Journal of the Korean Institute of Industrial Engineers
Vol. 36, No. 1, pp. 1-6, March 2010.

New Criteria for the Consistency in Reasonable Pairwise Comparison Matrices

Jae-Bum Kim[1] · Yong-Gon Cho[2] · Yun-Bae Kim[1] · Keun-Tae Cho[1†]

[1]Department of Systems Management Engineering, Sungkyunkwan University, Suwon, 440-746, South Korea
[2]Korea Evaluation Institute of Industrial Technology, Seoul, 135-080, South Korea

합리적 쌍대비교행렬에서 일관성에 대한 새로운 기준

김재범[1] · 조용곤[2] · 김윤배[1] · 조근태[1]

[1]성균관대학교 시스템경영공학과 / [2]한국산업기술평가관리원

The Analytic Hierarchy Process (AHP) has been applied widely in various decision making fields. One of advantages of the AHP is the consistency test. However, it has several problems such as the limit of its concept, the limit of 9 scales and stern criteria, contradictory pairwise comparison. In this paper, we propose new criteria for the consistency with more realistic and ideal conditions. To derive the criteria, we conduct the simulation and use the bootstrap method, which is one of resampling techniques in the simulation area.

Keywords: Analytic Hierarchy Process, Consistency Ratio, Reasonable Pairwise Comparison Matrices

1. Introduction

The Analytic Hierarchy Process (AHP), which was developed by Saaty in the early 1970s, has been applied widely in a variety of decision making fields because of simplicity, easy of application and generality of theory and also many researches is gone concern in the structure of theory itself.

In the specific literature of simulation for ranking and selection, to facilitate the relative criticality weighting, the AHP is commonly used in the multicriteria decision making field. The AHP enables pairwise comparisons of importance between the alternatives of systems, which are objects of simulation, and computes the weights based on the pairwise comparison values using methods such as an eigenvalue method (Morrice and Mullarkey, 1998).

One of advantages of the AHP is to have the mechanism that can check Decision Maker (after this, DM)'s consistency. Consistency is to measure contradictory of judgments that DM makes. For example, let's suppose that there are three elements A, B and C. If A is more important than B, B is more important C and C is more important that A, these DM's judgments are inconsistency. Of course, A should be more important than C. It is known to be 'transitivity' in the decision making theory. To measure such a transitive contradictory about DM's judgments and then remove it or re-examine their judgments is the concept of consistency in the AHP (Saaty, 1980).

The latest investigators presented problems about the consistency concept. Pairwise matrices which are not reasonable logically can be formed even if they pass the consistency test, i.e., a 'Contradictory Matrix' can be created (Kwiesielewicz and Uden, 2002). More fundamental problem is that DM's preferences are restricted

† Corresponding author : Keun-Tae Cho, Department of Systems Management Engineering, Sungkyunkwan University 300 Chonchon-dong, Changan-gu, Suwon, Kyonggido, 440-746, South Korea, Fax : +82-31-290-7610, E-mail : ktcho@skku.edu
Received August 15, 2009; Revision Received August 18, 2009; Accepted January 18, 2010.

by 9 point scales: although DM performed relative evaluations correctly, if the elements of a pairwise matrix exceed the range of 9 point scales, the matrix is to discard. That is, there is the debate that a pairwise matrix which is valid logically is revaluated due to limits of 9 point scales. There were many researches (Karapetrovic and Rosenbloom, 1999; Ko and Lee, 2001; Pelaez and Lamata, 2003) for such problems and the latest study (Finan and Hurley, 1999; Lekinen, 2000) about which verbal scales are mapped to DM's preference. In addition, researches about new consistency index (Pelaez and Lamata, 2003) and verification method of the AHP validity (Karapetrovic and Rosenbloom, 1999) were presented; geometric consistency index (Aguaron *et al*., 2003a; Aguaron *et al*., 2003b), probability measure (Laininen and Hamalainen, 2003), the empirical distribution of maximal eigenvalue (Ko and Lee, 2001) and the analysis of pairwise matrix through regression analysis (Laininen and Hamalainen, 2003).

