統計数理(2009)
第 57 巻 第 2 号 425–442

［原著論文］

多重母集団寸法指標のノンパラメトリック最尤推定

―2時点の個票データへの適用―

佐井 至道†

(受付 2008 年 9 月 26 日；改訂 2009 年 6 月 1 日；採択 6 月 16 日)

要　　旨

標本調査で得られた個票データの公開におけるリスク評価では，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．その推定のために，初期の頃にはポアソンガンマモデルが，最近ではピットマンモデルなどの超母集団モデルが用いられている．加えて，制約付きノンパラメトリック最尤推定法も提案され，計算時間上の問題も克服されつつある．

官庁統計の標本調査は継続調査として行われる場合が多いが，そのような場合にも，個々の時点の個票データに対して独立にリスク評価を行うのがこれまでは一般的であった．本論文では，継続調査で得られた複数の時点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，多重標本寸法指標からの多重母集団寸法指標のノンパラメトリック推定法を提案する．

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，多重母集団寸法指標の推定値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を検討する．また，推定を安定させるために，個々の時点の個票データからそれぞれ母集団寸法指標を推定した後に，その推定値を制約条件に組み込み，多重母集団寸法指標を推定する方法も提案する．

キーワード： 多重指標，寸法指標，個票データ，リスク評価，ノンパラメトリック推定，官庁統計．

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．寸法指標とは，いくつかの変数について値の組み合わせが同じである個体数を求めた場合の，個体数別の頻度分布である．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と呼ばれる母集団寸法指標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)とも呼ばれる)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，母集団寸法指標の推定値

† 岡山商科大学 経済学部：〒700–8601 岡山市北区津島京町 2–10–1

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，いったんサンプリングされた個体が数回に渡って調査対象となる場合と，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．

いったんサンプリングされた個体が数回に渡って調査対象となる場合，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，その個票データに対応する仮想的な母集団の寸法指標の推定値を利用することが可能である．

一方で，調査時点ごとに異なる個体がサンプリングされる場合，これまでは個々の時点の個票データからそれぞれ母集団寸法指標を推定することにより，リスク評価を行うのが一般的であった．

リスク評価によって個票データが危険と判断された場合，例えば年齢について，90 歳以上を実年齢でなく 90 歳以上とするトップコーディングや，5 歳階級ごとにまとめる丸めのように，カテゴリーを併合する秘匿措置を用いて個体を識別しにくくする．そのような秘匿措置を段階的に施しながら母集団寸法指標を推定し，推定値そのもの，あるいはそれから計算される指標が基準を満足するところで，個票データを安全と判断する．指標としては，変数の値の組み合わせが，母集団でも標本でも他のすべての個体と異なる（一意と呼ばれる）個体数の推定値などが用いられる．

ところが，複数の時点の個票データに対して同じ秘匿措置を施した場合，ある時点の公開リスクが大きく減少しても，他の時点の公開リスクが同じように減少するとは限らない．その構造を把握して効率的な秘匿を行うためには，複数の時点における変数の同じ値の組み合わせに対する個体数の組の情報が必要になる．

本論文では，継続調査で得られた複数の時点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，多重標本寸法指標からの多重母集団寸法指標のノンパラメトリック最尤推定法を提案する．その導入により，どのような秘匿措置を施せば，複数の時点の個票データのリスクを効率的に減少できるのか，あるいは個々の時点の個票データに対して異なる秘匿措置を施した方がよいのか，という秘匿措置の指針が得られるものと考えられる．

多重寸法指標とは，いくつかの変数の値の組み合わせが同じである個体数を複数の母集団について求めた場合の，個体数の組別の頻度分布である．定着しているとは言えないが，本論文では，母集団においては多重母集団寸法指標，標本においては多重標本寸法指標という用語を用いることにする．

母集団寸法指標を推定するための超母集団モデルを多重母集団寸法指標への推定に拡張するのは容易ではなく，これまで妥当なモデルは提案されていない．これに対して，ノンパラメトリック最尤推定の対象を母集団寸法指標から多重母集団寸法指標に拡張するのは容易である．

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，多重標本寸法指標を基にした多重母集団寸法指標のノンパラメトリック推定法を提案する．

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 3.1 節で紹介し，3.2 節ではその個票データを題材として，多重母集団寸法指標の推定値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．

3.3 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 3.4 節では 2 段階で推定を行う方法を提案する．

2. 多重母集団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，個体を特定することが可能な数種類のキー変数の値の組み合わせに基づいて K 個のセルのいずれか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 $(j=1,2,\ldots,K)$ 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 $(l=0,1,2,\ldots,L)$ 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．なお，本論文のノンパラメトリック推定では総セル数 K を用いない．

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 $(l=0,1,\ldots,L)$ 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 $\hat{S}_l$ が求まれば，例えば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．ある個体のキー変数の値の組み合わせが母集団でも標本でも他のすべての個体と異なれば，その個体は識別可能，すなわち母集団のどの個体であるかを特定することが可能である．例えば個票データに他人に知られたくない情報が含まれていれば，特定された個体についてその情報も分かってしまうため最も危険と考えられる．また母集団でサイズ 2 以上であっても，標本としてそのうちの 1 個以上の個体が取られると，識別されないまでも予測される可能性がある．このように小さいサイズの母集団寸法指標の推定が重要である．

