THE IMPACT ON SCALING ON THE PAIR-WISE COMPARISON OF THE ANALYTIC HIERARCHY PROCESS

Yuji Sato
Department of Policy Science, Matsusaka University
1846, Kubo, Matsusaka, Mie, 515-8511 – Japan
ysatoh@matsusaka-u.ac.jp

Keywords: linear scale, power scale, consistency index, rank reversal, discriminating sensitivity

Summary: *The objective of this study is to find the scale of the Analytic Hierarchy Process (AHP) appropriate for representing decision maker's perception. Specifically, two scales, linear scale and power scale, employed in the pair-wise comparison of the AHP are evaluated. The results offer some evidence that the power scale is preferable to the linear scale as the judgment scale.*

1. Introduction

One of the most popular methods for decision making has been the Analytic Hierarchy Process (AHP) developed by Thomas L. Saaty (Saaty, 1977). The advantage to the AHP is that data on decision-makers' (DMs) pair-wise comparisons are aggregated, and the degree of importance of each alternative is quantified. This quantification results in not only the identification of the most important alternative but also the ranking of all alternatives for each DM.

Two issues surround the use of the AHP. The first issue occurs because the allowable upper bound of consistency index (CI) is 10% of the RI. Pair-wise comparison matrices with CIs greater than this critical value are therefore not accepted due to their inconsistency. Consequently, in case applying the AHP to questionnaire on public opinion, the more samples whose CI exceed the limit, the fewer the samples available for analysis (henceforth, CI problem). The second issue surrounding the AHP involves the range of a_{ij}, the relative weight of alternative i (i for short) to j. When the range of a_{ij} is expanded (reduced) from 1-9 to 1-15 (1-5), the Frobenius root and its corresponding eigenvector changes, resulting in the alternation of the direction of the eigenvector and the rank of the priority of alternatives (Schenkerman, 1994) (henceforth, rank-reversal problem).

In addition to the above two traditional issues, one additional problem must be noted. If the range of a_{ij} were to be too reduced, then CI would converge to 0, making it impossible for the AHP to discriminate an important alternative from others. Since discriminating the most important alternative is one of the principal purposes of the AHP, verifying how sensitively the AHP discriminates an important alternative from others when a_{ij}s are scattered is essential (henceforth, discriminating-sensitivity problem).

In the AHP, a pair-wise comparison and a scale used in its procedure play a key role quantifying each DM's feeling; therefore, which scale should be used in the process of a pair-wise comparison is the most controversial issue concerning the refinement of this method. Indeed, the three aforementioned problems (the CI problem, the rank-reversal problem and the discriminating-sensitivity problem) are deeply related to the scale. The 1-9 linear scale, advocated by Saaty (Saaty, 1980) long considered the standard of the AHP, has been widely criticized in the literature, primarily because of the CI problem and the rank-reversal problem.

To overcome the deficiencies of Saaty's scale, various judgment scales for a pair-wise comparison have been proposed and evaluated to date. Aupetit and Genest proposed reducing the range of the linear scale to 1-5 (Aupetit and Genest, 1993); Harker and Vargas proposed extending the range of the linear scale to 1-13 and 1-50 (Harker and Vargas, 1987). In addition to these linear scales, Harker and Vargas also proposed two non-linear scales (quadratic and irrational) (Harker and Vargas, 1987). Lootsma claimed the superiority of the power scale to the 1-9 linear scale (Lootsma, 1989 and 1991). Saaty's and Lootsma's scales have often been compared and have generated much discussion.

Saaty verified the effectiveness of his 1-9 scale through many physical experiments, each of which had a theoretical value (Saaty, 1983). The AHP, however, deals with decision making on subjective issues, making it difficult to determine theoretically whether or not each DM's preference for alternatives derived from the AHP accurately represents each DM's feeling. Taking into account the fact that the AHP deals with not only objective issues that can be quantified but also subjective issues that do not have theoretical values, the effectiveness of a scale must be verified empirically through actual applications to subjective issues. In this study, we focus on comparing the effectiveness of Saaty's 1-9 linear scale and the power scale from the perspective of appropriateness for representing each DM's perception, where the criteria of appropriateness are consistency, robustness with respect to the change of the range of a scale, and discriminating sensitivity.

In this study, two types of data are used to evaluate the two scales. One is randomly generated data, five sizes of matrices (from 3×3 to 7×7) whose elements are generated in accordance with uniform randomness (random sample); the other is actual data, a matrix (4×4) whose elements are obtained from a 1997 survey of public opinion on the Japanese election results of October 1996 (actual sample). This election represented the first time the election law of the House of Representatives switched from a multi-member electorate system to a single-member electorate system. The survey was carried out four months following the election. The data obtained from this survey might thus be considered an accurate reflection of each respondent's subjective judgment and serve as an ideal actual sample for evaluating scales used in a pair-wise comparison.

