『岡山商大論叢』（岡山商科大学）
第43巻第３号　2008年２月
Journal of OKAYAMA SHOKA UNIVERSITY
Vol. 43　No. 3　February 2008

《論　　説》

多重寸法指標のノンパラメトリック推定

佐　井　至　道

Nonparametric Estimation for Multi Size Indexes

Shido Sai

1　はじめに

標本寸法指標からの母集団寸法指標の推定は，標本調査で得られた個票データを公開する場合のリスク評価のために，しばしば用いられる方法である。その推定については 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，母集団寸法指標への制約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。

本報告では，寸法指標を拡張した多重寸法指標に対するノンパラメトリック推定法を提案する。多重寸法指標については渋谷（2007），渋谷，佐井（2007）において紹介されており，EPSFの多重指標確率分割への拡張も検討されている。

まず2節において，本論文で提案する推定法の動機付けになった馬場，坂口（2006）の紹介した所属学会問題について，その概略を説明する。次に3節では，単純な設定の下で所属学会問題を考えるが，この場合，標本寸法指標からの母集団寸法指標の推定と同じ議論となる。4節では，より自然な設定の下で所属学会問題を考えるために多重寸法指標の概念を導入する。また多重寸法指標の推定も，ノンパラメトリック法を用いると通常の寸法指標の推定と同様の議論が可能となることを説明する。5節において，母集団多重寸法指標の推定を行うとともに，その安定性についても議論する。6節では，多重寸法指標を官庁統計などで得られる個票データのリスク評価に用いる可能性について考える。

2　所属学会問題

この節では，馬場，坂口（2006）が紹介した所属学会問題について説明する。

3つの学会G_1，G_2，G_3があり，不明者などを除いた所属会員はG_1：376名，G_2：996名，G_3：1395名である。この中には複数の学会に所属する者も含まれ，所属状況は図1のようになっている。3学会ののべ人数は2767名，実人数は2364名である。ここでは，各学会の所属会員数のみが既知であるものの，2学会，3学会に所属する人数は未知とする。

学会別に3つの会員リストが存在しており，それらのリストから個人レコードのサンプリングを行い，そのレコード同士をマッチングすることによって得られる所属学会状況を基に，母集団での所属学会状況を推定する。

サンプリング法としては，3つの会員リスト全体から非復元単純無作為

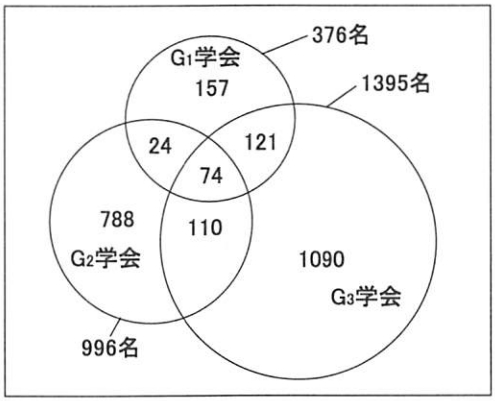


図１：学会への所属会員数

抽出する方法や，それぞれの会員リストから独立に非復元単純無作為抽出する方法などが考えられるが，いずれにしても個人レコードのサンプリングに限ることにする。もし３つの会員リストから同一個人に対する全レコードのサンプリングが可能であれば，推定は極めて簡単である。

3　単純な設定の下での寸法指標の導入

この節では，所属する学会名を無視し，所属学会数のみを考える単純化を行う。

３つの所属リストに含まれる全個人レコードを母集団とし，その大きさを N で表す。この母集団から大きさ n の標本を非復元単純無作為抽出し，抽出率を $\lambda = n/N$ と表す。なお，各所属リストからそれぞれ独立にサンプリングを行うと，以下の議論はできない。

このとき，母集団において所属学会数が l である人数を S_l $(l = 1,2,3)$，標本において所属学会数が l' である人数を $s_{l'}$ $(l' = 1,2,3)$ とすると，その関係は個票データのリスク評価で用いられる母集団寸法指標と標本寸法指標との関係と同じになる。個票データのリスク評価の場合，寸法指標はセル内に含まれる個体数別のセル数で，l 個の個体が含まれるセル数を母集団では S_l と表すが，ここでは，一つずつのセルを個人

に対応させており，セルに含まれる個体として3つの学会を考えていることになる。個票データのリスク評価については6節で説明する。

$s_{l'}$ が観測された下での S_l の尤度関数は

$$\mathrm{L}(S_1, S_2, S_3 \mid s_1, s_2, s_3) = \frac{1}{{}_N C_n} \sum_{C_1} \prod_{l=1}^{3} \frac{S_l!}{k_{l0}! k_{l1}! \cdots k_{ll}!} ({}_l C_0)^{k_{l0}} ({}_l C_1)^{k_{l1}} \cdots ({}_l C_l)^{k_{ll}} \tag{1}$$

