

Full Length Research Paper

Clinical effect of fuzzy numbers based on center of gravity

Rahim Saneifard^{1*} and Nader Hassasi²

¹Department of Applied Mathematics, Oroumieh Branch, Islamic Azad University, Oroumieh, Iran.

²Department of Mathematics, Malayer Branch, Islamic Azad University, Malayer, Iran.

Accepted 11 August, 2011

In this study, a model called “fuzzy reasoning model” was proposed for the case when the explanatory variables were crisp and the value of the binary response variable was reported as a number between zero and one. In this regard, the concept of possibilistic odds is also introduced. Then, the methodology and formulation of this model was used to explain the model parameters. Some goodness of fit criteria were proposed and a numerical example was given as an example.

Key words: Fuzzy number, defuzzification, centroid, clinical research.

INTRODUCTION

Since Jain, Dubis and Prade (Dubios et al., 1987) introduced the relevant concepts of fuzzy numbers, in classical set theory, an element either belongs to a set or it does not. In other words, the status of the element relative to the set is obvious. This property is related to the definition of that set. Some sets, such as the set of natural numbers equal or greater than 10 and the blood groups of the people, have such well defined and precise criteria that people with varied levels of knowledge are in total agreement regarding their members. Now, consider the set of tall men, the set of real numbers much greater than 10 or the set of patients with high blood pressure. There may be disagreement on the elements of these sets, because their definition contains imprecise or vague language. These imprecisely defined sets, called Fuzzy sets, play an important role in human thought. In many scientific researches, linguistic, rather than numerical terms are frequently used. For instance, in clinical research, to measure the severity of disease or pain in patients, linguistic terms like low, medium, high, etc, are used. These terms can be viewed as fuzzy sets. Moreover, the border line between these fuzzy sets is not crisp even if they are measured in numerical scale. In other words, cases in the neighborhood of the borderline have a vague status with regard to the disease.

A similar situation occurs in the definition of

hypertension. To model the relationship between these observations, an ordinary statistical model which is based on certain assumption and exact observations is not a good choice. We note that fuzzy models as compared to the usual statistical models consider possibilistic rather than probabilistic errors. In other words, there are some aspects of uncertainty that measure the vagueness of the phenomena and cannot be summarized in random terms. This kind of uncertainty is evaluated by a measure called possibility. Hence error terms are deleted in fuzzy regression models and are, in fact, hidden in the fuzzy coefficients. Almost all previous studies on fuzzy reasoning have focused on linear models and nonlinear models have been seldom considered.

A nonlinear reasoning model which is widely used in research, especially in classical clinical studies, is the reasoning model. In practice, there are many situations in which the ordinary reasoning method cannot be used. In this situation, the variations of the model error terms cannot be attributed wholly to the randomness of the phenomenon. Furthermore, the probabilistic assumptions of the reasoning model are not fulfilled. To deal with such situations, we combined the reasoning model with fuzzy set theory to present a new model which we called the fuzzy reasoning model.

Preliminaries

This was done with a simple definition of a fuzzy set, along with some definitions which are used in this paper

*Corresponding author. E-mail: srsaneifard@yahoo.com.

(Zimmerman, 1991; Tanaka et al., 1982).

Definition 1. The fuzzy set M of the real line \mathfrak{R} is called a fuzzy number if it is normal convex fuzzy set of \mathfrak{R} .

