Fuzzy Logistic Regression Analysis for Fuzzy Input and Output Data

Kazuhisa Takemura
Department of Psychology, Waseda University,
1-24-1 Toyama, Shinjuku-Ku, Tokyo, 162-8644 JAPAN
e-mail: kazupsy@ waseda.jp

Abstract - **In this paper, a logistic regression analysis for fuzzy data was proposed. In the proposed analysis , both input data, output data, and parameters are represent-ed by L-R fuzzy numbers. Two kinds of estimation method for the fuzzy logistic regression model were proposed. : Possibilistic logistic regression analysis and the least square logistic regression analysis. An application method of the proposed analysis was illustrated. Theoretical and practical implications for the proposed analysis were discussed.**

Ⅰ. INTRODUCTION

Human judgment in social situations often involves vagueness concerning the confidence . People may not be able to make judgments without representing some confidence intervals. In order to measure the vagueness of human judgment, the fuzzy rating method has been proposed and developed [1]. In the fuzzy rating method, respondents select a representative rating point on a scale and indicate lower or upper rating points if they wish depending upon the relative vagueness of their judgment as shown in Figure 1. For example, the fuzzy rating method would be useful for measuring perceived temperature indicating the representative value and the lower or upper values. This rating scale allows for assymetries, and overcomes the problem, identified by Smithson[4], of researchers arbitrarily deciding the most representative value from a range of scores. By making certain simplifying assumptions (not uncommon within fuzzy set theory), the rating can be viewed as a L-R fuzzy number as shown in Figure 2, hence making possible the use of fuzzy set theoretic operations [1,2].

To analyze fuzzy rating data, it would be useful to apply fuzzy linear regression analysis, in which observed values and estimated values are assumed to have fuzziness. Although the original version of fuzzy linear regression analysis proposed by Tanaka et al.[10] assumed that while output data are fuzzy numbers, input data are not fuzzy numbers, more recently, Sakawa and Yano[3] formulated three types of multiobjective programming problems for obtaining fuzzy linear regression models, and developed a possibilistic linear regression analysis, where both input data and output data are fuzzy numbers. The possibilistic linear regression analysis (which is one version of the multiobjective fuzzy linear analysis), is considered to be a generalized method of previous fuzzy regression models.

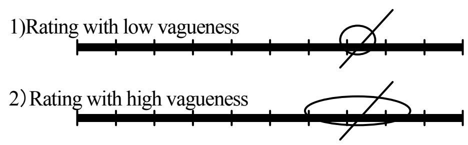


Figure 1. Example of fuzzy rating

This fuzzy linear regression method can be very effective for human related sciences such as psychology, sociology, and ergonomics, because most of the input data and output data for such sciences, are considered to be fuzzy.

Takemura [7,8,9] also developed the alternative fuzzy linear regression analysis using the least square method based on the extension principle in order to resolve the problems of the traditional fuzzy linear regression analysis [5,6] and to interpret the data more meaningfully from psychological point of view.

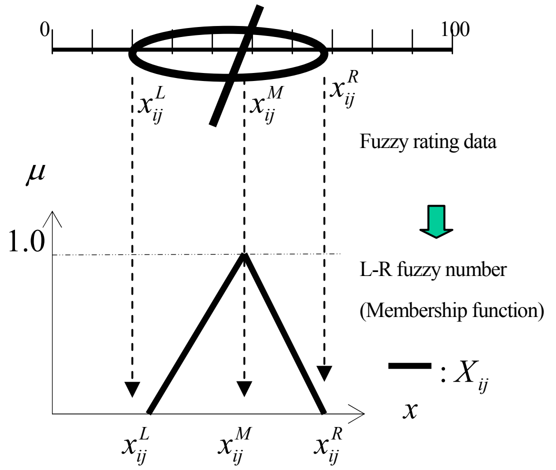


Figure 2. Fuzzy rating data and its representation by fuzzy number

However, these fuzzy regression analyses may fail to interpret psychological judgment data which have bounds of the psychological scales. For example, a perceived purchase probability has [0,1] interval and can not be greater than 1 or less than 0 shown in Figure 3. For such data, these fuzzy regression analyses may predict the values which are greater than 1 or less than 0.