Most researches did not escape the confined concept of the existing AHP consistency and were limited as some techniques in terms of methodologies with the approximation for statistical feature of the critical value or using the modified scales measurements. Thus, in this paper, we propose new criteria, which are expected to improve problems about the concept of consistency fundamentally. The criteria are to be drawn directly by generating Reasonable Pairwise Comparison Matrix (RPCM) using Monte-Carlo simulation rather than finding consistency indexes by the existing method.

The remainder of this paper is organized as follows. In section 2, we discuss several considerations in terms of concept and scale measure for new criteria for the consistency. In section 3, we propose new criteria and test procedure for consistency, Reasonable Pairwise Comparison Matrix (RPCM). In section 4, we calculate the range of new criteria, comparing with other critical values of the existing consistency test and provide an example to verify the availability of our procedure. In section 5, we discuss the results of our research and propose likely future direction.

2. Consideration

2.1 Considerations for the concept of consistency

In the existing literature of the AHP, random results of decision making processes, i.e., generating elements of pairwise comparison matrices by random number, are drawn and then the relative criteria of consistency test are made by the random index (RI). However, logical explanations for validity are insufficient (Ko and Lee, 2001).

The traditional eigenvector method (Saaty, 1980) for estimating weights in the AHP yields a way of measuring the consistency of a decision maker's preferences arranged in the form of a reciprocal pairwise comparison matrix. The consistency index (CI) is given by

$$CI = \frac{(\lambda_{\max} - n)}{(n-1)}$$

where $\lambda_{\max}$ is the largest eigenvalue of the n×n reciprocal pairwise comparison matrix. Saaty (1980) showed that if a decision maker is perfectly consistent (i.e., $a_{ik} = a_{ij} a_{jk}$ for all i, j, k = 1,⋯, n, $\lambda_{\max} = n$) and if the decision maker is not perfectly consistent, then $\lambda_{\max} > n$. To measure this consistency, he proposed a consistency ratio (CR) defined as

$$CR = \frac{CI}{RI} \times 100$$

where RI is the average value of CI obtained from 500 positive reciprocal pairwise comparison matrices whose entries were randomly generated using the 1 to 9 scale. He considers that a value of CR under 0.10 indicates that the decision maker is sufficiently consistent. <Table 1> gives values of the average RI for different values of n.

Table 1. Values of the random index for different matrix order

N	1	2	3	4	5	6	7	8	9
R.I	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

There are problems about the limit of the concept for consistency test like the same things motivated by Karapetrovic and Rosenbloom (1999). They noted that 'Paradox of consistency test', the existing consistency test of AHP tested the following hypothesis:

H0: the decision maker's choice is random
H1: the decision maker's choice is not random.

However, to be random is not to be mistake, illogical and not reasonable. The purpose of the consistency test should be to prevent the DM from being not reasonable or from making mistakes in pairwise comparisons. In addition, another problem is brought forward, called contradictory pairwise comparisons (Kwiesielewicz and Uden, 2002). In the following pairwise matrix,

$$\begin{pmatrix} 1 & 2 & 1/2 & 2 \\ 1/2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 1/2 & 1 & 1/2 & 1 \end{pmatrix}$$

In spite that the pairwise comparison matrix is contradictory, that is, not transitivity, the consistency test is success. In the above matrix, although A > B, B = C and A < C, which is "Contradictory", CR of the matrix is 0.07 and then the matrix is evaluated to be consistent.

In this paper, first of all, we suppose more realistic the AHP procedure in views of a DM's behavior to overcome above problems. Because the AHP is not the process calculated automatically preference but the process that a DM's subjective judgment is reflected entirely, we must observe results as such DM's normal actions in each procedures for decision making. That is, what we could infer rationally in the AHP process are as follows.

