Journal of Business Research 69 (2016) 2138–2144

Contents lists available at ScienceDirect

Journal of Business Research

Correlation coefficient evaluation for the fuzzy interval data☆

Chih-Ching Yang

Department of Statistic, Tamkang University, 5F.-3, No.75 Bitan Road, Xindian Dist., New Taipei City 231, Taiwan (ROC)

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form 1 October 2015
Accepted 1 November 2015
Available online 22 December 2015

Keywords:
Fuzzy correlation coefficient
Pearson's correlation coefficient
Fuzzy data
Fuzzy correlation with mean and standard deviation

ABSTRACT

The issue of evaluating an appropriate correlation with fuzzy data is an important topic in the econometrics and management science, especially when the data sets illustrate uncertainty, inconsistence, and incompleteness. This study extends the concept of Pearson's correlation coefficient to compute the correlation coefficient of the data sets that are fuzzy in nature. However, no common proposal for such extension exists. This study proposes several ways to evaluate the correlation coefficient when the fuzzy data are with interval types. Two empirical studies show that the methods that this study proposes for evaluating the coefficient of fuzzy correlations are useful and efficient from the perspective of econometrics and management.

1. Introduction

Econometrics' and management science's analyses draw primarily on statistical techniques. Investigating the relationship between two empirical evidences for management and econometrics is useful to evaluate the companies' efficiency. To be more specific, a strong correlation between companies' management capabilities and ability is useful for business management (Albert, Francisco, Jorge, & Nick, 2013; Huang & Huarng, 2015; Kristjuhan, Metsla, & Ling, 2013; Zsófia, Christoph, Stephan, & Henneberg, 2015). However, unlike physical systems, econometrics and management science do not have the luxury of having physical laws describing their dynamics. For instance, the price for a product or service usually varies depending on various market forces. Hence, finding some appropriate methods to conduct the evaluation of correlation coefficient of fuzzy data sets is necessary.

Traditional statistics reflects the results from a two-valued logic world, which often reduces the accuracy of inferential procedures. To investigate, for example, the population, people's opinions, economic phenomena, or the complexity of a subjective event more accurately, fuzzy logic allows accounting for the full range of possible values. In particular, when dealing with psychometric measures, fuzzy statistics serves as a powerful research tool. For example, Wu and Hsu (2002) develop a fuzzy time series model to overcome the bias of stock markets. Leu, Chen, and Lai (2008) develop the correlation coefficient between interview scores and academic scores for admission to a college with the approach of the alpha-cuts of fuzzy sample correlation coefficient at various alpha-values. De Carvalho and Tenório (2010) present partitioning fuzzy K-means clustering models for interval-valued data and furnish an additional interpretation tools for individual fuzzy clusters of interval-valued data.

2. Literature review

In general, the classical statistical analysis considers distribution of the observed variables. Logic terms such as “totally agree,” “partially agree,” “disagree completely,” or “don't know” explain these observations rather than equating them with quantitative values. Nevertheless, how to calculate the correlation for the fuzzy interval observations is a challenge to the traditional statistical analysis (Yang, Wu, & Sriboonchitta, 2012). Thus, several researchers develop different methods dealing with the correlation coefficient for fuzzy data sets (Hong, 2006; Hsu & Wu, 2010; Robinson & Amirtharaj, 2011). For example, Chen, Xu, and Xia (2013) propose some correlation coefficient formulas for hesitant fuzzy sets (HFSs). Under hesitant fuzzy environments, they apply the formula to business failure risk and demonstrate their application in clustering with interval-valued hesitant fuzzy information through a specific numerical example. Wei and Zhao (2013) apply the induced hesitant interval-valued fuzzy Einstein ordered weighted averaging operator and induced hesitant interval-valued fuzzy Einstein ordered weighted geometric operator to deal with multiple-attribute decision making under the hesitant interval-valued fuzzy environments. Chen (2012) takes the simple additive weighting method and the technique for order preference by similarity to an ideal solution as the main structure to deal with multiple-criteria decision analysis (MCDA) problems in the context of interval-valued fuzzy sets. In addition, comprehensive discussions exist on the influence of score functions and weight constraints, where the score function represents an aggregated effect of positive and negative evaluations in

☆ The author thanks Berlin Wu, National Chengchi University, and Yu-Ting Cheng for their careful reading and suggestions.

E-mail address: 96354502@nccu.edu.tw.

http://dx.doi.org/10.1016/j.jbusres.2015.12.021
0148-2963/

performance ratings and the weight constraint consists of the unbiased condition, positivity bias, and negativity bias. Saneifard and Saneifard (2012) develop an approach to evaluate the correlation coefficient for fuzzy data with central interval, which is useful to compare fuzzy numbers, namely, fuzzy correlations, in fuzzy environments and expert's systems.

On the other hand, several researchers propose the method of deriving the correlation coefficient for intuitionist fuzzy sets (Park, Kwun, Park, & Park, 2009; Wei, Wang, & Lin, 2011; Xu, 2006; Xu, 2010; Ye, 2010). Xu (2006) constructs a method to get the correlation coefficients of intuitionist fuzzy sets, which is a useful tool to deal with vagueness and uncertainty and has advantages over the existing methods. Ye (2010) generates a scheme of evaluation for two entropies by measuring interval-valued intuitionist fuzzy sets and establishing the entropy weighted model; this method is useful to determine the criteria weight on alternatives. Ye then proposes an evaluation formula of weighted correlation coefficient between an alternative and the ideal alternative. One can rank the alternatives and select the most desirable one(s) according to the values of the weighted correlation coefficients. Wei et al. (2011) build an optimal model building on the negative ideal solution and max-min operator, by which they can measure the attribute weights. Then the interval-valued intuitionist fuzzy weighted averaging operator aggregates the interval-valued intuitionist fuzzy information corresponding to each alternative, and then ranks the alternatives and selects the most desirable one(s) according to the correlation coefficient. Wei et al. use a mathematical-programming approach to derive fuzzy measures drawing on the classical definition of the correlation coefficient. This derivation is quite promising, but in order to employ their approach, the mathematical programming is necessary.

