Performance of Entropy-Based Criteria Weights in Solving Multi-Criteria Problems

M Mat Kasim and A A Jemain

1School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, Sintok, 06010 Kedah, Malaysia
2School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor, Malaysia

Email: 1maznah@uum.edu.my

Abstract. Weights of criteria are often predetermined prior to solving multi-criteria problems. One way to estimate the weights is by manipulating the intrinsic information contained in the criteria. The measure of entropy which is usually modelled by probability concept has been used to represent the criteria weight. The aim of this paper is to propose the use of density functions obtained by kernel-based method to represent the probability in estimating entropy of criteria. Furthermore, a simulation experiment was used to investigate the pattern of decisions provided by the proposed method in solving multi-criteria problems. The results showed that the performance of the proposed method is quite consistent. The use of probability density functions representing probability in computing entropy of criteria contributed to another development in entropy theory. Hence, the use of the resulted entropy-based criteria weights would lead to a more valid outcome in resolving multi-criteria decision-making problems.

1. Introduction
A multi-criteria problem usually comprises of a set of alternatives evaluated under several predefined criteria. The problem can be represented as a decision matrix as in the following figure.

| Alternatives/Attributes | $c_1$ | $c_2$ | ... | $c_n$ |
|-------------------------|-------|-------|-----|-------|
| $a_1$                   | $x_{11}$ | $x_{12}$ | ... | $x_{1n}$ |
| $a_2$                   | $x_{21}$ | $x_{22}$ | ... | $x_{2n}$ |
| ...                     | ...    | ...   |     | ...   |
| $a_m$                   | $x_{m1}$ | $x_{m2}$ | ... | $x_{mn}$ |

Figure 1: A general representation of a multi-criteria problem as a decision matrix

The final phase of solving this problem is to aggregate the local scores of the alternatives with the weights of the attributes or criteria by using certain mathematical aggregation operators which are usually referred to as multi-criteria decision making (MCDM) methods. Then, based on the overall scores, the alternatives are ranked or grouped, or the best alternative will be chosen. Since the weights of criteria have major role in computing the overall scores of the alternatives, the concept of weight of criteria is very important in the whole decision-making process [1,2] even though these weights do not
have any consensus meaning. Choo et al. [3] suggested several meanings such as marginal or relative contribution towards optimal unit and gradient of the overall value function.

There are two main approaches to find weights of criteria, which are subjective and objective approaches [4], where the first approach depends on evaluation given by evaluator(s), while the latter one relies on the intrinsic information contained in the criteria. Several subjective approaches are available such as Analytic Hierarchy Process (AHP) [5,6], rank-based [7], direct rating and point allocation methods [8]. Meanwhile, examples of objective approaches are CRITICs [9], and entropy-based method [10]. The concept of entropy was introduced by Shannon and Weaver [11] in which entropy of an event is defined in terms of the probability of the outcomes of the event. However, the probability was defined as the proportion of the observed value or outcome to its maximum value [10]. Besides that, Hwang and Yoon [12] defined entropy as the ratio of each observation to the sum of all observations, while Chen and He [13] looked at probability of an outcome as the proportion of the absolute value of the difference between the outcome or the observation and its average value to the sum of all differences. These definitions of probability are not accurate since probability is often defined as the proportion of the number of targeted outcomes to the number of all possible outcomes.

Nevertheless, many researches used the entropy-based weighting method in solving multi-criteria problems, for example, Hwang and Yoon’s entropy weighting method was used by Vatansaver and Akgul [14] to calculate attributes’ weight to measure the performance of airline companies’ websites, Poursarabi and Ghasemi [15] determined the importance of doing business indicators in Iran, while Deng et al. [16] used Zeleny’s entropy method in determining criteria weights in selecting apples of a food company that produces apple juice concentrate.

This paper aims to introduce one alternate definition of probability in calculating entropy, which is based on the density of the criteria that can later be used to estimate weights of the criteria. This proposed weighting method had been used to estimate criteria weights for construction of hospital index [18]. In order to investigate the performance of this new estimates of entropy-based weights of criteria, a simulation experiment was carried out where several multi-criteria problems were generated. For each generated multi-criteria problem, the proposed entropy-based criteria weights were used in the process of finding overall scores and ranking of the alternatives. The performance of the proposed method was compared with other existing methods.