推定にはピットマンモデルなどの超母集団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 $(s_1,s_2,\ldots,s_L)$ 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 $(S_1,S_2,\ldots,S_L)$ を求めるが，その尤度関数は抽出率が十分小さいときに

$$\mathrm{L_{ap}}(S_1,S_2,\ldots,S_L \mid s_1,s_2,\ldots,s_L)=\frac{1}{{}_NC_n\ \lambda^n(1-\lambda)^{N-n}}\cdot\prod_{l'=1}^{L}\frac{e^{-\mu_{l'}}{\mu_{l'}}^{s_{l'}}}{s_{l'}!} \tag{2.1}$$

と近似できる．ただし，

$$\mu_{l'}=\sum_{l=l'}^{L}S_l\cdot{}_lC_{l'}\lambda^{l'}(1-\lambda)^{l-l'} \tag{2.2}$$

であり，尤度関数の添字の ap は近似を意味する．

(2.1)式を最大にする母集団寸法指標を推定値とするのが自然ではあるが，推定値は非常に不安定で，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の推定値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) $S_l\geq 0$ $(l=1,2,\ldots,L)$
(b) $\sum_{l=1}^{L}l\cdot S_l=N$
(c) $S_l\geq S_{l+1}$ $(l=1,2,\ldots,L-1)$

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

制約条件(d)は母集団寸法指標の推定値の対数が下に凸であることを意味する．これまで扱った多くの寸法指標は，概ねこれらの制約条件を満たしていた．

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 $(S_1,S_2,\ldots,S_L)$ 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の推定値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 3.2 節で紹介する．

2.2 多重標本寸法指標と多重母集団寸法指標

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，多変量に拡張したピットマンモデルの適用を検討しているが，実データへの当てはまりは良くない．また，佐井(2007, 2008)では 馬場・坂口(2006)の紹介した学会所属データに対してノンパラメトリック推定を行っている．このデータは多重寸法指標においてサイズの最大値が 1 となる特殊なケースである．

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，キー変数の同じ値の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．複数時点の母集団 $(D_1,D_2,\ldots,D_M)$ においてサイズの組 $(l_1,l_2,\ldots,l_M)$ となるセル数を $S_{(l_1,l_2,\ldots,l_M)}$ と表し，多重母集団寸法指標と呼ぶ．標本でも同様に，$(d_1,d_2,\ldots,d_M)$ においてサイズの組 $(l_1,l_2,\ldots,l_M)$ となるセル数を $s_{(l_1,l_2,\ldots,l_M)}$ と表し，多重標本寸法指標と呼ぶ．L_i を第 i 時点でのサイズの最大値とするとき $l_i=0,1,\ldots,L_i$ として，ここではサイズとして 0 を含めて議論する．なお，キー変数のとる値の数の積として総セル数 K は得られるが，その中にはあり得ない組み合わせなどが含まれることが多く，実質的な総セル数が確定できないことが多い．本論文で提案する方法でも K を確定しないことにする．そのため，もし K が確定できれば，少なくとも 1 時点でサイズが正となるセル数を K から引くことによって求められる $s_{(0,0,\ldots,0)}$ も，ここでは求めない．また $S_{(0,0,\ldots,0)}$ は推定対象にしないことにする．

$(0,0,\ldots,0)$ を除く各時点におけるすべてのサイズの組み合わせに対する多重母集団寸法指標の組と多重標本寸法指標の組を，それぞれ

$$\boldsymbol{S}=(S_{(0,0,\ldots,0,1)},\ldots,S_{(0,0,\ldots,0,L_M)},\ldots, S_{(L_1,L_2,\ldots,L_{M-1},0)},\ldots,S_{(L_1,L_2,\ldots,L_{M-1},L_M)}), \tag{2.3}$$

$$\boldsymbol{s}=(s_{(0,0,\ldots,0,1)},\ldots,s_{(0,0,\ldots,0,L_M)},\ldots, s_{(L_1,L_2,\ldots,L_{M-1},0)},\ldots,s_{(L_1,L_2,\ldots,L_{M-1},L_M)}) \tag{2.4}$$

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 $\boldsymbol{s}$ が得られたときの $\boldsymbol{S}$ の尤度関数は

$$\mathrm{L}(\boldsymbol{S}\mid\boldsymbol{s})=\frac{1}{\prod_{i=1}^{M}{}_{N_i}C_{n_i}}\sum_{\mathcal{C}_1}\prod_{(l_1,l_2,\ldots,l_M)}\left\{\frac{S_{(l_1,l_2,\ldots,l_M)}!}{\prod_{(l'_1,l'_2,\ldots,l'_M)}k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}!}\right. \tag{2.5}$$

$$\cdot \prod_{(l'_1,l'_2,\ldots,l'_M)} \left(\prod_{i=1}^{M} {}_{l_i}C_{l'_i} \right)^{k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}} \Bigg\}$$