2. Set-up for the empirical test

2.1 Definition of the linear ant the power scales for a pair-wise comparison

A scale employed in a pair-wise comparison can be defined as a function mapping the result of a pair-wise comparison between two alternatives to reciprocal value. Let ***I*** be a **linearly symmetric interval**, whose center (set as 0) represents two alternatives of equal importance with the magnitude of importance symmetrically distributed on both sides (set, + and -). Further let $\boldsymbol{x_{ij}} \in \{-k, ..., -1, 0, 1, ..., k\} \subset I$ ($x_{ij} + x_{ji} = 0$ and $i, j = 1, ..., n$) be **pair-wise comparison data** representing each result of pair-wise comparisons between an alternative i (i, for short) and j. If we describe the **relative weight of *i* to *j* as $\boldsymbol{a_{ij}}$**, then a scale can be defined as the monotone, increasing function f: $x_{ij} \rightarrow a_{ij}$, which satisfies $f(x_{ij}) \cdot f(x_{ji}) = 1$.

Definition 1: Linear Scale

$$f_L(x_{ij};c) = \begin{cases} cx_{ij}+1 & x_{ij} \geq 0 \\ 1/(-cx_{ij}+1) & x_{ij} < 0 \end{cases} \quad (\text{for } \forall c>0) \tag{1}$$

Definition 2: Power Scale

$$f_P(x_{ij};m) = m^{x_{ij}} \quad (\text{for } \forall m>1) \tag{2}$$

Henceforth, we describe the pair-wise comparison matrix generated by the linear scale, $f_L(x_{ij};c)$, as $A_L = (f_L(x_{ij};c))$ and that generated by the power scale, $f_P(x_{ij};m)$, as $A_P = (f_P(x_{ij};m))$[1]. Saaty's linear scale corresponds to the case k=8 for x_{ij} and c=1 in Equation (1); Aupetit-Genest's linear scale corresponds to that of k=12 and c=1; and Harker-Vargas's linear scale corresponds to that of k=8 and c=0.5. However, in the AHP, since Saaty's 1-9 linear scale has been employed as a standard, we fix k=8 and c=1 for purposes of this study.

[1] Lootsma employed the similar definition of the power scale; the pair-wise comparison data x_{ij} was set as $x_{ij} \in \{2k \mid k=0, ...,4\}$ (Lootsma, 1991).

In the definition of the power scale, the size of m greatly affects the results. If $m \to 1$, then a weight vector converges to $1/n \cdot \mathbf{1}$[2], resulting in no useful information concerning the degree of importance of alternatives. In contrast, if m is set large, then the values of pair-wise comparisons diverge extremely or converge to 0, resulting in a deviation from individuals' feelings. One reasonable way of setting m is to set the maximum of $f_P(x_{ij};m)$ the same as that of $f_L(x_{ij};1)$, which results in $m^8=9$ ($m \cong 1.3161$); however, under this condition, $f_P(x_{ij};1.3161) < f_L(x_{ij};1)$ when $0 < x_{ij} < 8$ because f_P is a convex function. In this case, CI of $A_P=(f_P(x_{ij};m))$ tends to be smaller than that of $A_L=(f_L(x_{ij};c))$, making the comparison about the size of CI unfair. Therefore, in this study, we fix $\boldsymbol{m(c)}$, so as to satisfy the following equation:

$$\int_0^8 (cx+1)dx = \int_0^8 m^x dx. \qquad (3)$$

From Equation (3), the approximate value of $m(1)$ is calculated as $m(1)=1.3945$. Henceforth, the constant c and $m(c)$ are represented by 1 and $m(1)$ respectively, and the functions $f_L(x_{ij};c)$ and $f_P(x_{ij};m)$ are described as f_L and f_P, for short. Insofar as the calculation of weight is concerned, we follow the eigenvalue method, because the weight calculated by the log-least-square method and that calculated by the geometric-mean method coincide (Crawford and Williams, 1985), and that the latter coincides the weight calculated by the eigenvalue method.

2.2 Samples

In this study, we use two qualitatively different types of data to evaluate the appropriateness of the two above-defined scales for representing each DM's feeling. One type of data is randomly generated, five sizes of matrices (from 3×3 to 7×7) whose elements are generated in accordance with uniform randomness (random sample); the other is actual data, a matrix (4×4) whose elements are obtained from a 1997 survey of public opinion on the Japanese election results of October 1996 (actual sample).

2.2.1 Random sample

The random sample is generated in the following way. First, we select 20000 sets (5000 sets for 3×3 matrix) of **random pair-wise comparison data** $\boldsymbol{\{x_{ij}\}_r}$ (i, j =1, ..., h and h =3, ..., 7) from {-8, ..., -1, 0, 1, ..., 8} in accordance with uniform randomness, and next map $\{x_{ij}\}_r$ by f_L and f_P, generating five types (from 3×3 to 7×7) of reciprocal pair-wise comparison matrices A_L and A_P, respectively. In this study, these two sets of five types of 20000 matrices are defined as the random sample and used in the evaluation of f_L and f_P.