This paper is organized as follows. The next section discusses about the definition of entropy and the existing entropy-based weighting methods. It is followed by the proposed density-based entropy method, simulation experiment, its results and discussions, and the paper concludes with a conclusion section.

2. Definition of Entropy and the Existing Entropy-Based Weighting Methods

Entropy which characterizes uncertainty is traditionally used as the average measure of information contained in a message [11], is defined by probability theory as follows. Entropy, $E_n(H)$, of a discrete random event $H$ with possible outcomes, $i = 1, 2, ..., n$, is given as

$$E_n(H) = -\kappa \sum_{i=1}^{n} p_i \ln(p_i)$$  \hspace{1cm} (1)$$

where $\kappa$ is a constant or a normalized factor which usually equals to $1/\ln(n)$, and $p_i$ is the probability of outcome $i$. $E_n(H)$ value is between zero and one, and it reaches its maximum value of one, when all outcomes have the same chance to occur or $p_i = 1/n$ for all. Zeleny [10] has extended the use of entropy as an estimate of the degree of uncertainty among alternatives within each criterion.
Furthermore, the complement of uncertainty, that is, certainty is taken as the proxy measures of criteria weights. The weight of criterion $j$ is given as

$$w_j = \frac{1 - En_j}{\sum_{j=1}^{m} 1 - En_j}$$  \hspace{1cm} (2)

with $En_j$ is the entropy measure of criterion $j$, $j = 1, \ldots, m$. He also proposed that entropy value of criterion $j$ as

$$E^n_j(x_{1j}, \ldots, x_{nj}) = -\kappa \sum_{i=1}^{n} \frac{x_{ij}}{\bar{x}_j} \ln \left( \frac{x_{ij}}{\bar{x}_j} \right)$$  \hspace{1cm} (3)

with $\bar{x}_j$ the maximum value of $x_{ij}$, $i = 1, \ldots, n$.

Hwang and Yoon [12] also used entropy as a base to measure the weights of criteria. They proposed the following formula to measure entropy, $E^n_j^{hy}$ and is given as

$$E^n_j^{hy}(x_{1j}, \ldots, x_{nj}) = -\kappa \sum_{i=1}^{n} \left( \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \right) \ln \left( \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \right)$$  \hspace{1cm} (4)

Meanwhile, Chen and He [13] suggested the following formula to measure entropy of a criterion.

$$E^n_j^{ch}(x_{1j}, \ldots, x_{nj}) = -\kappa \sum_{i=1}^{n} \frac{|x_{ij} - \bar{x}_j|}{\sum_{i=1}^{n} |x_{ij} - \bar{x}_j|} \ln \left( \frac{|x_{ij} - \bar{x}_j|}{\sum_{i=1}^{n} |x_{ij} - \bar{x}_j|} \right)$$  \hspace{1cm} (5)

with $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$.

Based on formula (3), (4) and (5), the probability, $p_j$ are represented in terms of maximum observation, sum of observations, or absolute differences of observations from the mean of the criterion respectively which are not exactly represent the true meaning of probability. The following section focuses on the proposed method, called as density entropy where the probability is represented as the probability density function (pdf) of the criterion.

3. Proposed Methodology in the Probability Setting

As previously mentioned, to determine the entropy, the probability value, $p_j$, in (1) are stated as $\frac{x_{ij}}{\bar{x}_j}$ as in (3) or $\frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$ as in (4) where these two expressions are not exactly representing probability. However, probability density function (pdf) is a natural concept in probability theory. The pdf is usually used to indicate the probability of a random variable to be within a particular range of value as opposed to taking on any one value, and is very appropriate with the entropy definition given by Shannon and Weaver. Estimating the density of a data set can be done in two approaches. The first approach is by using the parametric density function based on certain statistical distribution function which is assumed to be suitable to the real data distribution. The estimate of the density is feasible if the estimate of the parameter is available. The second method is discussed by Silverman [19] and Scott [20]. This method is special since it can be used on any data set without any assumption about its distribution. Even though the calculation is very intensive, the advanced technology may help to smooth out the computation and able to produce robust estimate [20].

