Research of Weighting Method Based on Beta Distribution

ZHAO Xin-shuang
Air Force Early Warning Academy
Wuhan, China
xinsher@163.com

CAI Yi-chao
Air Force Early Warning Academy
Wuhan, China
yccai@163.com

QU Shi
Air Force Early Warning Academy
Wuhan, China
qushi@163.com

***Abstract*—This paper presents a method to get the weights based on the information of index data. Each index has many values and the values of each index should be standardized firstly. The Beta distribution is used to fit the parameters of each index because the Beta distribution can be approached many distributions when the parameters are adjusted. According to the parameters, the variance can be obtained, and the weights can be obtained by normalizing all of the variances. The process of this method is simpler than entropy coefficient method, but the results of the two methods are the same.**

Keywords-Beta distribution; weight method; evaluation;entropy coefficient

I. INTRODUCTION

Determining weight is an important process in synthetically evaluation. Whether the weight is reasonable directly affects the accuracy of decision. The result with reasonable weight can reflect the real condition of alternatives; otherwise, the decision will become nonsensical. At present, the methods of determining the weight can be divided into two categories: the subjective weighting and the objective weighting. Subjective weighting is a method of qualitative analysis by subjective scoring, such as Analytic Hierarchy Process (AHP), Delphi, etc. It determines the weight based on the subjective preference of experts. Objective weighting is a method of quantitative analysis, such as Principal Components Analysis (PCA), Entropy Coefficient and Grey Relational Analysis (GRA), etc. It determines the weight by the data and mathematic models, so it's accuracy is higher in most case [1][2].

This paper determines the weights based on the information of index data, and it has the similar idea with entropy coefficient method. It fits the parameters of each index through Beta distribution and takes the normalized variances as weights. This method can avoid complex calculation and obtain the similar result with entropy coefficient method.

II. SOME CHARACTERS OF BETA DISTRIBUTION

Beta distribution is a wide used statistical distribution and its probability density function (PDF) is

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}x^{\beta-1},\quad \alpha,\beta>0,\quad 0<x<1,\quad \Gamma \text{ is}$$

Gamma function. The shape of the curve of Beta distribution density function can be controlled by the parameters α,β .The shape of the curve distributes in the limited range, that from the uniform distribution to approximate normal distribution.

According to the different parameters, Beta distribution has several forms. It is normal distribution when $\alpha=\beta=4$ .It is trapezoid distribution when $\alpha=\beta=2$ and uniform distribution when $\alpha=\beta=1$ .It is symmetrical when $\alpha=\beta$. Beta distribution has strong universality for interval estimation: (1) The Beta distribution coverage from the uniform distribution to approximate normal distribution, and has wide application prospect in practice; (2) According to the values of the two parameters, the curve of the Beta distribution density changes greatly. As long as choosing the appropriate values of parameters, Beta distribution can be used to fit a variety of interval sequence distribution, especially for the "partial peak" distribution (3) The Beta distribution includes great random information for fitting and quantization to interval estimation, and is easy to carry out sensitivity analysis.

The common distributions, such as the uniform distribution, the normal distribution, the trapezoid distribution, the triangular distribution and the Rayleigh distribution all can be approximated by the Beta distribution through changing the parameters[3]. Furthermore, Beta distribution is bounded, and this feature conforms to the character of index data. For these reasons, it assumes that the index data can be fitted with Beta distribution [4].

III. THE PRINCIPLE AND CALCULATION OF WEIGHTING METHOD BASED ON BETA DISTRIBUTION

A. The priciple of weighting method based on Beta distribution

Each index has different effect to the evaluation because information of its values is different, so it should be given different weight. The idea of this method is similar to entropy coefficient method. If the difference of the values that in an index is small, it has little effect for making a distinction between the projects. So this index should be given a small weight. Conversely, if the difference of values is great, this index will play an important role in decision-making and should be given a big weight. Especially, if the values of index G_j is all the same, index G_j will has no effect to the evaluation and its weight should be zero [5] [6].

The difference of index values can be reflected by their variance. And the values of indexes usually present a particular distribution. But if we directly use the data's variance to reflect the diversity, the error will be great. The reason lies on that this variance can not reflect the distribution of the index values, but different distribution means different variance.

