© Robin Rivest, M.Sc. Business Analytics, robin.rivest@videotron.ca

BUILDING AN EVALUATION "COMMON GROUND" FOR RESEARCH ON AHP REFINEMENTS

ABSTRACT

From its inception in the late 70's/early 80's, the Analytic Hierarchy Process (AHP) has attracted the interest of multiple criteria decision aid (MCDA) scholars and practitioners all around the world. Thousands of reviewed papers on the AHP have been published through the years, and its reach is extensive, as one might consider that if you put "analytic hierarchy process" in Google Scholar, you get over 1.5 million entries.

Most AHP research papers report on specific applications of the method to various business decision contexts. However, there is a significant portion of these publications that explore the inner workings or theoretical aspects of the method, proposing improvements or bringing attention to potential drawbacks. A large proportion of those papers address one of the three following topics: choosing a numerical scale, reducing the number of necessary comparisons, or deriving the priority vector.

When one considers this overall body of work, the disconcerting finding is that none (or too few of them) share a common basis to illustrate or highlight the performance of their suggestion. There is a compelling need for an evaluation framework that would enable informative comparisons between the various approaches. This lack of a validation "common ground" is the subject of my master's thesis [Rivest, 2019] which introduces several foundational elements with the intent to launch a dialog with both scholars and practitioners about providing the AHP community with a shared evaluation framework for these areas of research and more.

Keywords: characterizing the priority vector, power of discrimination, limit of precision in preference quantification, simulation techniques, representative test cases, test domain coverage.

1. Introduction

The purpose of this study is twofold. First, it aims to raise the awareness of the absence of a "common ground" to evaluate and to compare results of experiments. Second, it also exposes some foundational orientations to circumscribe and resolve this deficiency.

For instance, let us start with the "reduction of comparisons" quandary. In his review of the state of research on the AHP, Brunelli (2014, p.40) specifically mentions about this subject that « *There are various research papers on methods for dealing with incomplete preferences, but very few investigated the relation between the number of missing comparisons and the stability of the obtained priority vector … It is safe to say that there is need and space for further investigation* ».

Evaluations of various suggestions to reduce the number of necessary comparisons are conducted in dissimilar frameworks. Many are only considering a few isolated test cases, while those which make use of more cases fail to establish why the implied coverage is suitable. Furthermore, in some instances, suggestions remain incomplete or undetermined.

Those evaluations that attempt to assess the validity of resulting priority vectors use a variety of irreconcilable proximity measures whose merit remains unstated. Therefore, it is nearly impossible to juxtapose results obtained from one study to the next.

Finally, one of the most important deficiency is the lack of a method to establish an objective threshold beyond which further gains would be unrequited. This prevents proper gauging of how far or how near the suggested approaches provide a practical satisfactory approximation.

Similar observations can be made with research on methods to derive the priority vector as well as the exploration of numerical scale alternatives.

2. Literature Review

Here are a few excerpts (and a high-level overview), from the literature review found in Rivest (2019) which illustrate some of the shortcomings mentioned in the introduction.

The reader will find that the disparity is exhibited for numerous aspects, such as: the dimensions of matrices (i.e. number of alternatives compared), the extent of cases tested and the use of simulation data. And, in all cases, no objective limit of precision is stated.

The following table contains a few examples of approaches to reduce the number of comparisons which illustrate the disparity of testing frameworks.

Source*	**Basic idea**	**Extent of cases tested**	**Proximity measure**	**Matrix size**	**Limit of precision**
Shen *et al*, 1992	Evaluate alternatives in multiple subsets, then combine by prorating results (single pivot)	Only one example	None	7	None
Ishizaka, 2012b	Evaluate alternatives in multiple subsets, then combine by prorating results (multiple pivots)	A few examples	None	12	None
Fedrizzi & Giove, 2013	Proceed by iteration until some condition of sufficiency (left to be determined) is attained.	Only one partial example	Incomplete	5, 9	None
Rezaei, 2015-2016	Make only (2n-3) comparisons using best and worst alternatives	46 participants	Rank variation; Total deviation; Euclidian distance	4, 5, 6	None
Pamučar *et al*, 2018	Use only (n-1) comparisons	A few examples	None	4, 5, 8	None
Abastante *et al*, 2019	Proceed with direct estimate of weights which are then calibrated (prorated) with priorities obtained for only a subset of alternatives	98 participants	MSE between non-normalized vectors	10	None

