Journal of Uncertain Systems
Vol.6, No.4, pp.256-262, 2012
Online at: www.jus.org.uk

B-Spline Method of Uncertain Statistics with Applications to Estimate Travel Distance

Xiaowei Chen[1], Dan A. Ralescu[2,*]

[1] *Department of Risk Management and Insurance, Nankai University, Tianjin 30071, China*

[2] *Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-0025, USA*

Received 2 June 2012; Revised 5 September 2012

Abstract

Uncertainty theory is a branch of mathematics for modeling uncertainty in human reasoning, and uncertain statistics is a methodology for collecting and interpreting experts' experimental data by uncertainty theory. In this paper, we will first discuss how to collect experts' data. We apply this method to a model for estimating the distance between Beijing and Tianjin. Based on the experts' experimental data, B-spline method is used to estimate the empirical uncertainty distribution. Finally, a numerical example is given.

Keywords: uncertainty theory, uncertainty distribution, uncertain statistics

1 Introduction

Classical mathematical statistics was proposed as a methodology for collecting and interpreting the test data with correlative information in a system by probability theory. Statistical methods have been found useful in many fields. It should be pointed out that classical statistics to a great extent depends on historical data. In many cases, it is impossible to get historical data or the access to historical data is costly. On the other hand, fuzzy statistics was introduced to deal with data using fuzzy set theory initiated by Zadeh [24]. Fuzzy statistics is a mathematical model dealing with vague qualitative data, frequently generated from natural language. Fuzzy statistics methods were studied such as fuzzy decision making [1], fuzzy point estimation[7, 8], fuzzy interval estimation [3], fuzzy hypotheses testing, fuzzy regression [18, 19], fuzzy Bayesian statistics by fuzzy probability measure [17], and fuzzy-Bayes decision rule [20].

It is undeniable that probability theory and fuzzy set theory are useful tools to deal with objective and subjective uncertainty. However, in our real life, human natural language statements like "about 100km", "approximately 39^0C", "roughly 80kg", "low speed", "middle age", and "big size" are typically used to express the imprecise information and knowledge. Usually, such imprecise quantities were modeled by subjective probability or fuzziness. However, many surveys showed that they are neither due to randomness nor to fuzziness. For details of these contents, please see Liu [16]. This fact provides a motivation for uncertainty theory founded by Liu [9] as a branch of axiomatic mathematics for modeling human uncertainty based on normality, duality, subadditivity and product axioms. As a model for dealing with human uncertainty, this theory has been successfully applied to many fields: uncertain programming [12], uncertain logic [15], uncertain control [14] and uncertain finance [10]. For exploring the recent developments of uncertainty theory, the readers may consult Liu [13].

In uncertainty theory, uncertainty distribution plays a central role. Uncertainty statistics was introduced in [13] to estimate experimental uncertainty distribution using linear interpolation method. In addition, the method of moments was proposed by Wang and Peng [22] to estimate the parameters in uncertain distribution. Considering the knowledge of more than one expert, the Delphi method [4, 21] was suggested to determine uncertainty distributions. Uncertain statistics is based on experts' experimental data rather than historical data. In this paper, we will first describe how to get the experts' data. Then we will study a specific model to estimate the distance between "Beijing" and "Tianjin". Based on the experts' experimental data, B-spline

*Corresponding author. Email: ralescd@ucmail.uc.edu (D.A. Ralescu).

method is proposed to estimate empirical uncertainty distribution. The rest of this paper is organized as follows: the next section will introduce some basic concepts of uncertainty theory. The method of collecting experts' data and estimating uncertainty experimental distribution via linear interpolation method is included in Section 3. B-spline method to estimate the distance between the two cities: "Beijing" and "Tianjin" is established in Section 4. Finally, a brief summary is given in Section 5.