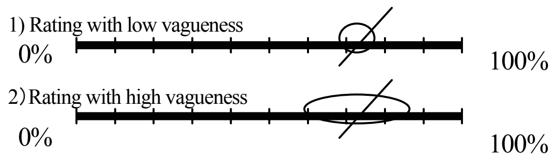


Figure 3. Fuzzy rating scale that has bounds

These events that the predicted values can be greater than highest bound or less than lowest bound can be problem if the predicted values are used in a subsequent analysis.

Therefore, the present study tried to solve this problem by setting predicted values that greater than the lowest value such as 0 and less than the highest value such as 1. The present study developed the concept of logistic regression for the crisp numbers, and then proposed the fuzzy version of logistic regression analysis for fuzzy input and output data. In the proposed analysis , both input data, output data, and parameters are represent-ed by L-R fuzzy numbers. Two kinds of estimation method for the fuzzy logistic regression model were proposed. : Possibilistic logistic regression analysis and the least square logistic regression analysis.

Ⅱ. A MODEL OF FUZZY LOGISTIC REGRESSION

A set of fuzzy input-output data for i-th observation is defined by:

$$\left(P_i; X_{i0}, X_{i1}, X_{i2}, \ldots, X_{im}\right),\ i = 1,2,\cdots,n \quad (1)$$

where P_i is a fuzzy dependent variable($P_i \in (0,1]$.), and X_{im} is a fuzzy independent variable represented by L-R fuzzy numbers. For simplicity, we assume that P_i *and* X_{im} are positive for any membership value, $\alpha \in (0,1]$.

The logistic function assumed in the fuzzy logistic regression analysis is represented in the Figure 4 , and the fuzzy logistic model is presented in the Figure 5. The fuzzy logistic regression model (where both input and output data are fuzzy numbers) is represented as follows:

$$\overline{\log(P_i \Theta (1-P_i))} = A_0 X_{i0} + A_1 \otimes X_{i1} + \ldots + A_m \otimes X_{im} \quad (2)$$

where $\overline{\log(P_i \Theta (1-P_i))}$ is estimated fuzzy log odds, Θ is the division operator based on the extension principle, $X_{i0} = 1$, $A_j\ (j = 0,1,\cdots,m)$ is fuzzy regression parameter represented by L-R fuzzy number, and $\otimes$ is the product operator based on the extension principle.

It should be noted that although the explicit form of the membership function of $\overline{\log(P_i \Theta (1-P_i))}$ can not be directly obtained, the α-level set of $\overline{\log(P_i \Theta (1-P_i))}$ can be obtained from the result of Nguyen's theorem [9].

Let $P_{i(\alpha)}^L$ be the lower bound of dependent fuzzy variable, $P_{i(\alpha)}^R$ be the upper bound of the fuzzy dependent variable. Then, α level set of the fuzzy dependent variable P_i can be represented as $P_{i\alpha} = \left[P_{i(\alpha)}^L, P_{i(\alpha)}^R\right]$, $\alpha \in (0,1]$.

Therefore, the α level set of the left term in (2) is as follow:

$$[\overline{\log(P \Theta (1-P))}]_\alpha$$

$$= [\min(\overline{\log(P_{i(\alpha)}^L / (1-P_{i(\alpha)}^L)}), \overline{\log(P_{i(\alpha)}^R / (1-P_{i(\alpha)}^R)}),$$

$$\max(\overline{\log(P_{i(\alpha)}^L / (1-P_{i(\alpha)}^L)}), \overline{\log(P_{i(\alpha)}^R / (1-P_{i(\alpha)}^R)})], \quad (3).$$

