

『岡山商大論叢』（岡山商科大学）

第45巻第1号 2009年6月

Journal of OKAYAMA SHOKA UNIVERSITY

Vol.45 No.1 June 2009

《論 説》

母集団多重寸法指標の ノンパラメトリック推定

佐井至道

Nonparametric Estimation of Population Multi Size Indices

Shido Sai

1 はじめに

個票データが標本調査で得られた場合、その公開のリスク評価のために標本寸法指標を基にして推定された母集団寸法指標を用いることが多い。Bethlehem et al. (1990) がポアソンガンマモデルを提案して以来、様々な超母集団モデル、あるいは確率分割モデルと呼ばれる推定のためのモデルが提案されてきており、その中で、Pitman (1995) が提案したピットマンモデル (Ewens-Pitman sampling formula (EPSF) と呼ばれる) の実データへの当てはまりの良いことが報告されている。一方、佐井 (2002, 2003) は、超母集団モデルを用いない制約付きノンパラメトリック最尤推定法を提案し、佐井 (2006) では、母集団寸法指標に対する制約条件をペナルティー関数で置き換えるなど、計算時間上の問題を克服するいくつか

の方法について提案を行った。

官庁統計の標本調査では、例えば労働力調査が毎月実施されるように、継続調査として行われる場合がほとんどである。継続調査では、パネル調査のように、同じ個体が数回に渡って連続してサンプリングされる場合と、原則的に毎回異なる個体がサンプリングされる場合がある。

同じ個体が連続してサンプリングされる場合、同一個体のレコードをリンクして1つのレコードに加工した個票データを作成すると、その公開のリスク評価としては上記の通常の方法を用いることができる。一方で、調査ごとに個票データを作成する場合、これまでは個々の個票データに対して独立にリスク評価を行うのが一般的であった。

本論文では、継続調査で得られた複数の個票データに対する同時リスク評価を行うために、寸法指標の概念を多重寸法指標に拡張し、標本多重寸法指標からの母集団多重寸法指標のノンパラメトリック推定法を提案する。

まず2.1節では寸法指標の概念と、標本寸法指標に基づく母集団寸法指標のノンパラメトリック推定法について簡単に紹介する。2.2節では、寸法指標の概念を多重寸法指標に拡張するとともに、標本多重寸法指標を基にした母集団多重寸法指標のノンパラメトリック推定法を提案する。

3.1節では、アメリカにおいて1990年と2000年に実施されたセンサスの1%抽出個票データを題材に、母集団多重寸法指標に対する制約条件と、その条件をペナルティー関数として対数尤度関数へ取り込む方法について提案する。3.2節では、種々の条件の下での推定結果を示すが、その改善のために3.3節では2段階のステップで推定を行う方法を提案する。

2 母集団多重寸法指標のノンパラメトリック推定

2.1 標本寸法指標と母集団寸法指標

この節では、寸法指標の定義と、標本調査で得られた個票データを公開する際のリスク評価について、その概略を述べる。

母集団の N 個の個体が、個体を特定するために用いられる数種類のキー変数の値の組み合わせに基づいて K 個のセルに分けられているものとして、第 j 番目のセルのサイズ（含まれる個体数）を F_j ($j = 1, 2, \dots, K$) とする。ここで、サイズ l のセル数、すなわち $F_j = l$ となるセル数を S_l ($l = 0, 1, 2, \dots, L$) とし、母集団寸法指標と呼ぶ。 L はセルのサイズの最大値である。

次に標本（個票データ）の大きさを n として、抽出率を $\lambda = n/N$ と表す。標本では F_j , S_l の代わりに f_j , s_l ($l = 0, 1, 2, \dots, L$) という表記を用いるが、後者が標本寸法指標である。

標本調査で個票データが得られている場合、計算された標本寸法指標を基にして母集団寸法指標を推定するのが、個票データのリスク評価としてしばしば用いられる方法である。母集団寸法指標の推定値 \hat{S}_l が求めれば、例えば最も識別されやすい個体数である母集団でも標本でもサイズ 1 のセル数の推定値として $\lambda \hat{S}_1$ を用いることができる。

推定にはピットマンモデルなどの超母集団モデル、あるいは確率分割モデルと呼ばれるモデルが用いられることが多いが、佐井（2002, 2003）では、標本寸法指標に基づく母集団寸法指標の制約付きノンパラメトリック推定法を提案した。ここではその要点のみを簡潔に述べる。

非復元単純無作為抽出によって得られた標本から標本寸法指標 (s_1, s_2, \dots, s_L) が計算されたときに、尤度を最大にするような非負の母集団寸法指標 (S_1, S_2, \dots, S_L) を求めるが、その尤度関数は抽出率が十分小さいときに

$$\begin{aligned} L_{\text{ap}}(S_1, S_2, \dots, S_L \mid s_1, s_2, \dots, s_L) \\ = \frac{1}{N C_n \lambda^n (1-\lambda)^{N-n}} \cdot \prod_{l'=1}^L \frac{e^{-\mu_{l'}} \mu_{l'}^{s_{l'}}}{s_{l'}!} \end{aligned} \quad (1)$$

と近似できる。ただし、

$$\mu_{l'} = \sum_{l=1}^L S_l \cdot {}_l C_{l'} \lambda^{l'} (1 - \lambda)^{l-l'} \quad (2)$$

であり、尤度関数の添字の ap は近似を意味する。

しかし、(1)式を最大にする母集団寸法指標は大きく増減を繰り返し、このままでは実用性に乏しい。そこで、佐井 (2003) では、推定の際に母集団寸法指標にいくつかの制約を置く提案を行い、実データを基に比較を行った。そのうち、下記の4つの制約を課す場合の結果が最も良かった。

$$(a) S_l \geq 0 \quad (l = 1, 2, \dots, L)$$

$$(b) \sum_{l=1}^L l \cdot S_l = N$$

$$(c) S_l \geq S_{l+1} \quad (l = 1, 2, \dots, L - 1)$$

$$(d) 2 \cdot \log S_l \leq \log S_{l-1} + \log S_{l+1} \quad (l = 2, 3, \dots, L - 1)$$

