

階層化意思決定法におけるスケール選択の影響に関する考察

佐藤祐司
松阪大学

(受理 1997 年 12 月 26 日 ; 再受理 1998 年 7 月 30 日)

和文概要 階層化意思決定法 (AHP) は, Thomas L. Saaty によって考案された意思決定法の一つで, 一対比較を通して選択肢に対する人間の主観的な価値判断を, それぞれの選択肢に対する重要度として定量化する点に特徴がある. 一対比較に用いるスケールは, 重要度や整合度 (C.I.) と密接な関係があり, 人間の感覚を如何にうまく捉えることができるかという観点から, これまでにさまざまなスケールが提案されているが, その有効性に関して理論的な評価を下すのは極めて困難である. そこで本論文では, 一対比較に用いるスケールとして線型スケールと指数スケールを取り上げ, ランダムに構成したサンプルと, 意思決定主体のバイアスがかかったサンプルの, 2 種類のサンプルを用いて, C.I. の値, 重要度の序列変動について両スケールを比較した. また, 最も重要性が高い項目を他の項目から判別することも AHP の重要な目的の一つであることから, この判別性能に関する比較も併せて行なった. その結果, 一対比較に用いるスケールとしては指数スケールの方が優れていることが検証された.

1. はじめに

AHP (Analytic Hierarchy Process) は, T. L. Saaty [10] によって考案された意思決定法の一つで, いくつかの評価項目 (n 種類) に対する人間の経験や勘に基づく主観的な価値判断を, それぞれの評価項目の重要度として定量化する手法であり, 経済問題, 政策決定, 都市計画問題等, さまざまな分野で広く適用されている [16, 19, 20]. AHP は問題の階層化, 一対比較, 重要度算出, という手順で進められ, 評価項目を対ごとに相対評価することによって両者の重み付けを行なう, いわゆる一対比較法を用いる点に大きな特徴がある. 評価項目 i (以下, 単に i) の j に対する相対的な重みを a_{ij} ($i, j = 1, \dots, n$) とすれば, $a_{ji} = 1/a_{ij}$ と置くことによって一対比較全体の結果は, $n \times n$ の reciprocal な一対比較行列 $A = (a_{ij})$ としてまとめられる (ここで対角項は全て $a_{ii} = 1$ である). この行列 A が完全に推移律を満たす (任意の i, j, k に対して $a_{ik} \cdot a_{kj} = a_{ij}$ が成立する) 場合には, 行列 A のランクは 1 で, その固有値は Frobenius 根のみが正 (各成分 a_{ij} の如何に拘わらず Frobenius 根の値はちょうど n), 他は全て 0 となる. このとき Frobenius 根に対応する非負固有ベクトルが, その意思決定主体の評価項目に対する重みを与えると自然に解釈できる. ただし, 一般に人間は, 対ごとの比較において必ずしも整合的な判断をするとは限らないことから, 行列 A が完全に推移律を満たす保証はない. この場合には行列 A の Frobenius 根は必ず n より大きくなり, n との差は Frobenius 根以外の全ての固有値の総和に等しいことが知られている. したがってこの差が小さいならば, その主体の各評価項目に対する価値判断はほぼ整合的であると期待できる. 実際 AHP においては, この差を $n - 1$ で割った値を整合度 (Consistency Index, 以下, C.I.) と定義し, C.I. がある一定値以下であるか否かを行列 A の有効性の判

定基準として用いる。

重要度や C.I. の値は一対比較に用いるスケールと密接な関係があり、人間の感覚を如何にうまく捉えることができるかという観点から、これまでにさまざまなスケールが考案され (例えば [1, 3, 6, 13]), またその有効性を巡って、多くの研究 [4, 8, 11, 12, 17] がなされてきている。C.I. に関しては、行列 A の有効性を判定する基準として、ランダム整合度 (Random Index. 以下, R.I.) の 10% を C.I. の許容上限とすることを Saaty は提唱しているが、その根拠は明確ではないとの指摘がある [9]¹。また、重要度に関しては、 a_{ij} のとり得る値の範囲を拡大 (1/15 から 15, 等), あるいは縮小 (1/5 から 5, 等) することによって、Frobenius 根の値のみならず、対応する固有ベクトルの向きが変動し、得られた評価項目に対する重要度の序列を覆してしまうことがある、という問題も指摘されている (例えば [15])。さらに、 a_{ij} のとり得る値の範囲を極端に縮小すると、C.I. の値は 0 に収束するが、重要度の判別もできなくなる。最も重要性が高い項目を他の項目から判別することも AHP の目的の一つとして挙げられるので、 a_{ij} の値がある程度分布した場合の、この判別性能も重要な問題の一つである。

そこで本論文では、一対比較に用いるスケールとして線型スケールと指数スケールを取り上げ、まず、ランダムに構成した一対比較行列を用いて、C.I. の値、重要度の序列変動、評価項目の判別性能の、3つの観点から比較・検証を行なった (4章)。

その結果、それぞれのスケールで測った R.I. の値はほぼ一致するにも拘わらず、重要度の序列変動に関して、指数スケールの方がより robust であることが示された。さらに、評価項目に対する判別性能も指数スケールの方が高く²、ランダムな一対比較行列を用いた分析の結果を見る限りでは、指数スケールの方が有効性が高い (4.1 節-4.3 節)。

しかし AHP の適用対象には、人間の主観を問う社会的、日常的な問題もあり、このような事例においては、人の感覚や意識などの、いわゆる“真の値”が確認できない。したがって、AHP における、人間の感覚を結果にうまく反映させるためのさまざまな工夫は、方法論としてその有効性を理論的に証明することが極めて困難であり、実際の適用事例をもとにした検証が不可欠である。Saaty 自身による物理実験 [14] 等の実証例もあるが、AHP の適用対象に鑑みると、検証に用いるサンプルとして、人工的に構成したものや物理的な現象から得られたもの以外に、例えば社会問題に対する意識調査等の、定量的な評価が困難な事象から得られたサンプルを用いる必要があると考えた。

そこで、次に本論文では、AHP を適用して行なった、一般有権者約 960 名を対象とする政治意識に関する面接調査 [18] から得られたサンプルを用いて、線型スケールと指数スケールの有効性の比較・検証を試みた (5章)。この調査は、選挙制度が従来の中選挙区制から小選挙区比例代表並立制へと改正され、新しい選挙制度のもとで行なわれた 1996 年 10 月の衆議院議員総選挙の数ヶ月後に実施されたもので、有権者の政治に対する関心が高まっていた時期の政治意識に関する調査であり、社会的なテーマに関する、意思決定主体の主観が強く反映されたサンプルとして、本検証にふさわしいものと判断した。

その結果、線型スケールを用いた場合と比較して、指数スケールを用いた場合には、C.I. の値がより 0 に近い値に偏って分布する²ことがこの検証において示された (5.1 節)。

¹文献 [9] では、一部の評価項目に対する選好順序が循環を起こしている非整合な一対比較行列でも C.I. の値が小さい場合があり、C.I. の大小だけからでは一対比較の有効性は判定できない、とも指摘されている。この点は C.I. の信頼性に関する問題の一つの側面ではあるが、本論文ではこの議論には立ち入らずに、C.I. の値によって行列 A の有効性を判定する。

² χ^2 検定を行なったところ 1% 水準で有意差があることが検証された。

また、重要度の序列変動に関しても、C.I. の値が小さなサンプルに限定すると、線型スケールを拡大・縮小した場合には重要度の序列が変動するケースがある一方で、指数スケールの底を変化させた場合にはこのような変動が全く起きず、ランダムな一対比較行列で確認された指数スケールの robust 性が、ここでも追認された (5.2 節).

さらに、最も重要性が高い項目を他の項目から判別することも AHP の目的の一つとして挙げられるので、最も重要と評価された項目と 2 番目に重要と評価された項目の重要度の差を、一対比較に用いるスケールの一種の判別性能として測ったところ、線型スケールより指数スケールの方が判別性能が高い²ことも検証できた (5.3 節).

