An Analysis of Possible Applications of Fuzzy Set Theory to the Actuarial Credibility Theory

Krzysztof Ostaszewski[1] and Waldemar Karwowski[2]
[1]Department of Mathematics
[2]Center for Industrial Ergonomics
Department of Industrial Engineering
University of Louisville
Louisville, Kentucky 40292, U.S.A.
Telephone: (502) 588-6829; (502) 588-7173
Fax: (502) 588-0150; (502) 588-7397

ABSTRACT

In this work we review basic concepts of the actuarial credibility theory from the point of view of introducing applications of fuzzy set-theoretic method. We show how the concept of actuarial credibility can be modeled through the fuzzy set membership functions, and how fuzzy set methods, especially fuzzy pattern recognition, can provide an alternative tool for estimating credibility.

INTRODUCTION

Credibility theory is one of the most fundamental tools of actuarial science applied to casualty and property insurance. Casualty and property insurance are characterized by high frequency of claims (even for the same individual or group), and significantly more variable patterns of both claim frequency and severity. On the other hand, the time until payment, or until a failure of a status, are of less importance, as claims arise so frequently.

THE CONCEPT OF ACTUARIAL CREDIBILITY

The simplest description of credibility can be as the measure that an actuary believes should be attached to a given body of data about risks considered for insurance for rate-making purposes. To say that data is *"fully credible"* means that the data is sufficient for setting the premium rates based on it, while the data concerning loss experience is *"too small to be credible"* if we believe that the future experience may well be very different, and that we have more confidence in the knowledge prior to data collection.

For example, data concerning personal automobile liability insurance loss experience in the state of Kentucky is *"fully credible"* if it is adequate for rate levels in the state without reference to any previous data, or other states or countries experience. The standard mathematical models of credibility produces a number Z between 0 and 1 which is a measure of credibility assigned to the data, while 1 - Z is treated as a measure of credibility assigned to the alternative (e.g., previous data, or other states' experience, in the case of personal automobile liability insurance in Kentucky). We then have

$$C = ZR + (1-Z)H$$

where R is the mean loss calculated from the current observation, it is the prior mean, and C is the compromise estimate used for setting the net premium.

DETERMINATION OF CREDIBILITY

Mathematical models of actuarial credibility assume generally that losses are generated randomly by the distribution of a variable of the form

$$Y = X_1 + X_2 + \ldots + X_N$$

where N is the random claim frequency, while each X_i, a random variable as well, corresponds to the individual claim severity. If N is assumed to have the Poisson distribution, the variables X_i are independent identically distributed, and we adopt the approach of interval estimation, the credibility Z can be estimated as

$$Z = \sqrt{\frac{N}{n_F}}$$

where N is the observed number of losses, and

$$n_F = \frac{y^2}{k^2}\left(1 + \left(\frac{\sigma}{m}\right)^2\right)$$

Here, k is the fluctuation limit away from the mean of total claims, y is the prescribed confidence interval boundary for the standard normal distribution, and σ/m is the coefficient of variation of the individual claim severity distribution. An alternative method (Herzog, 1992) is to evaluate the posterior total claim size distribution using the classical Bayesian approach. The third standard method is the Bühlman's (1967) credibility estimate

$$Z = \frac{n}{n + K}$$

where n is the number of exposure units in the experience and K is the ratio of the expected value of process variance to the variance of hypothetical means.

DETERMINATION OF CREDIBILITY WITH FUZZY PATTERN RECOGNITION

Ostaszewski (1992) gives an extensive discussion of applicability of fuzzy set theoretic methods in actuarial science. He points out that pattern recognition methods can be applied directly to classification of risks, thus creating an alternative rate-making approach. If

$$\underline{x}_1 = \left(x_1^{(1)}, x_2^{(1)}, \ldots, x_p^{(1)} \right)$$

$$\underline{x}_2 = \left(x_1^{(2)}, x_2^{(2)}, \ldots, x_p^{(2)} \right)$$

$$\ldots$$

$$\underline{x}_n = \left(x_1^{(n)}, x_2^{(n)}, \ldots, x_p^{(n)} \right)$$

is the data set representing the historical loss experience, and

$$\underline{y} = (y_1, \ldots, y_P)$$

represents data concerning the recent experience (vector coordinates represent risk characteristics and loss features), one can use a clustering algorithm (see Ostaszewski, 1992, for an example of such direct application and further references) to assign $\underline{y}$ to fuzzy clusters in data. If μ is the maximum membership degree of $\underline{y}$ in a cluster, the number $Z = 1 - \mu$ could be used as the credibility measure of the experience provided by $\underline{y}$, while μ gives the membership degree for the historical experience indicated by the cluster.

Using our previous automobile rate-making example, consider an insurer with historical experience in the states of Ohio, Pennsylvania and California, extending her business to Kentucky. The insurer can cluster new data from Kentucky into patterns from other states, and arrive at a credibility reading of her loss experience in Kentucky versus the historical net premiums from Ohio, Pennsylvania and California (or subsets of this three-element set, if clustering so indicates).