と表される．ただし，$k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}$ は母集団においてサイズの組 $(l_1,l_2,\ldots,l_M)$ のセルのうち，標本においてサイズの組 $(l'_1,l'_2,\ldots,l'_M)$ と観測される数を表し，$\mathcal{C}_1$ は $\boldsymbol{S}$ から $\boldsymbol{s}$ が生成されるような $k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}$ の組についてのすべての組み合わせを表す．なお，(2.5)式の積において，母集団におけるサイズの組み合わせ $(l_1,l_2,\ldots,l_M)$ は $(0,0,\ldots,0)$ を含まず，標本におけるサイズの組み合わせ $(l'_1,l'_2,\ldots,l'_M)$ は $(0,0,\ldots,0)$ 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 $\mathcal{C}_1$ の組み合わせの数は膨大になり，時間的に(2.5)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 $(i=1,2,\ldots,M)$ 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，母集団の大きさを N_i に限定すると，観測された標本の大きさが n_i のときに(2.5)式に対応する尤度関数は

$$\begin{aligned}
(2.6)\quad & \mathrm{L_B}(\boldsymbol{S}\,|\,\boldsymbol{s}) \\
&= \sum_{\mathcal{C}_1} \prod_{(l_1,l_2,\ldots,l_M)} \left\{ \frac{S_{(l_1,l_2,\ldots,l_M)}!}{\prod_{(l'_1,l'_2,\ldots,l'_M)} k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}!} \right. \\
&\qquad \left. \cdot \prod_{(l'_1,l'_2,\ldots,l'_M)} \left\{ \prod_{i=1}^{M} {}_{l_i}C_{l'_i} \lambda_i^{l'_i} (1-\lambda_i)^{l_i-l'_i} \right\}^{k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}} \right\} \\
&= \prod_{i=1}^{M} \lambda_i^{n_i} (1-\lambda_i)^{N_i-n_i} \cdot \sum_{\mathcal{C}_1} \prod_{(l_1,l_2,\ldots,l_M)} \left\{ \frac{S_{(l_1,l_2,\ldots,l_M)}!}{\prod_{(l'_1,l'_2,\ldots,l'_M)} k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}!} \right. \\
&\qquad \left. \cdot \prod_{(l'_1,l'_2,\ldots,l'_M)} \left\{ \prod_{i=1}^{M} {}_{l_i}C_{l'_i} \right\}^{k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}} \right\} \\
&= \prod_{i=1}^{M} {}_{N_i}C_{n_i} \lambda_i^{n_i} (1-\lambda_i)^{N_i-n_i} \cdot \mathrm{L}(\boldsymbol{S}\,|\,\boldsymbol{s})
\end{aligned}$$

と書ける．尤度関数の添え字の B はベルヌーイ抽出を意味する．多重母集団寸法指標によらず，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．従って，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多重母集団寸法指標は，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 $(i=1,2,\ldots,M)$ となる多重母集団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 $(l_1,l_2,\ldots,l_M)$ についてのすべての $k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}$ の組 $\boldsymbol{k}_{(l_1,l_2,\ldots,l_M)}$ 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

$$\begin{aligned}
(2.7)\quad & f(\boldsymbol{k}_{(l_1,l_2,\ldots,l_M)}) \\
&= \frac{S_{(l_1,l_2,\ldots,l_M)}!}{\prod_{(l'_1,l'_2,\ldots,l'_M)} k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}!} \\
&\qquad \cdot \prod_{(l'_1,l'_2,\ldots,l'_M)} \left\{ \prod_{i=1}^{M} {}_{l_i}C_{l'_i} \lambda_i^{l'_i} (1-\lambda_i)^{l_i-l'_i} \right\}^{k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}}
\end{aligned}$$

である．積において，標本におけるサイズの組み合わせ $(l'_1,l'_2,\ldots,l'_M)$ は $(0,0,\ldots,0)$ を含む．

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

$$
\begin{aligned}
&(2.8)\ f\bigl(k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}\bigr)\\
&=e^{-\{S_{(l_1,l_2,\ldots,l_M)}\cdot\prod_{i=1}^{M}{}_{l_i}C_{l'_i}\lambda_i^{l'_i}(1-\lambda_i)^{l_i-l'_i}\}}\\
&\quad\cdot\left\{S_{(l_1,l_2,\ldots,l_M)}\cdot\prod_{i=1}^{M}{}_{l_i}C_{l'_i}\lambda_i^{l'_i}(1-\lambda_i)^{l_i-l'_i}\right\}^{k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}}\Bigg/k_{(l_1,l_2,\ldots,l_M),(l'_1,l'_2,\ldots,l'_M)}!
\end{aligned}
$$

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

$$
f\bigl(s_{(l'_1,l'_2,\ldots,l'_M)}\bigr)=\frac{e^{-\mu_{(l'_1,l'_2,\ldots,l'_M)}}\,\mu_{(l'_1,l'_2,\ldots,l'_M)}^{s_{(l'_1,l'_2,\ldots,l'_M)}}}{s_{(l'_1,l'_2,\ldots,l'_M)}!}. \tag{2.9}
$$

ただし

$$
\mu_{(l'_1,l'_2,\ldots,l'_M)}=\sum_{(l_1,l_2,\ldots,l_M)(\geq(l'_1,l'_2,\ldots,l'_M))}\left\{S_{(l_1,l_2,\ldots,l_M)}\cdot\prod_{i=1}^{M}{}_{l_i}C_{l'_i}\lambda_i^{l'_i}(1-\lambda_i)^{l_i-l'_i}\right\} \tag{2.10}
$$