Definition 2. A fuzzy number N is called LR -type if it has the membership function as follows:

$$N(x) = \begin{cases} L\left(\frac{m-x}{\sigma}\right), & x \leq m, \\ R\left(\frac{x-m}{\beta}\right), & x > m. \end{cases} \quad \forall x \in \mathfrak{R}$$

Where, L and R are decreasing shape functions from \mathfrak{R}^+ to $0, 1$, with $L(0) = 1, L(x) < 1$ for all $x > 0, L(x) > 0$ for all $x < 1$ and $L(1) = 0$. Similar conditions hold for \mathfrak{R} . The real number m is called the mean value of N , while σ and β are called the left and the right spreads, respectively. Symbolically N is denoted by $(m, \sigma, \beta)_{LR}$. In special cases, $L(x) = R(x)$, N is called triangular fuzzy number and is denoted by $(m, \sigma, \beta)_T$. Its membership function is:

$$N(x) = \begin{cases} 1 - \frac{m-x}{\sigma}, & m - \sigma \leq x \leq m, \\ 1 - \frac{x-m}{\beta}, & m < x \leq m + \beta. \end{cases}$$

If in addition, $\sigma = \beta$, then N is denoted by $(m, \sigma)_T$ and is called a symmetric triangular fuzzy number.

Definition 3. Let $M = (m, \sigma, \beta)_{LR}$ be a fuzzy number and $\lambda \in \mathfrak{R}$. Then;

$$\lambda M = \begin{cases} (\lambda m, \lambda \sigma, \lambda \beta)_{LR}, & \lambda > 0, \\ (\lambda m, -\lambda \beta, -\lambda \alpha), & \lambda < 0. \end{cases}$$

Definition 4. Let $M = (m, \sigma, \beta)_{LR}$ and $N = (n, \gamma, \delta)_{LR}$ be two LR -type fuzzy number. Then;

$$M + N = (m, \sigma, \beta)_{LR} + (n, \gamma, \delta)_{LR} = (m + n, \sigma + \gamma, \beta + \delta)_{LR}$$

There are several defuzzification methods introduced in literatures. A common method which is used in this paper is as follows.

Definition 5. Let W be a fuzzy set, then defuzzification of W is defined as a real number as follows:

$$D(W) = \frac{\int_x xW(x)dx}{\int_x W(x)dx}$$

METHODOLOGY AND FORMULATION

Consider the data set $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = 1, 2, \dots, n$, where X_i is the vector of crisp observation on the independent variables like sex, age, marital status, weight, blood pressure and blood cholesterol for the i -th case. μ_i , the corresponding response observation, is a number in $[0,1]$ and indicates the possibility of i -th case having the relevant property i.e. $\mu_i = Poss(Y_i = 1)$. Therefore, the fuzzy reasoning model with fuzzy coefficients, are as follows:

$$P_i = b_0 + b_1x_{i1} + \dots + b_nx_{in}, \quad i = 1, \dots, m. \quad (1)$$

b_0, b_1, \dots, b_n are the model parameters which are treated as fuzzy numbers and $P_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$ is the estimator of the logarithmic transformation of possibilistic odds. To simplify the calculation, we assume that the fuzzy numbers $b_j = (a_{cj}, s_{Lj}, s_{Rj})_T$, $j = 1, \dots, n$ are triangular. Then $P_i, i = 1, \dots, m$ is the triangular fuzzy number. We have;

$$W_i = (f_{ci}(x), f_{Lis}(x), f_{Ris}(x))_T,$$

Where:

$$\begin{cases} f_{ci}(x) = a_{c0} + a_{c1}x_{i1} + \dots + a_{cn}x_{in}, \\ f_{Lis}(x) = s_{L0} + s_{L1}x_{i1} + \dots + s_{Ln}x_{in}, \\ f_{Ris}(x) = s_{R0} + s_{R1}x_{i1} + \dots + s_{Rn}x_{in}. \end{cases} \quad (2)$$

So, the membership function of the fuzzy estimated output can be shown as follows:

$$P_i(p_i) = \begin{cases} 1 - \frac{f_{ci}(x) - p_i}{f_{Lis}(x)}, & f_{ci}(x) - f_{Lis}(x) \leq p_i \leq f_{ci}(x), \\ 1 - \frac{p_i - f_{ci}(x)}{f_{Ris}(x)}, & f_{ci}(x) \leq p_i \leq f_{ci}(x) + f_{Ris}(x). \end{cases} \quad (3)$$