Let $z_{i(\alpha)}^L$ be a lower value of $\overline{\log(P_i \Theta (1-P_i))}_\alpha$, and $z_{i(\alpha)}^R$ be a upper value of $\overline{\log(P_i \Theta (1-P_i))}_\alpha$. Then,

$$Z_i = \left[z_{i(\alpha)}^L, z_{i(\alpha)}^R\right], \quad \alpha \in (0,1] \quad (4)$$

where

$$z^R_{i(\alpha)} = \sum_{j=0}^{m} \left\{ \max\left(a^R_{j(\alpha)} x^L_{ij(\alpha)}, a^R_{j(\alpha)} x^R_{ij(\alpha)} \right) \right\} \tag{5}$$

$$z^L_{i(\alpha)} = \sum_{j=0}^{m} \left\{ \min\left(a^L_{j(\alpha)} x^L_{ij(\alpha)}, a^L_{j(\alpha)} x^R_{ij(\alpha)} \right) \right\} \tag{6}$$

In above equation (4), $a^L_{j(\alpha)} x^L_{ij(\alpha)}$ is a product between lower value of α-level fuzzy coefficient for j -th attribute and α -level set of fuzzy input data Xij.

$a^L_{j(\alpha)} x^R_{ij(\alpha)}$, $a^R_{j(\alpha)} x^L_{ij(\alpha)}$ or $a^R_{j(\alpha)} x^R_{ij(\alpha)}$ is defined in the same manner respectively. It should be noted that $z^R_{i(\alpha)}$ and $z^l_{i(\alpha)}$, i =1,..,k, involve the minimization and maximization operators. In order to deal with these operators in $z^R_{i(\alpha)}$ and $z^l_{i(\alpha)}$, i =1,..,k, assume that the following relation hold for any fixed degree α :

$$a^L_{j(h)} \geq 0, \ \ j \in J_1 \tag{7}$$

$$a^L_{j(h)} \leq 0, \ \ a^R_{j(h)} \geq 0, \ \ j \in J_2 \tag{8}$$

$$a^R_{j(h)} \leq 0, \ \ j \in J_3 \tag{9}$$

$$\begin{aligned} &where \\ &j \in \{0, \cdots, m\} = J_1 \cup J_2 \cup J_3, \\ &J_1 \cap J_2 = \phi, \quad J_2 \cap J_3 = \phi, \quad J_3 \cap J_1 = \phi. \end{aligned} \tag{10}$$

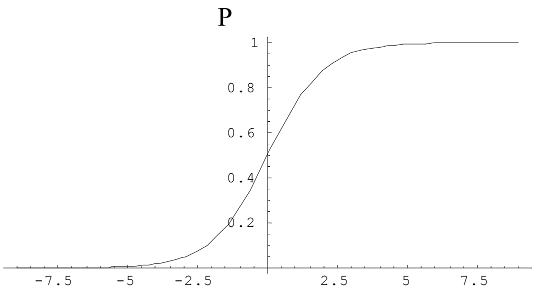


Figure 4. Logistic function assumed in the model

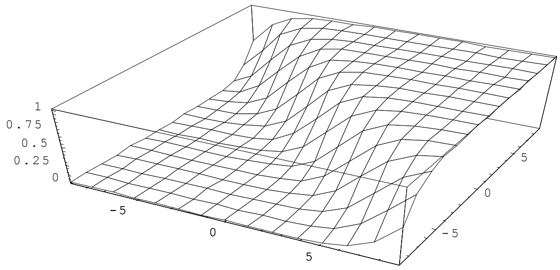


Figure 5. Fuzzy logistic function assumed in the model

Ⅲ. ESTIMAION METHOD USING MAXIMIZATION OF POSSIBILISTIC MEASURE

In order to estimate the fuzzy parameters of the fuzzy logistic regression model, two estimation methods can be proposed. One is an estimation method using maximization of possibilistic measure, and the other is the least square estimation method under constraints.

The former method is an application of the estimation of Sakawa and Yano's [3] possibilistic linear model where both input and output data are fuzzy numbers can be constructed . In the Sakawa & Yano' s [3] original formulation of the possibilistic linear regression analysis for fuzzy input-output data, it was assumed that input - output data and parameters were the symmetric fuzzy numbers. In the following, the formulation of the estimation method using maximization of possibilistic measure is slightly generalized by using the asymmetric numbers, because the fuzzy rating data can be often represented by asymmetric fuzzy numbers.