また、標本寸法指標を生成する可能性のあるすべての母集団寸法指標 (S_1, S_2, \dots, S_L) について網羅的に尤度を計算することは計算時間上困難であるため、最大尤度の数種類の探索法について比較を行った。

更に、佐井 (2006) では母集団寸法指標に対する制約条件をペナルティー関数で置き換えるなど、計算時間上の問題を克服する方法について提案を行った。ここではその詳細については省略し、ペナルティー関数については3.1節で紹介する。

2.2 標本多重寸法指標と母集団多重寸法指標

この節では、前節で述べた寸法指標の概念を多重寸法指標に拡張する。多重寸法指標については、渋谷、佐井 (2007)、Sibuya and Sai (2008) が紹介してモデルの適用を検討しており、佐井 (2007, 2008) では馬場、坂口 (2006) の紹介した多重寸法指標の特殊なケースである学会所属データに対して検討を行っている。

M 個の時点において同じ調査項目についての調査が行われたとして、第 i 時点における母集団を D_i 、個票データが得られる標本を d_i と表す。

また、第 i 時点における母集団の個体数を N_i 、標本の個体数を n_i 、抽出率を $\lambda_i = n_i/N_i$ とする。

ここで、各時点の母集団において前節と同様にセルを構成し、キー変数の同じ組み合わせのセルについて、各時点において含まれる個体数を考える。 (D_1, D_2, \dots, D_M) においてサイズの組 (l_1, l_2, \dots, l_M) となるセル数を $S_{(l_1, l_2, \dots, l_M)}$ と表し、母集団多重寸法指標と呼ぶ。標本でも同様に、 (d_1, d_2, \dots, d_M) においてサイズの組 $(l'_1, l'_2, \dots, l'_M)$ となるセル数を $s_{(l'_1, l'_2, \dots, l'_M)}$ と表し、標本多重寸法指標と呼ぶ。 L_i を第 i 時点でのサイズの最大値とすると $l_i, l'_i = 0, 1, 2, \dots, L_i$ として、ここではサイズとして 0 を含めて議論する。なお、 $s_{(0,0,\dots,0)}$ を観測することは一般に困難であるが、 $S_{(0,0,\dots,0)}$ は推定対象にしないことにする。

各時点における $(0, 0, \dots, 0)$ を除くすべてのサイズの組み合わせに対する母集団多重寸法指標の組と標本多重寸法指標の組を、それぞれ

$$S = (S_{(0,0,\dots,0,1)}, \dots, S_{(0,0,\dots,0,L_M)}, \dots, S_{(L_1, L_2, \dots, L_{M-1}, 0)}, \dots, S_{(L_1, L_2, \dots, L_{M-1}, L_M)}), \quad (3)$$

$$s = (s_{(0,0,\dots,0,1)}, \dots, s_{(0,0,\dots,0,L_M)}, \dots, s_{(L_1, L_2, \dots, L_{M-1}, 0)}, \dots, s_{(L_1, L_2, \dots, L_{M-1}, L_M)}) \quad (4)$$

と表す。

各時点において独立に非復元単純無作為抽出された標本から s が得られたときの S の尤度関数は

$$L(S | s) = \frac{1}{\prod_{i=1}^M N_i C_{N_i}} \sum_{C_1} \prod_{(l_1, l_2, \dots, l_M)} \left\{ \frac{S_{(l_1, l_2, \dots, l_M)}!}{\prod_{(l'_1, l'_2, \dots, l'_M)} k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}!} \cdot \prod_{(l'_1, l'_2, \dots, l'_M)} \left(\prod_{i=1}^M l_i C_{l'_i} \right)^{k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}} \right\} \quad (5)$$

と表される。ただし、 $k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}$ は母集団においてサイズの組

(l_1, l_2, \dots, l_M) のセルのうち、標本においてサイズの組 $(l'_1, l'_2, \dots, l'_M)$ と観測される数を表し、 C_1 は S から s が生成されるような $k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}$ の組についてのすべての組み合わせを表す。なお、(5)式の積において、サイズの組み合わせ (l_1, l_2, \dots, l_M) は $(0, 0, \dots, 0)$ を含まず、サイズの組み合わせ $(l'_1, l'_2, \dots, l'_M)$ は $(0, 0, \dots, 0)$ を含む。

時点の数 M 、母集団 D_i の大きさ N_i と最大サイズ L_i が増加するにつれ C_1 の組み合わせの数は膨大になり、時間的に(5)式の計算は困難となるため、佐井 (2002, 2003) と同様の手順によりポアソン分布の確率関数の積で近似を行う。

まず、各時点において独立に、しかも母集団 D_i ($i = 1, 2, \dots, M$) から各個体を他の個体とは独立に確率 λ_i でサンプリングするベルヌーイ抽出を考える。ベルヌーイ抽出では大きさ n_i の標本が得られていても、母集団の大きさは n_i 以上のすべての可能性を考えなければならないが、観測された標本の大きさが n_i のときに母集団の大きさを N_i に限定すると、(5)式に対応する尤度関数は

$$\begin{aligned}
 & L_B(S | s) \\
 &= \sum_{C_1} \prod_{(l_1, l_2, \dots, l_M)} \left\{ \frac{S_{(l_1, l_2, \dots, l_M)}!}{\prod_{(l'_1, l'_2, \dots, l'_M)} k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}!} \right. \\
 & \quad \cdot \left. \prod_{(l'_1, l'_2, \dots, l'_M)} \left\{ \prod_{i=1}^M C_{l'_i} \lambda_i^{l'_i} (1 - \lambda_i)^{L_i - l'_i} \right\}^{k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}} \right\} \\
 &= \prod_{i=1}^M \lambda_i^{n_i} (1 - \lambda_i)^{N_i - n_i} \\
 & \quad \cdot \sum_{C_1} \prod_{(l_1, l_2, \dots, l_M)} \left\{ \frac{S_{(l_1, l_2, \dots, l_M)}!}{\prod_{(l'_1, l'_2, \dots, l'_M)} k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}!} \right\}
 \end{aligned}$$

$$\begin{aligned}
 & \cdot \prod_{(l'_1, l'_2, \dots, l'_M)} \left\{ \prod_{i=1}^M {}_i C_{l'_i} \right\}^{k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}} \\
 & = \prod_{i=1}^M {}_{N_i} C_{n_i} \lambda_i^{n_i} (1 - \lambda_i)^{N_i - n_i} \cdot L(S | s) \tag{6}
 \end{aligned}$$