As it is known, P_i is the natural logarithm of possibilistic odds of getting or having the known property for the i -th case. According to the extension principle, if M is a fuzzy number with membership function P_i and $f(x) = e^x$, then $f(M) = e^M$ is a fuzzy number with the following membership function:

$$e^{M(x)} = \begin{cases} M(\ln x), & x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

So, after estimating the model coefficients, we can determine the membership function of the possibilistic odds $e^{P_i(x)}$, $x > 0$ as

follows:

$$e^{P_i(x)} = P_i(\ln x) = \begin{cases} 1 - \frac{f_{ci}(x) - \ln(x)}{f_{Lis}(x)}, & f_{ci}(x) - f_{Lis}(x) \leq \ln(x) \leq f_{ci}(x), \\ 1 - \frac{\ln(x) - f_{ci}(x)}{f_{Ris}(x)}, & f_{ci}(x) \leq \ln(x) \leq f_{ci}(x) + f_{Ris}(x). \end{cases} \quad (5)$$

Thus, for a new fuzzy case, our model can predict its possibilistic odds as a fuzzy number by the use of the crisp vector of input observation.

Estimation of fuzzy coefficients

In order to estimate the coefficients $b_j, j = 0, \dots, m$, we followed the possibilistic approach in fuzzy linear reasoning models with fuzzy output, fuzzy coefficients and non-fuzzy input vector. The basic idea is to minimize the fuzziness of the obtained model by minimizing the total support of the fuzzy coefficients. So, it is assumed that:

1) Each observation, p_i has a membership degree as big as h in the function of the fuzzy estimated output, P_i , that is;

$$P_i(p_i) \geq h \text{ and } w_i = \ln\left(\frac{\mu_i}{1 - \mu_i}\right), h \in (0,1). \quad (6)$$

2) The fuzzy coefficients $(b_j, j = 0, \dots, m)$ are such that the fuzziness of the model is minimized. Since the fuzziness of a fuzzy number increases with its spreads, minimizing the sum of the spread of fuzzy outputs leads to a minimum value of the fuzziness of the model.

The determination of fuzzy coefficient leads to a linear programming problem, in which the objective function is the sum of the spreads of the fuzzy outputs,

$$Z = m(s_{L0} + s_{R0}) + \sum_{j=1}^n \left[(s_{Lj} + s_{Rj}) \sum_{i=1}^m x_{ij} \right] \quad (7)$$

Where, x_{ij} is the value of the i -th observation for the j -th variable. On the other hand, from Equation (5) each constraint of the problem $(P_i(p_i) \geq h, i = 1, \dots, n)$ can be written as follows:

$$1 - \frac{f_{ci} - p_i}{f_{Lis}} \geq h \Rightarrow (1-h)s_{L0} + (1-h) \sum_{j=1}^n s_{Lj} x_{ij} - a_{c0} - \sum_{j=1}^n a_{cj} x_{ij} \geq -p_i,$$

$$1 - \frac{p_i - f_{ci}}{f_{Ris}} \geq h \Rightarrow (1-h)s_{R0} + (1-h) \sum_{j=1}^n s_{Rj} x_{ij} - a_{c0} - \sum_{j=1}^n a_{cj} x_{ij} \geq p_i. \quad (8)$$

Hence we have a total of $2m$ constraints. One can minimize the

objective function using linear programming algorithms such as the simplex method to estimate the mode value, and the left and right spreads of each coefficient.

Definition 6. Consider the fuzzy reasoning model which is derived based on m crisp observations. Then, the mean degree of memberships for observed values in the membership function of the estimated ones, that is

$$MDM = \frac{1}{m} \sum_{i=1}^m P_i(p_i) = \frac{1}{m} \sum_{i=1}^m e^{P_i\left(\frac{\mu_i}{1-\mu_i}\right)} \quad (9),$$

Which is used as an index for evaluating the model. Large membership degrees of the observed values confirm that the model constructed from these data supports the data well. The maximum value of the MDM is 1 and the minimum value is 0. So, a value near 1 indicates good model fitting.

