Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 2, pp. 445~449

A Note on Linear Regression Model Using Non-Symmetric Triangular Fuzzy Number Coefficients

Dug Hun Hong[1] · Kyung Tae Kim[2]

Abstract

Yen et al. [Fuzzy Sets and Systems 106 (1999) 167-177] calculated the fuzzy membership function for the output to find the non-symmetric triangular fuzzy number coefficients of a linear regression model for all given input-output data sets. In this note, we show that the result they obtained in their paper is invalid.

Keywords : Fuzzy regression analysis, Fuzzy triangular coefficients, Minimization of fuzziness, Non-symmetric coefficients

The following model shows the dependence of the output variable on the inputs variables,

$$\widetilde{Y}=f(\mathbf{x},\widetilde{A})=\widetilde{A}_0+\widetilde{A}_1x_1+\cdots+\widetilde{A}_nx_n, \tag{1}$$

where $\widetilde{Y}$ is the fuzzy output, $\mathbf{x}=[x_1,x_2,\cdots,x_n]^T$ is the real-valued input vector, and $\widetilde{A}=\{\widetilde{A}_0,\ \widetilde{A}_1,\cdots,\ \widetilde{A}_n\}$ is a set of fuzzy numbers.

The membership function for the set $\widetilde{Y}$ is defined by Zadeh's extension principle as follows:

$$\mu_{\widetilde{Y}}(y)=\begin{cases}\max_{(a_1,\cdots,a_n)=f^{-1}(y,x)}\{\min_j\mu_{\widetilde{A}_j}(a_j)\} & \text{if } f^{-1}(y,x)\neq\phi,\\ 0 & \textit{otherwise}.\end{cases}$$

Then the regression analysis problem is defined as: given a set of crisp data

1) First Author : Professor, Department of Mathematics, Myongji University, Kyunggi 449-728, Korea
E-mail : dhhong@mju.ac.kr

2) Professor, Department of Electronics and Electrical Engineering, Kyungwon University, Kyunggi 461-701, Korea
E-mail : ktkim@kyungwon.ac.kr

points $\langle x_1, y_1\rangle$, $\langle x_2, y_2\rangle$, $\cdots$, $\langle x_m, y_m\rangle$, we want to find a set of fuzzy parameters $\widetilde{A}_0$, $\widetilde{A}_1$, $\cdots$, $\widetilde{A}_n$ for the Eq.(1) which is the best fit to the given data points, according to some criteria of goodness of fit.

In (1), $\widetilde{A}_i$ is the fuzzy coefficient of the variable x_i in the regression model of additive form. If $\widetilde{A}_i$s have triangular membership functions, then each fuzzy number coefficient $\widetilde{A}_i$ can be uniquely defined by

$$\widetilde{A}_i = \{a_i^L, a_i^C, a_i^U\},$$

where a_i^L is the lower limit, a_i^U is the upper limit, and a_i^C is the point having the property that $\mu_{\widetilde{A}_i}(a_i^C) = 1$. The property of symmetry of the fuzzy coefficient $\widetilde{A}_i$ enables us to establish the following two relations:

$$a_i^C = (a_i^L + a_i^U)/2$$

and

$$a_i^S = a_i^C - a_i^L = a_i^U - a_i^C,$$

where a_i^C is the center and a_i^S the spread of $\widetilde{A}_i$.

For symmetric triangular fuzzy number coefficients, the membership function $\mu_{\widetilde{A}_i}$ for $\widetilde{A}_i$, $i=1,\cdots,n$ can be described as

$$\mu_{\widetilde{A}_i}(a_i) = \begin{cases} 1-(a_i^C - a_i)/a_i^S, & a_i^C - a_i^S \le a_i \le a_i^C, \\ 1-(a_i - a_i^C)/a_i^S, & a_i^C \le a_i \le a_i^C + a_i^S, \\ 0 & otherwise. \end{cases}$$