と書ける。尤度関数の添字の B はベルヌーイ抽出を意味する。母集団多重寸法指標によらず、ベルヌーイ抽出の尤度関数は非復元単純無作為抽出の尤度関数の定数倍となる。したがって、各時点において非復元単純無作為抽出された標本に対して最大尤度をとる母集団多重寸法指標は、ベルヌーイ抽出を想定した場合に、第 i 時点での母集団の大きさが N_i ($i = 1, 2, \dots, M$) となる母集団多重寸法指標の中で最大尤度をとるものに一致する。

ベルヌーイ抽出では母集団におけるサイズの組 (l_1, l_2, \dots, l_M) についてのすべての $k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}$ の組 $k_{(l_1, l_2, \dots, l_M)}$ が、他のサイズの組とは独立に多項分布に従い、その確率関数は

$$\begin{aligned}
 & f(k_{(l_1, l_2, \dots, l_M)}) \\
 & = \frac{S_{(l_1, l_2, \dots, l_M)}!}{\prod_{(l'_1, l'_2, \dots, l'_M)} k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}!} \\
 & \cdot \prod_{(l'_1, l'_2, \dots, l'_M)} \left\{ \prod_{i=1}^M {}_i C_{l'_i} \lambda_i^{l'_i} (1 - \lambda_i)^{l_i - l'_i} \right\}^{k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}} \tag{7}
 \end{aligned}$$

である。積において、サイズの組み合わせ $(l'_1, l'_2, \dots, l'_M)$ は $(0, 0, \dots, 0)$ を含む。

ここで、抽出率 λ_i ($i = 1, 2, \dots, M$) が十分小さいと仮定する。このとき、 $k_{(l_1, l_2, \dots, l_M)}$ のうち $k_{(l_1, l_2, \dots, l_M), (0, 0, \dots, 0)}$ を除くと、その期待値は小さく、互いの共分散も小さい。そこで $k_{(l_1, l_2, \dots, l_M), (l'_1, l'_2, \dots, l'_M)}$ ($(l'_1, l'_2, \dots, l'_M) \neq (0, 0, \dots, 0)$) を、それぞれ独立なポアソン分布の確率関数で

$$\begin{aligned}
& f\left(k_{(l_1, l_2, \dots, l_M)}, (l'_1, l'_2, \dots, l'_M)\right) \\
&= e^{-\left\{S_{(l_1, l_2, \dots, l_M)} \cdot \prod_{i=1}^M C_{l'_i} \lambda_i^{l'_i} (1 - \lambda_i)^{l_i - l'_i}\right\}} \\
&\quad \cdot \left\{S_{(l_1, l_2, \dots, l_M)} \cdot \prod_{i=1}^M C_{l'_i} \lambda_i^{l'_i} (1 - \lambda_i)^{l_i - l'_i}\right\}^{k_{(l_1, l_2, \dots, l_M)}, (l'_1, l'_2, \dots, l'_M)} \\
&\quad / k_{(l_1, l_2, \dots, l_M)}, (l'_1, l'_2, \dots, l'_M)! \tag{8}
\end{aligned}$$

と近似すると、ポアソン分布の再生性から $k_{(l_1, l_2, \dots, l_M)}, (l'_1, l'_2, \dots, l'_M)$ の (l_1, l_2, \dots, l_M) に関する和である標本多重寸法指標 $s_{(l'_1, l'_2, \dots, l'_M)}$ も、他の多重寸法指標とは独立に次のようにポアソン分布に従う。

$$f\left(s_{(l'_1, l'_2, \dots, l'_M)}\right) = \frac{e^{-\mu_{(l'_1, l'_2, \dots, l'_M)}} \mu_{(l'_1, l'_2, \dots, l'_M)}^{s_{(l'_1, l'_2, \dots, l'_M)}}}{s_{(l'_1, l'_2, \dots, l'_M)}!} \tag{9}$$

ただし

$$\begin{aligned}
& \mu_{(l'_1, l'_2, \dots, l'_M)} \\
&= \sum_{(l_1, l_2, \dots, l_M) \geq (l'_1, l'_2, \dots, l'_M)} \left\{ S_{(l_1, l_2, \dots, l_M)} \cdot \prod_{i=1}^M C_{l'_i} \lambda_i^{l'_i} (1 - \lambda_i)^{l_i - l'_i} \right\} \tag{10}
\end{aligned}$$

である。したがって尤度関数(5)式は

$$\begin{aligned}
& L_{\text{ap}}(S | s) \\
&= \frac{1}{\prod_{i=1}^M N_i C_{n_i} \lambda_i^{n_i} (1 - \lambda_i)^{N_i - n_i}} \\
&\quad \cdot \prod_{(l'_1, l'_2, \dots, l'_M)} \frac{e^{-\mu_{(l'_1, l'_2, \dots, l'_M)}} \mu_{(l'_1, l'_2, \dots, l'_M)}^{s_{(l'_1, l'_2, \dots, l'_M)}}}{s_{(l'_1, l'_2, \dots, l'_M)}!} \tag{11}
\end{aligned}$$

と近似される。積においてサイズの組み合わせ $(l'_1, l'_2, \dots, l'_M)$ は $(0, 0, \dots, 0)$ を含まない。

3 アメリカのセンサスデータに対する検討

3.1 母集団多重寸法指標に対する制約条件とペナルティー関数の導入

この節では、アメリカにおいて1990年と2000年に実施されたセンサスの1%抽出個票データ (U. S. Census Bureau (1993, 2003)) を用いて数値実験を行う。ワシントン州に在住している20歳以上の就業者の個人レコードを用いるが、個票データに含まれる項目のうち2回の調査において共通の10項目 (年齢, 実子の有無, 血縁の子の有無, 性別, 結婚, 通勤手段, 通勤時間, 職業, 労働週数, 週労働時間) をキー変数として選ぶ。なお、このデータには既に秘匿措置が施されており、ここで行う実験は提案する推定方法の評価を目的とするもので、この個票データのリスク評価を目的とするものではない。

