SENSITIVITY ANALYSIS OF THE TOPSIS METHOD IN RESPECT OF INITIAL DATA DISTRIBUTIONS

Rūta Simanavičienė[1], Vaida Petraitytė[2]

[1]Vilnius Gediminas Technical University. Address: Saulėtekio al. 11, 10223, Vilnius, Lithuania
[2]Claim Department, AB "Lietuvos draudimas". Address: J. Basanavičiaus St. 12, 03224 Vilnius, Lithuania
E-mail: [1] ruta.simanaviciene@vgtu.lt, [2]petraityte.vaidaa@gmail.com

Received: June 2016 Revised: August 2016 Published: November 2016

Abstract. The present article investigates the sensitivity of the multiple criteria decision-making method TOPSIS in respect of attribute probability distributions. To carry out research, initial data – attribute values – were generated according to a normal, log-normal, uniform, and beta distributions. Decision matrixes were constructed from the generated data. By applying the TOPSIS method to the matrixes generated, result samples were received. A statistical analysis was conducted for the results obtained, which revealed that the distributions of the initial data comply with the distributions of the results received by the TOPSIS method. According to the most common alternative rank value, it was ascertained that the TOPSIS method is the most sensitive for data distribution according to beta distribution, and the least sensitive for data distribution according to lognormal distribution.

Keywords: multiple criteria decision-making, TOPSIS method, sensitivity analysis, probability distribution.

1. Introduction

While analysing scientific works on sensitivity analysis methods, one can frequently find works dedicated to the analysis of the sensitivity of mathematical models. The sensitivity analysis of mathematical models studies the relationship between information flowing in and out of the model [9].The significance of the sensitivity analysis in analysing multiple criteria decision-making methods has been approved by scholars as an opportunity to increase the reliability of a multiple criteria decision. This analysis verifies whether slight alterations of initial data or preferences will change the final decision results [2]. Generally, in case of a multiple criteria task, the sensitivity analysis is carried out in respect of attribute significance values ([11], [12], [13], [6]) by using pseudorandom values evenly distributed in the range [0, 1]. R. Simanavičienė and L. Ustinovičius suggested the sensitivity analysis in respect of initial data distributions [10]. They studied a sensitivity attribute distribution according to a uniform and normal law of multiple criteria decision-making methods TOPSIS, SAW and COPRAS when attribute values change within the range ± 10 % of a fixed attribute value.

Commonly, uniform random sizes are generated for a statistical analysis of mathematical models [0, 1] ([8], [5], [13]). Nevertheless, when conducting statistical modelling, it is possible to generate samples of variables according to other distribution laws [7]. In particular, when carrying out statistical research in the field of medicine, it is noticeable that the variables of that field have various distributions as uniform, normal, log-normal, beta, gama, etc. [3].

The aim of this work is to identify the impact of initial data (attribute values) distributions on the results of the multiple criteria method TOPSIS. For the analysis of the TOPSIS method sensitivity, a mathematical model sensitivity analysis is used on the basis of samples. For the research, it was chosen to generate attribute values according to a uniform, log-normal, and beta distribution, and to observe the impact of certain distributions on the results of the TOPSIS method. During the research, it will be monitored whether the attribute distributions comply with alternative distributions, and towards which distribution the TOPSIS method is the most sensitive.

2. Sensitivity analysis on the basis of samples

The sensitivity analysis of a mathematical model in respect of the basis of samples is one of the techniques of sensitivity analysis when a model is designed repeatedly based on the combinations of values, which are made by using sampling distributions of known input factors [9]. The aforementioned sensitivity analysis method could be described in the following steps:

- An experiment is designed, and input factors requiring analysis are defined.
- Probability density function or deviation limits are determined for each factor.
- An input vector is generated by a chosen method.
- By using generated input data, model results are calculated.
- The impact of each input factor or relative importance for the outcome variable(s) is evaluated.

A sensitivity analysis could be conducted for a set of factors, which could be the following:

- A certain input characteristic;
- Distribution parameters describing a random process;
- A reason of a factor whose value determines the mechanism of alternation of a process selection.

In the case when a sample is provided through a model, model calculations are carried out repeatedly for each realization – in order to obtain the sample of variable results of the field under investigation. When having a sample of results, a statistical analysis of sample data is conducted: an empirical distribution function is formed; such numerical characteristics as an average, a standard deviation, confidence intervals, etc. are calculated [9].

3. Multiple criteria evaluation method TOPSIS

The TOPSIS method means Technique for Order Preference by Similarity to Ideal Solution [4]. An assumption of the TOPSIS method: an alternative which is the furthest from the "negative-ideal" alternative is the best. On the basis of this assumption, the alternatives examined are listed in a priority line.