* See full bibliography in Rivest (2019)

These observations are followed by similar findings in the research on methods to derive the priority vector as well as the exploration of numerical scale alternatives. The review concludes with several remarks including these:

- Use of simulation data is neither widespread, nor systematized.
- Simulated priority vectors are almost always generated using the *uniform* distribution, which will we see, provides a test coverage that is not properly aligned with the value domain.
- No characteristics have been proposed to assert the extent of test cases needed to ensure proper coverage
- There is no convention for the size of matrices to be verified
- There is no convention on the proximity measures to be used
- The undertaken limit of precision is either undefined or set arbitrarily

These findings motivate the identification of solution paths.

3. Challenges and solution paths

To overcome the shortcomings identified in the literature review the following challenges must be resolved : identifying a few key characteristics for priority vectors that can be used to define practical value domains from which to draw sufficiently diverse test cases; developing a line of reasoning for the selection of an all-around applicable proximity measure; and, another one to define an objective limit of precision for assessing how close two priority vectors really are.

Value domain for priority vectors

A self-evident characteristic of a priority vector is its size or dimension, given by the number of alternatives. A more subtle one is hinted at by Ishizaka *et al* (2012, p.4769) in the following sentence: « *A high difference of performances can also be highly discriminating even with a low weight of the criterion.* »

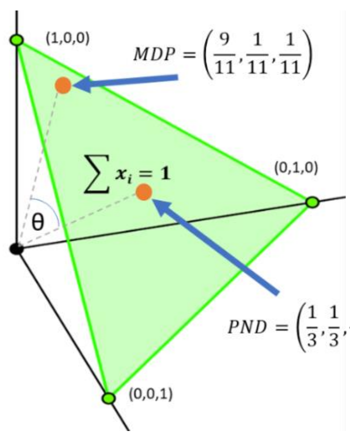

Fig.1 Geometric interpretation of the angle (θ) as a measure of the power of discrimination. (a.k.a. cosine distance)

One can consider that the priority vector with the less discriminating power is the one in which all weights are equal. Let us refer to it as the Point-of-No-Difference (PND). We can then entertain the thought that the one which gives the maximum weight to one option and the least weight to all others has the most discriminating power (MDP).

This concept lends itself to the following simple geometric interpretation which can be illustrated with a graph showing both points on a 2-simplex (see Fig.1, left): the angle between any vector and the PND can be used as a measure of its power of discrimination.

Note that the maximum value will be constrained by the highest relative importance or potential, ρ attributed for a given set of alternatives. In Fig.1, $\rho = 9$ gives the vector $\left(\frac{9}{11}, \frac{1}{11}, \frac{1}{11}\right)$ which is 45.9° from the PND, whereas the MDP when $\rho = 5$, the vector $\left(\frac{5}{7}, \frac{1}{7}, \frac{1}{7}\right)$, would be closer to the PND at 38.9°.

This gives us three attributes to characterize priority vectors that can be used to guide the generation of simulation data in such a way that the value domain is adequately covered.

Selecting a sensible proximity measure

In their Encyclopedia of distances, Deza & Deza (2009, p.298) state: « *There are many similarities used in Data Analysis; the choice depends on the nature of data and is not an exact science.* ».

Measuring the power of discrimination with the cosine distance leads intuitively to the idea of doing the same for the proximity measure. In this context, one might argue that it is approximately equivalent to the Euclidian distance (the usual "go to" measure of proximity between vectors). In the original AHP method, vectors are normalized to sum 1, but certain adaptations use different normalizations. For instance, Schoner et al (1993) explore the use of different normalizations which would make the Euclidian distance less judicious but would not deter the cosine distance.

Thus, the cosine distance has properties that can make it more comprehensive for the purpose of measuring the proximity of priority vectors and is also easier to interpret.

Establishing an objective limit of precision

Having an objective limit of precision is relevant to providing an upper bound for how far an approximation can deviate while remaining suitable.