One way to estimate the density of data set nonparametrically, is with a kernel estimation technique. Let $x_{1j}, \ldots, x_{nj}$ be the quality values of $n$ alternatives for criterion $j$ with the density function of
\( f_j(x_{1j}, \ldots, x_{nj}) \). So, the density function can be approximated for \( f_j(x_{1j}, \ldots, x_{nj}) \). Suppose the pdf is \( \hat{f}_j(x_{1j}, \ldots, x_{nj}) \), be given as

\[
\hat{f}_j(x_{1j}, \ldots, x_{nj}) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x_{ij} - \bar{x}_j}{h} \right) = f_j
\]  

(6)

with \( h = 1.06\sigma_j n^{-\frac{1}{5}} \) is the optimal bandwith or the smoothing parameter, \( \sigma_j \) is the standard deviation of criterion \( j \), and \( K(.) \) is any suitable kernel function. Please note that the value of \( h \) is not fixed, and this value was chosen as suggested by Silverman (1986) when Gaussian is used as the kernel function. In this research, the following kernel Gaussian function is used due to its efficency score is above 95\% (Silverman, 1986).

\[
K(t) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} t^2 \right) \text{ for } |t| < \infty
\]  

(7)

In general, a kernel function is any function that has the following properties.

\[
\int K(t)dt = 1, \quad \int tK(t)dt = 0, \quad \text{and} \quad \int t^2K(t)dt = c \neq 0
\]  

(8)

Based on the estimated density function, the entropy value, \( E_{n_j}^{den} \) can be estimated as follows.

\[
E_{n_j}^{den}(x_{1j}, \ldots, x_{nj}) = -\kappa \sum_{i=1}^{n} \left( \frac{f_{ij}}{\sum_{j=1}^{n} f_{ij}} \right) \ln \left( \frac{f_{ij}}{\sum_{j=1}^{n} f_{ij}} \right).
\]  

(9)

Then, the weight of criterion \( j \) can be estimated based on the complement value of the entropy as in (2), and the normalization process is performed on the values to assure that the sum of the weights is one.

### 4. Simulation Experiment

In this section, the discussion focuses on the performance of the three entropy-based weights as proxy measures of criteria weights in MC problems, with the assumption that all of the three methods are able to measure entropy within probability setting. A simulation experiment was conducted by using R programming language. Various MC methods were generated by changing the number of criteria: 3, 4, 5, 6, and 7 with two different numbers of alternatives (SZ): 5 and 20. Creation of the MC problems was done arbitrarily. The combination of different number of criteria and alternatives produces 20 different MC problems. Three consistency measures were used to compare the quality of decisions of the three criteria weights produced by each entropy method. The comparative measures used were average percentage of exact top ranked alternative (TR), average percentage of all exact alternatives’ rankings (AR), and average of absolute difference of all alternatives’ rankings (RD).

Five main steps in the simulation experiment were

i. Generation of ‘real’ entropy-based weights or ‘population’ entropy weights

ii. Generation of sample entropy-based weights

iii. Normalization of performances of alternatives

iv. Calculation of overall score of alternatives by using simple weighted average method by using ‘real’ entropy-based weights (RR) and ‘sample’ entropy-based weights (SR).

v. Giving ranks to the two sets of evaluated alternatives based on values obtained in part (iv)

vi. Comparison of top ranked alternatives (TR) between RR and SR.

vii. Comparison of all rankings of all alternatives (AR) between RR and SR.

viii. Calculation of average of absolute ranking difference (RD) between RR and SR.

The first step is generation of ‘real’ weights of each of the four entropy-based methods. It was done by generating 5000 numerical values randomly from each of the following distributions: 1) integer numbers valued between 250 and 5050; 2) Normal with mean =100 and standard deviation = 20; 3) Chi
with degree of freedom (d.f) = 60; 4) Gamma with mean = 8 and standard deviation = 2; 5) Normal with mean = 10 and standard deviation = 2; 6) Gamma with mean = 8 and standard deviation = 1; and 7) Chi with d.f = 6. These distributions were used with the assumption that each criterion was representing alternatives’ performances which was different from other criteria. The values would form a decision matrix as portrayed in Figure 1. If the number of criteria is 3, a matrix of size 3 x 5000 would be formed randomly. This step was repeated but with different number of criteria: 4, 5, 6, and 7 criteria and 5000 alternatives. Based on these ‘big’ matrices, four sets of weights of each method were obtained for each type of MC problem.

