原 著

不完備型嗜好テストにおける四者択一方式の有効性

談 小 健*・高 野 泰*・岸 野 洋 久*

* 東京大学大学院農学生命科学研究科

Efficiency of the Method of Four-fold Choice in Incomplete Preference Trials

Xiaojian Tan*, Yasushi Takano*, and Hirohisa Kishino*

*Laboratory of Biometrics, Graduate School of Agricultural and Life Sciences,
The University of Tokyo, 1-1-1, Yayoi Bunkyo-ku, Tokyo 113-8657, Japan

Paired comparison is popular in preference trials. The estimated Scores have high accuracy, even if the data include measurement errors. However, incomplete comparison becomes inevitable, especially when the number of objects to be scored is large. In such cases, high precision by paired comparison is not guaranteed. In this paper, we propose a maximum likelihood estimation of scores, based on four-fold choice data. An empirical study of flower preference showed that the method of four-fold choice takes about the same time as a paired comparison, and the estimated scores were almost the same for the two procedures. It was clearly shown, by numerical simulation, that the precision of the estimated score of the most preferred object does not decrease with an increased number of objects to be compared in the method of four-fold choice. This is in great contrast to the paired comparison method. The estimated scores of less preferred objects have larger variances in the method of four-fold choice, whereas those from paired comparisons have similar variances for all objects. Thus, the four-fold choice method is effective when the scores of Most preferred objects are matters of concern.

Key words: Four-fold choice; Paired comparison; Preference trial; Thurstone's model; Maximum likelihood estimation.

1. はじめに

一般に、いくつかの比較対象に関して、人々の評価や嗜好の度合いを調査する方法には、まず、「比較したい全ての対象を被検者に示して、それらの評点（絶対評点）、あるいは、それらの間の順位（絶対順位）を付けてもらう方法」が考えられるが、この方法は明らかに、比較対象の数が多くなると、被検者に対する負担が大きくなり、適切な評価が困難になる。

心理学の分野では、「比較したい対象の中から順次2つの対象を提示して相対的な評価をして

もらい、それらの評価を組合せて全体の評点や順位を推定する方法（一対比較法）：例えば、田中 (1977)、佐藤 (1985) など」が用いられている。この方法の特徴を以下に挙げておく。

- a. 1つの質問は「2つの対象のどちらがより好みか?」を聞くもので、選択自体は簡単で、被検者の負担は少ない。
- b. 質問方法によっては、「嗜好にどの程度の差があるか?」すなわち、比較の度合いを評点で答えてもらうこともできる。心理学では、伝統的に“5段階”または“7段階”のスコア表示がよく使われる。
- c. N人の被検者に対する一対比較試験で得られた頻度データに Thurstone (1927) のモデルを適用して、嗜好度を表すパラメータを推定することができる。
- d. 対の組合せごとに評価の基準が変化する可能性があり、いわゆる“組合せ効果”の生じる恐れがある。これは、3個以上の比較対象の評価順位がループになって、矛盾を引き起こす現象として現れる。完備型の場合については、たとえば、佐藤 (1985) に解説がある。
- e. 比較対象が K 個あるとき、全ての対組合せの数は $\binom{K}{2} = \frac{1}{2}K(K-1)$ である。すなわち、
 K が大きいときに、一人の被検者に全ての組合せについて質問をすること（完備型の調査）は、被検者への負担が大きくなり、一定の条件での判定が期待できなくなる恐れがある。

この中で、bの質問方法に関して、Saaty (1977) は完備型の比較で、評価の高い方の相対評点を、同等（評点 = 1）も含めて9段階とし、他方の評点をその逆数で与える“比尺度”の比較行列を作成し、その第1固有ベクトルによって、被検者個人の評価尺度を的確に表現できることを示した。

また、cの推定方法に関して、Mosteller (1951) は最小2乗法に関しては、Thurstone (1927) の case V（等分散で「全ての対の相関がゼロ」という条件）でも「相関が等しい」という条件でも同等な解が得られることを示した。さらに、Gulliksen (1956) は、不完備データに対して最小2乗法による解法を与えた。