According to the characters of Beta distribution, it can be known that the Beta distribution has fine properties. So the

978-1-4799-8646-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IHMSC.2015.31

Beta distribution is used to fit the parameters of each index, the variance can be obtained based on the parameters and the weights can be obtained by normalizing all of the variances. The reason lies in that using the variance directly can not reflect the distribution and character of index values.

In the interval of [0,1], Beta distribution has good properties. By adjusting the parameters α, β, it can fit all kinds of distribution of [0,1] interval. Therefore, as to the unknown distribution of data, it can be fitted by Beta distribution. That can avoid the trouble to judge the distribution of the data. At the same time, Beta distribution has the higher fitting precision.

B. The calculation of weighting method based on Beta distribution

The indexes present different physical implication, so it can not be compared directly. Before the evaluation, the index values must be normalized. The definition domain of the probability density function of Beta distribution is $(0,1)$, so before the index values are fitted by Beta distribution, they must be located in $(0,1)$. Index value can divide into cost type, benefit type and fixed type. These types can be normalized according to the following formulas. Assuming that there are m projects and n indexes, the index matrix is $R^{'} = (r_{ij}^{'})_{m\times n}$ and the normalized matrix is $R = (r_{ij})_{m\times n}$.

$$r_{ij} = \frac{(\max\limits_{i} r_{ij}^{'} - r_{ij}^{'})}{(\max\limits_{i} r_{ij}^{'} - \min\limits_{i} r_{ij}^{'})} \quad i \in I_1 \text{ (} I_1 \text{ is cost type index)} \tag{1}$$

$$r_{ij} = \frac{(r_{ij}^{'} - \min\limits_{i} r_{ij}^{'})}{(\max\limits_{i} r_{ij}^{'} - \min\limits_{i} r_{ij}^{'})} \quad i \in I_2 \text{ (} I_2 \text{ is benefit type index)} \tag{2}$$

$$r_{ij} = 1 - \frac{| r_{ij}^{'} - r_i |}{\max | r_{ij}^{'} - r_i |} \quad i \in I_3 \text{ (} I_3 \text{ is fixed type index)} \tag{3}$$

If $r_{ij} = 0$, it can be added 1/1000 and if $r_{ij} = 1$, it can be subtracted 1/1000, that will not affect the accuracy of the evaluation.

For the values of each index, it locates in $(0,1)$ through the above conversion. Using MATLAB to fit the normalized data of each index, we can obtain the parameters of Beta distribution. In addition, the two parameters can be calculated by the least square method. When the capability of samples large and the intervals are small, the unknown Parameters α and β can be deduced by the relationship between the frequency and the probability density function. The specific steps are as follows.

Step1. Estimating the probability density function

$$f(C_i, \gamma, \eta) = m_i / 2nh_i, i = 1,2,..., N \tag{4}$$

Where C_i is the middle value of the interval, and the range of data is divided into N intervals, m_i is the number of the sample data in the interval.

Step2. Using the least square method to solve the shape coefficients.

$$\delta_i = \ln\{m_i / 2nh_i f(C_i, \gamma, \eta)\} \tag{5}$$

Assumed that $R = \sum_{i=1}^{N} W_i \delta_i^2$, the coefficients γ and η can be calculated by the equations $\partial R/\partial\gamma = 0$ and $\partial R/\partial\eta = 0$. Where W_i is the weight coefficient, it can be selected according to the measured data. $f(C_i, \gamma, \eta)$ is the probability density function of Beta distribution.

At last, if the assumption test can be passed, the variance of each index values can be calculated through the following formula.

$$D = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{6}$$

Each index has a variance. Gathering all the variances and normalizing them, the result is the weight of corresponding index.

IV. ANALYSIS OF AN EXAMPLE

This paper uses the data of reference [7]. Reference [7] uses entropy coefficient method to do a comprehensive evaluation of the performance of 12 non-state-owned banks. It chooses 7 indexes to reflect the performance, they are capital profit rate (CPR), cost utilization rate (CUR), overtime loan rate (OLR), non-interest property rate (NIPR), flow ability rate (FR), property utilization rate (PUR), own property rate (OPR) respectively. In the 7 indexes, CPR, CUR, PUR and OPR is benefit type index, OLR and NIPR is cost type index and FR is fixed type index (the best value is 100). The data of judgment matrix comes from authoritative statistics.