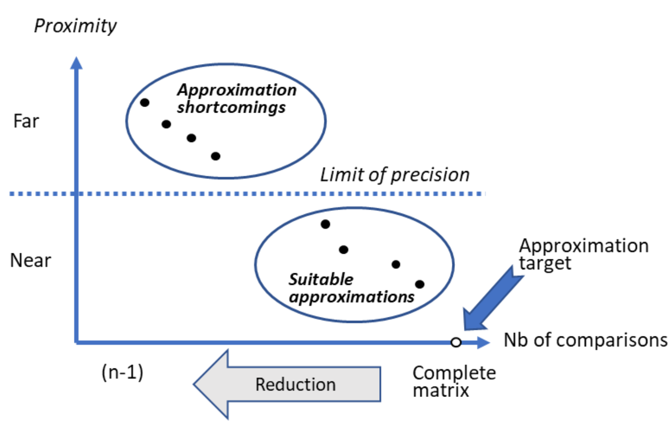

Fig.2 As the number of comparisons is reduced, the approximation error is likely to grow.

Fig.2 (left) illustrates the use of this concept with the reduction in the number of comparisons used.

As the number of comparisons is reduced the priority vector obtained may get farther away from the one that would have been obtained if all comparisons had been elicited.

One way to establish such a limit is to consider what Triantaphyllou & Mann (1990, p.297) refer to as the *forward error*, which is attributed to the use of a discrete numerical scale.

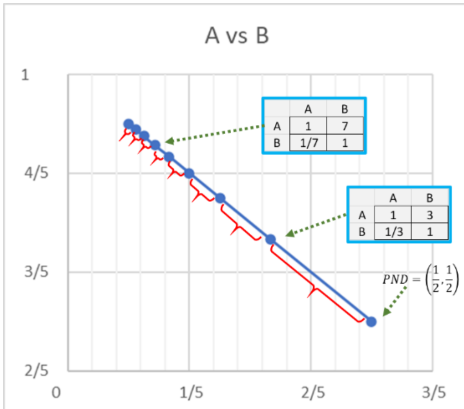

Fig. 3 Brackets (in red) show gaps between possible priority vectors when using the original linear scale (1 to 9).

Given the fact that a discrete scale is used to map elicited expressions of relative importance to numerical values, there are gaps which might be measured to represent the level of imprecision (or *discretization error*) that cannot be overcome.

Fig. 3 (left) illustrates those gaps for a vector of dimension 2.

4. Limit of precision

Combining the concepts described above and using an appropriate strategy to generate priority vectors, it is possible to establish a justifiable limit of precision in three steps.

<u>First, simulation data have to be produced.</u>

Fig.4a and 4b (below – left) illustrate the overall spectrum of values for the power of discrimination with the angle between the vector and the PND for vectors of dimensions 3 to 18 when using the linear scale (1 to 9). Blue triangles represent the upper bounds for each dimension. Orange diamonds represent the lower bound (the first step above the PND). They also highlight the limited coverage obtained when only drawing weight values from the uniform distribution (dotted grey and yellow lines).

Fig.4c and 4d (below – right) illustrate a series of 500 randomly generated points that more appropriately cover the spectrum of the power of discrimination for vectors of dimension 11. Note: Due to the restricted format required for this publication, these illustrations are limited to ρ =5 and 9. More details can be found in Rivest (2019).

<u>Second, discretization must be imparted.</u>

To this end, a process inspired from Triantaphyllou and Mann (1990, p.298) can be used to impart the effect of discretization and have pairs of points (before and after) from which proximity can be measured. Fig.5 (below) illustrates the steps of the discretization process.

<u>Last, consider the results and choose an appropriate value for the intended application.</u>

Fig.6 (below) illustrates the distribution of the discretization error obtained for 250 samples with $n = 11, \rho = 5$. The graph shows that all, but three points (~1%), are less than 6° away.

The average angle is 2.85° and the 95th percentile is 5°. So, one can reasonably make the interpretation that the limit of precision is about 5° to 6°.