上記のeで指摘したように、一対比較試験では対象の数が多くなると、質問数はほぼその2乗に比例して多くなり、被検者に対する負担から完備型の調査は不可能になる。花卉などの園芸品種の販売動向調査や生物育種の選抜試験などにおいては、多くの人に好まれる対象については詳しい情報を必要とするが、皆にあまり好まれない対象についての情報はあまり必要としない場合が多い。しかし、一対比較試験では、一方を選択することと他方を逆方向に選択することとが同値であり、好まれる対象と好まれない対象とが同じ精度で推定される。これに対して、たとえば、三者択一や四者択一などの方式の質問で選ばれたものについては、結果的に「他のいくつかの者との比較で選択された」という事実を反映していることから、何回も選択される対象に関する情報が多く蓄積される方向に偏ってくる。したがって、「嗜好性の高い対象に興味が集中している場合には、一対比較よりこのような多者択一の方が効率が良くなる」ことが予想される。

本研究では、このような予想を検証するために、パンジーの花に対する嗜好テストにおいて、“一対比較（二者択一）”方式に対する“四者択一”方式の有効性を比較してみた。ここで四者

択一方式が選ばれた理由は、同時に比較する花の画像を1枚のスライドに割り付ける画面分割のし易さを考慮したことによる。

なお、Gilford (1954) および Torgerson (1958) には、1回の質問で3個の刺激の相互間距離（あるいは類似度）の大小に関して、完全な順序で回答する場合（例えば、3刺激 A, B および C に対して、 $AB < BC < AC$ など）で回答する場合）に、Thurstone のモデルを用いた最小2乗法を紹介している。この場合、基本的に、1つの刺激を軸にして、他の2刺激との距離の大小についての情報が得られる。例えば「Bを軸にして、BはCよりAに近い($AB < BC$)」あるいは「BはAよりCに近い($BC < AB$)」という事象のデータが得られることになる。Torgerson (1958) では、このような距離組合せの全てについて頻度データが得られる場合には、一般に4個以上の刺激に対しても推定可能であることを示している。しかし、この方法は、我々の調査で扱う四者択一のような不完全なデータに関しては適用できない。

2. 嗜好テストの方法

花卉の市場における販売戦略を決めるとき、あるいは、花卉の育種における花の色や形を選抜するときなど、数多くの花の中から「どのような花が多くの人に好まれるか」を知っておく必要があり、そのために消費者を対象にした嗜好テストが行われることがある。ここでは、そのような状況を想定して調査を行った。

まず、調査の準備として、パンジー（*Viola* 属のいくつかの園芸品種）の20種類の花のカラー・スライドから画像処理によって、それぞれ1個ずつの花を取り出した画像を作成した。中には草姿の乱れた種類もあったが、この実験では花に注目していることから、茎葉の状態による影響を取り除くことにした。そのために、花の咲いていないあるパンジーの茎葉部分の画像を全ての花の背景として、その上に各花の画像をスーパーインポーズしたものを用いた。全体のアンケート調査にはいくつかの調査が組み合わせられたものであったが、ここでは、そのうちの“一対比較”と“四者択一”の2種類の調査を比較する。

一対比較：1枚のスライドに、2種類の花の画像を図1aのように並べたものを60組作成し、これらのスライドをスクリーンに映写しながら、被検者に好みの強い方の花に対応する回答欄に丸印を付けてもらう。

四者択一：1枚のスライドに、4種類の花の画像を図1bのように並べたものを20組作成し、被検者に4つの花のうち“最も好ましい”1つの花に対応する欄に丸印を付けてもらう。

両方の試験で、それぞれ、花の画像をどのように組み合わせるかが問題になるところであるが、今回の調査では、

- ・各花とも同じ回数だけ比較に登場すること、
- ・同じ組合せのスライドが登場しないこと、
- ・特定のグループ内での比較に片寄らないこと、

などを考慮しながら、乱数を用いて、組合せやスライド内の配置を決定した。

調査に用いた花の画像の数は20個で、1~60問が一対比較形式の質問で、61~80問が四者択一形式の質問であった。したがって、1つの花の画像は、一対比較で6回ずつ、四者択いで4回ずつ登場することになる。

図 1. 質問用スライドの例：(a) 一対比較方式 (2 枚組) (b) 四者択一方式 (4 枚組).