Definition 7. Consider the fuzzy reasoning model with crisp input

observations, $\frac{\mu_i}{1 - \mu_i}$, and fuzzy estimated outputs e^{P_i} , the mean of squares errors index of the model is defined as follows:

$$MSE = \frac{1}{m} \sum_{i=1}^m \left[D(e^{P_i}) - \frac{\mu_i}{1 - \mu_i} \right] \quad (10)$$

in which, $D(e^{P_i})$ is the defuzzification of e^{P_i} .

Application in clinical studies

In clinical diagnosis, the performance of all screening tests depends on the cut off point used to separate normal and abnormal individuals. A too high cut off point may cause abnormal individuals to be classified as normal and a too low cut off point may classify healthy individuals as abnormal. As an example, consider diabetes mellitus (DM) which is a common metabolic disorder that shares the phenotype of hyperglycemia. Several distinct types of DM exist and are caused by a complex interaction of genetics and environmental factors (Saneifard, 2009a; Saneifard et al., 2011). There are no widely accepted or rigorously validated cut off points for defining positive screening tests for diabetes in non pregnant adults (Saneifard, 2009b; Yen et al., 1999). The latest suggested cut off point for fasting plasma glucose is less than 100 for normal and higher than 126 mg/dl for abnormal cases, while for 2 h postprandial plasma glucose is less than 140 for normal and 200 mg/dl for abnormal cases (Ruoning, 1997). However, identical cut off point in all texts does not guarantee the crispness. In other words, in the neighborhood of the cut off point, a little increase or decrease in blood plasma glucose cannot change the individual's status from normal to abnormal. In this case, physicians do not rely on one result and repeat the examination to reach more reliability. Also, they consider any other clinical symptoms of diabetes to decide the classification. However, they admit that the status of patient 2 h postprandial plasma glucose is in the 140 to 200 mg/dl interval or whose fasting plasma glucose is in the interval 100 to 126 mg/dl, is unknown. Fuzziness believes that degree of vagueness in this interval is not the same.

Example

In order to determine the diabetic status of a community in a clinical survey, a sample of 2 h postprandial plasma glucose of each person was made available. Considering 200 mg/dl as cut off point, it was found that 15 cases fell in the interval of 140 to 200 mg/dl. In order to predict the possibilistic odds of diabetes for these vague

statues cases, additional information such as sex, age, body mass index (BMI), family history, and 2 h plasma glucose (mg/dl) which are shown to be significant risk factors in diabetes (Kim and Bishu, 1998), were used. We consulted with an expert to assign a possibility of the disease to each case. Then, the following possibilistic model was fitted:

$$P_i = b_0 + b_1(\text{sex}_i) + b_2(\text{Blood glucose}_i) + b_3(\text{Age}_i) + b_4(\text{BMI}_i) + b_5(\text{family history}_i), i = 1, \dots, 15.$$

Where, for simplicity in computations, the reasoning coefficients $b_j = (a_{cj}, s_{Lj}, s_{Rj})_T, j = 0, 1, \dots, 5$ are assumed to be triangular fuzzy numbers and the fuzziness of the variables relationships is hidden in these coefficients. Depending on the definition of the coefficients in fuzzy models, there are two types of models:

- 1) The model with symmetric coefficients: b_j in which $s_{Lj} = s_{Rj} = s_j$, for some $j = 0, 1, \dots, 5$.
- 2) The model with non symmetric coefficients: b_j in which $s_{Lj} \neq s_{Rj}$, for some $j = 0, 1, \dots, 5$.

Both models for different h values were fitted to our data and

results obtained were similar. So, for simplicity in computation and interpretation, we chose the symmetrical model to fit our data. Now, to decide about the h value, we fitted the symmetrical model for several values of h and observed its effect on the model coefficients.