2 Preliminaries

Let Γ be a nonempty set, and $\mathcal{L}$ a σ-algebra over Γ. The uncertain measure $\mathcal{M}$ [9] is a set function defined on $\mathcal{L}$ satisfying the following four axioms:

Axiom 1. (Normality) $\mathcal{M}\{\Gamma\} = 1$;

Axiom 2. (Duality) $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$ for any event Λ;

Axiom 3. (Subadditivity) For every countable sequence of events $\{\Lambda_i\}$, we have

$$\mathcal{M}\left\{\bigcup_{i=1}^{\infty} \Lambda_i\right\} \le \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}.$$

Axiom 4.(Product Axiom) Let $(\Gamma_k, L_k, \mathcal{M}_k)$ be uncertainty spaces for $k = 1, 2, \ldots$ The product uncertain measure $\mathcal{M}$ is an uncertain measure satisfying

$$\mathcal{M}\left\{\prod_{i=1}^{\infty} \Lambda_i\right\} = \bigwedge_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}$$

where Λ_k are arbitrarily chosen events from $\mathcal{L}_k$ for $k = 1, 2, \ldots$, respectively.

An uncertain variable is a measurable function from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers. The distribution of the uncertain variable ξ is $\Phi(x) = \mathcal{M}\{\gamma \in \Gamma | \xi(\gamma) \le x\}$. The expected value of an uncertain variable ξ was defined by Liu [9] as

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \ge x\}\mathrm{d}x - \int_{-\infty}^{0} \mathcal{M}\{\xi \le x\}\mathrm{d}x$$

provided that at least one of the two integrals is finite. The variance of an uncertain variable ξ is defined as $E\left[(\xi - E[\xi])^2\right]$. Some useful uncertainty distributions are shown in the following part.

Example 1. An uncertain variable ξ is called linear if it has the following uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x < a \\ (x-a)/(b-a), & \text{if } a \le x \le b \\ 1, & \text{if } x > b \end{cases}$$

denoted by $\mathcal{L}(a, b)$ where a and b are real numbers with $a < b$.

Example 2. An uncertain variable ξ is called zigzag if it has the following uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x < a \\ (x-a)/2(b-a), & \text{if } a \le x < b \\ (x+c-2b)/2(c-b), & \text{if } b \le x \le c \\ 1, & \text{if } x > c \end{cases}$$

denoted by $\mathcal{Z}(a, b, c)$ where a, b and c are real numbers with $a < b < c$.

Example 3. An uncertain variable ξ is called normal if it has the following uncertainty distribution

$$\Phi(x) = \left(1 + \exp\left(\frac{\pi(e-x)}{\sqrt{3}\sigma}\right)\right), \ -\infty < x < +\infty, \ \sigma > 0.$$

denoted by $\mathcal{N}(e, \sigma)$.

Definition 1 [13] *Let ξ be an uncertain variable with uncertainty distribution function Φ. Then the inverse function Φ^{-1} is called the inverse uncertainty distribution of ξ.*

It is obvious that the inverse functions of the linear uncertain variable $\mathcal{L}(a,b)$, zigzag uncertain variable $\mathcal{Z}(a,b,c)$ and normal uncertain variable $\mathcal{N}(e,\sigma)$ are

$$\Phi^{-1}(\alpha) = (1-\alpha)a + \alpha b,$$

$$\Phi^{-1}(\alpha) = \begin{cases} (1-2\alpha)a + 2\alpha b, & \text{if } \alpha < 0.5 \\ (2-2\alpha)b + (2\alpha-1)c, & \text{if } \alpha \geq 0.5 \end{cases}$$

and

$$\Phi^{-1}(\alpha) = e + \frac{\sigma\sqrt{3}}{\pi}\ln\frac{\alpha}{1-\alpha},$$

respectively. The expected value of an uncertain variable ξ can also be expressed by its inverse uncertainty distribution.

Theorem 1 [13] *Let ξ be an uncertain variable with uncertainty distribution Φ. If the expected value exists, then*

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha)\mathrm{d}\alpha.$$

Note that the expected values of linear uncertain variable $\mathcal{L}(a,b)$, zigzag uncertain variable $\mathcal{Z}(a,b,c)$ and normal uncertain variable $\mathcal{N}(e,\sigma)$ are $(a+b)/2$, $(a+2b+c)/2$ and e, respectively.