これらの個票データは標本調査で得られていると見なすことが可能で、計算された寸法指標を基にして母集団の寸法指標を推定するべきであるが、実際の母集団寸法指標が未知であるため推定結果を評価することができない。そこで数値実験では、これらの個票データを母集団と見なし、1990年のデータを D_1 、2000年のデータを D_2 とする。個体数はそれぞれ $N_1 = 24846$ 、 $N_2 = 30234$ である。

表1に母集団多重寸法指標 S を示す。例えば $S_{(0,1)} = 21971$ 、 $S_{(1,0)} = 18154$ 、 $S_{(1,1)} = 1008$ のように、サイズ l_1 を縦に、 l_2 を横にとっている。サイズの最大値は1990年が19、2000年が17であるが、最大サイズを統一するため $L_1 = L_2 = 19$ とする。

母集団と見なした個票データ D_i ($i = 1, 2$) から、それぞれ抽出率 $\lambda_i = 1/2$ で標本 d_i を非復元無作為抽出した。 $n_1 = 12423$ 、 $n_2 = 15117$ である。この標本から求められた標本多重寸法指標 s を表2に示す。

表1、2において、行和と列和はそれぞれ1990年と2000年の寸法指標を表すが、1回分のデータのみからはサイズ0の頻度は観測できないことに注意を要する。

標本多重寸法指標を基に、(11)式を最大とする母集団多重寸法指標の近似

- (A) $S_{(l_1, l_2)} \geq 0$ ($l_1 = 0, 1, \dots, L_1, l_2 = 0, 1, \dots, L_2$)
- (B) $\sum_{(l_1, l_2)} l_1 \cdot S_{(l_1, l_2)} + \sum_{(l_1, l_2)} l_2 \cdot S_{(l_1, l_2)} = N_1 + N_2$
- (C) $S_{(l_1, l_2)} \leq S_{(l_1-1, l_2)}$ ($l_1 = 1, 2, \dots, L_1, l_2 = 0, 1, \dots, L_2$),
 $S_{(l_1, l_2)} \leq S_{(l_1, l_2-1)}$ ($l_1 = 0, 1, \dots, L_1, l_2 = 1, 2, \dots, L_2$)
- (D) $2 \cdot \log S_{(l_1, l_2)} \leq \log S_{(l_1-1, l_2)} + \log S_{(l_1+1, l_2)}$ ($l_1 = 1, 2, \dots, L_1 - 1,$
 $l_2 = 0, 1, \dots, L_2$),
 $2 \cdot \log S_{(l_1, l_2)} \leq \log S_{(l_1, l_2-1)} + \log S_{(l_1, l_2+1)}$ ($l_1 = 0, 1, \dots, L_1,$
 $l_2 = 1, 2, \dots, L_2 - 1$)

なお、すべての制約条件で $S_{(0,0)}$ を除いて考える。また、制約条件(D)については $S_{(l_1, l_2)} \geq 2$ についてのみ適用する。

ここで佐井 (2006) に倣って、渋谷 (2005) が提案した次のような滑らかなペナルティー関数を導入する。

$$\text{Pnlt}(x; \varepsilon) = \varepsilon \log(e^{\frac{x}{\varepsilon}} + 1). \quad (12)$$

ただし、 ε はパラメータである。(12)式の微分は

$$\frac{d\text{Pnlt}(x; \varepsilon)}{dx} = \frac{e^{\frac{x}{\varepsilon}}}{e^{\frac{x}{\varepsilon}} + 1} \quad (13)$$

となり、ロジスティック分布の分布関数を横軸方向に ε 倍だけ拡大したものととなる。

図1に $\varepsilon=0.01, 0.001$ としたペナルティー関数 $y = \text{Pnlt}(x; \varepsilon)$ のグラフを示す。 $\varepsilon \rightarrow 0$ のとき、グラフは $y = 0$ ($x < 0$), $y = x$ ($x \geq 0$) に近づくため、 ε は原点付近での滑らかさを表すパラメータと考えることができる。

(B)を除く制約条件をペナルティー関数で表し、次のように対数尤度関数へ取り込む。

$$\begin{aligned}
& F(S | s) \\
&= \log L_{\text{ap}}(S | s) - c_1 \sum_{(l_1, l_2)} \text{Pnlt}(-S_{(l_1, l_2)}; \varepsilon_1) \\
&\quad - c_2 \sum_{(l_1, l_2)} \text{Pnlt}(-(S_{(l_1-1, l_2)} - S_{(l_1, l_2)}); \varepsilon_2) \\
&\quad - c_2 \sum_{(l_1, l_2)} \text{Pnlt}(-(S_{(l_1, l_2-1)} - S_{(l_1, l_2)}); \varepsilon_2) \\
&\quad - c_3 \sum_{(l_1, l_2)} \text{Pnlt}(-(\log S_{(l_1-1, l_2)} + \log S_{(l_1+1, l_2)} - 2 \cdot \log S_{(l_1, l_2)}); \varepsilon_3) \\
&\quad - c_3 \sum_{(l_1, l_2)} \text{Pnlt}(-(\log S_{(l_1, l_2-1)} + \log S_{(l_1, l_2+1)} - 2 \cdot \log S_{(l_1, l_2)}); \varepsilon_3). \quad (14)
\end{aligned}$$

(A), (C), (D)の制約条件に関するペナルティ関数において, ε_1 , ε_2 , ε_3 はそれぞれの滑らかさを表すパラメータであり, c_1 , c_2 , c_3 はそれぞれのペナルティ関数の関数 F への影響力を表す係数である。

関数 F を $S_{(0,0)}$ を除く各 $S_{(l_1, l_2)}$ ($l_1 = 0, 1, \dots, L_1$; $l_2 = 0, 1, \dots, L_2$) で偏微分し, 0 と置いた連立方程式を解くことによって関数 F の最大値を得ることが最善であるが, 数的に解くことは困難である。