である．したがって尤度関数(2.5)式は

$$
\begin{aligned}
&\mathrm{L_{ap}}(\boldsymbol{S}\,|\,\boldsymbol{s})\\
&\quad=\frac{1}{\prod_{i=1}^{M}{}_{N_i}C_{n_i}\lambda_i^{n_i}(1-\lambda_i)^{N_i-n_i}}\cdot\prod_{(l'_1,l'_2,\ldots,l'_M)}\frac{e^{-\mu_{(l'_1,l'_2,\ldots,l'_M)}}\,\mu_{(l'_1,l'_2,\ldots,l'_M)}^{s_{(l'_1,l'_2,\ldots,l'_M)}}}{s_{(l'_1,l'_2,\ldots,l'_M)}!}
\end{aligned} \tag{2.11}
$$

と近似される．積において，標本におけるサイズの組み合わせ $(l'_1,l'_2,\ldots,l'_M)$ は $(0,0,\ldots,0)$ を含まない．

3. 多重母集団寸法指標の推定方法

3.1 アメリカのセンサスデータ

3 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．数値実験では推定を安定させるために，多重母集団寸法指標の推定値に制約条件を課す検討を行う．

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．この場合，総セル数はキー変数の取り得るすべての値の組み合わせとして $K\doteqdot 2.739\cdot 10^{11}$ である．しかし，この中には電車通勤で年間の労働週数が 0 週のようなあり得ない組み合わせが含まれている．また，例えば週労働時間としては，記録可能な 0 時間から 99 時間(正確には 99 時間以上を意味する)までを用いて計算している．したがって，上記の値は名目上の総セル数であ

る．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 1% 抽出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，計算された多重寸法指標を基にして母集団の多重寸法指標を推定することができる．しかし，実際の多重母集団寸法指標は得られない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 に多重母集団寸法指標 $\boldsymbol{S}$ 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，推定では最大サイズを $L_1 = L_2 = 19$ と大きい方に合わせる．なお実際の推定では母集団における最大サイズ L は未知であるため，サイズの最大値として十分大きい値を設定するべきであるが，計算時間上の問題もあり，推定結果に影響を与えない程度の比較的小さい値を選ぶことになろう．

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 を非復元単純無作為抽出した．$n_1 = 12423$，$n_2 = 15117$ である．抽出率 1/2 は十分小さいとは言えず，(2.8) によるポアソン近似の精度に問題がある．しかし，本数値実験では母集団が小さいため，抽出率を小さくとった場合には多重標本寸法指標がほとんど情報を持たなくなってしまうので，ここでは抽出率としては 1/2 を用いた．

この標本から求められた多重標本寸法指標 $\boldsymbol{s}$ 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 時点のデータからはサイズ 0 の頻度は観測できない．例えば表 1 において，行和の値 23204 は 1990 年にサイズ 0 のセル数であるが，2000 年のデータがなければ観測することはできない．なお，これらの値は他の時点において個体が入っていたセルのみの数である．

3.2　多重母集団寸法指標の推定値に対する制約条件とペナルティー関数の導入

多重標本寸法指標を基に，(2.11) 式を最大とする多重母集団寸法指標の近似的な最尤推定値を求めるが，推定を安定させるために，寸法指標の推定と同様に次のような制約条件の一部，

表 1.　多重母集団寸法指標（ワシントン州の 20 歳以上の就業者）．出典：U.S. Census Bureau (1993, 2003)．

1990年 \ 2000年	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		21971	939	202	53	23	6	4	1	5											23204
1	18154	1008	254	85	58	33	7	9	5	1	2	1									19617
2	675	202	78	48	18	20	8	8	3	4	2	1	2	1							1070
3	104	72	32	34	7	7	12	5	5	2	1	2		2							285
4	19	31	28	21	12	8	4	7	3	4	1	2	1	2	2			1			146
5	10	16	7	10	17	9	4	1	3	3		1	2	1							84
6	4	8	4	5	8	9	2	5	2	2		1	1	1							52
7	2	1	6	8	3	1	4	2	1	4	1	1	1		1						36
8			4	4	4	6	3	2	3	1	2					1					30
9	1	1		2	2	1	1			2											10
10		2	1	2		2															7
11				3	2	3		1		1											10
12					1	1	2	1		1											6
13					1	1			3												5
14																					0
15																					0
16																					0
17																					0
18																					0
19								1													1
計	18969	23312	1353	424	186	124	53	46	29	30	9	9	7	7	3	1	0	1	0	0	

表 2. 多重標本寸法指標(抽出率：1/2).

1990年	2000年 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		11973	406	75	23	11	5	1		1											12495
1	9817	471	105	36	23	11	4	6	1			1									10475
2	277	97	41	23	18	8	3	2	1												470
3	49	45	22	15	10	4	1	2	1												149
4	11	6	15	6	6	2	1														47
5	5	5	7	3	4		2	1													27
6	2	6	1	5	4	3	2		1												24
7			1	3			1														5
8			2	1	3																6
9																					0
10																					0
11					1																1
12																					0
13																					0
14																					0
15																					0
16																					0
17																					0
18																					0
19																					0
計	10161	12603	600	167	92	39	19	12	4	1	0	1	0	0	0	0	0	0	0	0	

またはすべてを多重母集団寸法指標の推定値に課すことを検討する.