各質問の回答時間については、予備調査において被検者にとって負担の少ない回答時間を考慮した結果、一対比較の質問では1問当たり約6秒、四者択一の質問では約8秒が適当であると判断した。

被検者については、東京大学、明治大学、および愛国学院大学の3大学の学生、合計191名を対象にした。

3. 解析方法およびシミュレーション

3-1. 心理学的尺度モデル

各花に対する好み度合いの尺度を想定するために、Thurstone (1927) の心理学的尺度 (psychological scale value: case V) を基礎におく。すなわち、ここでの花の画像を心理学的刺激と考えて、その集合を

$$S = \{s_1, s_2, \dots, s_K\}$$

とするとき、例えば、 n -番の被検者がある花の画像 s_i を見たときにそれを好ましいとして選択する度合い X_{in} は、被検者ごとに、さらにその被検者がその花に接するごとに独立に変異し、その分布は正規分布で、分散は全ての花で等しいものとする。つまり、この被検者がその花に j 目に接したときの好ましさの度合いを X_{inj} とすれば、

$$X_{inj} \sim N(\mu_i, \sigma^2)$$

である。このモデルのかわりに、 n -番目の被検者の花 s_i に対する評価 X_{in} が、 j に関係なく、その花に接するときはいつでも一定の値になるようなモデルを考えることもできるが、その場合には、全ての花に対するその被検者の評価順位は、 $\{X_{1n}, X_{2n}, \dots, X_{Kn}\}$ のそれぞれの値の大きさの順に従うので、ある個人の判断する順位に矛盾は生じないことになる。しかし、実際には、24名の被検者に対して行った予備調査の結果、1名を除く他の23名の評価順位に矛盾が生じていることが分かった。そのため、各被検者がそれぞれの花を比較するたびに独立な X_{inj} の値を仮定することにした。

このような調査で、いくつかの質問に対する個人の回答が自己矛盾を引き起こすメカニズムは複雑であろう。例えば、それが被検者の単なる誤認や回答ミスによるものか、嗜好性の定義のあいまいさによるものか、あるいは、嗜好に対する評価の多面性が背後に隠れているものか、など、心理学的に重要な問題点が含まれていることも考えられる。しかし、この論文では分析の可能性に重点をおいて、被検者の回答に合わせて「ある程度の矛盾を含むモデル」を採用しておくことにする。

3-2. 一対比較モデル

これは、2つの花の画像 s_i と s_j を比較する場合である。上のモデルに従えば、ある被検者がそれぞれの花から受ける影響度を、それぞれ X_i および X_j としたとき、それらは、それぞれ独立に、等しい分散 σ^2 を持つ正規分布、 $N(\mu_i, \sigma^2)$ および $N(\mu_j, \sigma^2)$ に従って分布し、 $X_i > X_j$ のとき、「 s_i の方が s_j より好ましい」と判断することになる。すなわち、 $P[\cdot]$ を確率関数として、

$$P[s_i \text{ を } s_j \text{ より好ましいと思う}] = P[X_i > X_j]$$

ただし, $X_i \sim N(\mu_i, \sigma^2), X_j \sim N(\mu_j, \sigma^2)$

と定式化することができる. また, この確率は,

$$\begin{aligned} P[X_i > X_j] &= P[X_i - X_j > 0] = \int_0^{\infty} \frac{1}{\sqrt{4\pi\sigma^2}} e^{-\frac{(x - (\mu_i - \mu_j))^2}{4\sigma^2}} dx \\ &= \int_{-(\mu_i - \mu_j)}^{\infty} \frac{1}{\sqrt{4\pi\sigma^2}} e^{-\frac{u^2}{4\sigma^2}} du = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}\sigma}\right) \end{aligned} \quad (1)$$

と表すことができるが, パラメータを推定するための条件として, $\sum_{i=1}^K \mu_i = 0$, および $\sigma^2 = 1$ とする.

K 個の花の中から 2 つの花をとる組合せの数は $\frac{1}{2}K(K-1)$ で, $K=20$ のときには 190 になる.

この全ての「対」の集合を D_A とし, M 個の要素を持つ D_A の部分集合 $D \subset D_A$ の各要素に対応する対組合せで, 調査が行われるものとする.

N 名の被検者に対して行われた調査の結果, $(i, j) \in D$ の組合せに対応する花の組合せ (s_i, s_j) の比較で s_i を選んだ被検者数を, f_{ij}, s_j を選んだ被検者数を f_{ji} とすれば, $f_{ij} + f_{ji} = N$ である.