Suppose we have a decision matrix X where lines mark the alternatives examined (m – number of alternatives), columns – efficiency indicators (n – number of efficiency indicators).

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \tag{1}$$

where: x_{ij} – *i*-th alternatives, value of the *j*-th efficiency indicator.

By applying the TOPSIS method, a decision matrix X is normalized by making a vector normalization:

$$\bar{x}_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} . \tag{2}$$

Suppose known values of attribute significance $q_j, \left(j = \overline{1,n}\right)$, then a weighted normalized matrix is formed $\overline{X}^* = \left(v_{ij}\right) \quad \left(i = \overline{1,m},\ j = \overline{1,n}\right)$, whose elements are calculated according to the formula (3):

$$v_{ij} = \bar{x}_{ij} \cdot q_j . \tag{3}$$

The "ideal" variant (alternative) is determined according to the following formula:

$$A^+ = \{(\max_i v_{ij} \,|\, j \in J), (\min_i v_{ij} \,|\, j \in J') \,|\, i = \overline{1,m}\} = \{a_1^+, a_2^+, ..., a_n^+\}, \tag{4}$$

where J – a set of attribute indexes whose higher values are better; J' – a set of attribute indexes whose lower values are better. The "negative-ideal" variant is determined according to the following formula:

$$A^- = \{(\min_i v_{ij} \,|\, j \in J), (\max_i v_{ij} \,|\, j \in J') \,|\, i = \overline{1,m}\} = \{a_1^-, a_2^-, ..., a_n^-\}. \tag{5}$$

A distance between a comparative *i*-th and "best-ideal" A^+ variant is determined by calculating a distance in the *n-dimensional* Euclidean space upon the following formula:

$$L_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - a_j^+\right)^2}, \ (i = \overline{1,m}) \tag{6}$$

and between the i-th and "negative-ideal" A^-, upon following the formula:

$$L_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - a_j^-\right)^2}, \ (i = \overline{1,m}). \tag{7}$$

The criterion of the TOPSIS method is a relative distance of the i-th alternative to the "worst-ideal" variant:

$$K_i = \frac{L_i^-}{L_i^+ + L_i^-}, i = \overline{1,m}. \tag{8}$$

The best alternative out of the ones analysed will be that whose K_i value is the highest [4].

4. Proposed algorithm for the sensitivity analysis

Based on the sensitivity analysis of a mathematical model and sample, the main principles and the idea provided in the article [10], an algorithm is formed to examine a multiple criteria method sensitivity in respect of attribute distributions:

1. A multiple criteria decision-making task is formed, which consists of a set of analysed alternatives $A = \{A_i : i = \overline{1,m}\}$ and a set of evaluation attributes $R = \{R_j : j = \overline{1,n}\}$.
2. A decision matrix is made up: $X = (x_{ij}),\ (i = \overline{1,m},\ j = \overline{1,n})$, where x_{ij} is a quantitative estimation of the i-th alternative of j-th attribute.
3. The distribution law *F(x)* is chosen to generate attribute values.
4. The parameters of a selected distribution *F(x)* are set based on the values of valuation attributes $R_j : j = \overline{1,n}$.
5. According to a chosen distribution, it is generated after K values of each attribute $x_{ij}^k,\ (i = \overline{1,m},\ j = \overline{1,n},\ k = \overline{1,K})$, from which K decision matrixes are then formed: $X_k = (x_{ij}^k),\ (i = \overline{1,m},\ j = \overline{1,n},\ k = \overline{1,K})$.
6. In order to check the influence of attribute distributions on the results of the TOPSIS method, the significances of equal size attributes are chosen:
$$q_j = \frac{1}{n},\ (\forall j = \overline{1,n}). \tag{9}$$
7. Accordingly, by taking one generated matrix $X_k = (x_{ij}^k),\ (i = \overline{1,m},\ j = \overline{1,n},\ k = \overline{1,K})$, an evaluation of alternatives by the TOPSIS method is made. Calculation results are provided as samples $A^k = (a_1^k, a_2^k, \ldots, a_m^k),\ (k = \overline{1,K})$, whose elements are the values of the TOPSIS method criterion for each alternative.
8. Taking rationality values of alternatives into account, alternatives are ranked by assigning a rank *1* to the highest value, while rank *m* – to the lowest.
9. A statistical analysis of the results obtained is performed.

5. Statistical methods to evaluate the sensitivity of the TOPSIS method

1. Compatibility criteria. In order to check whether the distribution of the TOPSIS criterion values to i-th alternative complies with attribute distributions, nonparametric hypotheses about the distribution were formulated.