3 Experts' Data and Empirical Uncertainty Distribution

Different from classical statistics, uncertain statistics is based on experts' experimental data rather than historical data. The reason is simple. In many cases, it is impossible to get historical data or the access to historical data is costly. Then the experts' data becomes an effective alternative. For example, we want to establish a model to get "how far from Beijing to Tianjin". Assume that the real distance between them is not exactly known to us. Thus we think the distance is an imprecise quantity. The consultation process of a questionnaire survey for estimating the travel distance between Beijing and Tianjin is as follows
Q1: May I ask you how far is from Beijing to Tianjin? What do you think is the minimum distance?
A1: 100km. (an experts experimental data (100, 0) is acquired)
Q2: What do you think is the maximum distance?
A2: 150km. (an experts experimental data (150, 1) is acquired)
Q3: What do you think is a likely distance?
A3: 130km.
Q4: What is the degree of belief that the real distance is less than this distance?
A4: 0.6. (an experts experimental data (130, 0.6) is acquired)
Q5: Is there another number this distance may be?
A5: 140km.
Q6: What is the degree of belief that the real distance is less than this distance?
A6: 0.9. (an experts experimental data (140, 0.9) is acquired)
Q7: Is there another number this distance may be?
A7: 120km.
Q8: What is the degree of belief that the real distance is less than this distance?
A8: 0.3. (an experts experimental data (120, 0.3) is acquired)
Q9: Is there another number this distance may be?
A9: No idea.

Then this 1st expert's experimental data are

$$(100,0), \quad (120,0.3), \quad (130,0.6), \quad (140,0.9), \quad (150,1).$$

Using the same consultation process, the 2nd expert's and 3rd expert's data are

$$(110, 0), \quad (120, 0.3), \quad (130, 0.8), \quad (140, 1)$$

and

$$(100, 0), \quad (120, 0.2), \quad (130, 0.9), \quad (150, 1),$$

respectively.

Based on the such experimental data, Liu [13] introduced the linear interpolation method to obtain the empirical uncertainty distribution

$$(x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)$$

where (perhaps after arrangement)

$$x_1 < x_2 < \cdots < x_n, \ 0 \le \alpha_1 \le \alpha_2 \le \cdots \le \alpha_n.$$

The empirical uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x < x_1 \\ \alpha_i + \dfrac{(\alpha_{i+1} - \alpha_i)(x - x_i)}{x_{i+1} - x_i}, & \text{if } x_{i+1} \le x \le x_i \\ 1, & \text{if } x > x_n. \end{cases} \tag{1}$$

The three experts' empirical uncertainty distributions are denoted by $\Phi_1(x)$, $\Phi_2(x)$ and $\Phi_3(x)$, respectively. The graphs of uncertainty distributions $\Phi_1(x)$, $\Phi_2(x)$, and $\Phi_3(x)$ are listed in Figure 1.

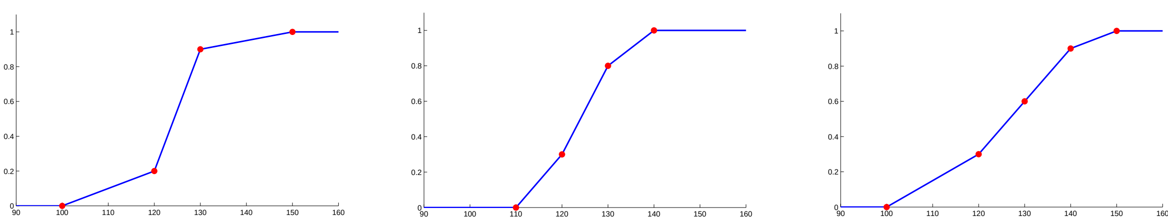

Figure 1: The uncertainty distributions $\Phi_1(x)$, $\Phi_2(x)$ and $\Phi_3(x)$

In order to combine the three domain experts' data, we will use the method introduced in [13] to aggregate domain experts' empirical uncertainty distribution

$$\Phi(x) = \frac{1}{3}(\Phi_1(x) + \Phi_2(x) + \Phi_3(x)).$$

The aggregated uncertainty distribution is shown in Figure 2.