そこで適当な初期値から, ステップごとに, F が増加する方向に各

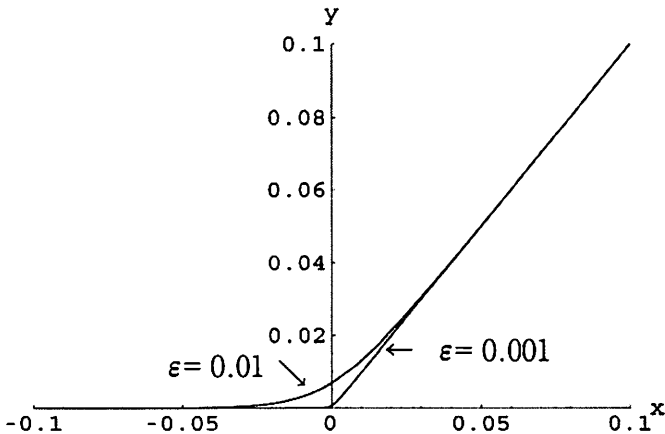


図1：滑らかなペナルティ関数

$S_{(l_1, l_2)}$ を独立に移動させる勾配法によって最大値を探索する。

$\hat{S}_{(l_1, l_2)(p)}$ を p 番目のステップにおける母集団多重寸法指標の推定値とするとき、

$$\hat{S}_{(l_1, l_2)(p+1)} = \hat{S}_{(l_1, l_2)(p)} + t_{(l_1, l_2)(p)} \cdot g \left(\frac{\partial F}{\partial \hat{S}_{(l_1, l_2)(p)}} \right) \quad (15)$$

によって、 $p+1$ 番目のステップの推定値 $\hat{S}_{(l_1, l_2)(p+1)}$ を求める。ただし、 $g(\cdot)$ は微係数を調整する関数である。また $t_{(l_1, l_2)(p)}$ は移動する距離を調整する係数で、 (l_1, l_2) と $\hat{S}_{(l_1, l_2)(p)}$ などを基に各ステップで変化させる。

なお、このようにして母集団寸法指標の推定値を変化させた場合、一般に $p+1$ 番目の推定値は制約条件(B)が表す超平面 $\sum_{(l_1, l_2)} l_1 \cdot S_{(l_1, l_2)} + \sum_{(l_1, l_2)} l_2 \cdot S_{(l_1, l_2)} = N_1 + N_2$ 上では得られないが、推定値と原点とを通る直線とこの超平面との交点へ移動させる補正をステップごとに行う。母集団の大きさに関する制約条件としては、 $\sum_{(l_1, l_2)} l_1 \cdot S_{(l_1, l_2)} = N_1$, $\sum_{(l_1, l_2)} l_2 \cdot S_{(l_1, l_2)} = N_2$ という2つの条件を用いるべきであるが、上記のような補正が容易ではないため、ここではやや緩い制約条件としている。

3.2 母集団多重寸法指標の推定

この節では、母集団多重寸法指標の推定を種々の方法で行い、それらの結果の比較を行う。

数値実験では、原則として(14)式で、 $c_1=1.0$, $c_3=0.0$, $\varepsilon_1=0.001$, $\varepsilon_2=\varepsilon_3=0.01$ とし、 c_2 については 0, 0.1, 1.0, 10.0, 100.0, 1000.0 の6通りを用いる。これに加えて、 $c_2=1000.0$, $c_3=10.0$ の場合も考える。また(15)式において

$$g(x) = \text{sign}(x) \cdot |x|^{0.5}, \quad (16)$$

$$t_{(l_1, l_2)(p)} = b(p) \cdot \hat{S}_{(l_1, l_2)(p)} / (l_1 + l_2)^{0.5} \quad (17)$$

として、 p 番目のステップにおける母集団多重寸法指標の推定値が大きいほど大きく変化させ、サイズの和 $l_1 + l_2$ が大きいほど小さく変化させる。また、 $b(p)$ は 10^{-4} を初期値として、3 ステップ連続して F が増加しなかった場合には 0.9 倍し、 F が増加した場合に 1.1 倍し、 10^{-7} を下回った時点で探索を終了する。

結果を表 3～9 に示す。なお、推定値は整数に四捨五入している。

表 3 は制約条件として (A)、(B) のみを課した場合の結果である。表 4～8 では、制約条件 (C) の影響を徐々に強くした推定結果を示しており、表 9 では制約条件 (D) も追加している。表 4 ($c_2=0.1$)、表 5 ($c_2=1.0$) の推定結果が実際の母集団多重寸法指標に近いが、 c_2 の値によって結果は大きく異なっている。

3.3 推定方法の改善

残念ながら多重寸法指標の推定に関する経験はそれほど多いとは言えず、適切な制約条件、係数、パラメータを選択できる保証はない。これに対して、寸法指標の推定に関しては経験的知識の蓄積があり、ある程度安

表 3：母集団多重寸法指標の推定値 ($c_2=0.0$, $c_3=0.0$)

		2000年																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計	
1 9 9 0 年	0		22118	761	340	1	0	8	28	0	0	0	0	0	0	0	0	0	0	0	0	0	23255
	1	18145	1031	437	1	15	71	0	0	0	16	0	0	0	0	0	0	0	0	2	1	0	19718
	2	845	0	0	0	0	23	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	899
	3	89	289	174	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	801
	4	2	0	0	0	0	101	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	124
	5	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29
	6	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92
	7	1	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33
	8	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24
	9	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	8
	10	0	0	0	0	0	0	23	0	0	0	13	0	0	0	0	0	0	0	0	0	0	36
	11	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
計	18915	23435	1396	374	17	224	100	30	0	16	69	0	0	0	0	0	0	0	2	1	0		

定した推定結果を得ることが可能である。母集団多重寸法指標のサイズ0を除く行和と列和は、それぞれ1990年と2000年の母集団寸法指標であるが、表3～9で推定されたこれらの値も c_2 の値によって大きく異なっている。

そこでこの節では、推定の改善のために、まず各年の母集団寸法指標を

表4：母集団多重寸法指標の推定値 ($c_2=0.1, c_3=0.0$)