In the possibilistic linear regression model, the α -level set of $[\overline{\log(P\Theta(1-P))}]_{\alpha}$ in (2) can be obtained through resolving the minimization problem of the summation of width for the predicted fuzzy variable under the condition:

$$Pos\left(Z_i = \log(P_i\Theta(1-P_i)\right) \geq h \tag{11}$$

That is,

Objective function:

$$Min \quad \sum_{i=1}^{n} \left\{ \left(z^L_{i(\alpha=1)} - z^L_{i(\alpha=0)} \right) + \left(z^R_{i(\alpha=0)} - z^R_{i(\alpha=1)} \right) \right\} \tag{12}$$

subject to :

$$-z^L{}_{i(h)} \geq -y^R_{i(h)} \tag{13}$$

$$z^R_{i(h)} \geq y^L_{i(h)} \tag{14}$$

$$a^L_{j(h)} \geq 0,\ j \in J_1 \tag{15}$$

$$a^L_{j(h)} \leq 0,\ a^R_{j(h)} \geq 0,\ \ j \in J_2 \tag{16}$$

$$a^R_{j(h)} \leq 0,\ j \in J_3 \tag{17}$$

$$-a^L_{j(h)} + a^R_{j(h)} \geq 0 \tag{18}$$

$$\begin{aligned} &j \in \{0,\cdots,m\} = J_1 \cup J_2 \cup J_3, \\ &J_1 \cap J_2 = \phi,\quad J_2 \cap J_3 = \phi,\quad J_3 \cap J_1 = \phi, \end{aligned} \tag{19}$$

where, $i = 1,\ldots,n$. In the case of asymmetric fuzzy parameters , we should solve linear programming (LP) problem adding conditions (20) and (21). Therefore , in the case of symmetric case, we can assume that $c^L_j = c^R_j$.

$$a^L_{j(h)} = a^M_j - (1-h)c^L_j \tag{20}$$

$$a^R_{j(h)} = a^M_j + (1-h)c^R_j \tag{21}$$

The above problem can be reduced to the linear programming problem. Therefore, the problem can be solved by the ordinary linear programming method.

Ⅲ. THE LEAST SQUARE ESTIMAION METHOD UNDER CONSTRAINTS

The latter case of the fuzzy logistic regression method is using the least square method under constraints. In order to estimate the fuzzy parameters in this method, we should define some dissimilarity measure.

To define the dissimilarity between a predicted value and an observed value of the dependent variable, we adopt the following indicator $D\,\mathrm{i}(\alpha)^2$.

$$D_i\,(\alpha)^2 = (y^L_{i(\alpha)} - z^L_{i(\alpha)})^2 + (y^R_{i(\alpha)} - z^R_{i(\alpha)})^2 \tag{22}$$

Definition by equation (22) can be applied to interval data as well as L-R fuzzy number. That is, the equation (22) represents a sum of squares for distance between interval data.

To generalize, a dissimilarity indicator representing a square of distance for L-R fuzzy numbers can be written as follows:

$$Di^2 = \sum_{j=0}^{k} \mathrm{w_j}\ (\ (y^L_{i(\alpha j)} - z^L_{i(\alpha j)})^2 + (y^R_{i(\alpha j)} - z^R_{i(\alpha j)})^2\ ) \tag{23}$$

where $\alpha_j =$ j h/k, j =0,..,k, h is an equal interval, and w_j is a weight for j-th level.

In the case of triangular fuzzy number with w_j=1, the above equation is approximately represented as:

$$Di^2 = (y^L_{i(0)} - z^L_{i(0)})^2 + (y^L_{i(1)} - z^R_{i(1)})^2 + (y^R_{i(0)} - z^R_{i(0)})^2 \tag{24}$$

The proposed method is to estimate fuzzy coefficients using minimization of sum of D_i^2 respecting *i. That is,*