2.2.2 Actual sample

The actual sample is obtained from a 1997 survey of public opinion on the Japanese election results of October 1996. The total number of responses is 796 and the response ratio is 83.1%. In this survey, the following three questions are formatted in the AHP system:

Q1: the criteria used in the choice of the candidate in the election
Q2: the reason why respondents were non-partisan
Q3: the criteria used in the choice of the candidate in the previous election in July 1993

Each question asks respondents about subjective issues that could not be measured theoretically. In these questions, respondents are asked to conduct a pair-wise comparison of each criterion in accordance with the segmental method, where I ={-8, ..., -1, 0, 1, ..., 8}.

The **actual pair-wise comparison data** $\boldsymbol{\{x_{ij}\}_d}$ (i, j =1, ..., 4, $x_{ij} \in I$) is mapped by f_L and f_P, generating a 4×4 reciprocal pair-wise comparison matrix A_L and A_P, respectively. A_L and A_P are defined as the actual sample and used in the evaluation of f_L and f_P in this study. The size of the actual sample is 1409, which is the total number of qualified answers[3] obtained from the above three questions.

[2] a vector whose elements are all 1
[3] the CIs of the answers are smaller than 0.15

2.3 Testing item

In this research, using the two types of samples defined in Section 2.2, we evaluated f_L and f_P from the following perspectives: distribution of CI, rank-reversal problem, and discriminating-sensitivity problem.

3. Comparison between the linear and the power scales using the random sample

3.1 CI

Let $\boldsymbol{r}\mathbf{CI}_{L(h)}$ and $\boldsymbol{r}\mathbf{CI}_{P(h)}$ denote the **CI obtained from the random sample**, A_L and A_P, whose matrix sizes are $h \times h$, respectively. Furthermore, let $\mathbf{M}(\boldsymbol{r}\mathbf{CI}_{L(h)})$ and $\mathbf{M}(\boldsymbol{r}\mathbf{CI}_{P(h)})$ denote the **mean of** $\boldsymbol{r}\mathbf{CI}_{L(h)}$ **and** $\boldsymbol{r}\mathbf{CI}_{P(h)}$, respectively, over the entire random sample. Table 1 summarizes $m(c)$, $\mathrm{M}(r\mathrm{CI}_{L(h)})$ and $\mathrm{M}(r\mathrm{CI}_{P(h)})$ corresponding to each c=0.5, 1, 1.5, 2, 3, 4 and 5 in accordance with Equation (3). As can be seen, each pair of $\mathrm{M}(r\mathrm{CI}_{L(h)})$ and $\mathrm{M}(r\mathrm{CI}_{P(h)})$ was nearly equal. However, $r\mathrm{CI}_{L(h)}$ and $r\mathrm{CI}_{P(h)}$ had different distributions for each random-sample size h. Therefore, f_L and f_P could not be considered equivalent scales just because $\mathrm{M}(r\mathrm{CI}_{L(h)})$ nearly equaled $\mathrm{M}(r\mathrm{CI}_{P(h)})$. Thus, we focused on the top 1%, 3%, 5%, 7%, 10%,15%, 20% and 25% values of $r\mathrm{CI}_{L(h)}$ and $r\mathrm{CI}_{P(h)}$, and defined the **threshold point** (**T-point**, for short) as the border value of each $r\mathrm{CI}_{L(h)}$ and $r\mathrm{CI}_{P(h)}$. As Table 2 shows, $r\mathrm{CI}_{L(h)} > r\mathrm{CI}_{P(h)}$ held at almost all the T-points for each sample size h, implying that the CI of a pair-wise comparison matrix generated by f_L was greater than that generated by f_P for each T-point. In this study, f_L and f_P were evaluated based on this T-point.

M(r CI$_{L(h)}$) and M(r CI$_{P(h)}$) **Table 1**

c	0.5	1	1.5	2	3	4	5
$m(c)$	1.2687	1.3945	1.4792	1.5438	1.6406	1.7132	1.7717
$\mathrm{M}(r\mathrm{CI}_{L(3)})$	0.2378	0.5270	0.8106	1.088	1.627	2.153	2.668
$\mathrm{M}(r\mathrm{CI}_{P(3)})$	0.2511	0.5382	0.804	1.052	1.513	1.938	2.339
$\mathrm{M}(r\mathrm{CI}_{L(4)})$	0.3925	0.8996	1.400	1.916	2.945	3.968	4.984
$\mathrm{M}(r\mathrm{CI}_{P(4)})$	0.4028	0.9013	1.400	1.880	2.788	3.652	4.484
$\mathrm{M}(r\mathrm{CI}_{L(5)})$	0.4711	1.111	1.776	2.451	3.825	5.193	6.563
$\mathrm{M}(r\mathrm{CI}_{P(5)})$	0.5051	1.168	1.832	2.487	3.767	5.008	6.219
$\mathrm{M}(r\mathrm{CI}_{L(6)})$	0.5301	1.257	2.019	2.793	4.358	5.933	7.512
$\mathrm{M}(r\mathrm{CI}_{P(6)})$	0.5719	1.340	2.125	2.909	4.461	5.991	7.501
$\mathrm{M}(r\mathrm{CI}_{L(7)})$	0.6160	1.457	2.336	3.229	5.033	6.847	8.665
$\mathrm{M}(r\mathrm{CI}_{P(7)})$	0.6762	1.620	2.618	3.640	5.725	7.843	9.983

r CI$_{L(h)}$ and r CI$_{P(h)}$ corresponding to each T-point (c =1, m =m (1)) **Table 2**