		2000年																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計	
1 9 9 0 年	0		221	04	848	256	34	13	13	13	2	0	0	0	0	0	0	0	0	0	0	23285	
	1	18179	1026	301	52	29	13	13	13	3	3	3	3	1	0	0	0	0	0	0	0	19640	
	2	576	141	85	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	0	0	891
	3	141	141	85	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	0	0	456
	4	13	13	13	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	0	0	117
	5	7	7	7	7	7	7	7	7	2	1	1	1	1	0	0	0	0	0	0	0	0	51
	6	7	7	7	7	7	7	7	2	2	1	1	1	1	0	0	0	0	0	0	0	0	47
	7	7	7	7	7	7	6	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	46
	8	1	2	2	2	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	17
	9	0	1	2	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0	0	0	0	16
	10	0	0	0	1	2	2	2	2	1	1	1	1	1	0	0	0	0	0	0	0	0	13
	11	0	0	0	0	2	2	2	2	1	1	1	1	1	0	0	0	0	0	0	0	0	12
	12	0	0	0	0	0	2	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	9
	13	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
計	18930	23448	1375	372	130	95	87	81	25	18	17	14	6	1	1	1	0	0	0	0	0	0	

表5：母集団多重寸法指標の推定値 ($c_2=1.0, c_3=0.0$)

		2000年																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計	
1 9 9 0 年	0		221	25	875	230	44	15	11	11	6	1	0	0	0	0	0	0	0	0	0	23322	
	1	18195	1045	303	55	28	15	11	11	6	1	1	1	1	0	0	0	0	0	0	0	19674	
	2	599	139	85	18	15	15	11	11	6	1	1	1	1	0	0	0	0	0	0	0	0	902
	3	127	127	85	18	15	15	11	11	6	1	1	1	1	0	0	0	0	0	0	0	0	417
	4	21	21	21	18	15	15	11	11	6	1	1	1	1	0	0	0	0	0	0	0	0	140
	5	7	7	7	7	7	7	7	5	2	1	1	1	1	0	0	0	0	0	0	0	0	62
	6	6	6	6	6	6	6	6	8	3	1	1	1	1	0	0	0	0	0	0	0	0	50
	7	6	6	6	6	5	5	5	5	3	1	1	1	1	0	0	0	0	0	0	0	0	46
	8	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	14
	9	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	11
	10	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	10
	11	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	9
	12	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	8
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
計	18961	23480	1391	356	141	100	80	72	43	16	12	8	4	3	2	1	1	1	0	0	0	0	

それぞれ推定した後に、その推定値によって行和と列和を固定しながら母集団多重寸法指標を推定する2段階の推定方法を提案する。

母集団寸法指標の推定では、2.1節で紹介した制約条件(a), (c), (d)を3.1節と同様にペナルティー関数で表して、次のように対数尤度関数に取り込

表6：母集団多重寸法指標の推定値 ($c_2=10.0, c_3=0.0$)

		2000年																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
1 9 9 0 年	0		22059	927	184	53	23	13	7	4	2	1	1	1	0	0	0	0	0	0	23278	
	1	18158	1063	263	62	28	21	13	7	4	2	1	1	1	0	0	0	0	0	0	19625	
	2	596	170	66	26	18	17	13	7	4	1	1	1	1	0	0	0	0	0	0	823	
	3	113	112	60	23	17	17	13	7	4	1	1	1	0	0	0	0	0	0	0	371	
	4	29	29	29	22	17	15	13	7	4	1	1	1	0	0	0	0	0	0	0	189	
	5	11	11	11	11	11	11	10	7	4	1	1	1	0	0	0	0	0	0	0	91	
	6	5	5	5	5	5	5	5	4	3	1	1	1	0	0	0	0	0	0	0	47	
	7	2	2	2	2	2	2	2	2	2	1	1	0	0	0	0	0	0	0	0	23	
	8	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	14	
	9	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	8	
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
計		18918	23456	1367	340	155	115	86	59	30	15	10	7	6	5	4	4	3	3	2	2	

表7：母集団多重寸法指標の推定値 ($c_2=100.0, c_3=0.0$)

		2000年																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
1 9 9 0 年	0		21793	889	153	48	21	11	6	4	2	1	1	1	1	1	1	1	1	1	23035	
	1	17945	889	220	64	26	15	9	6	3	2	1	1	1	1	1	1	1	1	1	19288	
	2	636	203	97	42	24	15	9	6	3	2	1	1	1	1	1	1	1	1	1	1044	
	3	86	74	51	28	19	13	9	6	3	2	1	1	1	1	1	1	1	1	1	307	
	4	26	26	28	19	13	10	7	5	3	2	1	1	1	1	1	1	1	1	0	144	
	5	10	9	9	9	8	7	5	3	2	1	1	1	1	1	1	1	1	0	0	72	
	6	4	4	4	4	4	4	3	2	1	1	1	1	1	1	1	1	0	0	0	36	
	7	2	2	2	2	2	2	1	1	1	1	1	1	1	1	0	0	0	0	0	21	
	8	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	14	
	9	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	12	
	10	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	11	
	11	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	10	
	12	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9	
	13	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	8	
	14	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	
	15	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	
	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
計		18726	23108	1405	328	152	92	62	41	27	18	14	12	11	10	9	8	8	7	6	6	

む。また、探索の各ステップで制約条件(b)の補正を行う。

$$F(S_1, S_2, \dots, S_L \mid s_1, s_2, \dots, s_L)$$

$$= \log L_{ap}(S_1, S_2, \dots, S_L \mid s_1, s_2, \dots, s_L)$$

表 8 : 母集団多重寸法指標の推定値 ($c_2 = 1000.0$, $c_3 = 0.0$)

		2000年																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
1 9 9 0 年	0		21230	1087	136	41	18	9	6	4	3	2	2	2	2	1	1	1	1	1	1	22547
	1	17508	876	188	58	22	12	7	4	3	2	2	2	1	1	1	1	1	1	1	1	18689
	2	709	188	80	40	20	11	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1084
	3	85	60	49	24	15	10	7	4	2	2	1	1	1	1	1	1	1	1	1	1	262
	4	22	22	21	14	10	8	5	4	2	1	1	1	1	1	1	1	1	1	1	1	120
	5	8	8	7	7	7	5	4	3	2	1	1	1	1	1	1	1	1	1	1	1	61
	6	4	3	3	3	3	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	36
	7	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	25
	8	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21
	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	18
	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	17
	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	15
	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	14
	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	13
	15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	12
	16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	11
	17	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	10
	18	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	9
	19	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9
計	18351	22399	1462	295	132	79	53	38	28	22	20	18	17	16	15	14	13	12	11	10		

