えば通常の最小二乗推定量(OLSE)はバイアスを持つなどの問題が生じる。この点について 2 節で検討する。次に、 $v_n = 0$ の場合を考える。この場合は、不均一分散のもとでの推定問題となる。このとき、 $\mu_B(n) = 1/\sigma_n^2(\mu(n))$ ならば、ファジィ数量

† Note on the Fuzzy Group in the Fuzzy Quantification Method 1

Norio WATANABE

* 中央大学理工学部管理工学科

Department of Industrial and Systems Engineering, Chuo University

化 I 類の解である重み付き最小二乗推定量 (WLSE) は、最良線形不偏推定量 (BLUE) となり、その妥当性が保証される。しかし $\mu_B(n) \neq 1/\sigma_n^2$ ($\mu(n)$) の場合は一般には最良でなくなる。このような場合であっても $\mu_B(n)$ を用いるのは妥当だと考えられる状況がある。3 節においてはこのような問題について議論する。

2. 重みと非線形性

たとえば y_n がグレードのように有限区間の値を取るような変数であるとき、真のモデルは非線形となり、線形モデルは近似モデルとなる。本節では、本来非線形であるものを線形モデルで近似するという状況を考える。理論モデルは次式で与えられる。

$$y_n = f_n(\mu(n)) + e_n \quad (n=1, \dots, N) \quad (1)$$

ただし

$$\begin{aligned} f_n(\mu(n)) &= \sum_{i=1}^p a_i \mu_i(n) + v_n(\mu(n)), \\ Ee_n &= 0, \quad Ee_n^2 = \sigma_n^2(\mu(n)), \quad Ee_n e_m = 0, \\ &\quad (n \neq m) \end{aligned}$$

第 1 の問題は、どのような意味を持つ a_i を推定しようとしているのかという点にある。 a_i あるいは v_n は一意的ではなく、目的によって推定すべき a_i は変化する。簡単のために単回帰を考える ($p=2$, $\mu_1(n)$: 一定)。さらに f_n は n によらず同じ関数とする。問題は $f(x)$ をどのような直線で近似するかということになる ($x=\mu_2(n)$ とする)。たとえば、区間 $[0, 1]$ 全体での近似を目的とするのか、あるいは $x=0.5$ の近辺でよい近似になるような直線を目的とするのかといったことを明確にする必要がある。この目的が明確になれば、そのような係数 a_i を推定するためにどのような重みを用いるべきかが定まってくる。全区間において近似したければ重みを一定とするのが適当であり、0.5 近辺の近似の場合は、0.5 近辺の重みを重くし、0 や 1 の近くで小さくすればよい。

図1 非線形モデルの線形近似

次に、シミュレーションによる例を示す。図 1 の実曲線を真の $f(x)$ とするモデルから得られた $N=20$ のデータに対し、(a), (b), (c) の 3 種類の WLSE を求める。これを 100 回繰り返したときの標本平均値から定まる回帰直線が図 1 に描かれている。重みによって回帰直線が大きく変化しており、異なる重みを用いると、推定誤差では説明できない違いを持つ回帰直線を推定することになる。また、局所的により近似を得るためには、かなり極端な重みを用いる必要があることもわかる。

以上のことから、ファジィ群 B の数値は、近似の目的を反映すべきであると言える。結果を解釈するときには、 a_i がどのような意味を持っているか、 μ_B の意味に基づいて十分検討する必要がある。

ところで、[2] で与えられている例においては (P.103)、目的変数の値が小さいところと大きいところで μ_B が小さくなっていることから、目的変数の平均値の周辺であてはまりのよい回帰直線を求めていると解釈することも可能である。ただし [2] の例においては非線形性を積極的に指示する根拠はないようであり、特に問題にはならないと思われるが、一般には解釈の際に注意が必要である。すなわち、データの信頼性が高いところの μ_B を大きくしたり、説明変数や目的変数の値の大小で μ_B を決定するような場合、局所的な線形化を行っている可能性のあることに留意する必要がある。たとえば、 μ_i の値によっては y_n の予測値の信頼性が非常に低くなるといったこともありうる。