Matrix Size	3×3		4×4		5×5		6×6		7×7	
T-point	$r\mathrm{CI}_{L(3)}$	$r\mathrm{CI}_{P(3)}$	$r\mathrm{CI}_{L(4)}$	$r\mathrm{CI}_{P(4)}$	$r\mathrm{CI}_{L(5)}$	$r\mathrm{CI}_{P(5)}$	$r\mathrm{CI}_{L(6)}$	$r\mathrm{CI}_{P(6)}$	$r\mathrm{CI}_{L(7)}$	$r\mathrm{CI}_{P(7)}$
1%	0	0	0.04741	0.03277	0.1752	0.1482	0.3486	0.3168	0.4834	0.4366
3%	0.000991	0	0.08662	0.06862	0.2446	0.2319	0.4819	0.4507	0.6574	0.5839
5%	0.003511	0.006150	0.1141	0.09807	0.3002	0.2936	0.5628	0.5169	0.7536	0.6709
7%	0.006296	0.006150	0.1448	0.1251	0.3472	0.3412	0.6293	0.5800	0.8285	0.7534
10%	0.01230	0.006150	0.1781	0.1648	0.4129	0.4077	0.7097	0.6494	0.9169	0.8377
15%	0.02681	0.02468	0.2318	0.2234	0.5178	0.4990	0.8123	0.7518	1.019	0.9601
20%	0.04701	0.02468	0.2865	0.2899	0.6186	0.5796	0.8991	0.8350	1.099	1.062
25%	0.06781	0.05581	0.3421	0.3458	0.7136	0.6610	0.9773	0.9088	1.171	1.159

3.2 Rank-reversal problem

With respect to the change of c and $m(c)$ defining the function f, the range of a_{ij} changes, resulting in the perturbation of the Frobenius root, and the concomitant alternation of the direction of the corresponding eigenvector. This alternation, in turn, reverses the preference order of alternatives, resulting in the rank-reversal problem. If a rank reversal occurs between primal alternatives, derived preference order cannot be trustworthy.

In the analysis of the rank-reversal problem, we randomly extracted 1000 samples from each set of 20000 random samples A_L and A_P from 3×3 to 7×7 defined in Section 2.2.1, and evaluated the robustness of f_L

and f_P with respect to the change of c and $m(c)$ as the following range:

$$c\text{: } 0 \le c \le 10, \quad m(c)\text{: } 1 \le m(c) \le 3. \tag{4}$$

The ranges of c and $m(c)$ in Equation (4), corresponding to the wider range of a_{ij} mapped by f_L and f_P than that mapped by the previously proposed scales f are wide enough for evaluating f_L and f_P. In the comparison of f_L and f_P, all samples are categorized by each 1%, 3%, 5%, 7%, 10%, 15%, 20% and 25%-T-point defined in Section 3.1, and in each category, f_L and f_P are evaluated as to whether or not rank reversal occurs. Among the above-mentioned 1000 samples, **A** and **A'** denote the **cumulative number of samples generated by f_L and f_P to each T-point**, respectively. Furthermore, among the A and A' samples in each category, **B** and **B'** respectively denote the **cumulative number of samples where the rank reversal occurs**[4] **with respect to the change of c and $m(c)$** over the range shown in Equation (4). Table 3 shows (A, B), (A', B') for h=4. As for the total number of rank-reversal samples, B ≥ B' for h=3, 4 and B ≤ B' for h=5, ..., 7; however, for samples with small CIs, B ≥ B' held in every matrix size h=3, ..., 7, implying that f_P was more robust than f_L for the change of c and $m(c)$. Figures 1.1 and 1.2 show rank-reversal examples whose CIs are small and the rank reversal occurs in the neighborhood of c=1.

Number of samples and their ratios: Rnak reversed **Table 3**

Matrix Size	3×3				4×4				5×5				6×6				7×7			
Scale	f_L		f_P		f_L		f_P		f_L		f_P		f_L		f_P		f_L		f_P	
T-point	A	B	A'	B'	A	B	A'	B'	A	B	A'	B'	A	B	A'	B'	A	B	A'	B'
1%	21	0	47	0	13	0	16	0	13	0	8	0	10	2	7	0	6	0	2	0
3%	41	0	47	0	36	4	37	0	23	0	30	0	23	3	22	0	20	4	13	0
5%	59	0	144	0	50	8	63	0	51	0	55	0	38	4	36	1	41	7	27	0
7%	75	0	144	0	84	15	79	0	72	5	77	1	58	6	60	2	60	11	58	0
10%	106	0	144	0	113	20	116	0	115	16	118	5	89	6	86	5	103	23	99	0
15%	169	0	231	0	169	26	154	0	180	22	173	8	125	12	123	11	159	30	156	13
20%	217	0	231	0	208	34	210	3	230	35	230	16	166	20	177	15	225	40	223	35
25%	275	7	336	0	260	42	265	3	272	44	278	35	228	33	220	16	288	46	308	45
100%	1000	36	1000	2	1000	291	1000	285	1000	127	1000	196	1000	161	1000	274	1000	201	1000	368