一対比較試験では, 例えば, f_{ij}/N の相対頻度を累積分布関数の値として持つ正規分布スコア z_{ij} を求め, この値を $(\mu_i - \mu_j)$ で近似する最小 2 乗法がよく使われる. しかし, ここでは, 後の四者択一の場合を考慮して最尤推定法を用いる.

$\{f_{ij} | (i, j) \in D\}$ を用いた $\mu_1, \mu_2, \dots, \mu_K$ の尤度関数は次のようになる.

$$L = \prod_{(i, j) \in D} \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}}\right)^{f_{ij}} \cdot \left\{1 - \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}}\right)\right\}^{(N - f_{ij})} \quad (2)$$

3.3. 四者択一モデル

一方, 四者択一の場合には, 4 つ組 (s_i, s_j, s_k, s_l) の中でどの 1 つが選ばれるか, を次のモデルで表現する.

ここでも, $\sum_{i=1}^K \mu_i = 0$, および, $\sigma^2 = 1$ の条件のもとで,

$$\begin{aligned} &P[s_i, s_j, s_k, s_l \text{ の中で } s_i \text{ を最も好ましいと思う}] \\ &= P[4 \text{ 個の変量 } X_i, X_j, X_k, X_l \text{ の中で } X_i \text{ が最大になる}] \\ &= P[X_i > X_j, X_i > X_k, X_i > X_l] \\ &= \int_{-\infty}^{\infty} \phi(x - \mu_i) \Phi(x - \mu_j) \Phi(x - \mu_k) \Phi(x - \mu_l) dx \\ &\equiv \Psi(\mu_i, \mu_j, \mu_k, \mu_l) \end{aligned} \quad (3)$$

などである. ここで, $\phi(x)$ は標準正規分布の密度関数である. また, $\Phi(x)$ は標準正規分布の分

布関数で、たとえば、 $P[X_j < X_i | X_i = x] = \Phi(x - \mu_j)$ などとなることを用いている。

K 個の花の中から 4 個をとる組合せの数は、 $\binom{K}{4} = \frac{K(K-1)(K-2)(K-3)}{4 \cdot 3 \cdot 2 \cdot 1}$ で、 $K=20$ のときには 4845 となる。この全ての“4 つ組”の集合を Q_A とする。M 個の要素を持つ Q_A の部分集合 $Q \subset Q_A$ の各要素に対応する 4 つ組で調査が行われる。

N 名の被検者に対して行われた調査の結果、 $(i, j, k, l) \in Q$ の組合せに対応する花の組合せ (s_i, s_j, s_k, s_l) の比較で s_i を選んだ被検者数を、

$$f_{i,jkl} = [s_i, s_j, s_k, s_l \text{ の中で } s_i \text{ を選んだ被検者の数}]$$

また、 s_j を選んだ被検者数を $f_{j,ikl}$ などとすれば、 $f_{i,jkl} + f_{j,ikl} + f_{k,ijl} + f_{l,ijk} = N$ となる。

このような表記法を用いれば、四者択一調査の尤度関数は、

$$L = \prod_{(ijkl) \in Q} \{\Psi(\mu_i; \mu_j, \mu_k, \mu_l)\}^{f_{i,jkl}} \cdot \{\Psi(\mu_j; \mu_i, \mu_k, \mu_l)\}^{f_{j,ikl}} \cdot \{\Psi(\mu_k; \mu_i, \mu_j, \mu_l)\}^{f_{k,ijl}} \cdot \{\Psi(\mu_l; \mu_i, \mu_j, \mu_k)\}^{f_{l,ijk}} \quad (4)$$

で表される。

これらの条件で、N 名の被検者での実験データから最尤法で、各花に対する $\mu_1, \mu_2, \dots, \mu_K$ を推定し、また、Fisher の情報行列を用いて、推定値の標準誤差を推定した。一対比較および四者択一の両方とも、尤度関数が複雑で、推定量を観測データの明示的な関数として表すことは困難であったので、ここでは、探索的に最適値を求めるポリトープ法 (J. A. Nelder and R. Mead: 1965) を用いて、(2) および (4) の対数をとった対数尤度を数値的に最大化した。