(A) $S_{(l_1,l_2)} \geq 0 \quad (l_1 = 0,1,\ldots,L_1,\ l_2 = 0,1,\ldots,L_2)$
(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) $S_{(l_1,l_2)} \leq S_{(l_1-1,l_2)} \quad (l_1 = 1,2,\ldots,L_1,\ l_2 = 0,1,\ldots,L_2)$,
$S_{(l_1,l_2)} \leq S_{(l_1,l_2-1)} \quad (l_1 = 0,1,\ldots,L_1,\ l_2 = 1,2,\ldots,L_2)$
(D) $2 \cdot \log S_{(l_1,l_2)} \leq \log S_{(l_1-1,l_2)} + \log S_{(l_1+1,l_2)} \quad (l_1 = 1,2,\ldots,L_1-1,\ l_2 = 0,1,\ldots,L_2)$,
$2 \cdot \log S_{(l_1,l_2)} \leq \log S_{(l_1,l_2-1)} + \log S_{(l_1,l_2+1)} \quad (l_1 = 0,1,\ldots,L_1,\ l_2 = 1,2,\ldots,L_2-1)$

制約条件(A)は多重母集団寸法指標の推定値が非負を表し，(B)は 2 時点での母集団の大きさの合計に関する制約条件で，課すことは妥当であろう．また(C), (D)については，2.1 節で述べた母集団寸法指標の推定値に対する制約条件(c), (d)を，各行と各列に適用したもので自然な拡張と考えられる.

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.

$$\mathrm{Pnlt}(x;\varepsilon) = \varepsilon \log(e^{\frac{x}{\varepsilon}} + 1). \tag{3.1}$$

ただし，ε はパラメータである．ペナルティー関数の導入により制約条件を尤度関数に取り込めるため，計算時間の短縮を図ることができる．また，このペナルティー関数はすべての x について微分可能であるため，後で述べる勾配法による探索も容易である．(3.1)式の微分は

$$\frac{d\mathrm{Pnlt}(x;\varepsilon)}{dx} = \frac{e^{\frac{x}{\varepsilon}}}{e^{\frac{x}{\varepsilon}} + 1} \tag{3.2}$$

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なる.

図 1 に $\varepsilon = 0.01,\ 0.001$ としたペナルティー関数 $y = \mathrm{Pnlt}(x;\varepsilon)$ のグラフを示す．$\varepsilon \to 0$ 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.

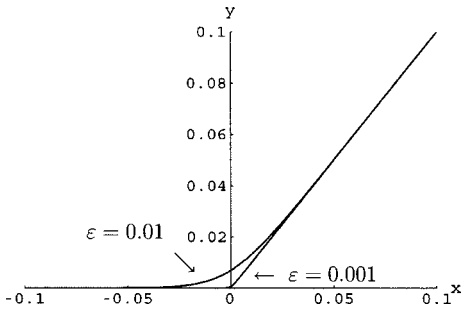

図 1. 滑らかなペナルティー関数 $\mathrm{Pnlt}(x;\varepsilon) = \varepsilon \log(e^{\frac{x}{\varepsilon}} + 1)$.

(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.

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

(A)，(C)，(D) の制約条件に関するペナルティー関数において，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ー関数の関数 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 $S_{(l_1,l_2)}$ $(l_1 = 0,1,\ldots,L_1;\ l_2 = 0,1,\ldots,L_2)$ 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，解析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 $S_{(l_1,l_2)}$ を独立に移動させる勾配法によって最大値を探索する．なお $S_{(l_1,l_2)}$ として実数の範囲を考える.

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

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

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，$\mathrm{g}(\cdot)$ 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，このようにして多重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．

3.3 アメリカのセンサスデータについての推定結果

この節では，多重母集団寸法指標の推定を種々の設定で行い，それらの結果の比較を行う．

数値実験では，原則として(3.3)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(3.4)式において

$$\mathrm{g}(x) = \mathrm{sign}(x) \cdot |\, x \,|^{0.5}, \tag{3.5}$$

$$t_{(l_1,l_2)(p)} = \mathrm{b}(p) \cdot \hat{S}_{(l_1,l_2)(p)} \;/\; (l_1 + l_2)^{0.5} \tag{3.6}$$

として，p 番目のステップにおける多重母集団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，$\mathrm{b}(p)$ 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，推定値と周辺頻度である行和，列和は，それぞれ整数に四捨五入しているため，合計が一致していない場合がある．また表中の 0 は実際には 0.5 未満の正の値である．

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．母集団寸法指標の推定において，2.1 節の制約条件(a), (b)のみを課した場合の結果と同様に，多くの推定値が 0 となる一方で，いくつかのサイズの組み合わせについては大きな値が得られている．

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多重母集団寸法指標に近いが，c_2 の値によって結果は大きく異なっている．

3.4 推定方法の改善 ～ 周辺母集団寸法指標を先行して推定する 2 段階法

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，ある程度安定した推定結果を得ることが可能である．多重母集団寸法指標の

表 3. 多重母集団寸法指標の推定値 ($c_2 = 0.0$, $c_3 = 0.0$).

1991年 \ 2000年	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		22116	761	340	1	0	8	28	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	2	1	0	19718
2	645	0	0	0	0	23	30	0	0	0	0	0	0	0	0	0	0	0	0	0	699
3	89	289	174	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	601
4	2	0	0	0	0	101	21	0	0	0	0	0	0	0	0	0	0	0	0	0	124
5	0	0	0	0	0	29	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	32
7	1	0	0	32	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	24
9	0	0	0	0	0	0	0	0	0	0	8	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	36
11	0	0	0	0	0	0	19	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
13	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
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	2	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
計	18915	23435	1396	374	17	224	100	30	0	16	69	0	0	0	0	0	0	2	1	0	

表 4. 多重母集団寸法指標の推定値 ($c_2 = 0.1,\ c_3 = 0.0$).