Objective function: $$Min \sum_{i=1}^{n} D_i{}^2 \tag{25}$$

Subject to :

$$a^L_{j(h)} \geq 0,\ j \in J_1 \tag{26}$$

$$a^L_{j(h)} \leq 0,\ a^R_{j(h)} \geq 0,\ \ j \in J_2 \tag{27}$$

$$a^R_{j(h)} \leq 0,\ j \in J_3 \tag{28}$$

$$-a^L_{j(h)} + a^R_{j(h)} \geq 0 \tag{29}$$

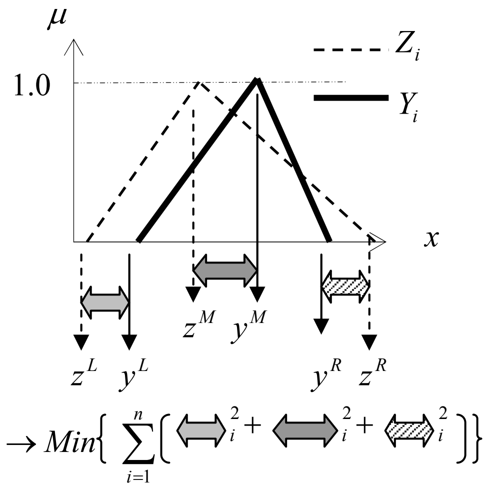
where

$$\begin{aligned} &j \in \{0,\cdots,m\} = J_1 \cup J_2 \cup J_3, \\ &J_1 \cap J_2 = \phi,\quad J_2 \cap J_3 = \phi,\quad J_3 \cap J_1 = \phi, \end{aligned} \tag{30}$$

$$z^L_{i(\alpha)} = \sum_{j \in J_1} a^L_{j(\alpha)} x^L_{ij(\alpha)} + \sum_{j \in J_2 J_3} a^L_{j(\alpha)} x^R_{ij(\alpha)} \tag{31}$$

$$z^R_{i(\alpha)} = \sum_{j \in J_1, J_{12}} a^R_{j(\alpha)} x^R_{ij(\alpha)} + \sum_{j \in J_3} a^R_{j(\alpha)} x^L_{ij(\alpha)} \tag{32}$$

The estimated coefficients can be derived through the quadratic programming method. The proposed fuzzy least square method is also shown in Figure 6.



Note. $Y_i = \log(P_i \Theta (1 - P_i))$, $Z_i = \overline{\log(P_i \Theta (1 - P_i))}$

Figure 6. Fuzzy least square method for fuzzy logistic regression analysis

Ⅴ. NUMERICAL EXAMPLE

Subject and Procedure The subjects rated ambiguous probability of preferring of a certain electric dictionary with seven different dictionaries. Three types of attributes information (Japanese dictionary 'Kojien' (Yes or No), Other dictionaries (4 types or 2 types), New roduct or Used Product) were manipulated . The subject answered of representative values, lower values, and upper values of their probabilities using fuzzy rating method. The fuzzy rating scale of desirability ranged from 0 point to 1 point such as a probability scale.

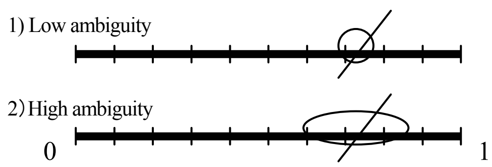


Figure 7. Example of fuzzy probability rating

Analysis and Results The fuzzy coefficients were obtained by the possibilistic logistic regression method as shown in Table 1. The fuzzy coefficients by the least square method were shown in Table 2.

Table 1. Results by Fuzzy Logistic Regression Analysis using the Possibilistic Regression Method

Attribute	Value	
Japanese Dictionary(L)	Lower	0.933
Japanese Dictionary (M)	Representative	1.113
Japanese Dictionary (R)	Upper	1.164
Other Dictionaries(L)	Lower	0.402
Other Dictionaries(M)	Representative	0.434
Other Dictionaries(R)	Upper	0.633
New or Used(R)	Lower	1..924
New or Used (M)	Representative	1. 924
New or Used(L)	Upper	3.201
Constant (L)	Lower	0.404
Constant (M)	Representative	0.628
Constant (R)	Upper	0.881

Table 2. Results by Fuzzy Logistic Regression Analysis using the Least Square Method