3-4. シミュレーションの方法

実験で採用した条件も含めてさらに多くの条件のもとで、一対比較方式と四者択一方式の精度を比較するために、次の 2 種類のシミュレーションを行った。

シミュレーション-1：花の数を $K=20$ とし、一対比較の 20 問と 60 問、および四者択一の 20 問と 40 問のそれぞれについて精度を評価して比較した。

シミュレーション-2：質問数を 40 問に固定して、花の数を $K=5 \sim 40$ で変えたときの精度の変化を追跡した。

その他の設定内容を以下に示す。

- 1) K 個の花に対して、嗜好度および選択確率については前小節までのモデルに従う。ただし、各花に対する嗜好度の平均値は、 $\mu_1=2.5$ と $\mu_K=-2.5$ の間を等分割する値とした。すなわち、 $\mu_i (i=1, 2, \dots, K)$ は次の式で与えられる。

$$\mu_i = \frac{5}{2} - \frac{5}{(K-1)}(i-1)$$

- 2) 各問における被検者の選択数の期待値には、与えられた嗜好度の平均値を使って、一対比較では (1) 式、四者択一では (3) 式で計算される確率そのものを代入した。すなわち、このシミュレーションでは被検者 1 名に対する嗜好度推定量の標準誤差の期待値を計算することになる。
- 3) 各回の質問で比較する花の画像の組合せは以下のように設定した。ここでは、一対比較および四者択一の両方とも、20 問の質問を 1 セットにして、20 個の花をランダムに並べ、

表 1. 嗜好度および標準誤差の推定値. 20 花, 191 人, のそれぞれの調査データから推定したもので, 各データセットごとに嗜好度の大きさにソートしてある.

a. 対比較 60 問

順位	花	嗜好度	標準誤差
1	1	0.943	0.056
2	12	0.714	0.055
3	7	0.659	0.057
4	15	0.354	0.056
5	8	0.197	0.054
6	14	0.129	0.055
7	19	0.071	0.054
8	2	0.063	0.054
9	10	0.021	0.054
10	18	0.000	0.054
11	4	-0.035	0.054
12	20	-0.071	0.055
13	11	-0.100	0.054
14	9	-0.113	0.054
15	13	-0.185	0.055
16	16	-0.226	0.054
17	3	-0.259	0.055
18	17	-0.469	0.054
19	5	-0.749	0.055
20	6	-0.942	0.059

b. 四者択一 20 問

順位	花	嗜好度	標準誤差
1	1	1.067	0.057
2	12	0.799	0.058
3	7	0.693	0.057
4	15	0.365	0.060
5	19	0.256	0.060
6	8	0.157	0.062
7	4	0.101	0.060
8	2	0.138	0.064
9	10	-0.013	0.063
10	14	-0.038	0.065
11	20	-0.067	0.061
12	13	-0.134	0.064
13	16	-0.136	0.064
14	3	-0.269	0.072
15	11	-0.306	0.065
16	5	-0.307	0.068
17	18	-0.433	0.068
18	9	-0.434	0.068
19	17	-0.455	0.074
20	6	-0.982	0.095

相次ぐ花を重複させながら 2 枚組または 4 枚組の組合せを作っていた. そして, 質問数が 20 より多い場合には, 必要な回数だけこの操作を反復した.

- 4) 花の数と質問数についてのそれぞれの設定に対して, 100 回のシミュレーションを行い, それぞれのシミュレーションでフィッシャーの情報行列から標準誤差を計算し, 平均をとった.

4. 結果

4-1. 嗜好テストの結果

191 名の全員のデータを用いた嗜好度の最尤推定値, およびその値を Fisher の情報行列に代入して計算された標準誤差の推定値を表 1 に示す. このなかで, 表 1a は一対比較調査のデータから, 表 1b は四者択一調査のデータから得られた推定値である. この表の結果を図にしたものが図 2 で, 横軸は一対比較での推定値, 縦軸は四者択一調査の推定値でプロットしたものである. 推定された相関係数は $r=0.922$ と高い値であった. 図中には 45 度の直線が示されている.