Example 1

$$A_L = \begin{pmatrix} 1 & 4 & 2 & 1 \\ 1/4 & 1 & 1/8 & 1/9 \\ 1/2 & 8 & 1 & 1 \\ 1 & 9 & 1 & 1 \end{pmatrix}, r\text{CI}_{L(4)} = 0.0659,\ c = 1 \qquad A_P = \begin{pmatrix} 1 & 2.71 & 1.39 & 1 \\ .369 & 1 & .0975 & .0699 \\ .717 & 10.3 & 1 & 1 \\ 1 & 14.3 & 1 & 1 \end{pmatrix}, r\text{CI}_{P(4)} = 0.126,\ m = m(1).$$

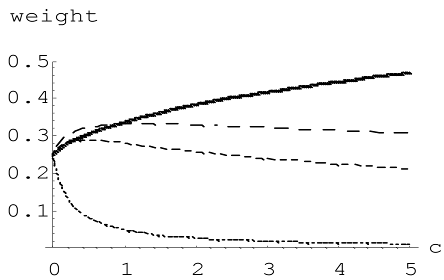

Figure 1.1: Change of the degree of importance derived from A_L

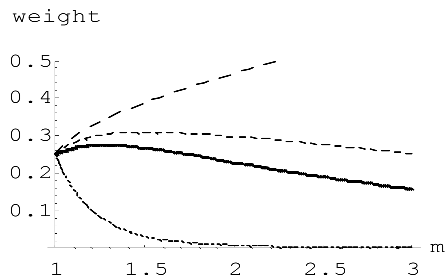

Figure 1.2: Change of the degree of importance derived from A_P

3.3 Discriminating sensitivity

One of the principal purposes of the AHP is to quantify the degree of importance for each alternative and

[4] As for the matrix whose size h=5 to 7, we counted the sample whose rank reversed between the highest and the second highest weighted alternatives.

to discriminate an important alternative from others. Let $\boldsymbol{v}^1$ and $\boldsymbol{v}^2$ respectively denote the **largest and second largest element of the eigenvector[5] corresponding to the Frobenius root**, and define $\Delta \boldsymbol{v} \equiv v^1 - v^2$ as the index of the discriminating sensitivity of scales f_L and f_P. Furthermore, let $\Delta \boldsymbol{v_L}$ and $\Delta \boldsymbol{v_P}$ denote the **$\Delta \boldsymbol{v}$ calculated by f_L and f_P**, respectively. Table 4 shows the result of comparisons between Δv_L and Δv_P for h=4. At 1% significant difference level in X^2-test, the number of samples holding $\Delta v_L < \Delta v_P$ was larger than that of $\Delta v_L \geq \Delta v_P$ for every matrix size, implying that f_L was inferior to f_P in discriminating sensitivity.

Comparison of discriminating sensitivity (4× 4) **Table 4**

	$f_L(x_{ij};1)$						$f_P(x_{ij};m(1))$					
	$v_L > v_P$		$v_L = v_P$		$v_L < v_P$		$v_L > v_P$		$v_L = v_P$		$v_L < v_P$	
T-point	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio
1%	99	44.4	2	0.900	122	54.7	83	31.3	1	0.380	181	68.3
3%	264	42.6	2	0.320	354	57.1	207	34.4	1	0.170	393	65.4
5%	406	40.3	2	0.200	599	59.5	369	36.0	2	0.200	653	63.8
7%	575	40.9	3	0.210	829	58.9	521	37.2	2	0.140	877	62.6
10%	813	40.6	3	0.150	1188	59.3	797	39.8	2	0.100	1206	60.2
15%	1237	41.2	3	0.100	1765	58.7	1207	40.2	3	0.100	1792	59.7
20%	1753	43.8	3	0.0700	2249	56.2	1656	41.1	3	0.0700	2367	58.8
25%	2216	44.3	4	0.0800	2787	55.7	2155	43.1	4	0.0800	2847	56.9
100%	7842	39.2	5	0.0300	12153	60.8	7842	39.2	5	0.0300	12153	60.8

4. Comparison between the linear and the power scales using the actual sample

4.1 CI

In judging the consistency of a pair-wise comparison matrix A, Saaty insists that the allowable upper bound of CI is 10% of RI. In contrast, E.F. Lane and W.A.Verdini advocate that the allowable upper bound of CI is 1%, 5% and 10% of RI for the case n=3, 4 and more than 5, respectively, because the size of CI depends on matrix size n (Lane and Verdini, 1989). However, these threshold values employing the certain ratio of RI as the allowable upper bound of CI have been criticized because their rationale is weak. Furthermore, since $r\mathrm{CI}_{L(h)}$ and $r\mathrm{CI}_{P(h)}$ have different distributions, comparing f_L and f_P with respect to the number of samples whose CI are smaller than the above-mentioned threshold value is not essential. Therefore, we evaluated f_L and f_P based on the cumulative number of samples with respect to each T-point of $r\mathrm{CI}_{L(4)}$ and $r\mathrm{CI}_{P(4)}$ as defined in Section 3.1.