表 9 : 母集団多重寸法指標の推定値 ($c_2 = 1000.0$, $c_3 = 10.0$)

		2000年																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
1 9 9 0 年	0		21230	1087	136	41	18	9	6	4	3	2	2	2	2	1	1	1	1	1	1	22547
	1	17508	876	188	58	22	12	7	4	3	2	2	2	1	1	1	1	1	1	1	1	18689
	2	709	188	80	40	20	11	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1084
	3	85	60	49	24	15	10	7	4	2	2	1	1	1	1	1	1	1	1	1	1	262
	4	22	22	21	14	10	8	5	4	2	1	1	1	1	1	1	1	1	1	1	1	120
	5	8	8	7	7	7	5	4	3	2	1	1	1	1	1	1	1	1	1	1	1	61
	6	4	3	3	3	3	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	36
	7	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	25
	8	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21
	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	18
	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	17
	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	15
	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	14
	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	13
	15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	12
	16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	11
	17	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	10
	18	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	9
	19	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9
計	18351	22399	1462	295	132	79	53	38	28	22	20	18	17	16	15	14	13	12	11	10		

$$\begin{aligned}
& -c'_1 \sum_{l=1}^L \text{Pnlt}(-S_l; \varepsilon'_1) - c'_2 \sum_{l=2}^L \text{Pnlt}(-(S_{l-1} - S_l); \varepsilon'_2) \\
& - c'_3 \sum_{l=2}^{L-1} \text{Pnlt}(-(\log S_{l-1} + \log S_{l+1} - 2 \cdot \log S_l); \varepsilon'_3). \tag{18}
\end{aligned}$$

c'_1, c'_2, c'_3 は係数, $\varepsilon'_1, \varepsilon'_2, \varepsilon'_3$ はパラメータであるが, これまで扱った母集団寸法指標の推定の多くでは, $c'_1=c'_2=10.0, c'_3=1.0, \varepsilon'_1=\varepsilon'_2=0.0001, \varepsilon'_3=0.001$ を標準の組み合わせとしてきた。ここでは, この組み合わせを含む c'_2, c'_3 のいくつかの組み合わせに対して母集団寸法指標の推定を行った。

1990年の推定結果を表10に, 2000年の推定結果を表11にそれぞれ示す。第1行がサイズを示し, その下に推定された母集団寸法指標を表示している。係数が変化しても安定した推定結果が得られている。

上記で標準と呼んだパラメータの組み合わせを用いて推定された1990年と2000年の母集団寸法指標を, それぞれ $\hat{S}_{1t}, \hat{S}_{2t}$ とするとき, 母集団多重寸法指標の最尤推定値の勾配法による探索における各ステップで, 制約条件(B)に変えて

表10：母集団寸法指標の推定値の比較（1990年）

c'_2	c'_3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0.1	0.0	19749	655	611	140	26	26	26	26	26	26	15	2	0	0	0	0	0	0	0
1.0	0.0	19748	654	612	140	26	26	26	26	26	26	15	2	0	0	0	0	0	0	0
10.0	0.0	19748	654	614	138	28	26	26	26	26	24	15	3	1	0	0	0	0	0	0
100.0	0.0	19742	669	597	142	28	27	26	26	26	25	13	3	1	1	1	1	0	0	0
1000.0	0.0	19782	764	478	169	44	27	25	24	23	22	12	5	1	1	1	1	0	0	0
10.0	1.0	19850	671	407	169	71	29	26	22	19	16	13	10	3	1	0	0	0	0	0
1000.0	10.0	19642	889	387	167	72	34	27	22	18	15	12	10	2	1	1	0	0	0	0

表11：母集団寸法指標の推定値の比較（2000年）

c'_2	c'_3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0.1	0.0	23374	1540	172	172	172	94	27	25	25	25	19	3	1	0	0	0	0	0	0
1.0	0.0	23375	1539	173	173	173	89	30	25	25	25	18	4	1	0	0	0	0	0	0
10.0	0.0	23367	1542	173	173	171	84	37	27	24	24	16	4	1	1	1	1	0	0	0
100.0	0.0	23374	1539	173	173	173	86	33	26	26	23	16	4	1	1	1	1	1	0	0
1000.0	0.0	23374	1539	173	173	173	88	33	26	26	23	16	4	1	1	1	1	1	1	0
10.0	1.0	23405	1454	262	177	119	80	54	38	25	17	12	9	0	0	0	0	0	0	0
1000.0	10.0	23405	1453	264	177	119	80	54	38	25	17	12	8	0	0	0	0	0	0	0

$$(B') \sum_{l_2=0}^{L_2} S_{(l_1, l_2)} = \hat{S}_{l_1} \quad (l_1 = 1, 2, \dots, L_1),$$

$$\sum_{l_1=0}^{L_1} S_{(l_1, l_2)} = \hat{S}_{l_2} \quad (l_2 = 1, 2, \dots, L_2)$$

を用いた補正を行う。ここでは上式の左辺と右辺との差が0.01 以下になるまで、各ステップにおいて補正を繰り返す。前節の検討で良い推定結果が得られなかった係数の組み合わせ $c_2=1000.0$, $c_3=10.0$ を用いた場合の、母集団多重寸法指標の推定結果を表12に示す。

推定結果は改善されている。この方法を用いることにより、係数などの決め方によらずに、ある程度、安定した推定が可能と考えられる。

4 おわりに

本論文では、継続調査で得られた複数の個票データに対する同時リスク評価を行うために、寸法指標の概念を多重寸法指標に拡張し、標本多重寸法指標に基づく母集団多重寸法指標の制約付きノンパラメトリック推定法を提案した。

母集団多重寸法指標の推定結果で、もし $S_{(1,1)}$ の値が相対的に大きいことがわかれば、第1 時点でサイズ1 のセルを減らす秘匿措置を行うと、

表12：周辺頻度を固定した場合の母集団多重寸法指標の推定値($c_2 = 1000.0, c_3 = 10.0$)