Data. Variable sampling of quantitative relationship scale is made of the TOPSIS criteria to i-th alternative, sampling volume $n = 100$.

Statistical hypothesis: $\begin{cases} H_0 : \textit{Sample data come from the stated distribution} \\ H_A : \textit{Sample data do not come from the stated distribution} \end{cases}$

To verify a hypothesis of the TOPSIS criterion values to i-th alternative distribution, the chi-squared and Kolmogorov–Smirnov criteria are used.

When calculating the chi-squared [1] and Kolmogorov–Smirnov [14] compatibility criteria, particular p-values are observed. If a p-value is less than a chosen alpha (α = 0.05), then reject the null hypothesis that the data come from that distribution.

2. The **distributions** which will be used to generate pseudorandom sizes are as follows:

a. Normal distribution. $X \sim N(\mu, \sigma^2)$. In order to generate a set of random sizes, having a normal distribution, you need to indicate the values of parameters μ and σ.

b. Uniform distribution $X \sim U(a,b)$. In order to generate a set of random sizes, having a uniform distribution, you need to specify the limits *a* and *b* of the interval in which there will be such sizes.

c. Lognormal distribution $X \sim LN(\mu, \sigma)$. In order to generate a set of random sizes, having a log-normal distribution, the values of the parameters $ln(\mu)$ and $ln(\sigma)$ should be indicated.

d. Beta distribution $X \sim Be(\gamma, \eta)$. In order to generate a set of random sizes, having a beta distribution, the values of the parameters γ and η should be indicated.

3. Numerical characteristic to determine a level of rank confidence. To identify the level of multiple criteria decision reliability, the results of the estimations of alternative ranks received by the TOPSIS method from generated decision matrixes are used. The most common rank estimation for each alternative is identified, and a relative frequency of that estimation is calculated – a rank reliability level which is expressed in per cent:

$$p(A_i) = \frac{n_i(l)}{K} \cdot 100\%, \qquad (10)$$

where $p(A_i)$ – confidence level of rank estimation l assigned to alternative A_i; K – sampling volume of results; $n_i(l)$ – frequency of rank estimation l assigned to the most frequent alternative A_i; [11].

6. Illustrative numerical examples

In order to illustrate the results of the analysis proposed, a multiple criteria task was formulated: "Which foreign country has the most favourable conditions for the business establishment?". Five countries (alternatives) were chosen: A_1 – Netherlands, A_2 – Switzerland, A_3 – Germany, A_4 – Norway, A_5 – Sweden. The alternatives were evaluated in respect of six attributes: R_1 – time necessary to establish a business (the number of calendar days necessary to complete procedures of legal business establishment), R_2 – time for tax preparation and payment (time necessary to prepare, fill in and pay taxes, in hours per year), R_3 – total number of taxes, R_4 – GDP per capita, R_5 – urban population (percentage of the total population), R_6 – initial procedures to register a business (procedures which are necessary to start a business, obtain all necessary documents and licences, and complete verification of all required documentation). Data for 2005–2013 attribute values have been taken from the World Bank Data website, at http://data.worldbank.org/indicator.

By using the indicator values, nine decision matrixes with equal indicator significances were made, and the evaluation of alternatives by the TOPSIS method was conducted. Priority lines of alternatives are shown in the diagram (Fig. 1).

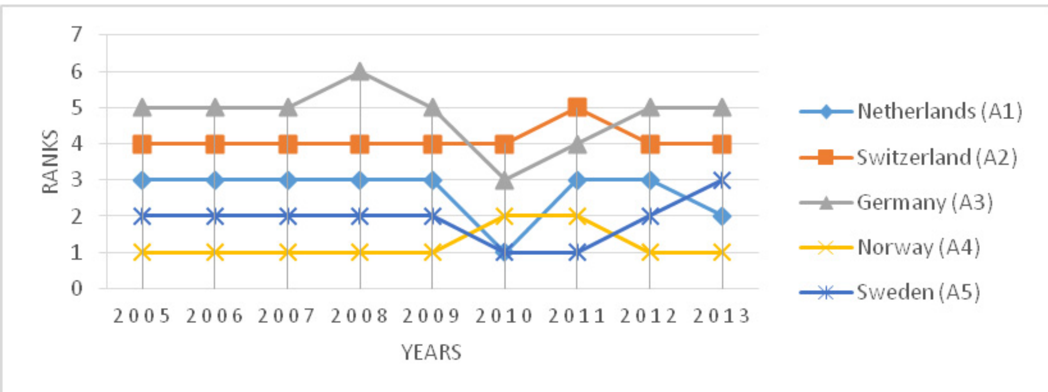

Fig. 1. Alternative ranking by the TOPSIS method

Based on the priority lines of alternatives, the conclusion suggests that the most favourable conditions to establish a business are in Norway (A_4) – rank 1.