Having established the membership function for each fuzzy coefficient $\widetilde{A}_i$, the fuzzy output from the linear model $f(\mathbf{x}, \widetilde{A})$ in (1) can be expressed according to the principle of extension and fuzzy arithmetic on fuzzy numbers [1] as

$$\widetilde{Y} = f(\mathbf{x}, \widetilde{A}) = (f^C(\mathbf{x}), f^S(\mathbf{x})),$$

where $f^C(x)$ is the center of the fuzzy linear model $f(\mathbf{x}, \widetilde{A})$ and has the form

$$f^C(\mathbf{x}) = a_0^C + a_1^C x_1 + \cdots + a_n^C x_n$$

and $f^S(\mathbf{x})$ is the spread of $f(\mathbf{x}, \widetilde{A})$ and defined as

$$f^S(\mathbf{x}) = a_0^S + a_1^S|x_1| + \cdots + a_n^S|x_n|.$$

where $[f(\mathbf{x}_j)]_h = [\tilde{A}_0]_h + [\tilde{A}_1]_h x_{j1} + \cdots + [\tilde{A}_n]_h x_{jn}$ such that $[\cdot]_h$ represents the h-level set of a fuzzy number.

In regression , the goal is to find the fuzzy coefficients that minimize the above-mentioned spread of fuzzy output for all the data sets. The cost function, Z , to be minimized can be written as

$$Z = a_0^S + \sum_{i=1}^{n}\left[a_i^S \sum_{j=1}^{m} |x_{ji}| \right]$$

which can also be expressed as

$$\text{Minimize} \quad Z = f^S(\mathbf{x}_1) + f^S(\mathbf{x}_2) + \cdots + f^S(\mathbf{x}_m)$$

subject to the set of constraints

$$y_j \in [f(x_j)]_h.$$

However, if these triangles are not symmetric , we need minimum three parameters to uniquely describe each. For example, $\tilde{A}_i$ can be described by the triplet $\{a_i^L, a_i^P, a_i^U\}$ or by $\{s_i^L, a_i^P, s_i^R\}$, where a_i^P is the point at which $\mu_{\tilde{A}_i}(a_i^P) = 1$, s_i^L is the left-side spread from the peak point a_i^P, and s_i^R represents the right-side spread. The membership function for each $\tilde{A}_i$ has the form

$$\mu_{\tilde{A}_i}(a_i) = \begin{cases} 1 - \dfrac{a_i - a_i^P}{s_i^R}, & a_i^P \le a_i \le a_i^P + s_i^R, \\ 1 - \dfrac{a_i^P - a_i}{s_i^L}, & a_i^P - s_i^L \le a_i \le a_i^P, \\ 0 & otherwise. \end{cases}$$

Following the principle of extension, Yen et al.[2] used the fuzzy membership function for the output by

$$\mu_{\tilde{Y}}(y) = \begin{cases} 1 - \dfrac{y - \sum_i a_i^P x_i - a_0^P}{s_0^R + \sum_i s_i^R |x_i|}, & a_0^P + \sum a_i^P x_i \le y \le a_0^P + \sum_i a_i^P x_i + \left(s_0^R + \sum_i s_i^R |x_i|\right), \\ 1 - \dfrac{a_0^P + \sum_i a_i^P x_i - y}{s_0^L + \sum_i s_i^L |x_i|}, & a_0^P + \sum_i a_i^P x_i - \left(s_0^L + \sum_i s_i^L |x_i|\right) \le y \le a_0^P + \sum_i a_i^P x_i, \\ 0 & otherwise. \end{cases} \tag{2}$$

i.e, $\tilde{Y} = \{S_0^L + \sum_i S_i^L |x_i|,\ a_0^P + \sum_i a_i^P x_i,\ S_0^R + \sum_i S_i^R |x_i|\}$.