実際の解析においては線形近似の必要性は高く、ファジィ群のひとつの役割として認識する必要がある。しかし、重みに対し線形近似におけるこのような役割を積極的に与えることは、メンバーシップ関数値で与えられる「ファジィ群」本来の用い方と矛盾を引き起こす可能性がある。メンバーシップ関数値は、推定すべき真のモデルにあてはまる度合と解釈するのが自然であるが、この度合は、線形性が満たされる度合と対応するとは限らないからである。解析の際にはこの点にも注意する必要がある。

本節では、 σ_n^2 は考慮に入れなかった。これは ν_n が小さくないときには ν_n が主として問題となるからである。次節では、 $\nu_n=0$ の場合について、 σ_n^2 との関係について検討する。

3. 重みと誤差分散

次式で与えられる理論モデルを仮定する。

$$y_n = \sum_{i=1}^p a_i \mu_i(n) + e_n \quad (n=1, \dots, N) \quad (2)$$

ただし

$$Ee_n = 0, \quad Ee_n^2 = \sigma_n^2(\mu(n)), \quad Ee_n e_m = 0, \\ (n \neq m)$$

ファジィ数量化 I 類の解は、 $\sum_{n=1}^N \mu_B(n) \cdot \{y_n - \sum_{i=1}^p a_i \mu_i(n)\}^2$ を最小化する WLSE となる。ある定数 c が存在して $\mu_B(n) = c/\sigma_n^2$ となると、WLSE が線形不偏推定量のクラスの中かで最良である。従って、与えられたファジィ群 B のデータが、誤差項の分散に反比例しているとみなせる場合は、ファジィ数量化 I 類の方法の理論的妥当性が保証されることになる。しかし、このような条件を満たしているのは稀であり、一般には μ_B が c/σ_n^2 から少なからずずれているのが普通である。このような μ_B を用いてファジィ数量化 I 類を適用して得られる WLSE が、最良の WLSE に比較し、どの程度悪くなるのかを調べる必要がある。これらはともに不偏推定量になるので、分散(二乗平均誤差)によってその良し悪しを比較することができる。推定量の性質は μ_i や σ_n^2 に依存するため一般的に議論することは困難である。ここでは

シミュレーションによって検討を行う。

あらかじめ説明変数 $\mu_1(n), \dots, \mu_p(n)$ を $[0, 1]$ 上の一様乱数によって、また、 σ_n^2 を、逆数が $[0.4, 4]$ 上の一様分布に従うように乱数によって決めておく ($n=1, \dots, N$)。そして目的変数 y_n を次のモデルから生成する。

$$y_n = \mu_1(n) + \dots + \mu_p(n) + e_n \quad (3) \\ e_n \sim N(0, \sigma_n^2)$$

μ_B としては次の 5 つの場合を考え、5 種の WLSE を計算する。

$$(i) \quad \mu_B(n) = 1 \quad (\text{OLSE})$$

$$(ii) \quad \mu_B(n) = 1/\sigma_n$$

$$(iii) \quad \mu_B(n) = 1/\sigma_n^2 \quad (\text{BLUE})$$

$$(iv) \quad \mu_B(n) = 1/\sigma_n^4$$

$$(v) \quad \mu_B(n) \sim [0, 1] \text{ 上の一様分布}$$

一組の $\mu_1(n), \dots, \mu_p(n), \sigma_n$ については、 N 組のデータを 100 回発生し、各 WLSE の二乗平均誤差を求めた。各 WLSE について、 p 個の回帰係数(真値はすべて 1)についての二乗平均誤差の和を r として、 $R = r_{\text{BLUE}}/r$ と定義する。ここで r_{BLUE} は (iii) の BLUE の二乗平均誤差の和を表わす。このような各 R の計算を 100 回繰り返したときに得られる R のヒストグラムを図 2 に示す(ここでは見やすさのために折れ線グラフを用いた)。(ii) と (iv) は、最適な重みの大小関係を保存している重みになる。(i) は重みを用いない OLSE で、(v) はランダムに重みをつけた場合である。(ii) や (iv) の結果より、少々の重みの差はそれほど問題にならないが、(i) や (v) より、重みによって効率がかなり変化することがわかる。従って、 μ_B が分散の逆数のある程度反映しているとみなせるようなときには μ_B を用いて重み付き最小二乗法を用いるのが適当と言える。しかし、 μ_B が分散と関連付けられない場合、 μ_B を重みとして利用する際は慎重になる必要がある。従って、 μ_B と分散を結びつける根拠がない場合や分散の不均一の程度が大きくない場合には、効率をある程度犠牲にしても重みを用いない通常の最小二乗法のほうが適当と考えられる。