		2000年																				
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
1 9 9 0 年	0		21953	1038	129	56	27	14	8	5	3	2	1	1	0	0	0	0	0	0	0	23238
	1	18120	1205	188	44	33	20	13	8	5	3	2	1	1	0	0	0	0	0	0	0	19656
	2	586	104	82	23	23	19	13	8	5	3	2	1	1	0	0	0	0	0	0	0	671
	3	138	88	82	23	23	19	13	8	5	3	2	1	1	0	0	0	0	0	0	0	407
	4	31	30	30	20	19	14	9	6	4	3	2	1	1	0	0	0	0	0	0	0	169
	5	10	10	10	9	9	7	5	3	2	2	1	1	1	0	0	0	0	0	0	0	71
	6	3	3	3	3	3	3	3	2	2	1	1	1	1	0	0	0	0	0	0	0	29
	7	3	3	3	2	2	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	26
	8	2	2	2	2	2	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	22
	9	2	2	2	2	2	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	19
	10	2	2	2	2	2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	16
	11	2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	13
	12	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	10
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
計		18899	23405	1454	262	177	119	80	54	36	25	17	12	9	0	0	0	0	0	0	0	

第2時点においてもサイズ1のセルを減らす効果の高いことがわかる。逆に各時点では S_1 の値が大きいにも関わらず $S_{(1,1)}$ の値が小さければ、各時点の個票データに対して独立に秘匿措置を行った方が効果的であろう。

本論文では、アメリカにおいて1990年と2000年に実施されたセンサスの1%抽出個票データを題材として、提案した推定方法を適用し、その有効性について検討も行った。母集団多重寸法指標の推定についてはこれまでの経験が少ないことを勘案し、初めに母集団多重寸法指標の行和と列和である母集団寸法指標を推定し、次にその値を制約条件として母集団多重寸法指標を推定するという2段階の推定法も提案した。これらの数値実験は抽出率が1/2と高かったこともあり、結果は概ね満足できるものであったが、一般の官庁統計で用いられている1/1000程度の抽出率で得られたデータへ適用するためには、更なる検討が必要と考えられる。

多重寸法指標の形状についても多様なパターンが存在する。同じセンサスの1%抽出個票データのうち、ワシントン州に在住している20歳以上の全個人レコードについて、3.1節で用いた10項目に、就業、年間収入の2項目を追加して多重寸法指標を求めると表13のようになる。

上部と左部に延びている非負の頻度のセルの多くには、非就業者が入っている。このような形状の母集団多重寸法指標を推定するためには、

$$(E) \quad l_1 \geq l_2 \text{ のとき } S_{(l_1, l_2)} \leq S_{(l_1+1, l_2-1)} \quad (l_1 + l_2 = 2, 3, \dots, \min(L_1, L_2)),$$

$$l_1 \leq l_2 \text{ のとき } S_{(l_1, l_2)} \leq S_{(l_1-1, l_2+1)} \quad (l_1 + l_2 = 2, 3, \dots, \min(L_1, L_2))$$

$$(F) \quad 2 \cdot \log S_{(l_1, l_2)} \leq \log S_{(l_1-1, l_2+1)} + \log S_{(l_1+1, l_2-1)}$$

$$(l_1 + l_2 = 2, 3, \dots, \min(L_1, L_2))$$

のような制約条件を追加しなければ、良い推定結果は得られない。母集団多重寸法指標に課す制約条件についても、今後検討していく必要がある。

なお、本論文では2時点の個票データを扱ったが、時点数が多い場合

には、母集団多重寸法指標に対してより強い制約条件を用いなければ、推定が不安定になると想像される。この点についても今後の課題になる。

謝 辞

本論文で扱った多重寸法指標に関して、慶應義塾大学名誉教授の渋谷政昭先生からは数々の有益な助言をいただきました。また、金沢大学の星野伸明先生、東京大学の竹村彰通先生からは本研究に対するアドバイスをいただきました。ここに感謝いたします。

本論文は科学研究費補助金（課題番号19300098, 18200019）、統計数理研究所共同利用研究プログラム（20-共研-2029）の研究成果に基づくものである。

参考文献

- [1] 馬場康維, 坂口尚文 (2006). 複数名簿のマッチングによる共通集合の推定, 2006年度統計関連学会連合大会予稿集, 117-117.
- [2] Bethlehem, J. G., Keller, W. J. and Pannekoek, J. (1990). Disclosure control of microdata, *Journal of the American Statistical Association*, 85, 38-45.
- [3] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions, *Probability Theory and Related Fields*, 102, 145-148.
- [4] 佐井至道 (2002). サイズインデックスの制約付き最尤推定, 岡山商大論叢, 37 (3), 61-79.
- [5] 佐井至道 (2003). 母集団寸法指標のノンパラメトリック推定, 統計数理, 51 (2), 183-198.
- [6] 佐井至道 (2006). ペナルティ関数を利用した母集団寸法指標の制約付きノンパラメトリック推定, 岡山商大論叢, 42(1), 1-21.
- [7] 佐井至道 (2007). 多重寸法指標のノンパラメトリック推定, 2007年度統計関連学会連合大会予稿集, 186-186.
- [8] 佐井至道 (2008). 多重寸法指標のノンパラメトリック推定, 岡山商大論叢, 43 (3), 1-18.
- [9] 渋谷政昭 (2005). 滑らかな罰金関数, 科学研究費補助金研究会資料, 1-3.
- [10] 渋谷政昭, 佐井至道 (2007). 多重指標の確率分割, 応用統計学会2007年度年会講演予稿集, 13-18.
- [11] Sibuya, M. and Sai, S. (2008). Analysis of a Dataset for Statistical Disclosure Control by Random Partition of a Multi-index, *Cherry Bud Workshop 2008*, 1-13.
- [12] U.S. Census Bureau (1993). 1990 Census of Population and Housing, Public Use Microdata Samples (microdata), Washington, D. C.
- [13] U.S. Census Bureau (2003). Census 2000 Public Use Microdata Sample (microdata), Washington, D. C.