At the second stage, random sampling from the ‘populations’ at first stage were done where for each sample of matrix, the sample entropy-based weights were computed for each criterion. This process was repeated for each sample size and for each type of MC problem. Quality of the alternatives or the values within each criterion were normalized, by dividing each observation with its maximum value, as given in the following formula.

\[ z_{ij} = \frac{x_{ij}}{\sum_{j=1}^{N} x_{ij}} \]  

(10)

At stage 4, the final score of each alternative was calculated by using ‘real’ weights and ‘sample’ weights. Here, simple weighted average method,

\[ pres_i = \sum_{j=1}^{N} z_{ij} w_j \]  

(11)

where \( j = 1, 2, \ldots, N \), \( i = 1, 2, \ldots, SZ \), and \( N = 3, 4, 5, 6, 7 \), and \( SZ = 5 \) and 20 was used, and the rankings of the alternatives were determined based on the index values. There were two rankings, one based on ‘real’ weights, denoted as RR, and another one based on ’sample’ weights, SR.

The next step is to compare the RR and SR in terms of all rankings (AR), top ranked alternatives (TR), and the average of absolute differences of rankings (RD) between RR and SR. All generated cases of MC problems were replicated 100 times. The following section provides the results and related discussions.

5. Results of the Simulation Experiment

Table 1 shows the percentage of all equivalent of rankings (AR) between RR and SR. The value 100 means rankings of alternatives between RR and SR are equivalent, whereas the value less than 100 means that not all rankings between RR and SR are equivalent. When the number of criteria is 3, and the number of alternatives is 5, all methods did not give equal ranking except the \( den \) method. When the number of criteria increase to 4, \( den \) performance is equal to \( ze \) and \( hy \). All methods had the best performance when the number of criteria is 7 and the number of alternatives is 5.

| No. of alternatives | No. of criteria | No. of alternatives | No. of criteria |
|---------------------|----------------|---------------------|----------------|
|                     | 3   | 4   | 5   | 6   | 7   | 3   | 4   | 5   | 6   | 7   |
| 5                   | ze  | 90  | 98  | 100 | 100 | 100 | 57  | 51  | 90  | 90  |
|                     | hy  | 89  | 98  | 100 | 100 | 100 | 84  | 48  | 90  | 90  |
|                     | ch  | 89  | 97  | 100 | 100 | 100 | 57  | 53  | 90  | 80  |
|                     | den | 100 | 98  | 98  | 100 | 100 | 90  | 63  | 59  | 71  |

Table 1: AR Results
Method *den* outperformed other methods when the number of alternatives is 20 and the number of criteria is 3 or 4. However, when the number of criteria is increased, from 4 to 5 or 6 with 20 alternatives, the other three methods are better than *den* method. All methods have same performance when the number of criteria is 7 with 20 alternatives.

![AR values vs different number of criteria with 5 alternatives](image1)

![AR values vs different number of criteria with 20 alternatives](image2)

**Figure 2**: AR values with different number of criteria

Figure 2 shows the performances of the four methods under study in terms of AR values with different number of criteria and alternatives.

| No of alternatives | No. of Criteria | 3 | 4 | 5 | 6 | 7 |
|--------------------|----------------|---|---|---|---|---|
| ze                 | 99             | 100| 99| 100| 100| 100|
| hy                 | 99             | 100| 99| 100| 100| 100|
| ch                 | 99             | 100| 99| 100| 100| 100|
| den                | 99             | 100| 94| 100| 100| 100|

Table 2: TR Results

| No of alternatives | No. of Criteria | 3 | 4 | 5 | 6 | 7 |
|--------------------|----------------|---|---|---|---|---|
| ze                 | 100            | 50 | 100| 100| 100| 100|
| hy                 | 100            | 50 | 99 | 100| 100| 100|
| ch                 | 100            | 50 | 99 | 100| 100| 100|
| den                | 100            | 100| 94 | 100| 100| 100|

Table 2 shows the percentage of top position (TR) of the alternatives in various types of MC problems. When the number of alternatives is 5, all methods seem to have about the same level of performance except when the number of criteria is 5, *den* method is at the lowest performance between the four methods under study. But when the number of alternatives was increased to 20, all methods gave 100% equal top rankings between RR and SR. When the number of criteria is 4, *den* method seems to be better, but with number of criteria is 5, *ze* method has the perfect equivalent performance. All methods are perfect equivalent of TR when the number of criteria was increased to 6 and 7.
Figure 3: TR values with different number of criteria

Figure 3 shows the performance of TR when the number of alternatives is 5 or 20 with 5 different number of criteria.