ファジィ数量化理論を適用するような状況では、

図2 2乗平均誤差の比 R の分布

μ_B は主観的に与えられることが多いと思われるが、このようなときには、最適な重みから大きくずれてしまう可能性が大きくなる。ファジィ数量化理論の適用結果の信頼性を考える場合には、 μ_B と分散の関係の検討が不可欠である。しかし μ_B が主観的に与えられている場合でも、その役割が重要であれば、適用結果の信頼性の保障は得られなくともファジィ数量化理論による結果の意義は失われなと言えよう。

なお、分散の均一性の検定法として、たとえば Goldfeld and Quandt [4] のノンパラメトリックな手法が応用できる ($1/\mu_B$ と分散の大小関係が同じという対立仮説を考えればよい)。

4. 考察

2, 3節の結果は、ファジィ数量化I類を適用する際に次のような点に留意しなければならないことを示している。

第一に、説明変数と目的変数間に非線形の関係があるかどうかをチェックする必要がある。非線形性がある場合、推定しようとしているものが重み μ_B に依存してくるので、与えられた μ_B のデータによって何が推定されるのかを明らかにしなければならない。

第二に、 μ_B が誤差分散と関わりがあるかどうかを考える必要がある。関わりがない場合、理論的妥当性は保証されないので、探索的な目的に利用するにとどめるか、別な観点から結果の妥当性について検討する必要がでてくる。一般に、一組のデータから個々の分散 σ_n^2 についての推論は行えないので、たとえば他のデータセットを用いて相互検証を行うなどの方法をとればよいであろう。理論的妥当性が保証されていない場合、恣意的結果を導くために重みを利用するという可能性も生じるため注意が必要である。

ファジィ数量化理論は、本来は質的な変数であっても、グレードというすでに数量化されたデータを用いた解析なので、通常重回帰分析の枠組みのなかで議論することは自然である(通常数量化理論においては本質的に質的な変数のままで扱っている)。従って、重回帰モデルに基づいて得られた以上の結果は妥当なものと言えよう。

参考文献

- [1] 江口、影井、若松：ファジィ数量化法の応用例、第9回ファジィシステムシンポジウム、797-800, 1993.
- [2] 寺野、浅居、菅野編著：「ファジィシステム入門」、オーム社、1987.
- [3] 和多田、田中、浅居：ファジィ数量化理論I類、行動計量学、11, 1, 66-73, 1983.
- [4] Goldfeld, S. M. and Quandt, R. E.: Some test for homoscedasticity, Journal of American Statistical Association, 60, 539-547, 1965.

(1994年9月16日 受付)
(1995年1月28日 再受付)

[問い合わせ先]

〒112 東京都文京区春日1-13-27

中央大学理工学部管理工学科

渡辺 則生

☎ : 03-3817-1949

☎ : 03-3817-1943

E-mail : watanabe@indsys.chuo-u.ac.jp

—— 著 者 紹 介 ——

渡辺 則生 (わたなべ のりお)

中央大学理工学部管理工学科

昭和55年 東京工業大学理学部数
学科卒業。昭和57年 東京工業大学
大学院総合理工学研究科システム科学
専攻修士課程修了。昭和60年 同博士
後期課程退学。昭和60年 自治医科大
学助手。平成元年 中央大学理工学部
講師。平成4年 同助教授、現在に至
る。理学博士。時系列解析、統計一般
理論の研究に従事。日本統計学会、日
本ファジィ学会などの会員。

Note on the Fuzzy Group in the Fuzzy Quantification Method 1

by

Norio WATANABE

Abstract :

Roles of the fuzzy group in the fuzzy quantification method 1 are studied using the linear regression model. First, relationship between the fuzzy group and linear approximation of a nonlinear model is discussed. Secondly, it is pointed out that the fuzzy group should be related to error variances of the linear regression model, and effect of fuzzy group data on regression analysis is studied by computer simulation.

Keywords : Fuzzy Quantification Method 1, Weighted Least Squares Method

Contact Address : **Norio Watanabe**

*Department of Industrial and Systems Engineering, Chuo University
Kasuga, Bunkyo-ku, Tokyo 112*

TEL : 03-3817-1949

FAX : 03-3817-1943