Let $\boldsymbol{a}\mathbf{CI}_{\boldsymbol{L}}$ and $\boldsymbol{a}\mathbf{CI}_{\boldsymbol{P}}$ denote the **CI obtained from the actual sample**, A_L and A_P, respectively. Table 5 summarizes each $r\mathrm{CI}_{L(4)}$ and $r\mathrm{CI}_{P(4)}$ corresponding to the T-point defined in Section 3.1, and the cumulative number of samples whose $a\mathrm{CI}_L$ and $a\mathrm{CI}_P$ are equal to or less than each $r\mathrm{CI}_{L(4)}$ and $r\mathrm{CI}_{P(4)}$. As can be seen in the table, $r\mathrm{CI}_{L(4)} > r\mathrm{CI}_{P(4)}$ from the 1% to 15%-T-point; however, the cumulative number of samples generated by f_L is smaller than that generated by f_P, at 1% significant difference level in X^2-test. This result showed that the CI of A_P tended to cluster nearer to 0 than that of A_L when the pair-wise comparison matrix A_L and A_P were generated by f_L and f_P from common $\{x_{ij}\}_d$, respectively. Implications that arise from these results are: (i) f_P quantified individuals' feelings more consistently than did f_L, and (ii) the number of samples with a large CI not being worthy of analysis may be reduced in applying the AHP to social research.

4.2 Rank-reversal problem

Table 6 shows (A, B), (A', B'), where A and A' are the cumulative number of samples generated by f_L and f_P respectively to each T-point among all (1409) actual samples, and among A and A' samples in each category, B and B' are the cumulative number of samples where the rank reversal occurred with respect to the change of c and $m(c)$ over the range shown in Equation (4). As can be seen, the number of rank-reversal samples was 57 among all A_L, and 43 among all A_P; this difference seemed to be small.

[5] standardized 1 by l_1-norm

Distribution of $a\,\mathrm{CI}_L$ and $a\,\mathrm{CI}_P$ **Table 5**

	$f_L(x_{ij};1)$			$f_P(x_{ij};m(1))$		
T-point	$r\,\mathrm{CI}_{L\,(4)}$	Obs.	Ratio	$r\,\mathrm{CI}_{P\,(4)}$	Obs.	Ratio
1%	0.04741	302	21.4%	0.03277	379	26.9%
3%	0.08662	430	30.5%	0.06862	549	39.0%
5%	0.1141	599	42.5%	0.09807	717	50.9%
7%	0.1448	657	46.6%	0.1251	731	51.9%
10%	0.1781	787	55.9%	0.1648	932	66.2%
15%	0.2318	949	67.4%	0.2234	1016	72.1%
20%	0.2865	1029	73.0%	0.2899	1106	78.5%
25%	0.3421	1102	78.2%	0.3458	1195	84.8%
100%	--	1409	100.0%	--	1409	100.0%

Number of samples and their ratios: Rnak reversed **Table 6**

	$f_L(x_{ij};1)$			$f_P(x_{ij};m(1))$		
T-point	A	B	B/A	A'	B'	B'/A'
1%	302	0	0%	379	0	0%
3%	430	1	0.233%	549	0	0%
5%	599	2	0.334%	717	0	0%
7%	657	3	0.457%	731	0	0%
10%	787	4	0.508%	932	0	0%
15%	949	4	0.422%	1016	6	0.591%
20%	1029	5	0.486%	1106	16	1.45%
25%	1102	9	0.817%	1195	18	1.51%
100%	1409	57	4.05%	1409	43	3.05%

However, insofar as the samples whose $a\mathrm{CI}_L$ and $a\mathrm{CI}_P$ were smaller than the 5%-T-point, which is nearly equal to 0.1, no rank reversal occurred among A_P, while two occurred among A_L, as shown in Example 2. Furthermore, among A_L, rank reversal occurred even though such samples' CIs were small enough to be included in the group of "consistent" sample; in contrast, among A_P, rank reversal was never seen in the consistent group.

Figures 2.1, 2.2 shows rank-reversal examples whose CI is small and whose reversals occurred in the neighborhood of c=1. In these cases, f_L is approximately equivalent to Saaty's 1-9 linear scale. Furthermore, in Example 2, the rank reversed between primal alternatives, rendering preference order derived by f_L untrustworthy.