The empirical uncertainty distribution Φ with the form (1) has an expected value

$$E[\xi] = \frac{\alpha_1 + \alpha_2}{2} x_1 + \sum_{i=2}^{n-1} \frac{\alpha_{i+1} - \alpha_{i-1}}{2} x_i + \left(1 - \frac{\alpha_{n-1} + \alpha_n}{2}\right) x_n.$$

Through the above formula, we get the estimated distance from Beijing to Tianjin as 124km. The real distance is 127km in google earth. The estimation error is 2.3%.

4 B-Spline Method

Besides Liu's method, Gao [4] introduced other techniques like quadratic-spline interpolation, cubic-spline interpolation, $\sin x$-spline interpolation, etc. In this section, we will introduce a new method called B-spline interpolation to fit a continuous uncertainty distribution. The reasons why we choose this method is that the fitting distribution via B-spline interpolation needs minimal support with respect to a given degree,

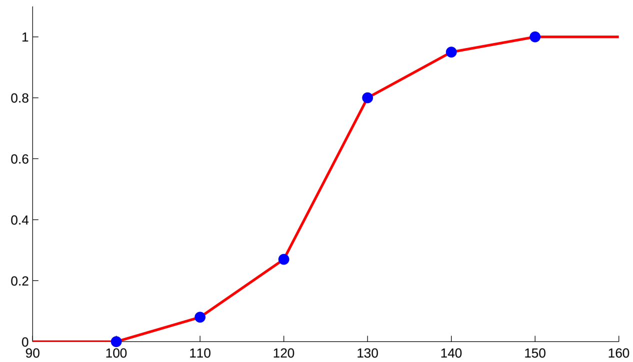

Figure 2: The figure of uncertainty distribution Φ

smoothness, and domain partition. A B-spline is simply a generalization of a Bezier curve. The term B-spline refers to a spline curve parameterized as a linear combination of B-splines. The uniform cubic B-spline is

$$N_i^3(x) = \begin{cases} 0, \quad \text{if } x \in (-\infty, x_i) \\ (x-x_i)^3, \quad \text{if } x \in [x_i, x_{i+1}), \\ h^3 + 3h^2(x - x_{i+1}) + 3h(x - x_{i+1})^2 \\ -3(x - x_{i+1})^3, \quad \text{if } x \in [x_{i+1}, x_{i+2}), \\ h^3 + 3h^2(x_{i+3} - x) + 3h(x_{i+3} - x)^2 \\ -3(x_{i+3} - x)^3, \quad \text{if } x \in [x_{i+2}, x_{i+3}) \\ (x_{i+4} - x)^3, \quad \text{if } x \in [x_{i+3}, x_{i+4}) \\ 0, \quad \text{if } x \in [x_{i+4}, +\infty). \end{cases}$$

We use the above uniform cubic B-spline to fit the uncertainty distribution of the collected three experts' data

$$(100, 0),\ (120, 0.3),\ (130, 0.6),\ (140, 0.9),\ (150, 1),$$

$$(110, 0), \quad (120, 0.3), \quad (130, 0.8), \quad (140, 1)$$

and

$$(100, 0), \quad (120, 0.2), \quad (130, 0.9), \quad (150, 1).$$

The fitting uncertainty distribution of

$$\Phi(x) = \frac{1}{3}(\Phi_1(x) + \Phi_2(x) + \Phi_3(x)).$$

is shown in Figure 3. Compared to the linear interpolation method, the uncertainty distribution is now smoother. The expected value of the fitting distribution is 124.2km while the real distance is 127km in google earth. The estimation error is 2.2%.

5 Conclusion

In this paper, a questionnaire survey for collecting experts' data was proposed. Using the collected experts' data, we applied uncertain statistics to estimate the distance between the two cities: Beijing and Tianjin. Besides, the B-spline method was introduced to estimate empirical uncertainty distribution.

Acknowledgments

This work was supported by National Natural Science Foundation of China Grant No.61273044.