As shown, changing h values do not change the coefficient centers (a_{cj}) but affects the spreads (s_j) and objective function values such that the vagueness of the fuzzy outputs increased with the h values. So, based on the Z values and the vagueness of the outputs, it seems that the value 0.6 is the rational choice for h . Now the symmetrical possibilistic reasoning model with $h = 0.6$ is fitted to our data. As aforementioned, to fit the model by linear programming methods and in order to determine the coefficients $b_j, j = 0, \dots, n$, the objective function should be minimized in such a way that two constraints for each observation are satisfied. The objective function in our example is:

$$Z = 2 \left(15s_0 + s_1 \sum_{i=1}^{15} \text{Sex}_i + s_2 \sum_{i=1}^{15} \text{Blood glucose}_i + s_3 \sum_{i=1}^{15} \text{Age}_i + s_4 \sum_{i=1}^{15} \text{BMI}_i + s_5 \sum_{i=1}^{15} \text{Family history}_i \right) = 2(15s_0 + 8s_1 + 2576s_2 + 744s_3 + 427s_4 + 8s_5) \quad (11)$$

This function should be minimized under 30 constraints. Using Lingo software, the above linear programming problem was solved and coefficients were as follows: $a_{c0} = -15.88, a_{c1} = 0.48, a_{c2} = 0.09, a_{c3} = 0.07, a_{c4} = -0.11, a_{c5} = 0.49, s_0 = 0.0,$

$s_1 = 0.58, s_2 = 0.0, s_3 = 0.0, s_4 = 0.01, s_5 = 1.13$. The minimized value of the objective function was $Z = 44.02$, and the optimal model was obtained as:

$$P = -15.88 + (0.48, 0.58)_T \text{Sex} + 0.09 \text{Blood glucose} + 0.07 \text{Age} + (-0.11, 0.019)_T \text{BMI} + (0.49, 1.13)_T \text{Family history},$$

This formula can estimate the possibility odds of diabetes for a case that is suspected in a diabetic status. Note that the estimated possibility odds for each case is a fuzzy output. Now, suppose there is a new person with $x_1 = 0, x_2 = 165, x_3 = 40, x_4 = 25,$

$x_5 = 0$. Then, according to model 11, the possibilistic odd of diabetes for this person is obtained as:

$$P_{new} = -15.88 + (0.48, 0.58)_T \times (0) + 0.09 \times 165 + 0.07 \times 40 + (-0.11, 1.13)_T \times 25 + (0.49, 1.13)_T \times 1 = (-15.88, 0)_T + (15.29, 0)_T + (2.90, 0)_T + (-2.85, 0.48)_T + (0.49, 1.13)_T = (-0.04, 1.61)_T = (0.04, 1.61)_T,$$

$$e^{P_{new}(x)} = \begin{cases} 1 - \frac{-0.04 - \ln x}{1.62}, & -1.66 \leq \ln x \leq -0.04, (0.19 \leq x \leq 0.96), \\ 1 - \frac{\ln x + 0.04}{1.62}, & -0.04 \leq \ln x \leq 1.58, (0.96 \leq x \leq 4.85). \end{cases}$$

Table 1. Distance between the defuzzified estimate and its related observed one.

Id	$\frac{\mu_i}{1-\mu_i}$	$D(e^{P_i})$	$D\left[e^{P_i} - \frac{\mu_i}{1-\mu_i}\right]$
1	0.11	0.22	0.11
2	0.18	0.27	0.09
3	0.54	0.93	0.39
4	0.72	0.68	0.04
5	0.96	3.30	2.34
6	1.00	1.78	0.78
7	1.50	2.14	0.64
8	1.50	1.30	0.20
9	2.33	3.22	0.89
10	5.67	4.96	0.71
11	9.00	16.5	7.56
12	32.3	36.1	3.86
13	19.0	16.03	2.97
14	19.0	16.31	2.69
15	99.0	108.8	9.88