以上のことから、一対比較において人間の感覚を定量評価する際には、線型スケールと比較して、指数スケールの方がより有効であることが実証された。

2. 一対比較に用いるスケール

一対比較の進め方としては、重要視する程度を表す言葉、数字を用いる方法や数直線を用いる方法等、いくつかの方法が既に確立されている。その中でも、中央を無差別 (0) とし、左右対称 (-, +) に重要視する程度が振り分けられている、数直線上の原点对称な区間 \mathcal{I} を用いる方法は、視覚に訴える効果があり、直観的に回答できることから、社会調査等に AHP を適用する場合に有効な手段である。一対比較に用いるスケールは、評価項目に対する対比較の結果を reciprocal な一対比較値に変換する関数と捉えることができるが、上記の区間 \mathcal{I} を用いる場合、項目 i と j の対比較データを $x_{ij} \in \{-k, \dots, -1, 0, 1, \dots, k\} \subset \mathcal{I} (x_{ij} + x_{ji} = 0)$ とすると、 $f(x_{ij}) \cdot f(x_{ji}) = 1$ を満たす単調増加関数 $f: x_{ij} \rightarrow a_{ij}$ として次のように定義できる。

定義 1: 線型スケール

$$f_L(x_{ij}; c) = \begin{cases} c \cdot x_{ij} + 1 & x_{ij} \geq 0 \\ 1/(-c \cdot x_{ij} + 1) & x_{ij} < 0 \end{cases} \quad (c: \text{任意の正の定数}). \quad (1)$$

定義 2: 指数スケール³

$$f_P(x_{ij}; m) = m^{x_{ij}} \quad (m: 1 \text{ 以上の任意の定数}). \quad (2)$$

以下、線型スケール $f_L(x_{ij}; c)$ を用いて構成した一対比較行列を $A_L = (f_L(x_{ij}; c))$ 、指数スケール $f_P(x_{ij}; m)$ を用いて構成した一対比較行列を $A_P = (f_P(x_{ij}; m))$ とする。

(1) において Saaty の線型スケール [13] は $c = 1, k = 8$ の場合、Aupetit, Genest の線型スケール [1] は $c = 1, k = 12$ の場合、Harker, Vargas の線型スケール [3] は $c = 0.5, k = 8$ の場合に、それぞれ対応する。しかし、AHP においては、通常、Saaty の 1-9 の線型スケールが用いられることから、本論文においても習慣に従い $k = 8$ とし、また c についても、 $c = 1$ の場合を中心に議論を進める。

これに対して、一対比較に指数スケールを用いる際には、Saaty [11] も述べているように底 m の決め方が結果に大きな影響を与える。Holder [4] は指数スケールを提案する中で、底の選択について具体的には言及していないが、底の値を 1 に近づけるとウェイトベクトルは $\frac{1}{n} \cdot \mathbf{1}$ (ここで $\mathbf{1}$ は全ての成分が 1 のベクトル) に収束し、評価項目に対する重要性の判別

³Lootsma も文献 [7] において、対比較データ x_{ij} を、 $x_{ij} \in \{2k \mid k = 0, \dots, 4\}$ とする、本論文と同様な指数スケールの定義を行なっている。

ができなくなる一方で、値を大きくとるにつれて、Saaty が指摘するように一対比較値が急速に増加・減少するために、人間の感覚とはかけ離れたものになる。底 m の選択には、一対比較に用いる数値の最大を $f_L(x_{ij}; 1)$ とそろえて、 $m^8 = 9$ ($m = 1.3161$) とする考え方もあるが、 f_P は凸関数であることから、 $0 < x_{ij} < 8$ のとき $f_P(x_{ij}; 1.3161) < f_L(x_{ij}; 1)$ となる。したがって、C.I. の値も小さくなる可能性があることから、C.I. の大きさに関する両スケールの公平な比較ができないため、本論文では、 $f_L(x_{ij}; 1)$ を 1 から 9 (すなわち x_{ij} を 0 から 8) まで変化させたときの、 $f_L(x_{ij}; c)$ と $f_P(x_{ij}; m)$ 、それぞれのスケールで測った一対比較値の平均的な評価が等しくなるように、次式を満たす底を用いる。

$$\int_0^8 (c \cdot x + 1) dx = \int_0^8 m^x dx \quad (3)$$

(3) より $c = 1$ としたときの底の値の近似値として $m^* = 1.3945$ が得られ、以下、断りのない限り、定数 c, m はそれぞれ $1, m^*$ とし、 $f_L(x_{ij}; c), f_P(x_{ij}; m)$ は、それぞれ単に f_L, f_P と表す。

なお本論文においては、重要度は全て固有値法に基づいて算出し、直接的近似解法である対数最小二乗法や、幾何平均を用いる方法は採らなかった⁴。

3. 比較・検証に用いるサンプル

本論文では、ランダムに構成したサンプル (以下単に、ランダムサンプル) と意思決定主体のバイアスがかかったサンプル (以下単に、データサンプル) の、性質の異なる 2 種類のサンプルを用いて、一対比較における、線型スケールと指数スケールの有効性の比較・検証を行なった。一対比較において評価項目間の相違を同時に判断できる項目数は、通常、 7 ± 2 個程度であると言われている。そこで、ランダムサンプルとして、サイズが 3×3 から 7×7 までの、5 種類の一対比較行列を用いた。また、データサンプルは AHP の実際の運用事例から得られたもので、一対比較の対象となる評価項目が 4 項目であることから、本論文において用いるデータサンプルは、サイズが 4×4 の一対比較行列のみである。

3.1 ランダムサンプル

$\{-8, \dots, -1, 0, 1, \dots, 8\}$ から一様分布に従って抽出された対比較データ $\{x_{ij}\}_r$ ($i, j = 1, \dots, h, h = 3, \dots, 7$) を、 f_L と f_P 、それぞれのスケールを用いて構成した 3×3 から 7×7 までの 5 種類の reciprocal な一対比較行列、 A_L と A_P (それぞれ 20000 個ずつ。ただし、 3×3 は 5000 個) をランダムサンプルとして用いる。

3.2 データサンプル

本論文で用いるデータサンプルは、次の調査から得られたものである。

・調査の趣旨と内容

1993 年の衆議院議員総選挙以来、政界再編成が進む中、1996 年 10 月には新しい選挙制度 (小選挙区比例代表並立制) が導入され、総選挙が実施された。このような混沌とした政治状況において、有権者がどのような政治的意識をもち、また、新しい選挙制度にどのような反応を示したかを解明するために、1997 年 2 月、筆者が荒木 (松阪大学) と共同で、三重県選挙管理委員会の協力のもとに「三重県民の選挙及び政治意識に

⁴referee より、固有値法以外に対数最小二乗法を用いてはどうか、との有益な助言を戴いたが、対数最小二乗法から得られる重要度と、一対比較行列の行の幾何平均から得られる重要度は一致することが知られており [2]、また後者は、固有値法から得られる重要度とほぼ等しい値をとることから、本論文では慣習に従い固有値法を用いた。

関する実態調査」を実施した(詳細については [18] を参照されたい). この調査は, 三重県全域 (69 市町村) を調査地域とし, 三重県選挙管理委員会が 1996 年 9 月に作成した選挙人名簿登録者から, 各市町村を単位に 69 に層化する層化二段無作為抽出法で選んだ 958 名を調査対象として行なわれたものであり, 年齢, 性別, 地域などの偏りがない母集団から得られたサンプルである. 調査方法は, 市町村の選挙管理委員と松阪大学の学生による面接訪問調査で, 回収サンプル数は 796, 回収率は 83.1%であった. この調査においては, 次の 3 つの一対比較法を用いた質問を設けた.

- ・設問 1: 小選挙区制度のもとで行なわれた 1996 年 10 月の総選挙, 小選挙区選挙での投票において, 候補者を選ぶ際の判断基準を問う質問
- ・設問 2: 支持する政党がない理由を問う質問
- ・設問 3: 中選挙区制度のもとで行なわれた 1993 年 7 月の総選挙での投票において, 候補者を選ぶ際の判断基準を問う質問

いずれの設問も, 表 1 に示したような, 抽象的で有権者個人の主観が強く反映される項目について尋ねる内容となっている.