Example 2

$$A_L=\begin{pmatrix} 1 & 3 & 1/3 & 1 \\ 1/3 & 1 & 1/3 & 1/7 \\ 3 & 3 & 1 & 1 \\ 1 & 7 & 1 & 1 \end{pmatrix},\ a\mathrm{CI}_L=0.0819,\ c=1.$$

$$A_P=\begin{pmatrix} 1 & 1.94 & .514 & 1 \\ .514 & 1 & .514 & .136 \\ 1.94 & 1.94 & 1 & 1 \\ 1 & 7.35 & 1 & 1 \end{pmatrix},\ a\mathrm{CI}_P=0.0949,\ m=m(1).$$

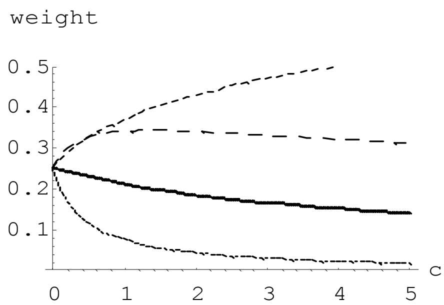

Figure 2.1: Change of the degree of importance derived from A_L

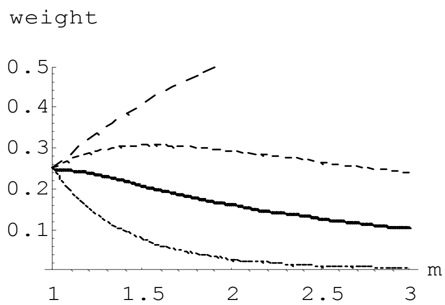

Figure 2.2: Change of the degree of importance derived from A_P

In addition to the robustness of f_P in the consistent group, the abruptness in the increase in the number of rank-reversal samples in the group with CIs larger than the 15%-T-point coincided with the number of rank-reversal samples. Leaving aside the discussion about the upper bound of acceptable CI, this coincidence might serve a good criterion for establishing the upper bound.

4.3 Discriminating sensitivity

Table 7 shows the results of comparisons between Δv_L and Δv_P for the 1409 actual samples in the same way as the analysis for the random sample in Section 3.3. As can be seen, both the 143 cases out of 182 samples of f_L and the 143 cases out of 194 samples of f_P are **1** matrix at the intersection of the 1%-T-point row and the $\Delta v_L = \Delta v_P$ column. Additionally, the 235 cases out of the 247 samples of both f_L and f_P are $v^1 = v^2$, resulting in $\Delta v_L = \Delta v_P = 0$ at the intersection of the 100%-T-point row and the $\Delta v_L = \Delta v_P$ column. These results, contrary to those of the random sample shown in Table 4, imply that the actual sample obtained from the survey on public opinion includes responses whose pair-wise comparison data $\{x_{ij}\}_d$ are all 0 or responses with more than two "tie" alternatives weighted most. On the contrary, in the comparison of f_L and f_P, the remaining samples, except for the responses whose pair-wise comparison data $\{x_{ij}\}_d$ are all 0, show that the number of samples satisfying $\Delta v_L < \Delta v_P$ is larger than that satisfying $\Delta v_L \geq \Delta v_P$, at 1% significance level in X^2-test. Thus, f_L is inferior to f_P in discriminating sensitivity.

Comparison of discriminating sensitivity **Table 7**

	$f_L(x_{ij};1)$						$f_P(x_{ij};m(1))$					
	$v_L > v_P$		$v_L = v_P$		$v_L < v_P$		$v_L > v_P$		$v_L = v_P$		$v_L < v_P$	
T-point	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio	Obs.	Ratio
1%	46	15.2	182	60.3	74	24.5	47	12.4	194	51.2	138	36.4
3%	64	14.9	194	45.1	172	40.0	77	14.0	197	35.9	275	50.1
5%	87	14.5	211	35.2	301	50.3	99	13.8	212	29.57	406	56.6
7%	105	16.0	212	32.3	340	51.8	100	13.7	214	29.27	417	57.1
10%	146	18.6	218	27.70	423	53.8	160	17.2	221	23.71	551	59.1
15%	205	21.6	225	23.71	519	54.7	189	18.6	221	21.75	606	59.7
20%	231	22.5	226	21.96	572	55.6	244	22.1	229	20.71	633	57.2
25%	285	25.9	226	20.51	591	53.6	314	26.3	229	19.16	652	54.6
100%	469	33.3	247	17.530	693	49.2	469	33.3	247	17.530	693	49.2

5. Concluding remarks

The AHP is a support system for decision making with subjective data as input and quantified data as output. To date, various judgment scales for a pair-wise comparison have been proposed and evaluated from the perspective of appropriateness for representing each DM's perception. In particular, Saaty's linear scale and Lootsma's power scale have often been compared and have generated much discussion. However, the weight obtained from a pair-wise comparison matrix cannot be verified theoretically, making evaluation of these two scales quite difficult. What is worse, this evaluation has not been able to be easily conducted because the CI of each scale, whose size is one of the criteria of evaluation, cannot simply be compared because of the difference of the properties of these two scales.