	2000年																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
	0		22104	848	256	34	13	13	13	2	0	0	0	0	0	0	0	0	0	0	23285
1	1	18179	1026	301	52	29	13	13	13	3	3	3	3	1	0	0	0	0	0	0	19640
9	2	576	141	95	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	891
9	3	141	141	95	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	456
0	4	13	13	13	13	13	13	13	13	3	3	3	3	1	0	0	0	0	0	0	117
年	5	7	7	7	7	7	7	7	2	1	1	1	0	0	0	0	0	0	0	0	51
	6	7	7	7	7	7	7	2	2	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	46
	8	1	2	2	2	2	2	2	2	1	1	1	0	0	0	0	0	0	0	0	17
	9	0	1	2	2	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	16
	10	0	0	0	1	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	13
	11	0	0	0	0	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	12
	12	0	0	0	0	0	2	2	2	1	1	1	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	3
	14	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
	16	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
	18	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
	計	18930	23448	1375	372	130	95	87	81	25	18	17	14	6	1	1	1	0	0	0	0	

表 5. 多重母集団寸法指標の推定値 ($c_2 = 1.0,\ c_3 = 0.0$).

	2000年																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
	0		22125	875	230	44	15	11	11	6	1	0	0	0	0	0	0	0	0	0	23322
1	1	18195	1045	303	55	28	15	11	11	6	1	1	1	0	0	0	0	0	0	0	19674
9	2	599	139	85	16	15	15	11	11	6	1	1	1	1	0	0	0	0	0	0	902
9	3	127	127	85	16	15	15	11	11	6	1	1	1	1	0	0	0	0	0	0	417
0	4	21	21	21	16	15	15	11	11	6	1	1	1	0	0	0	0	0	0	0	140
年	5	7	7	7	7	7	7	7	5	2	1	1	1	0	0	0	0	0	0	0	62
	6	6	6	6	6	6	6	6	3	1	1	1	1	0	0	0	0	0	0	0	50
	7	6	6	6	5	5	5	5	3	1	1	1	1	0	0	0	0	0	0	0	46
	8	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	0	0	0	0	0	0	0	0	11
	10	0	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	0	0	0	0	0	0	0	0	9
	12	0	0	0	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	3
	14	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	1
	16	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
	18	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
	計	18961	23480	1391	356	141	100	80	72	43	16	12	8	4	3	2	1	1	1	0	0	

表 6. 多重母集団寸法指標の推定値 ($c_2 = 10.0,\ c_3 = 0.0$).

	2000年																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
	0		22059	927	184	53	23	13	7	4	2	1	1	1	0	0	0	0	0	0	23278
1	1	18158	1063	263	62	28	21	13	7	4	2	1	1	1	0	0	0	0	0	0	19625
9	2	596	170	66	26	18	17	13	7	4	1	1	1	1	0	0	0	0	0	0	923
9	3	113	112	60	23	17	17	13	7	4	1	1	1	0	0	0	0	0	0	0	371
0	4	29	29	29	22	17	15	13	7	4	1	1	1	0	0	0	0	0	0	0	169
年	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	53	30	15	10	7	6	5	4	4	3	3	2	2	

表 7. 多重母集団寸法指標の推定値 ($c_2 = 100.0$, $c_3 = 0.0$).

2000年

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		21793	989	153	48	21	11	6	4	2	1	1	1	1	1	1	1	1	1	1	23035
1	17945	989	220	64	26	15	9	6	3	2	1	1	1	1	1	1	1	1	1	1	19288
9 2	636	203	97	42	24	15	9	6	3	2	1	1	1	1	1	1	1	1	1	0	1044
9 3	96	74	51	28	19	13	9	6	3	2	1	1	1	1	1	1	1	1	1	0	307
0 4	26	26	26	19	13	10	7	5	3	2	1	1	1	1	1	1	1	1	0	0	144
年 5	10	9	9	9	9	7	5	3	2	1	1	1	1	1	1	1	1	0	0	0	72
6	4	4	4	4	4	4	3	2	1	1	1	1	1	1	1	1	0	0	0	0	36
7	2	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	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	0	12
10	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	11
11	1	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	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	0	8
14	1	1	1	1	1	0	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	0	7
16	1	0	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	0	6
18	0	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	0	4
計	18726	23108	1405	328	152	92	62	41	27	18	14	12	11	10	9	8	8	7	6	6	

表 8. 多重母集団寸法指標の推定値 ($c_2 = 1000.0$, $c_3 = 0.0$).

2000年

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		21230	1087	136	41	18	9	6	4	3	2	2	2	2	1	1	1	1	1	1	22547
1	17508	876	196	56	22	12	7	4	3	2	2	2	1	1	1	1	1	1	1	1	18699
9 2	709	186	90	40	20	11	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1084
9 3	85	60	43	24	15	10	7	4	2	2	1	1	1	1	1	1	1	1	1	1	262
0 4	22	22	21	14	10	8	5	4	2	1	1	1	1	1	1	1	1	1	1	1	120
年 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	0	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	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	0	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	1	1	0	0	0	0	0	0	0	0	11
17	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10
18	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9
19	1	1	1	1	1	1	1	1	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. 多重母集団寸法指標の推定値 ($c_2 = 1000.0$, $c_3 = 10.0$).