表 1: 調査の設問で取り上げた 4 つの評価項目

	評価項目			
設問 1	選挙公約	支持政党	候補者の人柄	候補者の実績
設問 2	政治は難解	政治腐敗	政党の意図不明	どの政党を支持しても変化がない
設問 3	支持政党	選挙公約	候補者の実績	候補者の人柄

AHP における一対比較の進め方は従来より議論の対象となっており, 重要性の程度を表す言葉や数値を用いて回答させる方法や, 数直線を用いて回答させる方法などの, 回答方法に関するものの他に, 同じ階層における比較順序や, 階層図が多層構造をなしている場合の階層間での比較順序などの, 比較順序に関するものなど, さまざまな議論がなされてきている. 本論文では, アンケート調査などに AHP を応用する場合の実用性の観点から, 上記設問において一対比較は, 中央を無差別とし, 左右均等に 8 段階で重要視する程度を振り分けた数直線を用いて行なうように設計し, その結果を f_L と f_P , それぞれのスケールを用いて構成した, 4×4 の reciprocal な一対比較行列, A_L と A_P を本論文ではデータサンプルとして用いる.

なお総サンプル数は, それぞれの設問に対する有効な回答ののべ数, 1409 である.

4. ランダムサンプルを用いた線型スケールと指数スケールの比較

本章では, ランダムな対比較データ $\{x_{ij}\}_r$ から, f_L と f_P を用いて構成した 2 通りの一対比較行列, A_L と A_P に関して両スケールの比較を行なう.

4.1 検証: C.I. に関して

ランダムサンプル, A_L と A_P から計算される C.I. を rCI_{L_h} , rCI_{P_h} とすると, c の値を 0.5, 1, 1.5, 2, 3, 4, 5 とするとき, (3) に基づく, 対応する $m(c)$ の値及び, 全ランダムサンプルに関するこれらの平均値, $\overline{rCI_{L_h}}$ と $\overline{rCI_{P_h}}$ は, それぞれ表 2 に示したようにほぼ等しくなる. しかし, rCI_{L_h} と rCI_{P_h} は異なるスケールから算出されたものであり, また, いずれのサイズの行列の場合も異なる分布を示していることから, 値の大小を単純に比較することはできない. したがって, 平均値がほぼ等しくなるという結果だけから, f_L と f_P を

等質のスケールと結論付けることはできない。図1は $c = 1$ のときの 4×4 の一対比較行列から得られる C.I. の分布を示したものである⁵。

表 2: rCI_{L_h} と rCI_{P_h} の平均値

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
$\overline{rCI_{L_3}}$	0.2378	0.5270	0.8106	1.088	1.627	2.153	2.668
$\overline{rCI_{P_3}}$	0.2511	0.5382	0.8040	1.052	1.513	1.938	2.339
$\overline{rCI_{L_4}}$	0.3925	0.8996	1.400	1.916	2.945	3.968	4.984
$\overline{rCI_{P_4}}$	0.4028	0.9013	1.400	1.880	2.788	3.652	4.484
$\overline{rCI_{L_5}}$	0.4711	1.111	1.776	2.451	3.825	5.193	6.563
$\overline{rCI_{P_5}}$	0.5051	1.168	1.832	2.487	3.767	5.008	6.219
$\overline{rCI_{L_6}}$	0.5301	1.257	2.019	2.793	4.358	5.933	7.512
$\overline{rCI_{P_6}}$	0.5719	1.340	2.125	2.909	4.461	5.991	7.501
$\overline{rCI_{L_7}}$	0.6160	1.457	2.336	3.229	5.033	6.847	8.665
$\overline{rCI_{P_7}}$	0.6762	1.620	2.618	3.640	5.725	7.843	9.983

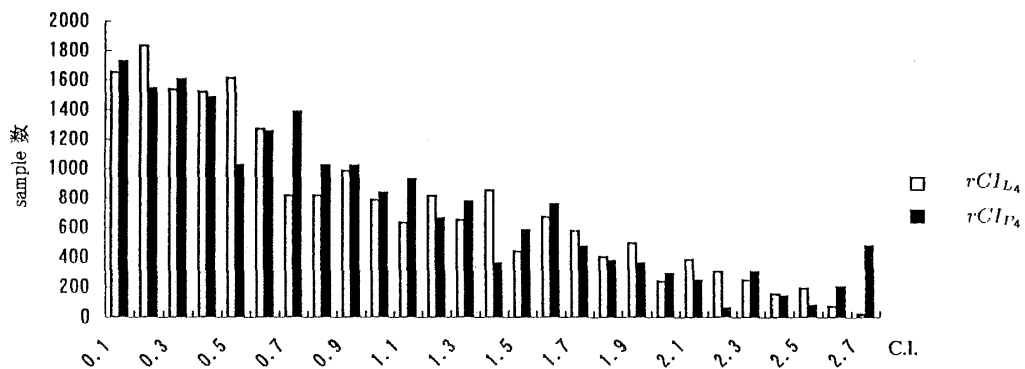


図 1: rCI_{L_4}, rCI_{P_4} の分布

そこで本論文では、 rCI_{L_h} と rCI_{P_h} それぞれを、昇べきの順でみたときの、上位 1%, 3%, 5%, 7%, 10%, 15%, 20%, 25% のサンプルの C.I. を基準として f_L と f_P の比較を行なうことにする。表 3 は、各基準点に対応する rCI_{L_h} , rCI_{P_h} の値を示したもので、ほとんどの箇所で $rCI_{L_h} > rCI_{P_h}$ となっており、同じ基準点で比較したとき、 f_L を用いた場合よりも f_P を用いた場合の方が C.I. の値が小さくなる傾向があることが判る。

表 3: 各基準点に対応する rCI_{L_h} と rCI_{P_h} の値 ($c = 1, m = m^*$)

行列サイズ	3 × 3		4 × 4		5 × 5		6 × 6		7 × 7	
	rCI_{L_3}	rCI_{P_3}	rCI_{L_4}	rCI_{P_4}	rCI_{L_5}	rCI_{P_5}	rCI_{L_6}	rCI_{P_6}	rCI_{L_7}	rCI_{P_7}
1%	0	0	0.04741	0.03277	0.1752	0.1482	0.3486	0.3168	0.4834	0.4366
3%	0.0009908	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

⁵ 4×4 以外のサイズに関する分布図は傾向が全く同じであったので、煩雑さを避けるために割愛した。

4.2 検証: 重要度の序列変動に関して

スケール $f_L(x_{ij}; c)$, $f_P(x_{ij}; m)$ を定義する際の定数, c や m の値を変えることによって, 一対比較値のとり得る値の範囲も変化する. 一対比較値のとり得る値の範囲を変化させると, Frobenius 根の値のみならず, 対応する固有ベクトルの向きが変動し, 得られた評価項目に対する重要度の序列を覆してしまうことがある. 特に, 重要度が高い評価項目間での序列変動があった場合, 得られた重要度に対する信頼性は低いと言わざるを得ない.

本論文では, 4.1 で用いた, サイズが 3×3 から 7×7 の一対比較行列, 各 20000 個のランダムサンプルの中からそれぞれ任意に 1000 個ずつを選び, c と m を次の範囲で変化させたときの重要度の変化を検証した.

$$c: 0 \leq c \leq 10, \quad m: 1 \leq m \leq 3. \quad (4)$$

このとき, f_L と f_P を用いた一対比較値のとり得る値の範囲は, これまでに提案されているさまざまなスケールの範囲をほぼ網羅することから, 今回の検証にあたっては充分であると考え⁶.

表 4-8 は, 1000 個のサンプルの rCI_{L_h} と rCI_{P_h} ($h = 3, \dots, 7$) の分布の累積数 (A, A') と, c と m を (4) に示した範囲で変化させたときに重要度の序列が変動⁷したサンプルの累積数 (B, B'), 及びその比率をまとめたものである. いずれのサイズの一対比較行列においても, 線型スケールを用いた場合と比較して, 指数スケールを用いた場合の方が総数は多いものの, C.I. の値が小さなサンプルに関しては指数スケールの方が少ない.

図 2.1.1-2.5.2 は, 序列が変動したサンプルの中で, C.I. の値が小さく, c の値が 1 前後で変動が起きている例について, その変動の様子を示したものである⁸. これらの例において, 序列変動が起こるパラメーター c の値を c' とすると, 線型スケールを用いて得られる各項目の重要度の $0 < c < c'$ における序列と, 指数スケールを用いて得られる各項目の序列とが一致する傾向がある点が特徴として挙げられる.