This study evaluated Saaty's linear scale and the power scale using two types of data, non-biased random sample and biased actual samples, from the point of view of appropriateness for representing each DM's feeling, where the criteria of appropriateness are the size of CI, robustness with respect to the change of the range of each scale, and discriminating sensitivity of each scale. In order to make comparisons between the two scales fair, the T-point was defined based on the relative size of the CI among all random samples rather than on its absolute value. The results of this study provide some evidence that, as a judgment scale, the power scale is preferable to Saaty's linear scale.

Nevertheless, several issues remain. (*i*) How do we determine the base of a power scale?—Which base m is the most appropriate for representing individuals' perceptions? (*ii*) How should we generate random samples?—On which interval I should they be based, and in accordance with which distribution should they follow? (*iii*) How would actual samples in a matrix other than 4×4 behave?

The first issue is which base to use for the power scale. In this study, we set $m=m(1)$ ($\cong$ 1.3945); therefore,

$f_L(x_{ij};1)>f_P(x_{ij};m(1))$ holds when $0 \leq x_{ij} \leq 6$ (corresponding to $1 \leq a_{ij}<7$ for Saaty's scale). This setting of m is likely to make the distribution of the CI of A_P clustered nearer to 0 than that of A_L. Additionally, the discriminating sensitivity of the power scale is superior to Saaty's linear scale. Furthermore, concerning the robustness with respect to the change of the range of each scale, the power scale was also superior to Saaty's linear scale. Thus, we might conclude that the power scale functioned better than Saaty's linear scale as the representation of each DM's perception. Whereas $m(1)$ seems to be a good option for defining the power scale in comparison with Saaty's linear scale, we cannot conclude that $m(1)$ is the best base for representing each DM's perception in the process of a pair-wise comparison. Determining the best base requires further analysis; other criteria besides CI, rank reversal and discriminating sensitivity must be investigated.

The second issue is how to generate the random sample. In this study, the random sample was generated by selecting numbers from the set {-8, ..., -1, 0, 1, ..., 8} in accordance with uniform randomness. In addition, we compared two scales through either (*i*) a random sample generated on the basis of a set {-5, -3, -1, 0, 1, 3, 5} in accordance with uniform randomness, or (*ii*) a random sample generated on the basis of a set {-8, ..., -1, 0, 1, ..., 8} whose distribution follows N(0, 4.2231[6]). Leaving out the detail of the results, we note that, in either case, results similar to those of the analyses in Section 3 were obtained. However, both the random sample defined in Section 2.2.1 and the two aforementioned random samples satisfy weak single-peakedness and symmetry, possibly affecting the results. On the other hand, samples obtained from the application of the AHP to a realistic decision-making process do not always satisfy these two properties. Thus, further analysis using a random sample following a different distribution is necessary.

The third issue would be the size of actual sample. In this study, the matrix size of the actual sample was only 4×4; therefore, further investigations from at least 3×3 to 7×7 matrices are needed.

[6] standard deviation of the $\{x_{ij}\}_d$

References

Aupetit, B. and Genest, C. (1993) "On Some Useful Properties of the Perron Eigenvalue of a Positive Reciprocal Matrix in the Context of the Analytic Hierarchy Process," *European Journal of Operational Research*, 70, 263-268.

Crawford, G. and Williams, C. (1985) "A Note on the Analysis of Subjective Judgment Matrices," *Journal of Mathematical Psychology*, 29, 387-405.

Harker, P. T. and Vargas, L. G. (1987) "The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process," *Management Science*, 33, 1383-1403.

Van Den Honert, R. C. and Lootsma, F. A. (1997) "Group Preference Aggregation in the Multiplicative AHP, The Model of the Group Decision Process and Pareto Optimality," *European Journal of Operational Research*, 96, 363-370.

Kumar, N. V. and Ganesh, L. S. (1996) "A Simulation-based Evaluation of the Approximate and the Exact Eigenvector Methods Employed in AHP," *European Journal of Operational Research*, 95, 656-662.

Lane, E. F. and Verdini, W. A. (1989) "A Consistency Test for AHP Decision Makers," *Decision Science*, 20, 575-590.

Lootsma, F. A. (1989) "Conflict Resolution via Pairwise Comparison of Concessions," *European Journal of Operational Research*, 40, 109-116.

Lootsma, F. A. (1991) "Scale Sensitivity and Rank Preservation in a Multiplicative Variant of the Analytic Hierarchy Process," Delft University of Technology, Report of the Faculty of Technical Mathematics, 91-20.

Saaty, T. L. (1977) "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, 15, 234-281.

Saaty, T. L. and Bennett, J. P. (1977) "A Theory of Analytical Hierarchies Applied to Political Candidacy," *Behavioral Science*, 22, 237-245.

Saaty, T. L. (1983) "Priority Setting in Complex Problems," *IEEE Transactions on Engineering Management*, EM-30, 140-155.

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

Schenkerman, S. (1994) "Avoiding Rank Reversal in AHP Decision-support Models," *European Journal of Operational Research*, 74, 407-419.

Vargas, L. G. (1994) "Reply to Schenkerman's Avoiding Rank Reversal in AHP Decision Support Models," *European Journal of Operational Research*, 74, 420-425.