2000年

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		21230	1087	136	41	18	9	6	4	3	2	2	2	2	1	1	1	1	1	1	22547
1	17508	876	196	56	22	12	7	4	3	2	2	2	1	1	1	1	1	1	1	1	18699
9 2	709	186	90	40	20	11	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1084
9 3	85	60	43	24	15	10	7	4	2	2	1	1	1	1	1	1	1	1	1	1	262
0 4	22	22	21	14	10	8	5	4	2	1	1	1	1	1	1	1	1	1	1	1	120
年 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	0	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	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	0	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	1	1	0	0	0	0	0	0	0	0	11
17	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	10
18	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	9
19	1	1	1	1	1	1	1	1	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	

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多重母集団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，3.2 節と同様に，2.1 節で紹介した制約条件(a)，(c)，(d) をペナルティー関数で表して，(3.7) 式のように対数尤度関数に取り込んだうえで，勾配法による最大値の探索を行う．なお，探索の各ステップの推定値は一般に制約条件(b) を満たさないため，3.2 節と同様に，推定値と原点とを通る直線と(b) が表す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．

$$
\begin{aligned}
(3.7)\qquad & \mathrm{F}(S_1,S_2,\ldots,S_L \mid s_1,s_2,\ldots,s_L) \\
&= \log \mathrm{L}_{\mathrm{ap}}(S_1,S_2,\ldots,S_L \mid s_1,s_2,\ldots,s_L) \\
&\quad - c_1' \sum_{l=1}^{L} \mathrm{Pnlt}(-S_l;\ \varepsilon_1') - c_2' \sum_{l=2}^{L} \mathrm{Pnlt}(-(S_{l-1}-S_l);\ \varepsilon_2') \\
&\quad - c_3' \sum_{l=2}^{L-1} \mathrm{Pnlt}(-(\log S_{l-1} + \log S_{l+1} - 2\cdot \log S_l);\ \varepsilon_3').
\end{aligned}
$$

c_1'，c_2'，c_3' は係数，ε_1'，ε_2'，ε_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．例えば S_1 が表 10 では 19642～19782，表 11 では 23367～23405 と推定されており，係数 c_2'，c_3' 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 $\hat{S}_{l_1}$，$\hat{S}_{l_2}$ とするとき，多重母集団寸法指標の最尤推定値の勾配法による

表 10. c_2'，c_3' の組み合わせに対する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	19748	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	136	28	26	26	26	26	24	15	3	1	0	0	0	0	0	0
100.0	0.0	19742	669	597	142	29	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	0	0	0	0
10.0	1.0	19656	871	407	169	71	29	26	22	19	16	13	10	3	1	0	0	0	0	0
1000.0	10.0	19642	899	387	167	72	34	27	22	18	15	12	10	2	1	1	0	0	0	0

表 11. c_2'，c_3' の組み合わせに対する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	86	33	26	26	23	16	4	1	1	1	1	1	0	0
10.0	1.0	23405	1454	262	177	119	80	54	36	25	17	12	9	0	0	0	0	0	0	0
1000.0	10.0	23405	1453	264	177	119	80	54	36	25	17	12	8	0	0	0	0	0	0	0

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

1990年 \ 2000年	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	計
0		21953	1036	129	56	27	14	8	5	3	2	1	1	0	0	0	0	0	0	0	23238
1	18120	1205	198	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	871
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	1	1	1	1	0	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	1	0	0	0	0	0	0	0	0	0	13
12	1	1	1	1	1	1	1	1	1	1	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	

探索における各ステップで，制約条件(B)に変えて

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

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，多重母集団寸法指標の推定結果を表 12 に示す．制約条件(B')による補正において誤差を許しているため，表 12 の周辺頻度は表 10, 11 の値とは一致していない場合がある．

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，(3.3)式におけるパラメータと係数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時点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，多重標本寸法指標に基づく多重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．

多重母集団寸法指標の推定結果で，例えば $S_{(1,1)}$ 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，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．

これまで扱った官庁統計などの個票データでは，個体数と比較してセル数が極めて多い場合がほとんどである．寸法指標では観測サイズ 0 のセルは観測できないものの，サイズ 1 以上のセル数と比較して数多く存在し，多重寸法指標ではその一部が $S_{(l_1,0)}$, $S_{(0,l_2)}$ として観測される．$S_{(1,1)}$ の評価では各時点の S_1 との比較が基本と考えられるが，多重寸法指標内で比較する場合には $S_{(l_1,0)}$, $S_{(0,l_2)}$ ではなく $S_{(l_1,l_2)}$ $(l_1, l_2 \neq 0)$ と比較するべきと考える．

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．多重母集団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，初めに多重母集団寸法指標

表 13.　多重母集団寸法指標(ワシントン州の 20 歳以上のすべての個人, 就業と年間収入の項目を含む). 出典：U.S. Census Bureau (1993, 2003).