表 4: 重要度の序列が変動したサンプル数と比率: 3×3

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	21	0	0%	47	0	0%
3%	41	0	0%	47	0	0%
5%	59	0	0%	144	0	0%
7%	75	0	0%	144	0	0%
10%	106	0	0%	144	0	0%
15%	169	0	0%	231	0	0%
20%	217	0	0%	231	0	0%
25%	275	7	2.55%	336	0	0%
100%	1000	36	3.60%	1000	2	0.20%

⁶ $c = 10$ のとき (3) から導かれる底の値は $m = 1.9654$ であるが, ここではさらに大きな値の範囲までとった.

⁷ 5×5 から 7×7 の一対比較行列に関しては, 重要度が最も高い項目と 2 番目に高い項目の間で序列が変動したサンプルのみを集計した.

⁸グラフは, 重要度の変動が起きている部分を拡大して表示した. また各行列は, $c = 1, m = m^*$ のときのものである.

・サンプル1

$$A_L = \begin{pmatrix} 1 & 1/8 & 1/2 \\ 8 & 1 & 1/2 \\ 2 & 2 & 1 \end{pmatrix}$$

$$rCI_{L_3} = 0.25$$

$$A_P = \begin{pmatrix} 1 & 0.0975 & 0.717 \\ 10.3 & 1 & 0.717 \\ 1.39 & 1.39 & 1 \end{pmatrix}$$

$$rCI_{P_3} = 0.316$$

weight

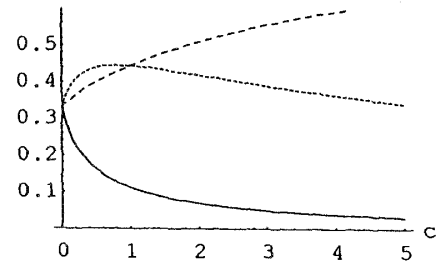


図 2.1.1 A_L の重要度の変化: サンプル1

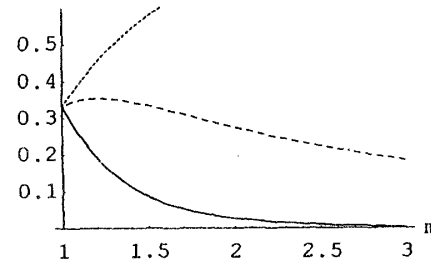


図 2.1.2 A_P の重要度の変化: サンプル1

表 5: 重要度の序列が変動したサンプル数と比率: 4×4

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	13	0	0%	16	0	0%
3%	36	4	11.11%	37	0	0%
5%	50	8	16.00%	63	0	0%
7%	84	15	17.86%	79	0	0%
10%	113	20	17.70%	116	0	0%
15%	169	26	15.38%	154	0	0%
20%	208	34	16.35%	210	3	1.43%
25%	260	42	16.15%	265	3	1.13%
100%	1000	291	29.10%	1000	285	28.50%

・サンプル2

$$A_L = \begin{pmatrix} 1 & 4 & 2 & 1 \\ 1/4 & 1 & 1/8 & 1/9 \\ 1/2 & 8 & 1 & 1 \\ 1 & 9 & 1 & 1 \end{pmatrix}$$

$$rCI_{L_4} = 0.0659$$

weight

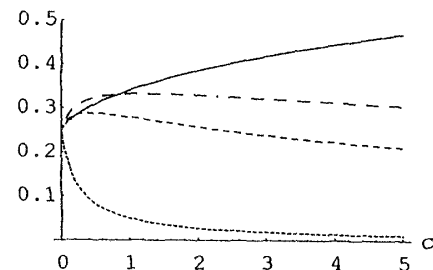


図 2.2.1 A_L の重要度の変化: サンプル2

$$A_P = \begin{pmatrix} 1 & 2.71 & 1.39 & 1 \\ 0.369 & 1 & 0.0975 & 0.0699 \\ 0.717 & 10.3 & 1 & 1 \\ 1 & 14.3 & 1 & 1 \end{pmatrix}$$

$rCI_{P_4} = 0.126$

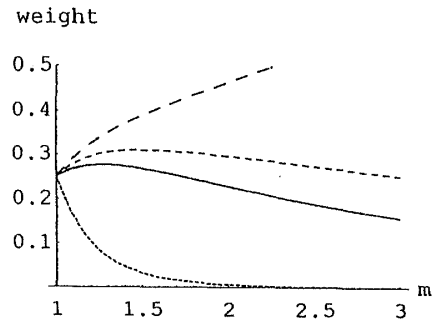


図 2.2.2 A_P の重要度の変化: サンプル 2

表 6: 重要度の序列が変動したサンプル数と比率: 5×5

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	13	0	0%	8	0	0%
3%	23	0	0%	30	0	0%
5%	51	0	0%	55	0	0%
7%	72	5	6.94%	77	1	1.30%
10%	115	16	13.91%	118	5	4.24%
15%	180	22	12.22%	173	8	4.62%
20%	230	35	15.22%	230	16	6.96%
25%	272	44	16.18%	278	35	12.59%
100%	1000	127	12.70%	1000	196	19.60%

・ サンプル 3

$$A_L = \begin{pmatrix} 1 & 1/7 & 1/5 & 6 & 1/5 \\ 7 & 1 & 6 & 2 & 1/2 \\ 5 & 1/6 & 1 & 5 & 1/3 \\ 1/6 & 1/2 & 1/5 & 1 & 1/4 \\ 5 & 2 & 3 & 4 & 1 \end{pmatrix}$$

$rCI_{L_5} = 0.365$

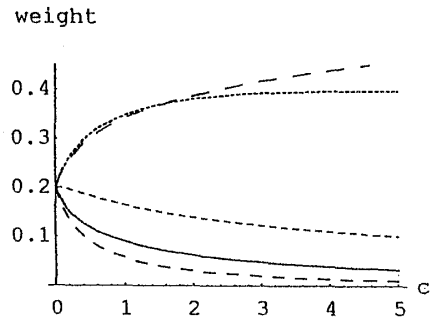


図 2.3.1 A_L の重要度の変化: サンプル 3

$$A_P = \begin{pmatrix} 1 & 0.136 & 0.264 & 5.27 & 0.264 \\ 7.35 & 1 & 5.27 & 1.39 & 0.717 \\ 3.78 & 0.19 & 1 & 3.78 & 0.514 \\ 0.19 & 0.717 & 0.264 & 1 & 0.369 \\ 3.78 & 1.39 & 1.94 & 2.71 & 1 \end{pmatrix}$$

$rCI_{P_5} = 0.353$

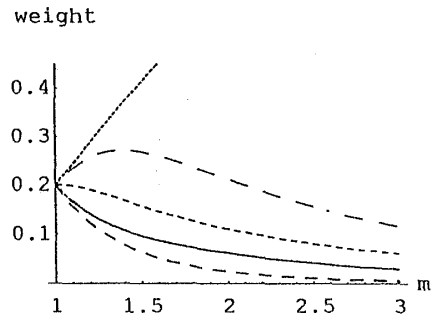


図 2.3.2 A_P の重要度の変化: サンプル 3

表 7: 重要度の序列が変動したサンプル数と比率: 6 × 6

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	10	2	20.00%	7	0	0%
3%	23	3	13.04%	22	0	0%
5%	38	4	10.53%	36	1	2.78%
7%	58	6	10.34%	60	2	3.33%
10%	89	6	6.74%	86	5	5.81%
15%	125	12	9.60%	123	11	8.94%
20%	166	20	12.05%	177	15	8.47%
25%	228	33	14.47%	220	16	7.27%
100%	1000	161	16.10%	1000	274	27.40%