(1) When a DM receives evaluation items, he or she, as a specialist for the problem, has prior knowledge for each alternative.
(2) The DM already drops direct evaluation for each alternative. The alternative to give the best among each alternative exists and the points of other alternatives, sum of which may be near to 1, are graded approximately by the DM.
(3) Because a DM may make mistakes in the evaluation and calculation, errors are likely to be inserted in each preference.
(4) Evaluation is the process fit for verbal scales.
(5) The DM is likely to try to reduce errors of (3) and do relative evaluation of each alternative by using the results of (1) and (2).

2.2 Considerations for the measure scale of evaluation : verbal scale

It happens that actual points of each relative evaluation pass over the range of 9 scales. Even if the previous comparisons are 'true', the pairwise comparisons are discarded or revaluated in the case (Finan and Hurley, 1999; Karapetrovic and Rosenbloom, 1999). Due to the limit of 9 scales and stern criteria, the serious problem is that some matrix can not even be used. For example, note the following pairwise matrix

$$\begin{pmatrix} 1 & 5 & ? \\ 1/5 & 1 & 5 \\ 1/? & 1/5 & 1 \end{pmatrix}$$

Even if the value of the question mark in the matrix is 9, the consistency test will be failed and the matrix will be discarded.

The problem arises from that the meaning of scales is confined to only quantitative numerical values. It is worth discussing that preference of each alternative may not be simply represented by the product of elements of pairwise comparisons. As other example, let's suppose the following pairwise matrix for three alternatives, A, B and C.

$$\begin{pmatrix} 1 & 3 & ? \\ 1/3 & 1 & 3 \\ 1/? & 1/3 & 1 \end{pmatrix}$$

The relationship among alternatives shows that A is little more important than B, B is little more important than C. For this matrix to be perfectly consistent, we need to replace the question mark with 9 which requires that A is very much more important than C. However, there may be other results that have validity intuitionally, indeed, 4, 5, or 7. This problem is caused by an inappropriate evaluation with the transitivity of 9 scales (Finan and Hurley, 1999). In Finan and Hurley's study (1999), when they put the questions to an executive MBA class at Queen's University, their answers were summarized in the following <Table 2>.

Table 2. The survey about the assessment in MBA class at Queen's University

Section	"more"	"much more"	"very much more"
total	84.5%	11.5%	4%

84.5% of the students were comfortable with the assessment that A is more important than C. Considering these results, points actually used in the pairwise comparisons should be verbal numerical values (Finan and Hurley, 1999; Lekinen, 2000). Verbal Scales is the measure that adjusts each 9 points measure to verbal range, when relative differences about stimulus are slight. We defined the matched relationship in <Table 3>. All measurements of pairwise comparisons will be mapped by this measure in this paper.

Table 3. The relationship between verbal scale and 9 scales

Verbal statement	9 scales	Sign
Equal	1	=
-	2	≥
Little more important	3	>
-	4	>/
More important	5	>>
-	6	>>/
Much more important	7	>>>
-	8	>>>/
Very much more important	9	>>>>

3. New Criteria for the Consistency

On the basis of more realistic suppositions about consistency, new criteria should be able to examine rational judgments by verbal scales and be required for the DM's reasonable pairwise comparisons. The hypothesis of consistency test becomes the following:

H0: the DM's choice is reasonable
H1: the DM's choice is not reasonable

We define the following concepts to represent these criteria.

Definitions: Given the positive reciprocal matrix (All pairwise matrices in the AHP are positive reciprocal) $A=[a_{ij}], A \subset R^2$, the matrix is said to be Reasonable Pairwise Comparison Matrix (RPCM) if and only if

i) $a_{ik}, a_{kj} > 1 \quad \mathrm{I}\max(a_{ik}, a_{kj}) < a_{ij} \le \min(a_{ik}a_{kj}, 9)$
ii) $a_{ik}, a_{kj} < 1$
$$\max(a_{ik}^{-1}, a_{kj}^{-1}) < a_{ij}^{-1} \le \min(a_{ik}^{-1}a_{kj}^{-1}, 9)$$
iii) $a_{ik}, a_{kj} = 1 \text{ or } = 9 \qquad \max(a_{ik}, a_{kj}) = a_{ij}$
iv) $a_{ik}, a_{kj} > 1,\ a_{kj} < 1 \text{ or } a_{ik} < 1,\ a_{kj} > 1$

In the relationship of i), ii), a_{ij} is derived.