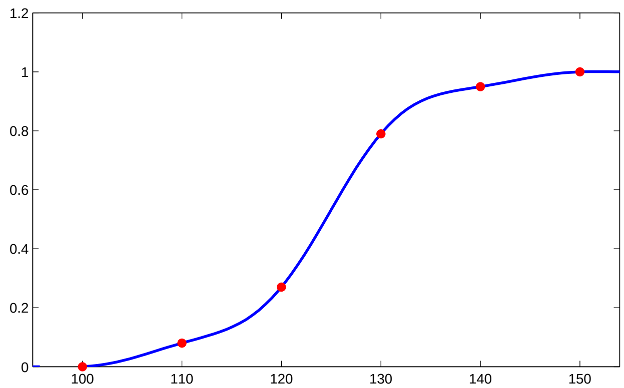

Figure 3: B-spline interpolation

References

[1] Buckley, J., Fuzzy decision making with data: applications to statistics, *Fuzzy Sets and Systems*, vol.16, no.2, pp.139–147, 1985.

[2] Casals, M., M. Gil, and P. Gil, The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information, *European Journal of Operational Research*, vol.27, pp.371–382, 1986.

[3] Corral, N., and M. Gil, A note on interval estimation with fuzzy data, *Fuzzy Sets and Systems*, vol.28, no.2, pp.209–215, 1988.

[4] Gao, J., Determine uncertainty distribution via Delphi method, *Proceedings of the First International Conference on Uncertainty Theory*, pp.291–297, 2010.

[5] Gao, X., Some properties of continuous uncertain measure, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol.17, no.3, pp.419–426, 2009.

[6] Gao, X., Y. Gao, and D. Ralescu, On Liu's inference rule for uncertain systems, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol.18, no.1, pp.1–11, 2010.

[7] Gil, M., Fuzziness and loss of information in statistical problems, *IEEE Transactions on Systems Man and Cybernetics*, vol.17, pp.1012–1025, 1987.

[8] Gil, M., N. Corral, and P. Gil, The fuzzy decision problem: an approach to the point estimation problem with fuzzy information, *European Journal of Operational Research*, vol.22, pp.26–34, 1985.

[9] Liu, B., *Uncertainty Theory*, 2nd Edition, Springer-Verlag, Berlin, 2007.

[10] Liu, B., Fuzzy process, hybrid process and uncertain process, *Journal of Uncertain Systems*, vol.2, no.1, pp.3–16, 2008.

[11] Liu, B., Some research problems in uncertainty theory, *Journal of Uncertain Systems*, vol.3, no.1, pp.3–10, 2009.

[12] Liu, B., *Theory and Practice of Uncertain Programming*, 2nd Edition, Springer-Verlag, Berlin, 2009.

[13] Liu, B., *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*, Springer-Verlag, Berlin, 2010.

[14] Liu, B., Uncertain set theory and uncertain inference rule with application to uncertain control, *Journal of Uncertain Systems*, vol.4, no.2, pp.83–98, 2010.

[15] Liu, B., Uncertain logic for modeling human language, *Journal of Uncertain Systems*, vol.5, no.1, pp.3–20, 2011.

[16] Liu, B., Why is there a need for uncertainty theory?, *Journal of Uncertain Systems*, vol.6, no.1, pp.3–10, 2012.

[17] Piasecki, K., On the Bayes formula for fuzzy probability measures, *Fuzzy Sets and Systems*, vol.18, no.2, pp.183–185, 1986.

[18] Tanaka, H., S. Uejima, and K. Asai, Fuzzy linear regression model, *IEEE Transactions on Systems Man and Cybernetics*, vol.10, pp.2933–2938, 1980.

[19] Tanaka, H., S. Uejima, and K. Asai, Linear regression analysis with fuzzy model, *IEEE Transactions on Systems Man and Cybernetics*, vol.12, pp.903–907, 1982.

[20] Uemura, Y., A decision rule on fuzzy events, *Japanese Journal Fuzzy Theory and Systems*, vol.3, pp.291–300, 1991.

[21] Wang, X., Z. Gao, and H. Guo, Delphi method for estimating uncertainty distributions, *Information: An International Journal*, vol.12, no.2, pp.449–460, 2012.

[22] Wang, X., and Z. Peng, Method of moments for estimating uncertainty distributions, http://orsc.edu.cn/online/100408.pdf.

[23] You, C., Some convergence theorems of uncertain sequences, *Mathematical and Computer Modelling*, vol.49, nos.3-4, pp.482–487, 2009.

[24] Zadeh L., Fuzzy sets, *Information and Control*, vol.8, pp.338–353, 1965.