Provided we know the indicator distributions and will be solving this multi-objective task for one year each day, would the alternative ranking change? In order to answer this question, a research is conducted during which indicator value K samples (K = 100) are generated according to a selected probability distribution. Decision matrixes are made from generated values for which, applying the TOPSIS method with equal size indicator significances, TOPSIS criterion values to alterations were obtained.

The following probability distributions were chosen for the research: uniform, normal, log-normal, and beta. Having conducted alteration ranking, rank mode estimation was calculated for each alternative, and a reliability level was calculated for this estimation according to the formula (10). The most frequently obtained rank estimations of alternatives and their reliability levels are provided in tables 1–2.

Table 1. Values of the most frequent ranks and reliability levels in cases of uniform, normal and log-normal distributions

Alternative	A1	A2	A3	A4	A5
Most frequent rank value	*3*	*4*	*5*	*1*	*2*
Reliability level in case of a uniform distribution	36%	56%	62%	59%	41%
Reliability level in case of a normal distribution	41%	69%	73%	65%	40%
Reliability level in case of a log-normal distribution	*55%*	*73%*	*75%*	*73%*	*50%*

Table 2. Values of the most frequent ranks and reliability levels in case of beta distribution

Alternative	A1	A2	A3	A4	A5
Most frequent rank value	*1*	*4*	*5*	*1*	*3*
Reliability level in case of beta distribution	32%	41%	62%	45%	31%

It may be observed that in case of uniform, normal and log-normal distributions, the values of the most frequent alternative ranks coincided and their reliability levels correlated. In case of beta distribution, ranks of A1 and A5 alternatives changed, and the reliability level of most frequent ranks is significantly lower than in cases of uniform, normal and log-normal distributions.

Verifications were made whether the TOPSIS criterion value samples of alternatives have the same distribution as indicator value samples. Significance level $\alpha = 0.05$. Chi-squared and Kolmogorov–Smirnov's criteria are used to verify the hypotheses, TOPSIS criterion value distribution of alternatives.

- Initial data were distributed according to a uniform distribution – do alternative samplings have a uniform distribution?

Table 3. Goodness-of-fit test estimation with $\alpha = 0.05$ that the result samples have a uniform distribution

Alternative	*Kolmogorov–Smirnov*		*Chi-squared*	
	p-value	H_0	*p-value*	H_0
A_1	0.33519	Accepted	N/A	
A_2	0.85131	Accepted	N/A	
A_3	0.50527	Accepted	N/A	
A_4	0.89737	Accepted	N/A	
A_5	0.18979	Accepted	N/A	

- Initial data were distributed according to a normal distribution – do alternative samplings have a normal distribution?

Table 4. Goodness-of-fit test estimation with α = 0.05 that the result samples have a normal distribution

Alternative	*Kolmogorov–Smirnov*		*Chi-squared*	
	p-value	H_0	*p-value*	H_0
A_1	0.99848	Accepted	0.8261	Accepted
A_2	0.19538	Accepted	0.21146	Accepted
A_3	0.20197	Accepted	0.63488	Accepted
A_4	0.16856	Accepted	0.11207	Accepted
A_5	0.38183	Accepted	0.92826	Accepted

- Initial data were distributed according to a lognormal distribution – do alternative samplings have a lognormal distribution?

Table 5. Goodness-of-fit test estimation with α = 0.05 that the result samples have a log-normal distribution

Alternative	*Kolmogorov–Smirnov*		*Chi-squared*	
	p-value	H_0	*p-value*	H_0
A_1	0.63119	Accepted	0.59078	Accepted
A_2	0.774656	Accepted	0.63072	Accepted
A_3	0.62623	Accepted	0.31217	Accepted
A_4	0.17572	Accepted	0.19329	Accepted
A_5	0.94423	Accepted	0.96434	Accepted

- Initial data were distributed according to a beta distribution – do alternative samplings have a beta distribution?

Table 6. Goodness-of-fit test estimation with α = 0.05 that the result samples have a beta distribution

Alternative	*Kolmogorov–Smirnov*		*Chi-squared*	
	p-value	H_0	*p-value*	H_0
A_1	0.58559	Accepted	0.21565	Accepted
A_2	0.8652	Accepted	0.94988	Accepted
A_3	0.99418	Accepted	0.90772	Accepted
A_4	0.80777	Accepted	0.07861	Accepted
A_5	0.86443	Accepted	0.28172	Accepted

Having conducted a verification of distribution compatibility, it was noticed that the distributions of the TOPSIS criterion values of the alternatives comply with the indicator value distributions.