・ サンプル 4

$$A_L = \begin{pmatrix} 1 & 9 & 5 & 2 & 6 & 2 \\ 1/9 & 1 & 1/4 & 1/3 & 1/4 & 1/6 \\ 1/5 & 4 & 1 & 1/8 & 1/7 & 1/6 \\ 1/2 & 3 & 8 & 1 & 6 & 9 \\ 1/6 & 4 & 7 & 1/6 & 1 & 4 \\ 1/2 & 6 & 6 & 1/9 & 1/4 & 1 \end{pmatrix}$$

$$rCI_{L_6} = 0.341$$

$$A_P = \begin{pmatrix} 1 & 14.3 & 3.78 & 1.39 & 5.27 & 1.39 \\ 0.0699 & 1 & 0.369 & 0.514 & 0.369 & 0.19 \\ 0.264 & 2.71 & 1 & 0.0975 & 0.136 & 0.19 \\ 0.717 & 1.94 & 10.3 & 1 & 5.27 & 14.3 \\ 0.19 & 2.71 & 7.35 & 0.19 & 1 & 2.71 \\ 0.717 & 5.27 & 5.27 & 0.0699 & 0.369 & 1 \end{pmatrix}$$

$$rCI_{P_6} = 0.396$$

weight

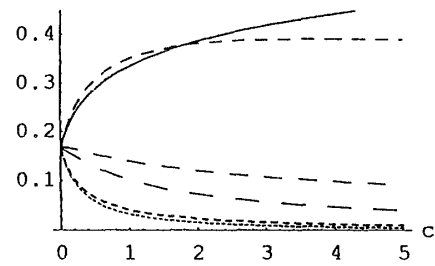


図 2.4.1 A_L の重要度の変化: サンプル 4

weight

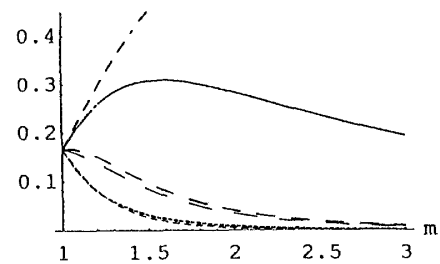


図 2.4.2 A_P の重要度の変化: サンプル 4

表 8: 重要度の序列が変動したサンプル数と比率: 7 × 7

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	6	0	0%	2	0	0%
3%	20	4	20.00%	13	0	0%
5%	41	7	17.07%	27	0	0%
7%	60	11	18.33%	58	0	0%
10%	103	23	22.33%	99	0	0%
15%	159	30	18.87%	156	13	8.33%
20%	225	40	17.78%	223	35	15.70%
25%	288	46	15.97%	308	45	14.61%
100%	1000	201	20.10%	1000	368	36.80%

・サンプル5

$$A_L = \begin{pmatrix} 1 & 4 & 5 & 3 & 1/2 & 1/5 & 1 \\ 1/4 & 1 & 2 & 1/3 & 1/4 & 4 & 1/4 \\ 1/5 & 1/2 & 1 & 3 & 3 & 1/3 & 1 \\ 1/3 & 3 & 1/3 & 1 & 1/4 & 2 & 1/6 \\ 2 & 4 & 1/3 & 4 & 1 & 1 & 2 \\ 5 & 1/4 & 3 & 1/2 & 1 & 1 & 1/2 \\ 1 & 4 & 1 & 6 & 1/2 & 2 & 1 \end{pmatrix}$$

$rCI_{L7} = 0.575$

$$A_P = \begin{pmatrix} 1 & 2.71 & 3.78 & 1.94 & 0.717 & 0.264 & 1 \\ 0.369 & 1 & 1.39 & 0.514 & 0.369 & 2.71 & 0.369 \\ 0.264 & 0.717 & 1 & 1.94 & 1.94 & 0.514 & 1 \\ 0.514 & 1.94 & 0.514 & 1 & 0.369 & 1.39 & 0.19 \\ 1.39 & 2.71 & 0.514 & 2.71 & 1 & 1 & 1.39 \\ 3.78 & 0.369 & 1.94 & 0.717 & 1 & 1 & 0.717 \\ 1 & 2.71 & 1 & 5.27 & 0.717 & 1.39 & 1 \end{pmatrix}$$

$rCI_{P7} = 0.271$

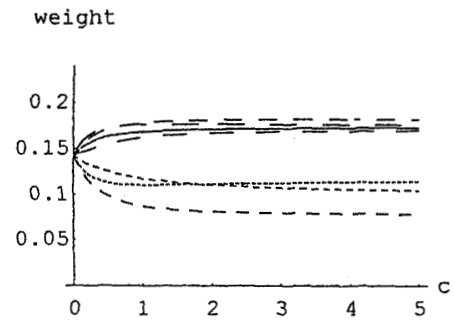


図 2.5.1 A_L の重要度の変化: サンプル5

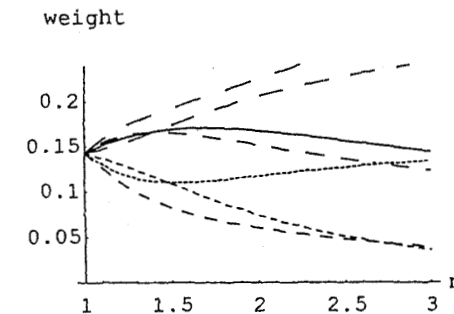


図 2.5.2 A_P の重要度の変化: サンプル5

4.3 検証: 判別性能に関して

AHP の目的には、各評価項目に対する価値判断を重要度として定量化し、最も重要性が高い項目を他の項目から判別することがある。重要度は、Frobenius 根に対応する固有ベクトルを l_1 -ノルムで 1 に正規化したものであることから、本論文ではその最大成分を v^1 、2 番目に大きい成分を v^2 とするとき、 f_L と f_P 、それぞれのスケールで測った $v^1 - v^2$ ($\equiv v$) の大小を比較することで、最重要と評価された項目が他の項目に対してどの程度の優位性をもっているかを、一種の判別性能として測った⁹。

表 9-13 は f_L を用いた場合の v の差 v_L と、 f_P を用いた場合の v の差 v_P の大小を、4.1 で述べた各基準点ごとに比較した結果をまとめたものである。いずれのサイズの一对比較行列においても $v_L < v_P$ となるサンプルの数が多い¹⁰ことから、 f_L と f_P を比較すると、 f_P の方が評価項目の重要度に対する判別性能が高いと言える。

表 9. 判別性能の比較: 3 × 3

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	14 (14.58%)	30 (31.25%)	52 (54.17%)	51 (22.17%)	31 (13.48%)	148 (64.35%)
3 %	70 (31.53%)	30 (13.51%)	122 (54.95%)	51 (22.17%)	31 (13.48%)	148 (64.35%)
5 %	105 (33.44%)	30 (9.55%)	179 (57.01%)	182 (27.41%)	31 (4.67%)	451 (67.92%)
7 %	136 (34.17%)	30 (7.54%)	232 (58.29%)	182 (27.41%)	31 (4.67%)	451 (67.92%)
10 %	171 (31.32%)	30 (5.49%)	345 (63.19%)	182 (27.41%)	31 (4.67%)	451 (67.92%)
15 %	257 (30.31%)	30 (3.54%)	561 (66.16%)	307 (28.22%)	31 (2.85%)	750 (68.93%)
20 %	320 (30.30%)	30 (2.84%)	706 (66.86%)	307 (28.22%)	31 (2.85%)	750 (68.93%)
25 %	396 (29.57%)	30 (2.24%)	913 (68.19%)	439 (29.21%)	32 (2.13%)	1032 (68.66%)
100 %	2343 (48.86%)	42 (0.84%)	2615 (52.30%)	2343 (46.86%)	42 (0.84%)	2615 (52.30%)

⁹ この判別性能をそれぞれのスケールで測った“比” (v^1/v^2) の大小で比較する方法も考えられるが、本論文の“差”を用いた比較とほぼ同様な結果が得られたため割愛した。

¹⁰ χ^2 検定を行なったところ、1%水準で有意差があることが検証された。

表 10. 判別性能の比較: 4×4

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	99 (44.39%)	2 (0.90%)	122 (54.17%)	83 (31.32%)	1 (0.38%)	181 (68.30%)
3 %	264 (42.58%)	2 (0.32%)	354 (57.10%)	207 (34.44%)	1 (0.17%)	393 (65.39%)
5 %	406 (40.32%)	2 (0.20%)	599 (59.48%)	369 (36.04%)	2 (0.20%)	653 (63.77%)
7 %	575 (40.87%)	3 (0.21%)	829 (58.92%)	521 (37.21%)	2 (0.14%)	877 (62.64%)
10 %	813 (40.57%)	3 (0.15%)	1188 (59.28%)	797 (39.75%)	2 (0.10%)	1206 (60.15%)
15 %	1237 (41.16%)	3 (0.10%)	1765 (58.74%)	1207 (40.21%)	3 (0.10%)	1792 (59.69%)
20 %	1753 (43.77%)	3 (0.07%)	2249 (56.15%)	1656 (41.13%)	3 (0.07%)	2367 (58.79%)
25 %	2216 (44.26%)	4 (0.08%)	2787 (55.66%)	2155 (43.05%)	4 (0.08%)	2847 (56.87%)
100 %	7842 (39.21%)	5 (0.03%)	121453 (60.77%)	7842 (39.21%)	5 (0.03%)	12153 (60.77%)