Proof: Generally, the DM has his or her absolute evaluations for each alternative with verbal scales. So, $w_i/w_j -> a_{ij}$ when '->' indicates mapping verbal scales with 9 scales.

$$\begin{aligned} a_{ij} &= f(w_i/w_j) \quad & w_i/w_j < 9\,. \\ &= 9 & w_i/w_j \ge 9 \end{aligned}$$

We assume that unknown f( ) is an incremental function, so that if x > y, f(x) > f(y). Let $w_i/w_j = r_{ij} > 1$, $r_{ij} = r_{ik}r_{kj}$ and if and only if $r_{ik} > 1, r_{kj} > 1$ then

$$\mathrm{f}(\mathrm{r_{ij}}) > \mathrm{f}(\mathrm{r_{ik}}),\ \mathrm{f}(\mathrm{r_{ij}}) > \mathrm{f}(\mathrm{r_{kj}}).$$

Thus, we have that

$$\begin{aligned} &\mathrm{f}(\mathrm{r_{ij}}) > \max(\mathrm{f}(\mathrm{r_{ik}}),\ \mathrm{f}(\mathrm{r_{kj}})) \\ &\mathrm{a_{ij}} > \max(\mathrm{a_{ik}},\ \mathrm{a_{kj}}) \end{aligned}$$

Also, to be perfectly consistent is reasonable. Therefore

$$\mathrm{f}(\mathrm{r_{ij}}) \le \mathrm{f}(\mathrm{r_{ik}})\mathrm{f}(\mathrm{r_{kj}}) \text{ or } \mathrm{f}(\mathrm{r_{ij}}) \le 9.$$
$$\text{Hence } \max(\mathrm{f}(\mathrm{r_{ik}}),\ \mathrm{f}(\mathrm{r_{kj}})) < \mathrm{f}(\mathrm{r_{ij}}) \le \min(\mathrm{f}(\mathrm{r_{ik}})\mathrm{f}(\mathrm{r_{kj}}),\ 9).$$

Additionally, as being the reasonable pairwise matrix, when $a_{ij} > 1$, $a_{ij} > a_{ik}$ and $a_{ij} > a_{kj}$ must be satisfied. That is, the matrix which is "row dominant" so that keep ranks of each alternatives, is the most reasonable pairwise comparisons. In this research, we put some restrictions: the rank preservation must be conserved and then the properties of "row dominant" should be assured for RPCM. Under this restriction, new criteria for RPCM, like the largest eigenvalue, are calculated by numerical experiments. Once the criteria are calculated, we can evaluate whether a DM's judgments are reasonable and logical through more advisably appraisable criteria of consistency.

4. Simulation for Critical Value and Application

In order to capture the range of the consistency criteria for RPCM, we conduct the Monte Carlo simulation. We then show an example which applied the RPCM procedure on the basis of experiment results to examine availability of the new criteria. In the Monte Carlo simulation, we use the bootstrap method, which resamples the pseudo data generated randomly during several times of iterations, in order to avoid computational intensive problems. The bootstrap is a method of computational inference that simulates the creation of new data by resampling from a single data set. Although developed first for cross sectional data, in recent years several bootstrap methods for dependent data have been developed (Efron and Tibshirani, 1993). We assume that the largest eigenvalue of matrix generated randomly are independent and can be resampled with the same probability.

The steps of numerical experiments are as follows.

<Step 1> Generate RPCM. First of all, elements of a pairwise matrix, of which all diagonal elements are 1, should be generated from 9 scale points [1/9, 1/8, ⋯ 1, 2, ⋯ 9] with probability 1/17.