Table 3: RD Results

| No of Alternative | No of Criteria | Method | No of Alternative | No of Criteria |
|------------------|---------------|--------|------------------|---------------|
| 5                | 3             | ze     | 20               | 0.10          |
|                  | 4             | 0.03   |                  | 1.38          |
|                  | 5             | 0.00   |                  | 0.13          |
|                  | 6             | 0.00   |                  | 0.10          |
|                  | 7             | 0.00   |                  | 0.10          |

Table 3 displays the absolute difference in ranking (RD) of all alternatives for different types of MC problems. For these problems, performances of the methods are better when the values of RD are smaller. With 5 alternatives and different number of criteria, den method is the best when the number of criteria is 3. When the number of criteria increases to 4, 5, 6 and 7, all methods have the same performance. Increasing the number of alternatives to 20, den method is the best when the number of criteria is 3 and 4. Increasing the number of criteria to 5 and 6, the performances of the other three methods outperformed den method. However, when the number of criteria is 7, all methods have the same level of performance. The corresponding graph is in Figure 4.
6. Discussion and Conclusion

This paper concerns about the use of entropy as proxy or estimate measure of criteria weights in MC problems. This paper introduces another way of calculating entropy by changing how the probability is represented. It is really based on the concept of probability specifically as probability density function. Furthermore, a simulation experiment is conducted in order to study its consistency by comparing with other three established methods when the estimates are used in solving multi-criteria problems. The simulation results show that the performance of the introduced method is quite consistent. So, estimates of criteria weights by using kernel-based density method is another option in solving multi-criteria problems.

7. Acknowledgements

The presentation of this paper is funded by Universiti Utara Malaysia under the High Impact Group Research Grant (PBIT) with S/O Code 12870.

8. References

[1] Kashim R, Kasim M M and Rahman R A 2018 J. of Inf and Comm Tech 17 4 569-585
[2] Ahmad N, Kasim M M and Ibrahim, H 2017 J of Supply Chain Management 6 4 289-293
[3] Choo E U, Schoner B and Wedley W C 1999 Comp & Ind Eng 37 3 527–541
[4] Ma J, Fan Z and Huang L 1999 Euro J of Ope Res 112 2 397-404
[5] Saaty T L 1980 The analytic hierarchy process: planning, priority setting, resource allocation (Pittsburgh, PA: RWS Publication)
[6] Saaty T L 1996 Decision making with dependence and feedback: The analytic network process (Pittsburgh, PA: RWS Publication)
[7] Barron F H and Barrett B E 1996 Management Sci 42 11 1515-1523
[8] Bottomley P and Doyle J R 2001 Omega 29 6 553-560
[9] Diakoulaki D, Mavrotas G and Papayannakis L 1995 Comp & Oper Res 22 7 763–770
[10] Zeleny M 1982 Multiple Criteria Decision Making (New York: McGraw-Hill Book Company)
[11] Shannon C E and Weaver W 1949 The mathematical theory of communication (Urbana: The University of Illinois Press)
[12] Hwang C -L and Yoon K 1981 Multiple Attribute Decision Making (New York: Springer-Verlag)
[13] Chen J J G and He Z 1997 Fuzzy Sets and Systems 88 11-22
[14] Vatansever K and Akgül Y 2018 Dec Sci Lett 7 2 119-130
[15] Poursarabi M M and Ghasemi R 2016 Int Conf on Entrepreneurship (Tehran: Iran)
[16] Deng H, Yeh C H and Willis R J 2000 Comp & Oper Res 2710 963-973
[17] Işık A T and Adali E A 2017 Eur J of Interdisciplinary Studies 3 3 80-86
[18] Desa N M, Jemain A A and Kasim M M 2015 Sains Malaysiana 44 2 239-247
[19] Silverman B W 1986 Density Estimation for Statistics and Data Analysis (London: Chapman and Hall)
[20] Scott D W 1992 Multivariate Density Estimation: Theory, Practice, and Visualization (New York: John Wiley)