In this paper, we calculate the weight of each index according the above method firstly, and then compare the weights with the result of entropy coefficient method. Through analysis the difference of the results of two methods, we can judge whether the method of this paper is feasible. Table 1 is the judgment matrix and the detailed values are the same with reference [7].

TABLE I. THE DATA OF BANK'S PERFORMANCE

	CPR	***CUR***	***OLR***	***NIPR***	***FR***	***PUR***	***OPR***
JT	0.343	8.909	20.681	3.521	79.598	4.557	5.002
ZX	0.483	13.268	0	1.952	101.481	4.364	5.107
GD	0.403	13.491	39.013	3.349	78.404	3.615	5.500
HX	0.837	17.40	0	3.211	138.618	5.918	5.202
MS	0.838	25.059	8.896	2.867	84.741	4.565	4.906
GF	0.148	3.132	0	3.043	99.187	4.923	3.500
SF	1.210	30.95	25.286	6.879	93.013	5.643	6.324
ZS	0.592	16.024	13.293	3.040	79.502	4.453	6.591
XY	0.636	18.854	12.586	2.984	105.632	4.504	7.226
PF	0.898	25.777	9.051	4.238	120.791	4.892	7.534
YT	0.19	1.656	13.873	3.938	100.865	11.659	2.557
BB	0.122	3.879	0	6.893	65.452	3.572	4.957

Firstly, we normalize the above data according formula (1)-(3), and then fit the normalized data through MATLAB to get the parameters. The data and parameters of each index can pass the assumption test. The parameters are in Table 2.

TABLE II. THE PARAMETER VALUES OF INDEX

	CPR	***CUR***	***OLR***	***NIPR***	***FR***	***PUR***	***OPR***
α	0.3803	0.4064	0.3898	0.4045	0.5161	0.2994	0.4946
β	0.4787	0.4585	0.2016	0.3742	0.4735	0.5398	0.4327

According to the formula (4), the normalized index weights is (0.1440,0.1449,0.1532,0.1523,0.1361,0.1354,0.1401). The final weight based on entropy coefficient method in reference [7] is (0.1374, 0.1397, 0.1545, 0.1587, 0.1348, 0.1297, 0.1462). The difference of the result which obtained by the two methods is little, and maximal difference of corresponding weights is 0.07. In addition, the normalization method of reference [7] is different with this paper; it may have a little effect to the difference. The results of these two methods reflect the validity of this method. The reason of the difference lies in that the method is different when the index data is normalized. If the data does not be fitted by Beta distribution, the normalized variances (means weight) will have great difference in the weight through the two methods.

V. CONCLUSION

The weighting method based on Beta distribution calculates the weight through index information, and it has the similar idea with entropy coefficient method. But comparing with entropy coefficient method, this method is simpler in comprehension and calculation. The result can be obtained and the assumption can be tested by MATLAB.

The results of the two methods are very near and this condition presents that this method has great feasibility and reliability, especially for the enormous data size. The weighting method based on Beta distribution provides a new train of thought, and it is a simple and feasible method in assigning weight.

REFERENCES

[1] LI Yin-guo,LI Xin-chun, “Weight Determination of Comprehensive Evaluation Model” in Journal of Liaodong University, No.2, 2007.

[2] CHEN Zhi-ping, HUA Ye-di, “New approach for determining index weights in the venture enterprise evaluation index system” in Journal of Systems Engeering, Vo1.24 ,No.3,Jun 2009.

[3] L i Yun, Zhu Yong, “Analysis of Beta Distribution in Data Processing” in Chinese Journal of scientific instrument, No 8,2004

[4] Wang Yingming, “Using the Method of Maximizing Deviations to Make Decision for Multi-indicies” in System Engineering and Electronics, No.7,1998.

[5] TAN Shou-lin,YAN Shuang-ka, Chen Xue-song, “Fighting Efficiency Evaluation Model of Cruise Missile based on Index Method” in Fire Control& Command Control, Vo1.35, No.5, May 2010.

[6] Wang Jing-min,Sun Yan-fu,Kang Jun-jie, “Evaluation of competitive of power enterprises based on entropy and improved TOPSIS” in Journal of NorthChina Electric Power University, Nov,2010

[7] Xie Chi, “Entropy Method and Its A pplication in Comprehensive Evaluation of Bank’s Performance” in China Soft Science, No.9,2002.