<Step 2> Determine the marginal elements. Calculate the unknown elements and determine the value of $[a_{ij(low)}, a_{ij(high)}]$ using the definition, i.e., $\max(a_{ik}, a_{kj}) < a_{ij} \le \min(a_{ik}a_{kj}, 9)$.

$$\begin{pmatrix} 1 & a_{12} & & & \\ & \cdots & & & \\ & & 1 & a_{ik} & \\ & & & \cdots & a_{kj} \\ & & & & \cdots \ 1 \end{pmatrix}$$

<Step 3> Calculate the largest eigenvalue, $\lambda_{\max,kl}$ where k is the number of iterations and l is the number of matrices generated.

<Step 4> Repeat from the step 1 to step 3 and get the pseudo set of the largest eigenvalues by using bootstrap resampling. Finally, we can calculate the largest eigenvalue, $\overline{\lambda}^*_{\max}$, and then set the range of eigenvalue for RPCM as follows.

$$\lambda_{RPCM} = [n,\ \overline{\lambda}^*_{\max}].$$

Once we can yield the confidence interval of the largest eigenvalue of RPCM through such a simulation experiment, these amounts become the criteria that can verify whether an arbitrary pairwise matrix is reasonable under the condition referred before.

The following is the comparisons with the largest eigenvalue of existing researches as shown in <Figure 1>. In the graph, the x-axis means the size of the matrix and the lines means consistency criteria values depending on the size of the matrix changes.

The critical value of RPCM is more flexible than anything of the existing consistency index on the whole, because it includes the pairwise comparison matrices, which are failed in the existing test but may be reasonable. Also, RPCM's has similar values as the K exponent.

The K exponent is the critical value satisfied that $\Pr(\lambda_{\max} \le k) = K$ based on the distribution of the largest eigenvalue. The K exponent is known to have good attribute comparing with other indicators (Ko and Lee, 2001).

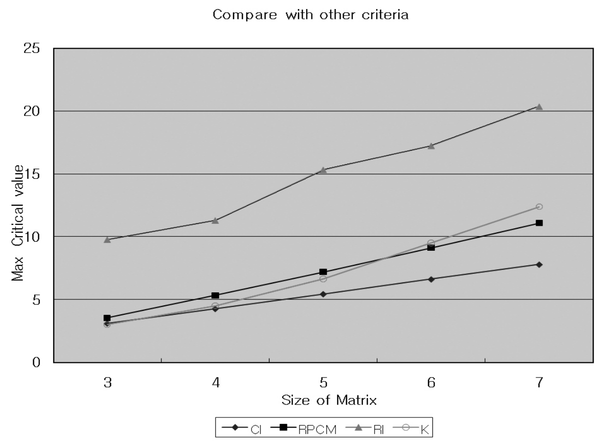

Figure 1. Comparisons of each critical points of consistency

Table 4. The criteria of RPCM obtained from simulation with 300 matrices and 10 iterations

Matrix size	RI	CI	K	RPCM
3	(3, 9.7691)	(3, 3.116)	(3, 3.0291)	(3, 3.5608)
4	(4, 11.3006)	(4, 4.27)	(4, 4.5249)	(4, 5.3333)
5	(5, 15.3237)	(5, 5.448)	(5, 6.6518)	(5, 7.2010)
6	(6, 17.2361)	(6, 6.62)	(6, 9.5084)	(6, 9.1187)
7	(7, 20.3545)	(7, 7.792)	(7, 12.3827)	(7, 11.0657)

We propose an example on how our RPCM criteria are applied in an arbitrary pairwise matrix in order to show their availability.

Example 1: Let's evaluate whether the following matrix is reasonable.

$$\begin{pmatrix} 1 & 3 & 8 \\ 1/3 & 1 & 4 \\ 1/8 & 1/4 & 1 \end{pmatrix}$$

We inspect whether the matrix satisfies conditions for RPCM, using the following relation

$$\max(3, 4) < 8 \le \min(12, 9).$$

Then, we calculate the largest eigenvalue of the matrix and compare it with the range of the critical value in <Table 4>. Thus, $\lambda_{\max}$= 3.01829. This pairwise matrix is consistent and RPCM.