2000年

1990年	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	計
0		37983	526	70	17	12	9	3	1	4	1	8	3	3	2		1	1		4	1			1					1	2		1	3	1	1	1	1	1		2										38664
1	31551	272	46	16	6	3	5	1		1						3	1		1	2							1	1	1	1	1		1	1	2					1					1					31919
2	224	27	11	4	1	1	1	1	1	1	1							1	2	1		1									1	1	1									1								282
3	42	3	4	2	2		2	1			1																																						57
4	9	2	4			2	2		1																																								20
5	6	4	1			1			1		1																																						14
6	4		1	2						1						1																																		9
7	1	1	1						1																																									4
8	2	2		1							1																																							6
9	2	2	2	2		1			1																																								10
10		2																																																2
11	2	1													1																																				4
12	4		1												1																																				6
13	2																																																	2
14	2		1	2																																															5
15		2																																																2
16		1																																																1
17		2	1																																															3
18	2			1																																													3
19																																																		0
20	2				1																																												3
21	1		2	2																																														5
22		1																																																1
23	2		2																																															4
24		1			1																																														2
25	1																																																		1
26		1		1																																															2
27		1			1																																														2
28																																																		0
29			1																																															1
30	1	1																																																2
31		1																																																1
32																																																		0
33		1																																																1
34	1																																																	1
35																																																		0
36			1																																														1
37	1		1																																															2
38			1																																															1
39			1																																															1
40		1																																																1
41						1																																												1
42		1																																																	1
43		1																																																	1
44							1																																												1
45			1																																																1
46																																																		0
47																																																		0
48			1																																															1
49																																																			0
50																																																			0
計	31862	38314	610	103	29	21	20	6	6	7	5	8	3	3	4	4	2	2	3	7	1	1	0	1	0	0	0	1	2	3	1	2	4	3	3	3	1	1	0	2	1	0	0	0	1	1	0	0	0	0	0	

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多重母集団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，多重母集団寸法指標の推定値に対する制約を強くするなど，更なる検討が必要と考えられる．ただし，官庁統計では一般に標本の大きさもここでの実験よりは大きい．また，ここではキー変数の値についてトップコーディングや丸めなどの秘匿措置を行わなかったが，通常は公開のリスクを小さくするため秘匿措置を行うことになる．その場合，多重母集団寸法指標の各時点のサイズの最大値は大きくなり，多重標本寸法指標の各時点のサイズの最大値もまた大きくなる．母集団寸法指標の推定では，標本寸法指標のサイズの最大値が大きいほど安定した推定を行うことができることが経験的に知られている．そのため抽出率の減少に比べて精度の低下は小さい可能性もある．

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寸法指標を推定するためには，

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

(F) $2\cdot\log S_{(l_1,l_2)} \leq \log S_{(l_1-1,l_2+1)} + \log S_{(l_1+1,l_2-1)}$ $(l_1+l_2=2,3,\ldots,\min(L_1,L_2))$

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．多重母集団寸法指標の推定値に課す制約条件についても，今後検討していく必要があろう．

なお，本論文では M 時点の多重母集団寸法指標に対する最尤推定法を提案したが，数値実験では 2 時点の個票データしか扱っていない．時点数が多い場合には，多重母集団寸法指標の推定値に対して，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．特に，相関構造を表現できるような制約条件が必要となるかもしれない．

また時点数が多い場合には，推定目標を多重母集団寸法指標全体ではなく，$S_{(1,1,\ldots,1)}$ や $S_{(1,0,\ldots,0)}$ などに絞ることや，各時点におけるサイズが 2 以下の個体数のように，特定の範囲に含まれるサイズの組み合わせに対する個体数を推定目標とすることも必要になろう．

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．2 名の査読者と編集委員の方々からも貴重なご意見をいただきました．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

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，p. 117.

Bethlehem, J. G., Keller, W. J. and Pannekoek, J. (1990)． Disclosure control of microdata, *Journal of the American Statistical Association*, **85**, 38–45.

Pitman, J. (1995)． Exchangeable and partially exchangeable random partitions, *Probability Theory and Related Fields*, **102**, 145–148.

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.

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.

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.

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.

佐井至道(2008)． 母集団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.

渋谷政昭(2005)． 滑らかな罰金関数，日本学術振興会科学研究費補助金研究会資料，1–3.

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.

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.

U. S. Census Bureau (1993)． 1990 Census of Population and Housing, Public Use Microdata Samples (microdata), Washington, D. C.

U. S. Census Bureau (2003)． Census 2000 Public Use Microdata Sample (microdata), Washington, D. C.

Nonparametric Maximum Likelihood Estimation of Multi-population Size Indices —Application to Microdata Sets of Two Occasions—

Shido Sai

Faculty of Economics, Okayama Shoka University

Population size indices estimated from observed sample size indices are often used to assess the disclosure risk of a microdata set sampled from a population. Parametric methods with superpopulation models such as the Poisson gamma model or the Pitman model are usually used to estimate population size indices. A restricted nonparametric maximum likelihood estimation method has also been proposed, and its problem of computing time is being improved.

Though most sample surveys of official statistics are repeated periodically, it is usual to carry out risk assessment for each microdata set independently. In order to assess the disclosure risk for several microdata sets simultaneously, the size indices are extended to those of the multi-indices, and a nonparametric method for estimating multi-population size indices from multi-sample size indices is proposed.

Restrictions for the estimates of multi-population size indices are expressed as penalty functions and are taken into a log likelihood function. To improve the estimation, a two-stage method is also proposed, in which the multi-population size indices are estimated under the condition that their marginal frequencies are fixed on the population size indices estimated in advance. These proposed methods are applied to 1990 and 2000 1-Percent Public Use Microdata Sample Files in the U. S.

Key words: Multi-indices, size indices, microdata, risk assessment, nonparametric estimation, official statistics.