表 11. 判別性能の比較: 5×5

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	85 (42.08%)	0 (0.00%)	117 (57.92%)	84 (41.58%)	0 (0.00%)	118 (58.42%)
3 %	281 (46.29%)	0 (0.00%)	326 (53.71%)	251 (41.69%)	0 (0.00%)	351 (58.31%)
5 %	487 (48.31%)	0 (0.00%)	521 (51.69%)	417 (41.70%)	0 (0.00%)	583 (58.30%)
7 %	673 (47.08%)	0 (0.00%)	735 (52.20%)	598 (42.68%)	1 (0.07%)	802 (57.24%)
10 %	930 (45.79%)	1 (0.05%)	1100 (54.16%)	875 (43.45%)	1 (0.05%)	1138 (56.50%)
15 %	1414 (44.02%)	1 (0.03%)	1797 (55.95%)	1314 (43.29%)	2 (0.07%)	1719 (56.64%)
20 %	1853 (43.91%)	1 (0.02%)	2366 (56.07%)	1714 (42.84%)	2 (0.05%)	2285 (57.11%)
25 %	2271 (42.99%)	1 (0.02%)	3011 (56.99%)	2069 (41.36%)	2 (0.04%)	2931 (58.60%)
100 %	7063 (35.32%)	3 (0.02%)	12934 (64.67%)	7063 (35.32%)	3 (0.02%)	12934 (64.67%)

表 12. 判別性能の比較: 6×6

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	116 (47.35%)	1 (0.41%)	128 (52.24%)	113 (46.89%)	0 (0.00%)	128 (53.11%)
3 %	343 (47.64%)	1 (0.14%)	376 (52.22%)	283 (41.37%)	1 (0.15%)	400 (58.48%)
5 %	505 (45.74%)	1 (0.09%)	598 (54.17%)	439 (43.55%)	1 (0.10%)	568 (56.35%)
7 %	687 (45.65%)	1 (0.07%)	817 (54.29%)	597 (41.75%)	2 (0.14%)	831 (58.11%)
10 %	879 (43.02%)	1 (0.05%)	1163 (56.93%)	763 (38.15%)	2 (0.10%)	1235 (61.75%)
15 %	1207 (40.19%)	3 (0.10%)	1793 (59.71%)	1108 (36.92%)	2 (0.07%)	1891 (63.01%)
20 %	1519 (37.90%)	4 (0.10%)	2485 (62.00%)	1444 (36.05%)	4 (0.10%)	2557 (63.85%)
25 %	1803 (36.05%)	4 (0.08%)	3195 (63.87%)	1798 (35.95%)	5 (0.10%)	3199 (63.95%)
100 %	6993 (34.97%)	5 (0.03%)	13002 (65.01%)	6993 (34.97%)	5 (0.03%)	13002 (65.01%)

表 13. 判別性能の比較: 7×7

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	38 (18.63%)	1 (0.49%)	165 (80.88%)	39 (19.31%)	0 (0.00%)	163 (80.69%)
3 %	228 (37.94%)	2 (0.33%)	371 (61.73%)	210 (34.37%)	0 (0.00%)	401 (65.63%)
5 %	422 (41.54%)	3 (0.30%)	591 (58.17%)	371 (37.03%)	0 (0.00%)	631 (62.97%)
7 %	557 (39.12%)	3 (0.21%)	864 (60.67%)	507 (36.19%)	1 (0.07%)	893 (63.74%)
10 %	790 (39.19%)	3 (0.15%)	1223 (60.66%)	719 (35.77%)	2 (0.10%)	1289 (64.13%)
15 %	997 (31.73%)	3 (0.10%)	2142 (68.17%)	928 (29.90%)	4 (0.13%)	2172 (69.97%)
20 %	1579 (39.42%)	6 (0.15%)	2421 (60.43%)	1257 (31.28%)	5 (0.12%)	2756 (68.50%)
25 %	1961 (39.21%)	9 (0.18%)	3031 (60.61%)	1801 (35.95%)	8 (0.16%)	3201 (63.89%)
100 %	7565 (37.83%)	13 (0.07%)	12422 (62.11%)	7565 (37.83%)	13 (0.07%)	12422 (62.11%)

5. データサンプルを用いた線型スケールと指数スケールの比較

本章では, 3.2 に示したデータサンプルの対比較データ $\{x_{ij}\}_d$ から, f_L と f_P を用いて構成した 2 通りの一対比較行列 A_L と A_P に関して両スケールの比較を行なう。

5.1 検証: C.I. に関して

一対比較行列の有効性の判定に, Saaty は R.I. の 10% を C.I. の許容上限とすることを提唱している. 一方, C.I. の値は, 行列のサイズ, すなわち評価項目の数 n に依存することから, E.F. Lane, W.A. Verdini [5] は, $n = 3, 4$ のときはそれぞれ R.I. の 1%, 5% を, $n \geq 5$ のときは R.I. の 10% を上限とするべきであるとしており, 通常, このような R.I. の一定割合を C.I. の許容上限として用いる. しかし 4.1 で述べたように, rCI_L と rCI_P は異なる分布を示すことから, 単に R.I. の 10% (あるいは 5%) を C.I. の上限¹¹として, C.I. の値がその上限を越えないサンプル数の多寡で両スケールを比較するのは公平ではない.

そこで本論文では, 4.1 で述べた各基準点に対応する rCI_L , rCI_P 以下のサンプルの累積数によって f_L と f_P の比較を行なうことにする. データサンプルをもとにした一対比較行列 A_L, A_P から計算される C.I. を dCI_L, dCI_P とする.

表 14: dCI_L, dCI_P の分布

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	rCI_L	累積サンプル数	比率	rCI_P	累積サンプル数	比率
1%	0.04741	302	21.43%	0.03277	379	26.90%
3%	0.08662	430	30.52%	0.06862	549	38.96%
5%	0.11413	599	42.51%	0.09807	717	50.89%
7%	0.14475	657	46.63%	0.12513	731	51.88%
10%	0.17805	787	55.86%	0.16475	932	66.15%
15%	0.23180	949	67.35%	0.22340	1016	72.11%
20%	0.28650	1029	73.03%	0.28985	1106	78.50%
25%	0.34205	1102	78.21%	0.34583	1195	84.81%
100%	-	1409	100.00%	-	1409	100.00%

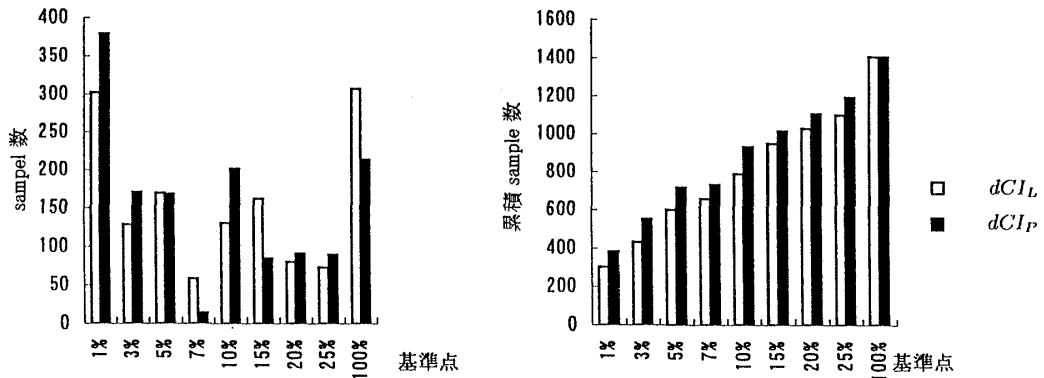


図 3: dCI_L, dCI_P の分布

表 14 は, 上で述べた各基準点に対応する rCI_L と rCI_P の値, 並びに dCI_L と dCI_P の値が, 各基準点に対応する rCI_L と rCI_P 以下の値となるサンプルの累積数をまとめたものである. 図 3 より, 同一のサンプルを A_L と A_P , 2 通りの一対比較行列にまとめた場合, C.I. の分布は A_L よりも A_P の方が, 0 に近い値に偏って分布する傾向があることが判る. これは人間の価値判断を定量化するにあたって, f_P の方がより人間の感覚を統合的に一対比較行列に反映していることを意味するだけでなく, 社会調査等の実際の AHP の適用の場

¹¹本論文では, C.I. の許容上限についてはふれない.