Attribute	Value	
Japanese Dictionary(L)	Lower	0.968
Japanese Dictionary (M)	Representative	0.968
Japanese Dictionary (R)	Upper	1.113
Other Dictionaries(L)	Lower	0.382
Other Dictionaries(M)	Representative	0.647
Other Dictionaries(R)	Upper	0.721
New or Used(R)	Lower	2.003
New or Used (M)	Representative	2.398
New or Used(L)	Upper	3.138
Constant (L)	Lower	0.381
Constant (M)	Representative	0.647
Constant (R)	Upper	0.721

According to Table 1 and 2, the stated probability was greater influenced by whether the target dictionary was new or used . The impact of Japanese dictionary ' Kojien' was slightly greater than that of other dictionaries . Both the estimation methods indicate similar results although the width of the fuzzy coefficients were slightly greater in the possibilisitic method than in the least square method.

Ⅵ. CONCLUSION

In the present paper, a fuzzy logistic regression analysis, where both input data and output data are represented by L-R fuzzy numbers the shapes of which are basically asymmetric, was proposed. Two kinds of estimation method for the fuzzy logistic regression model were proposed. : Possibilistic logistic regression analysis and the least square logistic regression analysis. The solutions of estimated parameters can be easily obtained by linear programming method for the posssibilistic regression method and ordinal quadratic programming method for the least square method. Numerical examples for both the methods were shown.

However, this study has a methodological restriction on statistical inferences for the fuzzy parameters. The future work on fuzzy logistic regression analysis of social judgment should be directed to statistical study of fuzzy logistic regression analysis such as statistical tests of parameters, outlier detection, and step-wise method for variable selection.

ACKNOWLEDGMENTS

This work was supported in part by Grant in Aids for Scientific Researchers, Japanese Ministry of Education, Science, and Culture. The author is grateful to S. Miyamoto, K. Nakamura, and D. Wakayama for their helpful comments.

REFERENCES

[1] R. Hesketh, Pryor, M. Gleitman, and T. Hesketh, Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale, in T.Zetenyi, Ed., *Fuzzy sets in psychology*, North-Holland, Amsterdam, 425-454, 1988

[2]M.Mizumoto, Fuzzy number, in Japan Society for Fuzzy Theory and Systems, Ed., *Fuzzy set*. Nikkan Kohgyo Shinbunsha, Tokyo, 157-185, 1992 (in Japanese).

[3]M. Sakawa and H.Yano, Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems*, **47**, 173-181, 1992.

[4]M.Smithson, *Fuzzy set analysis for behavioral and sciences*. Springer-Verlag, New York, 1987.

[5]K.Takemura, Prediction of behavioral intention by fuzzy multi-attribute attitude model: An application of possibilistic linear regression analysis to fuzzy rating data, *Proceedings of the 22nd Annual Conference of the Behaviormetric Society of Japan*, 122- 125, 1994. (in Japanese).

[6]K.Takemura, T.Hou, and J. Wo, Study on consumer's multi-attribute attitude in Taiwan: Using possibilistic linear regression analysis for fuzzy rating data, *Proceedings of the 11 th Fuzzy System Symposium,* 413-416, 1995.(in Japanese)

[7]K.Takemura.. A fuzzy linear regression analysis for fuzzy input-output data using the least squares method under linear constraints and its application to fuzzy rating data. *Journal of Advanced Computational Intelligence*, **3**(1), 36－40, 1999.

[8]K.Takemura, Vagueness in human judgment and decision making. In Z.Q. Liu and S.Miyamoto (Ed.s.), *Soft computing for human centered machines*, Springer Verlag, Tokyo, 249-281, 2000.

[9]K. Takemura, Fuzzy regression analysis for fuzzy input-output data using the least square method based on the extension principle, *Proceeding of the 17th Fuzzy System Symposium,* 799-804, 2001 (In Japanese).

[10]H.Tanaka, S.Uejima, and K.Asai, Linear regression analysis with fuzzy model, *IEEE Trans. Systems Man Cybernetics.* **12**, 903-907, 1982.