Hence, this matrix is accepted, that is, this matrix is RPCM and reflected in decision-making.

Example 2: Let`s evaluate the following Matrix

$$\begin{pmatrix} 1 & 6 & 2 & 3 \\ 1/6 & 1 & 4 & 3 \\ 1/2 & 1/4 & 1 & 8 \\ 1/3 & 1/3 & 1/8 & 1 \end{pmatrix}$$

This matrix is not dominant, using the following relation, $\max(6,4) > 2$. This pairwise matrix is not consistent because $\lambda_{\max}$= 5.45822 > 5.33333. Hence, the matrix is rejected. In dominant inspection step, this matrix is already not RPCM.

5. Conclusion

We embodied the problems about the existing consistency concept and proposed the new criteria to be able to reflect decision matrix flexibly and proposed a new criteria to improve problems about the Contradictory transitivity and tried to draw the criteria by generating Reasonable Pairwise Comparison Matrix using Monte-Carlo simulation based on the bootstrap method rather than finding only the consistency problem comparing with other existed methods : CI, K-exponents.

The K exponent, which is the critical value based on the parametric estimation, is known to have good attribute comparing with other indicators. However, it can not find the pairwise comparison matrices which may be not reasonable and its estimated distribution of the largest eigenvalue may have the exact goal what we expected,

In such a proposition, the bootstrap method may be useful at deduction of key criteria value of the AHP process. In order to compare it with other criteria of consistency and to avoid computer intensive and get more efficient results, we use the bootstrap in the numerical experiment.

We expect that the RPCM procedure could lead more reasonable results in the consistency test and correct the problems, we mentioned in this paper, with the process of dual verification.

References

Aguaron, J., Escobar, M. T., and Moreno-Jimenez, J. M. (2003a), Consistency stability intervals for a judgment in AHP decision support systems, *European Journal of Operational Research*, **145**(2), 382-393.

Aguaron, J. and Moreno-Jimenez, J. M. (2003b), The geometric consistency index : Approximated thresholds, *European Journal of Operational Research*, **147**(1), 137-145.

Carmone, F. J., Ali Karab., S., and Zanakis, H. (1999), A Monte Carlo investigation of incomplete pairwise comparison matrices in AHP, *European Journal of Operational Research*, **102**(3), 538-553.

Efron, B. and Tibshirani, R. J. (1993), An introduction to the bootstrap, Chapman and Hall/CRC, New York.

Finan, J. S. and Hurley, W. J. (1999), Transitive calibration of the AHP verbal scale, *European Journal of Operational Research*, **112**(2), 367-372.

Karapetrovic, S. and Rosenbloom, E. S. (1999), A quality control approach to consistency paradoxes in AHP, *European Journal of Operational Research*, **119**(3), 704-718.

Ko, K. and Lee, K. (2001), Statistical characteristics of response consistency parameters in the AHP, *KORMS*, **26**(4), 71-82.

Kwiesielewicz, M. and Ewa van Uden (2002), Inconsistent and contradictory judgments in pairwise comparison method in the AHP, *Computers and Operations Research*, **31**(5), 713-719.

Laininen, P. and Hamalainen, R. P. (2003), Analyzing AHP-matrices by regression, *European Journal of Operational Research*, **148**(3), 514-524.

Lekinen, P. J. (2000), Measurement scales and scale Independence in the Analytic Hierarchy Process, *Multi-Criteria Decision Analysis*, **9**(4), 163-174.

Morrice, D. J. and Mullarkey, P. W. (1998), An approach to ranking and selections for multiple performance measures, *Proceedings of the 1998 Winter Simulation Conference*, 719-726.

Pelaez, J. I. and Lamata, M. T. (2003), A new measure of consistency for positive reciprocal matrices, *Computer and Mathematics with Applications*, **46**, 1839-1845.

Saaty, T. L. (1980), *The Analytic Hierarchy Process*, McGraw-Hill.