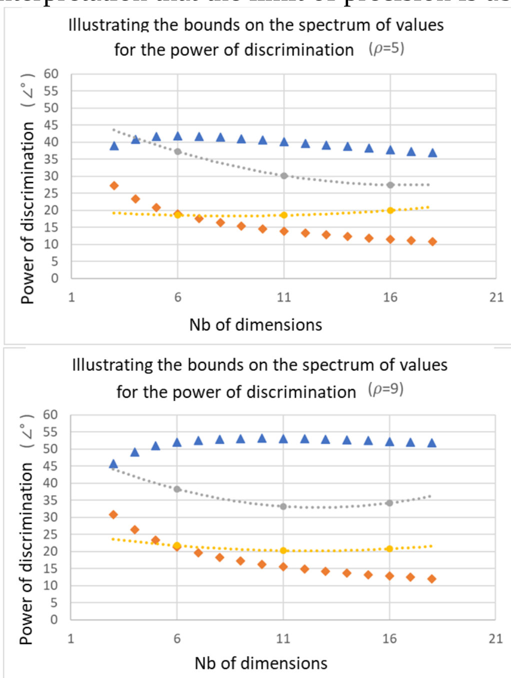

Fig.4 a) and b) required coverage

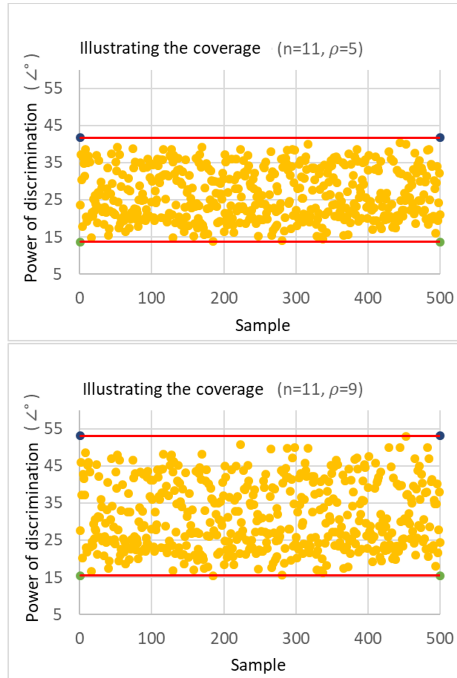

Fig.4 c) and d) generated coverage

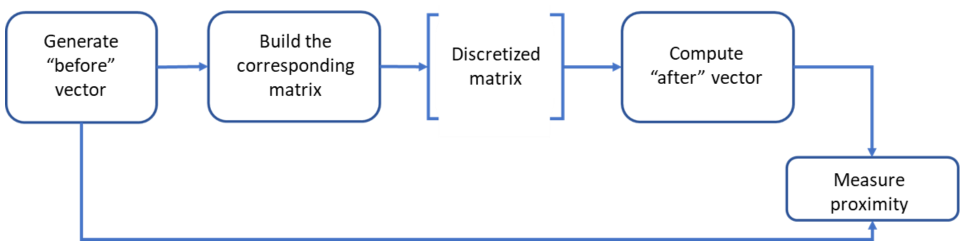

Fig.5 Process to measure the discretization error

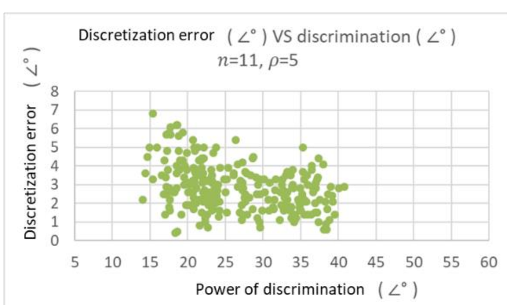

Fig.6 Sample results for discretization error

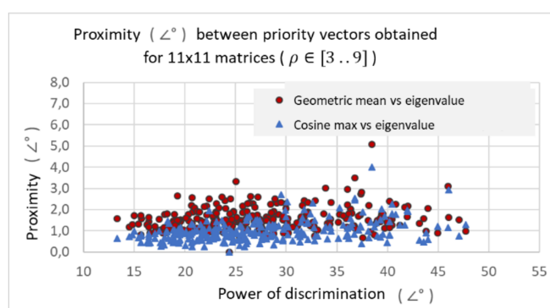

Fig.7 Application of the concepts to evaluate the proximity of the priority vectors with three different methods from the same matrices.