In sum, Pearson correlations cannot use the relationship between two variables for interval fuzzy data in the econometrics or management fields. In addition, using the mathematical programming in the analysis of the interval fuzzy data really limits the access of some researchers with no strong mathematical background. This study proposes several simple solutions of a fuzzy correlation coefficient without programming or the aid of computer resources. These solutions draw on the classical definition of Pearson correlation. The definitions that this study provides are also useful for interval-valued fuzzy data.

When considering the correlation for fuzzy data, two aspects need consideration: centroid and data shape. If the two centroids of the two fuzzy data sets are close, the correlation should be high. In addition, if the data shape of the two fuzzy sets is similar, the correlation should also be high.

This study provides an approach to dealing with these two aspects simultaneously, but before illustrating the approach of calculating fuzzy correlations, the next section presents a review of the fuzzy theory and fuzzy data sets.

3. Methods

3.1. Fuzzy theory and fuzzy data

In general, fuzzy theory can incorporate subjective uncertainties in mathematical equations and seems more tied to reality than traditional statistics. In addition, in an environment of imperfect information, traditional statistics may not be an appropriate candidate for such a representation. In other words, traditional statistics often ignores the intriguing and complicated yet sometimes conflicting nature of human logic and feelings.

Consider, for example, a fuzzy set of favorite topics for a person (see Table 1). Fuzzy logic could break down the degree of perceptions of the favorite topics into percentages, which gives a clear weighted representation of the topics. By doing so, the degree of perceptions of the favorite topic can be assigned the weight as a_1 (for "like") and the weight as a'_1 (for "dislike"), $a_1 + a'_1 = 1$, and so on. Binary logic expresses the degree of perceptions simply as 1 or 0, that is, "like" or "dislike."

Table 1
Comparing fuzzy numbers with integral numbers in favorite topics.

Degree of perceptions	Fuzzy logic	Binary logic		
Favorite topics	A_1 = like	A_2 = dislike	A_1 = like	A_2 = dislike
Politics	a_1	a'_1		ˇ
Culture	a_2	a'_2	ˇ	
Religions	a_3	a'_3	ˇ	
Finance	a_4	a'_4	ˇ	
Recreation	a_5	a'_5		ˇ

Drawing on the analysis of binary logic, the conclusion is that this person likes culture, religions, and finance but dislikes politics and recreation. The binary logic cannot realistically indicate the exact thought for favorite topics. Fuzzy logic, however, can describe the degree of perception for the favorite topic. According to Table 1, the fuzzy logic for like, μ_{A1}, and dislike, μ_{A2}, are as follows:

$$\mu_{A_1} = a_1 I_{\text{politics}}(x) + a_2 I_{\text{culture}}(x) + a_3 I_{\text{religions}}(x) + a_4 I_{\text{finance}}(x) + a_5 I_{\text{recreation}}(x),$$

$$\mu_{A_2} = a'_1 I_{\text{politics}}(x) + a'_2 I_{\text{culture}}(x) + a'_3 I_{\text{religions}}(x) + a'_4 I_{\text{finance}}(x) + a'_5 I_{\text{recreation}}(x),$$

where $I_c(x)$ is an indicator function, such that $I_c(x) = 1$ if $x \in C$, and $I_c(x) = 0$ if $x = 0$. This representation means that the person likes the topic of politics a_1, culture a_2, religion a_3, etc. The person dislikes the topic of politics a'_1, culture a'_2, religion a'_3, etc. The percentages for each category represent the degree of this person's perception towards these topics.

This result shows that, indeed, fuzzy logic is more useful to describe feelings, perceptions, and preferences because this method allows for a more descriptive and comprehensive explanation.

3.1.1. Continuous fuzzy data

Many applications use continuous fuzzy data. Fuzzy data comprise several types such as interval-valued numbers, triangular numbers, trapezoid numbers, exponential numbers, etc. Typically, the nomenclature depends on the shape of the membership function. Even though various types of fuzzy numbers exist, this study limits the discussion to three usual types: interval-valued numbers, triangular numbers, and trapezoidal numbers.

Definition 1. A fuzzy number is a trapezoidal fuzzy number if its membership function is

$$u_A(x) = \begin{cases} \dfrac{x-a}{b-a} & , \quad a \le x \le b \\ 1 & , \quad b \le x \le c \\ \dfrac{d-x}{d-c} & , \quad c \le x \le d \\ 0 & , \quad \text{otherwise} \end{cases}$$

when $c = d$, A is a triangular fuzzy number; when a $= b$ and $c = d$, A is an interval-valued fuzzy number.

3.1.2. Collecting continuous fuzzy data

In traditional sampling surveys, respondents choose one single answer or certain range of the answer. However, traditional methods are not able to truly reflect the complex conceptualization of each respondent. If respondents can express the degree of their feelings by using membership functions, the answer will be closer to real human conceptualization. Unfortunately, scholars disagree on the construction of continuous fuzzy data. Many studies use continuous fuzzy data without describing the construction method. The main issue is how to determine the membership function, which is quite subjective. To reflect