また, 図 3 は, 推定された嗜好度に対して標準誤差の値をプロットしたものである. 一対比較の標準誤差がほぼ一定なのに対して, 四者択一では嗜好度の低い花ほど誤差が大きくなっている

図 2. 一対比較および四者択一で推定された嗜好度の関係：横軸（一対比較方式での推定値）と縦軸（四者択一方式での推定値）の位置にその花の画像をプロットしたもの。45 度の直線を示してある。相関係数は 0.922 と高い値であった。

ことが分かる。これらのことは当初からある程度予想されていたことであり、高い嗜好度得点の花に対する精度がどうなるかが問題であった。この調査では、そこでも一対比較の方がわずかに精度が良いことになったが、質問の数は一対比較の 60 問に対して、四者択一では 20 問と少ない場合であった。そこで、質問数および花の数を変えたときの精度を比較するために、いくつかのモデルに対してシミュレーションを行って推定精度を比較した。

4-2. シミュレーションの結果

いくつかの設定を変えたシミュレーションの結果のうち、精度の違いを理解しやすいものを選んで図 4 に示した。この図の横軸は各花に設定した嗜好度の期待値で、最も好まれるものから順に $\mu_1 > \mu_2 > \dots > \mu_K$ の値である。縦軸は一人当たりの嗜好度の推定量に関する標準誤差である。縦軸の高さを表 1 の値と比較するには、人数分を考慮して $\sqrt{191}$ で割ることの他に、嗜好度の広がり $[-1, 1]$ と $[-2.5, 2.5]$ で異なること、および小節 3-1 で述べた「自己矛盾」の頻度などが問

図3. 推定された嗜好度と標準誤差の関係：一対比較 60 問からの推定値は三角印，四者択一 20 問からの推定値は丸印で示してある．花の番号は表 1 と対応している．

題になる．

図 4 の中で，一対比較 60 問の場合と四者択一 20 問の場合を比較してみると，嗜好度の高いものではわずかに一対比較 60 問の方が精度が高く，一対比較の方は嗜好度が低いものでもあまり変化しないのに比べて，四者択一の方は嗜好度が低いものでは急速に精度が悪くなる傾向があることなど，図 3 の調査データとよく似た結果であることが分かる．一対比較 20 問の場合は極端に精度が悪いが，これは，1 個の刺激に対してわずかに 2 回の質問だけなので，たまたま比較された相手によってその刺激の評価が大きく変化することを反映している．

その他，いずれの調査も質問数が増えると精度は高くなるが，ある程度以上では，質問数の増加に比べて精度はあまり変化しなくなる傾向にある．また，嗜好度が上位のものでは，一対比較 60 問に対して四者択一 40 問の方が，質問数が少ないにもかかわらず，高精度になっている．一対比較では，嗜好度が中間のものが，最も精度が高くなっている．この傾向は，微かではあるが，図 3 でも見ることができる．

図 5 は，質問数を 40 問で一定にして花の数を変えたとき，嗜好度が最も高い花について，精度の変化を示したものである．花の数が 5 個のときにはほぼ等しい精度であったが，花の数が 40 個では一対比較の方が約 4 倍ほど誤差が大きくなっている．すなわち，比較する対象が多くなるに従って，一対比較の精度は急速に悪くなってしまうことが分かる．

5. 考察

以上の結果から，比較対象の数が多く，嗜好度に関して下位のものの精度を犠牲にすれば，一対比較に比べて四者択一の方が，同じ規模の調査で効率の良い推定値を得られることが示唆さ

図4. シミュレーション-1の結果：花の数は20で，一対比較は20問および60問，四者択一では20問および40問のそれぞれの設定でシミュレーションを行ったもので，標準誤差の期待値を縦軸に示してある．横軸は設定した嗜好度の値で， $\mu_1=2.5$, $\mu_{20}=-2.5$ の間を19等分する位置である．

図5. シミュレーション-2の結果：質問数を40問に固定して花の数を変えて実行したシミュレーションで，一対比較と四者択一の推定精度を標準誤差の期待値で比較したもの．