But this result is **wrong**. Let $\widetilde{A}_i = \{S_i^L,\ a_i^P,\ S_i^R\}$, $i = 1, 2$ and $\lambda \in R$ be a real number. By the extension principle, the following rules for addition and real multiplication can be represented as

$$\lambda \widetilde{A} = \begin{cases} \{\lambda S^L,\ \lambda a^P,\ \lambda S^R\} & \text{if } \lambda \geq 0, \\ \{|\lambda| S^R,\ \lambda a^P,\ |\lambda| S^L\} & \text{if } \lambda < 0, \end{cases}$$

$$\widetilde{A}_1 + \widetilde{A}_2 = \{S_1^L + S_2^L,\ a_1^P + a_2^P,\ S_1^R + S_2^R\}.$$

This result can be generalized to linear combinations of fuzzy numbers as follows:

$$\begin{aligned} \widetilde{Y} &= \widetilde{A}_0 + \widetilde{A}_1 x_1 + \cdots + \widetilde{A}_n x_n \\ &= \{S_0^L + \sum_{x_i \geq 0} |x_i| S_i^L + \sum_{x_i < 0} |x_i| S_i^R,\ a_0^P + \sum_i a_i^P x_i,\ S_0^R + \sum_{x_i \geq 0} |x_i| S_i^R + \sum_{x_i < 0} |x_i| S_i^L\}, \end{aligned}$$

i.e.

$$\mu_{\widetilde{Y}}(y) = \begin{cases} 1 - \dfrac{y - \sum_i a_i^P x_i - a_0^P}{S_0^R + \sum_{x_i \geq 0} |x_i| S_i^R + \sum_{x_i < 0} |x_i| S_i^L}, \\ \quad \text{if } a_0^P + \sum a_i^P x_i \leq y \leq a_0^P + \sum_i a_i^P x_i + \left(S_0^R + \sum_{x_i \geq 0} |x_i| S_i^R + \sum_{x_i < 0} |x_i| S_i^L\right), \\ 1 - \dfrac{a_0^P + \sum_i a_i^P x_i - y}{S_0^L + \sum_{x_i \geq 0} |x_i| S_i^L + \sum_{x_i < 0} |x_i| S_i^R}, \\ \quad \text{if } a_0^P + \sum_i a_i^P x_i - \left(S_0^L + \sum_{x_i \geq 0} |x_i| S_i^L + \sum_{x_i < 0} |x_i| S_i^R\right) \leq y \leq a_0^P + \sum_i a_i^P x_i, \\ 0, \qquad \textit{otherwise}. \end{cases} \tag{3}$$

So $\mu_{\widetilde{A}}(y)$ in (2) is not true but $\mu_{\widetilde{A}}(y)$ in (3) is true.

We consider the following example.

Example. Let $\widetilde{A}_0 = \{S_0^L = 1,\ a_0^P = 2,\ S_0^R = 3\}$,
$\widetilde{A}_1 = \{S_1^L = 1,\ a_1^P = 4,\ S_1^R = 6\}$, and $x_1 = -2$. Then by (3), $\widetilde{Y} = \widetilde{A}_0 + (-2)\ \widetilde{A}_1 = \{$ left-side spread $= 1 + (|-2|)6 = 13$, peak point $= 2 + (-2)4 = -6$, right-side spread $= 3 + (|-2|)1 = 5\} = \{13,\ -6,\ 5\}$. But by (2), we have $\widetilde{Y} = \{3,\ -6,\ 15\}$ which is not true.

From the expression of (3), we get the constraints of the regression as

$$1-\frac{y-\sum_i a_i^P x_i - a_0^P}{S_0^L+\sum_{x_i\geq 0}|x_i|s_i^L+\sum_{x_i<0}|x_i|s_i^R}\geq h$$

and

$$1-\frac{a_0^P+\sum_i a_i^P x_i - y}{S_0^R+\sum_{x_i\geq 0}|x_i|s_i^R+\sum_{x_i<0}|x_i|s_i^L}\geq h$$

if we have taken membership function value at h-cut.

References

1. Tanaka, H. Vejima, S. and Asia ,K. (1982). Linear regression analysis with fuzzy model, IEEE Trans. Systems Man Cybernet. 12(6), 903-907.
2. Yen, K. K. Ghoshray, S. and Roig, G. (1999). A Linear regression model using triangular fuzzy number coefficients, *Fuzzy Sets and Systems* 106, 167-177.

[received date : Dec. 2004, accepted date Apr. 2005]