Assume, hypothetically, that the mean claims and the standard deviations of claims for Ohio, Pennsylvania, and California are:

Ohio: $\mu_1 = 100, \sigma_1 = 25$;

Pennsylvania: $\mu_2 = 125, \sigma_2 = 30$;

California: $\mu_3 = 175, \sigma_3 = 50$.

Let Kentucky experience be $\mu_4 = 200, \sigma_4 = 40$. Assuming equal probability for each of the three historical states, and using Bühlman's (1967) actuarial credibility formula we get:

$$K = \frac{\text{Expected value of process variance}}{\text{Variance of hypothetical mean}} =$$

3

$$= \cfrac{\frac{1}{3}(25)^2 + \frac{1}{3}(30)^2 + \frac{1}{3}(50)^2}{\sqrt{\frac{1}{3}\left(\left(100 - \frac{400}{3}\right)^2 + \left(125 - \frac{400}{3}\right)^2 + \left(175 - \frac{400}{3}\right)^2\right)}}$$

$$\approx 1.38$$

and
$$Z = \frac{n}{n+K} = \frac{3}{3+1.38} = 0.6849$$

We have, therefore:

$$C = ZR + (1 - Z)H =$$

$$0.6849.200 + (1 - 0.6849)\frac{400}{3} \approx 179.$$

On the other hand, if we consider just the means and standard deviations as features, and treat the data from the four states as four feature vectors:

$$\underline{x}_1 = \begin{bmatrix} 100 \\ 25 \end{bmatrix}, \underline{x}_2 = \begin{bmatrix} 125 \\ 30 \end{bmatrix}, \underline{x}_3 = \begin{bmatrix} 175 \\ 50 \end{bmatrix}, \underline{x}_4 = \begin{bmatrix} 200 \\ 40 \end{bmatrix}$$

Then we can use clustering methods to analyze them. We will use the classical Bezdek's (1981) clustering algorithm specified by a matrix

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},$$

parameter m = 2, initial partition

$$\tilde{U}^{(0)} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and the stopping parameter $\varepsilon = 0.3$.

The first step cluster centers are

$$v_1^{(0)} = \begin{bmatrix} 133.33 \\ 31.67 \end{bmatrix}, v_2^{(0)} = \begin{bmatrix} 200 \\ 40 \end{bmatrix}.$$

This results in a new partition

$$\tilde{U}^{(1)} = \begin{bmatrix} 0.8956 & 0.9870 & 0.251 & 0 \\ 0.1044 & 0.0130 & 0.7479 & 1 \end{bmatrix}.$$

Using the standard matrix norm we get

$$\|\tilde{U}^{(0)} - \tilde{U}^{(1)}\| \approx 1.068 > 0.3.$$

The second step cluster centers are

$$v_1^{(1)} = \begin{bmatrix} 115.828 \\ 28.511 \end{bmatrix}, \ v_2^{(1)} = \begin{bmatrix} 190.393 \\ 45.457 \end{bmatrix}.$$

The second step partition is:

$$\tilde{U}^{(2)} = \begin{bmatrix} 0.9697 & 0.9811 & 0.0695 & 0.0169 \\ 0.0303 & 0.0189 & 0.9305 & 0.9831 \end{bmatrix}$$

and $\|\tilde{U}^{(2)} - \tilde{U}^{(1)}\| \approx 0.28 < 0.3$, resulting in stopping.

At this point, we see that a cluster of Pennsylvania and Ohio rates differs significantly from the cluster of California and Kentucky rates. Due to such difference, one can use the membership of 0.9831 for Kentucky in its cluster as a new credibility rating Z, resulting in

$$C = 0.9831(200) + 0.0169\left(\frac{400}{3}\right) \approx 199.$$

Alternatively, one can propose to give the membership 0.9831 the meaning of credibility of the mean of Kentucky and California cluster, thus producing a new mean:

$$C = 0.9831\left(\frac{200 + 175}{2}\right) + 0.0169\left(\frac{100 + 125}{2}\right) \approx 186.$$

We believe this procedure, being a natural extension of the meaning of cluster membership and a modification of classical credibility, to be a potentially significant new development in our understanding of actuarial credibility.

CONCLUSIONS

Our paper provides a relatively simple idea for extending the fuzzy clustering methods to credibility theory models. Further empirical investigations are needed in order to determine which clustering algorithms are most appropriate for the purpose of credibility measurement.

REFERENCES

Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, Plenum Press, 1981.

Bühlman, H., Experience rating and credibility, *The ASTIN Bulletin*, 4, 199-207, 1967.

Herzog, T. N, *An Introduction to Bayesian Credibility Theory and Related Topics*, Casualty Actuarial Society, in Part 4 Study Notes, 1992.

Hossack, I. B., Pollard, J. H. and Zehnwirth, B., *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, 1983.

Ostaszewski, K., *An Investigation into Possible Applications of Fuzzy Set Methods in Actuarial Science*, Society of Actuaries, 1992, (to appear).