において、C.I. が大きいため調査の分析対象とならない無効なサンプルを減らすことができるという、実用上の利点もある。なお、表 14 の累積サンプル数に関して χ^2 検定を行なったところ χ^2 値は 20.16 で、自由度が 8 であることから $\chi^2 > \chi_{0.01}^2 = 20.09$ となり、1% の有意水準で表の累積サンプル数には有意な差があることが確かめられた。

5.2 検証: 重要度の序列変動に関して

表 15 は、1409 個全てのデータサンプルの dCI_L と dCI_P の分布の累積数 (A, A') と、 c と m を (4) に示した範囲で変化させたときに、重要度の序列が変動したサンプル数 (B, B')、及びその比率をまとめたものである。

表 15: 重要度の序列が変動したサンプル数と比率

基準点	$f_L(x_{ij}; 1)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	A	B	B/A	A'	B'	B'/A'
1%	302	0	0%	379	0	0%
3%	430	1	0.2326%	549	0	0%
5%	599	2	0.3339%	717	0	0%
7%	657	3	0.4566%	731	0	0%
10%	787	4	0.5083%	932	0	0%
15%	949	4	0.4215%	1016	6	0.5906%
20%	1029	5	0.4859%	1106	16	1.447%
25%	1102	9	0.8167%	1195	18	1.506%
100%	1409	57	4.045%	1409	43	3.052%

その結果、重要度の序列が覆ったサンプルは、 A_L の場合、57 個、 A_P の場合、43 個あった。ただし、 dCI_L, dCI_P の値が、それぞれ上位 5% 以内¹²の、C.I. が小さなサンプルに限定すると、 A_P では、序列の変動が起こった例は 1 つもないが、 A_L の場合、以下に示すように、2 つのサンプルにおいて序列の変動が起こった。特にサンプル 6 は、重要度が最も大きい項目と 2 番目に大きい項目間での序列の逆転であり、しかもサンプル 6, 7, いずれの例においても、 c が 1 前後の、Saaty の 1-9 の線型スケールとほぼ同じスケールのときに順序の変動が起こっていることから、 f_L を用いた場合の重要度の信頼性は高いとは言えない。また、序列変動が起こるパラメーター c の値を c' とすると、4.2 で述べたランダムサンプルの場合と同様、線型スケールを用いて得られる各項目の重要度の $0 < c < c'$ における序列と、指数スケールを用いて得られる各項目の序列とが一致している点の特徴として挙げられる。

・サンプル 6

$$A_L = \begin{pmatrix} 1 & 3 & 1/3 & 1 \\ 1/3 & 1 & 1/3 & 1/7 \\ 3 & 3 & 1 & 1 \\ 1 & 7 & 1 & 1 \end{pmatrix}$$

$$dCI_L = 0.0819$$

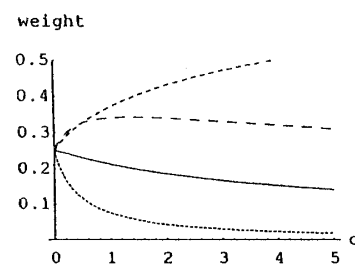


図 4.1.1 A_L の重要度の変化: サンプル 6

¹²Saaty が基準とする C.I. の上限、0.1 とほぼ同等の基準として選んだ。

$$A_P = \begin{pmatrix} 1 & 1.94 & 0.514 & 1 \\ 0.514 & 1 & 0.514 & 0.136 \\ 1.94 & 1.94 & 1 & 1 \\ 1 & 7.35 & 1 & 1 \end{pmatrix}$$

$$dCI_P = 0.0949$$

・サンプル7

$$A_L = \begin{pmatrix} 1 & 1/7 & 1/3 & 1/5 \\ 7 & 1 & 1 & 1/3 \\ 3 & 1 & 1 & 1 \\ 5 & 3 & 1 & 1 \end{pmatrix}$$

$$dCI_L = 0.0911$$

$$A_P = \begin{pmatrix} 1 & 0.136 & 0.514 & 0.264 \\ 7.35 & 1 & 1 & 0.514 \\ 1.94 & 1 & 1 & 1 \\ 3.78 & 1.94 & 1 & 1 \end{pmatrix}$$

$$dCI_P = 0.0949$$

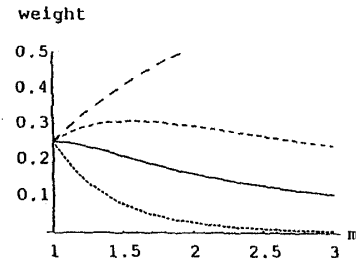


図 4.1.2 A_P の重要度の変化: サンプル6

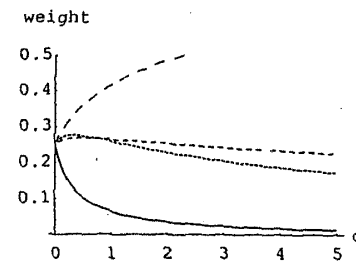


図 4.2.1 A_L の重要度の変化: サンプル7

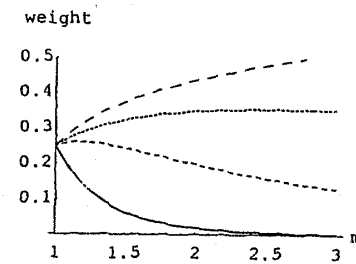


図 4.2.2 A_P の重要度の変化: サンプル7

5.3 検証: 判別性能に関して

表 16 は、1409 個全てのデータサンプルについて、4.3 と同様の検証を行なった結果をまとめたものである。表の $v_L = v_P, 1\%$ の欄は、 f_L の場合 182 例中 143 例、 f_P の場合 194 例中 143 例は、全ての対比較値が 1 のサンプル¹³である。また、表の $v_L = v_P, 100\%$ の欄は、247 例中 235 例は $v^1 = v^2$ となっているサンプルで、 $v_L = v_P = 0$ である。これらのことは、表 9-13 のランダムサンプルのケースとは対称的に、 $v_L = v_P = 0$ となっているサンプルが多く、現実の意識調査においては、対比較データ $\{x_{ij}\}_d$ に、全ての項目に対して無判別とするサンプルや、最も重要と判断する項目が 2 つ以上存在するサンプルが多く含まれていることを示している。

その一方で、全ての対比較値が 1 のサンプルを除いて f_L と f_P を比較すると、 $v_L < v_P$ となるサンプルの数が多い¹⁴ことから、ランダムデータを用いた場合の検証結果がここでも追認され、指数スケールの方が評価項目の重要度に対する判別性能が高いと言える。

6. 結論

AHP は人の主観を入力とし、重要度という評価項目に対する重みを出力とする意思決定手法である。そのため、人間の感覚を如何にうまく捉えることができるかという観点から、

¹³この場合、 f_L, f_P 、いずれのスケールを用いても対比較行列は 1 行列となり、C.I. は自明に 0 である。

¹⁴ χ^2 検定を行なったところ、1%水準で有意差があることが検証された。

表 16. 判別性能の比較

基準点	累積サンプル数 (比率)					
	$f_L(x_{ij}; l)$ (線型スケール)			$f_P(x_{ij}; m^*)$ (指数スケール)		
	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$	$v_L > v_P$	$v_L = v_P$	$v_L < v_P$
1 %	46 (15.23%)	182 (60.26%)	74 (24.50%)	47 (12.40%)	194 (51.19%)	138 (36.41%)
3 %	64 (14.88%)	194 (45.12%)	172 (40.00%)	77 (14.03%)	197 (35.88%)	275 (50.09%)
5 %	87 (14.52%)	211 (35.23%)	301 (50.25%)	99 (13.81%)	212 (29.57%)	406 (56.62%)
7 %	105 (15.98%)	212 (32.27%)	340 (51.75%)	100 (13.68%)	214 (29.27%)	417 (57.05%)
10 %	146 (18.55%)	218 (27.70%)	423 (53.75%)	160 (17.17%)	221 (23.71%)	551 (59.12%)
15 %	205 (21.60%)	225 (23.71%)	519 (54.69%)	189 (18.60%)	221 (21.75%)	606 (59.65%)
20 %	231 (22.45%)	226 (21.96%)	572 (55.59%)	244 (22.06%)	229 (20.71%)	633 (57.23%)
25 %	285 (25.86%)	226 (20.51%)	591 (53.63%)	314 (26.28%)	229 (19.16%)	652 (54.56%)
100 %	469 (33.29%)	247 (17.53%)	693 (49.18%)	469 (33.29%)	247 (17.53%)	693 (49.18%)