So, we can say that, the possibilistic odd of diabetes for this new case is about 0.96. Finally, to evaluate the model, we used the two criteria proposed earlier to find MDM index, we calculated the membership degree of each observed odds in the membership function of its related fuzzy output. Finally we obtain,

$$MDM = \frac{1}{m} \sum_{i=1}^m e^{P_i \left(\frac{\mu_i}{1-\mu_i} \right)} = \frac{1048}{15} = 0.7.$$

The MDM value was much greater than 0.5, indicating a good fit. To obtain the index MSE, we defuzzified each output and then calculated its distance to the corresponding observation (Table 1). For example, in case 6, by the COG defuzzification method, we have:

$$D(e^{P_6}) = \frac{\int_{e^{0.35-0.97}}^{e^{0.35}} x \left(1 - \frac{0.35-\ln x}{0.97} \right) dx + \int_{e^{0.35}}^{e^{0.35+0.97}} x \left(1 - \frac{\ln x - 0.35}{0.97} \right) dx}{\int_{e^{0.35-0.97}}^{e^{0.35}} \left(1 - \frac{0.35-\ln x}{0.97} \right) dx + \int_{e^{0.35}}^{e^{0.35+0.97}} \left(1 - \frac{\ln x - 0.35}{0.97} \right) dx} = 1.78$$

Table 1 shows the results of the calculations. The MSE value of the model was obtained as:

$$MSE = \frac{1}{m} \sum_{i=1}^m \left[D \left(e^{P_i} - \frac{\mu_i}{1-\mu_i} \right) \right]^2 = 12.91.$$

As aforementioned, this index shows the mean distance between the observed and the estimated response values. Its small value

confirms a good fit. Unfortunately, there are not any critical values with which to compare our indices. It seems that the indices, like MSE, are useful when we are interested in comparing several fuzzy models for the same data set and choosing the best one. The model with the smallest MSE value is chosen.

Conclusion

In this paper, the details of a fuzzy reasoning are discussed and a numerical example of its application to clinical studies is given. The proposed model is allied when the observations of the binary response variable are vague, but the observations of the explanatory variables are precise. Since the binary response variable has a vague status the $P(Y) = 1$ is not definable and the probability odds cannot be calculated. The value of μ_i detects the degree of adjustment to the category 1 criteria of the response variable for i -th case and is determined by a clinical expert. In our model it is a number between 0 and 1. The proposed model can therefore be used in other research areas with similar situations.

REFERENCES

Dubios D, Prade H (1987). Logistic regression in the medical literature. J. Clin. Epidemiol. 24: 279-300.
 Kim B, Bishu RR (1998). Evaluation of Fuzzy Linear Regression Models by Comparing Membership Functions. Fuzzy Sets Syst. 100: 343-352.
 Ruoning X (1997). S-curve Regression Model in Fuzzy Environments. Fuzzy Sets Syst. 90: 317-326.

- Saneifard R (2009a). Ranking L-R fuzzy numbers with weighted averaging based on levels. *Int. J. Ind. Mathematics*, 2: 163-173.
- Saneifard R (2009b). A method for defuzzification by weighted distance. *Int. J. Ind. Mathematics*, 3: 209-217.
- Saneifard R, Saneifard R (2011). Evaluation of Fuzzy Linear Regression Models by Parametric Distance. *Aust. J. Basic Appl. Sci.* 5: 261-276.
- Tanaka H, Uejima S, Asay K (1982). Linear Regression Analysis With Fuzzy Model. *IEEE Trans. Syst. Man Cybernet.* 12: 903-907.
- Yen KK, Ghosray S, Roig G (1999). A Linear Regression Model Using Triangular Fuzzy Number Coefficient. *Fuzzy Sets Syst.* 106: 167-177.
- Zimmerman HJ (1991). *Fuzzy sets theory and its applications*. Kluwer Academic Press. Dordrecht.