れた。

ここで、今回の花の嗜好に関する調査で分かった四者択一方式の特徴を、1 節での一対比較方式の場合と対応させて示す。

- a. 「4つの比較対象の中でどれが最も好ましいか?」を聞くもので、一対比較と比べてもそれほど難しくない。すなわち、1問当たりの所用時間は、約8秒で、一対比較の場合(約6秒)とあまり変わらない。全体で考えると、一対比較方式の60問と四者択一方式の45問がどちらも360秒で対応している。また、今回の実験のように、同じ被検者に2つの方式の調査を続けて行う限りでは、推定結果に高い相関がある(図2)。推定精度に関しては、嗜好度の上位のものは高精度で、下位のものは低精度になる(図4)。
- b. 嗜好の程度を評点として答えることは難しいものと思われる。4つの比較対象を順位づけることも同様に困難が予想される。これらは、可能であるとしても、被検者に対する負担が大きくなり、所用時間が変化するので、結局は効率が悪くなることが予想される。
- c. Thurstone (1927) は一対比較を想定してモデルを作成したのであるが、尤度関数を書き下すことで、小節3-3で示したように、最尤法を用いて推定することができる。ただし、計算式は複雑になる。
- d. 組合せ効果の現れ方については微妙な問題が含まれる。すなわち、同じ2つの比較対象を含んでいる質問があれば、この2つの質問に対する選択が矛盾する可能性もある。しかし、一対比較のときとは異なり、2つの同じ比較対象をできるだけ含まないような組合せを考えることができるので、その場合には2個の選択で矛盾することは少なくなる。ただし、その場合でも、3個以上の質問で矛盾を引き起こす可能性があることは一対比較の場合と同様である。そして、このような自己矛盾については、小節3-1で述べたように、その原因に嗜好の多面性なども考えられるので、それ自体が分析対象になることもある。
- e. 全ての組合せの数は、 $K=10$ で210、 $K=20$ では4845となり、一対比較の場合よりはるかに多い。しかし、たとえば、 $K=20$ の場合には同じ3個組合せが、もう1個の相手を変えながら17回出てくることになる。それを1回ずつにするような組合せがもしあれば、 $4845/17=285$ となる。このように、四者択一では組合せを選択することが重要な意味を持つてくる。

なお、この調査では、4個の花の画像を組み合わせで1枚のスライドにして提示することから必然的に四者択一になったが、食味試験などでは三者択一、あるいは五者択一などの方式を同様に採用することもできる。これらの場合でも、精度に関しては今回の四者択一と同様の結果になることが予想される。しかし、選択肢が多くなるに従って被検者の負担は累乗的に増加するので、実験の種類ごとに適当な規模を選択する必要がある。

比較する組合せの選択は推定精度に大きく影響する。ここでのシミュレーションで作成した組合せ方式は、20個の花をランダムに並べて、一対比較のときは2個ずつ、四者択一のときは4個ずつを1花ずつ巡回的にずらしながら20組を作成する方法であった。この方法は、各花の出現回数ができるだけ公平になるように、直感的に工夫したものであったが、結果的には特に良い組合せというわけではなかった。この組合せ方法をさらに改良して実験を設定することも考えられるが、それは今後の検討にゆだねることにした。

ここでは、比較対象の数、質問数を変えて、1問当たり被検者の負担を考慮しながら一対比較と四者択一の推定精度について検討したが、一般に、比較調査の方法は他の環境によって決め

られることもある。例えば、食味試験などでは、味覚や臭覚の特に優れた専門のパネリストを起用することもあり、このような場合には、比較対象の数を多くしたり、比較のスコアを答えてもらうことなど、1回の回答当たりの情報量を多くすることができる。しかし、一般消費者を対象にした嗜好テストなどでは、被検者もある集団から無作為にとられることが多く、反射的あるいは直感的な反応を数多く調査することが公平な推定値を得るために必要になる。このような場合に一対比較も含めた択一型の質問が採用されることが多い。

そして、比較対象の数が多く、質問数が限られているときに、四者択一を代表とする多数中の択一方式が有効である。

謝辞

文部省統計数理研究所の吉野諒三氏には、度重ねて文献を紹介していただきましたことに感謝しております。また、本稿の審査員A氏およびB氏には、それぞれ、詳細にわたる適切なコメントをいただきました。あわせてここに謝意を表します。

引用文献

- Guilford, J. P. (1954). *Psychometric Methods*. 2nd ed. McGraw-Hill, New York.
- Gulliksen, H. (1956). A least squares solution for paired comparisons with incomplete data. *Psychometrika*, **21**, 125–134.
- Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, **16**, 3–9.
- Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–313.
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, **15**, 234–281.
- 佐藤 信 (1985). 統計的官能検査法, 日科技連出版社.
- 田中良久 (1977). 心理学的測定法. 第2版, 東京大学出版会.
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, **34**, 273–286.
- Torgerson, W. S. (1958). *Theory and Methods of Scaling*. J. Wiley, New York.