一対比較に用いるスケールとしてさまざまなスケールが提案されてきており、特に Saaty の線型スケールと指数スケールは比較の対象となってきた。しかしこの両スケールの比較は、性質の異なるスケールの比較であることから、判断の整合性の判定基準である C.I. の大きさを単純に比較することができないため、議論は複雑なものとなっていた。また、算出された重要度に関する理論的な検証が不可能であることから、一対比較に用いるスケールとしての有効性について評価を下すのは極めて困難であった。

そこで本論文では、一対比較に用いるスケールとして Saaty 型の線型スケールと Lootsma 型の指数スケールを取り上げ、ランダムに構成した一対比較行列と、意思決定主体のバイアスがかかった一対比較行列の、2 種類サンプルを用いた実証的な比較・評価を行なった。両スケールの比較にあたっては、ランダムに構成したサンプルを基に C.I. に関する基準点を設けることによって両スケールの整合性の有無に関する相対比較を可能とし、評価においては、C.I. の分布、重要度の序列変動の有無、評価項目に対する判別性能の、3 つの側面から分析を行なった。

その結果、次のことが明らかになった。

ランダムに構成したサンプル (3 × 3 から 7 × 7 の 5 種類の行列) に関して

- ・一対比較値の評価面積をそろえると、いずれのサイズの一対比較行列においても、線型・指数、それぞれのスケールで測った整合度の平均値はほぼ等しくなる。
- ・3 × 3 から 7 × 7 までの各サイズのランダムサンプルの中から、それぞれ任意に選び出した 1000 個ずつのサンプルに関して、関数 f のパラメーターの変化に対して重要度の序列が変動するサンプルの数を線型・指数、両スケールについて比較した。その結果 C.I. の値が小さなサンプルに関しては、いずれのサイズの一対比較行列においても線型スケールよりも指数スケールの方が少ない。
- ・いずれのサイズの一対比較行列においても、判別性能は指数スケールの方が高い。

意思決定主体のバイアスがかかったサンプル (4 × 4 の行列) に関して

- ・C.I. の分布は、線型スケールを用いた場合よりも指数スケールを用いた場合の方が有意な差 (1%水準) をもって 0 に近い値に偏る。
- ・全サンプル中、関数 f のパラメーターの変化に対して重要度の序列が変動するサンプルの数は、線型スケールを用いた場合 57 個、指数スケールを用いた場合 43 個、その中で、それぞれのスケールで測った C.I. の値が上位 5% 以内のサンプルに限定すると、序列変動を起こすサンプルの数は、線型スケールを用いた場合 2 個、指数スケールを用いた場合 0 個。
- ・判別性能は指数スケールの方が高い。

本論文で用いた指数スケールは、底の値を 1.3945 としていることから、 $0 \leq x_{ij} \leq 6$ (Saaty のスケールでは、1-7) の範囲では $f_L > f_P$ となる。そのため、同一の対比較データからそれぞれのスケールを用いて一対比較行列を構成する際に、指数スケールを用いた場合の方が、C.I. の値が良好な分布を示す結果となった可能性もある。しかし、最重要と評価された項目と 2 番目に重要と評価された項目の重要度の差を、一対比較に用いるスケールの、一種の判別性能として評価したところ、この判別性能の検証においても指数スケールの方が優れている結果となった。

これらのことを併せて考えると、線型スケールと比較して指数スケールは、評価項目間の重要性の差を判別した上で、C.I. の値の分布においても良好な結果を示しており、加えて、C.I. の値が小さいサンプルに限定すると重要度の序列も変動を起こさないことから、一対比較に用いるスケールとして、より有効に機能していることが判る。

なお、本論文で行なった、AHP の実際の適用例に関する分析は 4×4 の一対比較行列に関するものである。 4×4 以外のサイズの一対比較行列に関する、本論文と同様な追検証は今後の研究課題としたい。

参考文献

- [1] B. Aupetit and C. Genest: 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** (1993) 263–268.
- [2] G. Crawford and C. Williams: A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, **29** (1985) 387–405.
- [3] P.T. Harker and L.G. Vargas: The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science*, **33** (1987) 1383–1403.
- [4] R.D. Holder: Some comments on the analytic hierarchy process. *Journal of Operational Research Society*, **41** (1990) 1073–1076.
- [5] E.F. Lane and W.A. Verdini: A consistency test for AHP decision makers. *Decision Science*, **20** (1989) 575–590.
- [6] F.A. Lootsma: Conflict resolution via pairwise comparison of concessions. *European Journal of Operational Research*, **40** (1989) 109–116.
- [7] F.A. Lootsma: 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*, no. 91-20 (1991).
- [8] C.K. Murphy: Limits on the analytic hierarchy process from its consistency index. *European Journal of Operational Research*, **65** (1993) 138–139.
- [9] K. Nishizawa: A consistency improving method in binary AHP. *Journal of the Operations Research Society of Japan*, **38** (1995) 21–33.
- [10] T.L. Saaty: A scaling method for priorities in hierarchy structures. *Journal of Mathematical Psychology*, **15** (1977) 234–281.
- [11] T.L. Saaty: Highlights and critical points in the theory and application of the analytic hierarchy process. *European Journal of Operational Research*, **52** (1994) 426–447.
- [12] T.L. Saaty: How to make a decision: the analytic hierarchy process. *European Journal of Operational Research*, **48** (1990) 9–26.

- [13] T.L. Saaty: *The Analytic Hierarchy Process* (McGraw-Hill, 1980).
- [14] T.L. Saaty: Priority setting in complex problems. *IEEE Transactions on Engineering Management*, **EM-30** (1983) 140–155.
- [15] S. Schenkerman: Avoiding rank reversal in AHP decision-support models. *European Journal of Operational Research*, **74** (1994) 407–419.
- [16] L.G. Vargas: An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research*, **48** (1990) 2–8.
- [17] S.A. Webber, B. Aposutolou and J.M. Hassell: The sensitivity of the analytic hierarchy process to alternative scale and cue presentations. *European Journal of Operational Research*, **96** (1997) 351–362.
- [18] 荒木義修: 三重県民の選挙及び政治意識に関するパネル調査結果報告. 地域社会研究所報, **9** (1997) 41–61.
- [19] 刀根薫, 真鍋龍太郎: AHP 事例集 (日科技連, 1990).
- [20] 中西昌武, 木下栄蔵: AHP によるファジィ数量化理論 III 類の提案. 1997 年度日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, 1997, 228–229.

佐藤 祐司

〒 515-8511 三重県松阪市久保町 1846

松阪大学政治経済学部

e-mail: ysatoh@matsusaka-u.ac.jp

ABSTRACT

AN EVALUATION OF JUDGMENT SCALE
IN THE ANALYTIC HIERARCHY PROCESS

Yuuji Satoh
Matsusaka University

The Analytic Hierarchy Process (AHP) of Thomas L. Saaty is a popular method for decision making. In the AHP, data on decision maker's (DM's) pair-wise comparison are aggregated, and the degree of importance of each alternative is quantified. This results in not only the identification of the most important alternative but also the ranking of all alternatives for each DM. Various judgment scales have been proposed and put to use in the process of pair-wise comparison. Saaty advocated 1-9 linear scale and claimed that this scale is consistent to the laws of physics. This linear scale is invited some criticisms in the literature. The objections mainly stem from rank reversal and inconsistency. The objective of this paper is to find a scale appropriate for representing DM's perception. Specifically, we compare two scales: linear and power scales. The criteria of appropriateness are robustness with respect to scale change and consistency. Two types of data are used in the paper: (1) randomly generated data, and (2) data on political consciousness (survey carried out over Mie prefecture in February 1997). The results offer some evidence that power scale is preferable to linear scale as the judgment scale.