5. Application: an example

There has been a long on-going debate over which method is best suited to derive the priorities from a comparison matrix. Some arguments are made from a purely mathematical perspective while others are made from other standpoints (e.g. practical issues). Here it is examined with a different kind of rationale, which can be summed up as: « What do the actual numbers tell us? ».

To illustrate how the evaluation framework described above can be applied, three methods for obtaining the priority vector are compared: the original right eigenvector, the geometric mean (Williams and Crawford, 1980, p.22) and the cosine maximization (Kou and Lin, 2014).

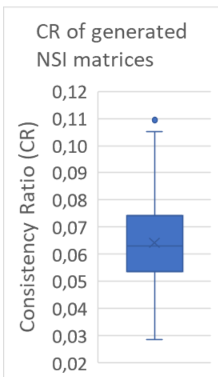

Fig.8 CR of NSI matrices

Fig.7 (above) illustrates that all three methods give results which are well within the limit of precision of each other. Thereby, one can state that, despite the theoretical or conceptual significance of various arguments, in the end, it might not make much of a difference from a practical point of view.

Note: Herman and Koczkodaj (1996, p.26) describe a method to impart some inconsistency in a generated matrix which they refer to as NSI (or *Not-So-Inconsistent*) matrices. For this application example, Fig.8 (left) shows the distribution of the consistency ratio (CR) for the set of matrices from which priority vectors were derived.

6. Limitations

The following elements must be taken into consideration before adopting the approach described here to conduct evaluations of potential refinements to the AHP process: 1) The limit of precision is dependent on the numerical scale to be used, e.g. when using the geometric scale with parameter $\sqrt{2}$, the limit of precision falls to ~ 4°; 2) All simulations and test runs where conducted using the more advanced functions of Microsoft Excel® and require a fair amount of manual interventions. Having a shareable integrated test environment implementing this instrumentation would make adoption easier and more straightforward.

7. Conclusions

The three main contributions of this study are: 1) A way to characterize priority vectors such that their value domain can be defined appropriately in order to provide an orientation to generate simulation data that ensures proper coverage of test cases; 2) A proximity measure with a rational interpretation to determine the proximity between a priority vector and its approximation obtained via an alternative method (e.g. reduction of comparisons); and, 3) An approach to establish an objective limit of precision for priority vectors.

Further information can be found in Rivest (2019).

8. Key References

Brunelli, M. (2014). Introduction to the analytic hierarchy process. Springer.

Deza, M. M., & Deza, E. (2009). Encyclopedia of distances. In *Encyclopedia of distances* (pp. 1-583). Springer, Berlin, Heidelberg.

Herman, M. W., & Koczkodaj, W. W. (1996). A Monte Carlo Study of Pairwise Comparison. *Inf. Process. Lett., 57(1)*, 25-29.

Ishizaka, A., Pearman, C., & Nemery, P. (2012). AHPSort: an AHP-based method for sorting problems. *International Journal of Production Research*, 50(17), 4767-4784.

Kou, G., & Lin, C. (2014). A cosine maximization method for the priority vector derivation in AHP. *European Journal of Operational Research, 235(1)*, 225-232.

Rivest, R. (2019). Techniques de simulation pour la recherche sur le perfectionnement de la méthode AHP. Mémoire de maîtrise [translation : Simulation techniques for research on AHP refinement, Masters Thesis], *Bibliothèque H.E.C. Montréal*. http://biblos.hec.ca/biblio/memoires/m2019a612764.pdf

Schoner, B., Wedley, W. C., & Choo, E. U. (1993). A unified approach to AHP with linking pins. *European Journal of Operational Research, 64(3)*, 384-392.

Triantaphyllou, E., & Mann, S. H. (1990). An evaluation of the eigenvalue approach for determining the membership values in fuzzy sets. *Fuzzy Sets and Systems, 35(3)*, 295-301.

Williams, C., & Crawford, G. (1980). Analysis of Subjective Judgment Matrices (No. RAND/R-2572-AF). RAND CORP SANTA MONICA CA.