7. Conclusions

1. Following the research, it was observed that the TOPSIS method is sensitive to indicator distributions.
2. Taking the distributions analysed into consideration, the TOPSIS method is the most sensitive for a beta distribution and the least sensitive for a log-normal distribution.
3. Having conducted a research on the compliance of initial data with result distributions, it was observed that the distributions of the TOPSIS criterion values of alternatives comply with the indicator distributions.

References

1. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai I.* Vilnius, TEV, 2006.
2. Giupponi C., Sgobbi A., Mysiak J., Camera R., Fassio A. NETSYMOD – an integrated approach for water resources management, *In Meire, P., Coenen, M., Lombardo, C., Robba, M., Sacile, R. (eds) Integrated Water Management: Practical Experiences and Case Studies.* p. 69–93, 2007.
3. Hoare, A., Regan, D.G., Wilson, D. P. Sampling and sensitivity analyses tools (SaSAT) for computational modelling, *Theoretical Biology and Medical Modelling,* 5(4) 2008. doi: 10.1186/1742-4682-5-4.
4. Hwang C. L., Yoon K. Multiple Attribute Decision making – methods and applications. *A State of the Art Survey*. Springer Verlag, Berlin, Heidelberg, New York, 1981.
5. Mun, J. *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Stochastic Forecasting, and Optimization*. 2006.
6. Podvezko V. Sudėtingų dydžių kompleksinis vertinimas. *Verslas: teorija ir praktika.* Vilnius: Technika, 9(3), p. 160–168, 2008.
7. Rubinstein, R., Reuven Y., Kroese, Dirk P. *Simulation and the Monte Carlo Method. 2nd edition*. 2008.
8. Sakalauskas, L., Žilinskas, K. Application of statistical criteria to optimality testing in stochastic programming. *Ūkio technologinis ir ekonominis vystymas (Technological and economic development of economy).* 12(4), p. 314–320, 2006
9. Saltelli A., Chan K., Scott E. M. *Sensitivity Analysis.* John Willey & Sons, Ltd. 2000.
10. Simanavičienė R., Ustinovičius L. Daugiatikslių sprendimo priėmimo metodų jautrumo analizė taikant Monte Karlo modeliavimą. *Informacijos mokslai*, Vilnius, Vilniaus universiteto leidykla, p. 182–190, 56, 2011,
11. Simanavičienė R. Statistinių metodų taikymas daugiatikslių sprendimų patikimumui įvertinti, *Informacijos mokslai,* 65, p. 120–126, 2013.
12. Triantaphyllou, E. A sensitivity analysis approach for some deterministic multi-criteria decision making methods, *Decision Sciences*, 28(1), p. 151–194, 1997.
13. Triantaphyllou E. *Multi-criteria decision making methods a comparative study*. Boston: Kluwer Academic Publishers, 2000.
14. Wah B. Y., Razali M. N. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling test, *Journal of Statistical Modelling and Analytics*, 2(1), p. 21–33. 2011.

METODO TOPSIS JAUTRUMO ANALIZĖ PRADINIŲ DUOMENŲ SKIRSTINIŲ ATŽVILGIU

Rūta Simanavičienė, Vaida Petraitytė

Santrauka. Straipsnyje nagrinėjamas daugiakriterinio sprendimo priėmimo metodo TOPSIS jautrumas rodiklių tikimybinių skirstinių atžvilgiu. Tyrimui atlikti, pradiniai duomenys – rodiklių reikšmės, buvo generuojamos pagal normalųjį, lognormalųjį, tolygųjį ir beta skirstinius. Iš sugeneruotų duomenų buvo konstruojamos sprendimo matricos. Taikant TOPSIS metodą sugeneruotoms matricoms, gautos rezultatų imtys. Buvo atliekama gautų rezultatų statistinė analizė, kuri parodė, jog pradinių duomenų skirstiniai nebūtinai sutampa su TOPSIS metodu gautų rezultatų skirstiniais. Pagal dažniausiai pasitaikančią alternatyvos rango reikšmę nustatyta, jog metodas TOPSIS yra labiausiai jautrus duomenų pasiskirstymui pagal beta skirstinį, mažiausiai jautrus duomenų pasiskirstymui pagal lognormalųjį skirstinį.

Reikšminiai žodžiai: daugiakriterinis sprendimų priėmimas, TOPSIS metodas, jautrumo analizė, tikimybinis pasiskirstymas.