と書くことができる。ただし，$k_{ll'}$ $(l = 0,1,2,3; l' = 0,1,...,l; l' \leq l)$ を母集団においてサイズ l のセルのうちで，標本においてサイズ l' となるものの数，すなわち，母集団において l 個の学会に所属しており，標本では l' 個の学会に所属する人数とするとき，C_1 は非負整数 $k_{ll'}$ の図2に示すような行和と列和に関する制約条件を満たすすべての組み合わせであり，その計算時間が尤度を求める上での障害となる。なお，図2において空欄はすべて0である。

このとき抽出率 λ が十分小さければ，佐井（2002，2003）で示したように，

$$\mathrm{L}(S_1, S_2, S_3 \mid s_1, s_2, s_3) = c_1 \cdot \prod_{l'=1}^{3} \frac{e^{-\mu_{l'}} \mu_{l'}{}^{s_{l'}}}{s_{l'}!} \tag{2}$$

母集団寸法指標の候補

標本寸法指標	S_0	S_1	S_2	S_3
s_0	k_{00}	k_{10}	k_{20}	k_{30}
s_1		k_{11}	k_{21}	k_{31}
s_2			k_{22}	k_{32}
s_3				k_{33}

図2：条件 C_1（母集団サイズから標本サイズへの移動パターン）

と近似できる。ただし

$$c_1 = \frac{1}{{}_N C_n \lambda^n (1-\lambda)^{N-n}}, \tag{3}$$

$$\mu_{l'} = \sum_{l=l'}^{3} S_l \cdot {}_l C_{l'} \lambda^{l'} (1-\lambda)^{l-l'} \tag{4}$$

であり，c_1 は S_l を含まない定数である。(2)式を最大にする (S_1, S_2, S_3) が近似的な最尤推定値となる。

以上の議論について，学会数を非負整数 L とする一般化は容易である。

ところで，個票データのリスク評価を目的に母集団寸法指標をノンパラメトリック推定する場合，上記の推定量は一般に非常に不安定であることが知られている。一方で，個票データから計算される寸法指標は単調減少などの共通の性質を持っていることが経験的に知られている。

佐井（2006）では，最大サイズが L の母集団寸法指標に対する制約条件として

(a)　$S_l \geq 0$（非負）

(b)　$\sum_{l=1}^{L} l \cdot S_l = N$（母集団の大きさが N）

(c)　$S_1 \geq S_2 \geq \cdots \geq S_L$（単調減少）

(d)　$2 \cdot \log S_l \leq \log S_{l-1} + \log S_{l+1}$（対数が下に凸）

を用い，これらをペナルティー関数に置き換えて，対数尤度関数に取り込む探索法を提案した。ただし，(b)はペナルティー関数とせずに，母集団の大きさを N とする補正を探索の各ステップにおいて行った。

ここで，のべ2767名分の個人レコードから1383レコードを非復元単純無作為抽出し，得られた標本寸法指標を基に母集団寸法指標の推定を行う。抽出率は $\lambda = 1383/2767 \doteqdot 0.4998$ である。

図３に，ピットマンモデルを用いた推定法，母集団寸法指標に対して制約(a)～(d)を課すノンパラメトリック推定法，(b)以外の制約を置かないノンパラメトリック推定法で，独立な10個の標本から推定された母集団寸法指標の推定値を示す。横軸はサイズ l，縦軸は $\log S_l$ を示す。また各図にお

いて太線は実際の母集団寸法指標 $(S_1, S_2, S_3) = (2035, 255, 74)$ を，細線はその推定値を示す。

ピットマンモデルは Pitman（1995）によって提案された確率分割モデルで，最大サイズを N とするモデルであるため，サイズ3での打ち切りにより特に S_3 が過小推定されている。この補正は容易ではない。

ノンパラメトリック推定法では推定誤差がやや大きいものの，偏りは小さく見える。この性質は佐井（2006）などにおける結果でも認められている。また，母集団寸法指標に対する制約条件を置く場合と置かない場合の結果に大きな違いは認められない。これは抽出率が大きいため，10回の推定のうち7回において，制約条件を置かなくとも推定値が制約条件を満たす領域で得られたためである。

図4，5，6に，図3の推定結果をそれぞれ3次元空間内にプロットした図を示す。3つの軸はそれぞれ S_1, S_2, S_3 を表し，塗りつぶされた部分は制約条件を満たす領域である。四角い点が真の母集団寸法指標を，他の10個の丸い点が推定値を示す。なお，各図において右側の図は，左の図の母集団寸法指標付近を拡大したものである。

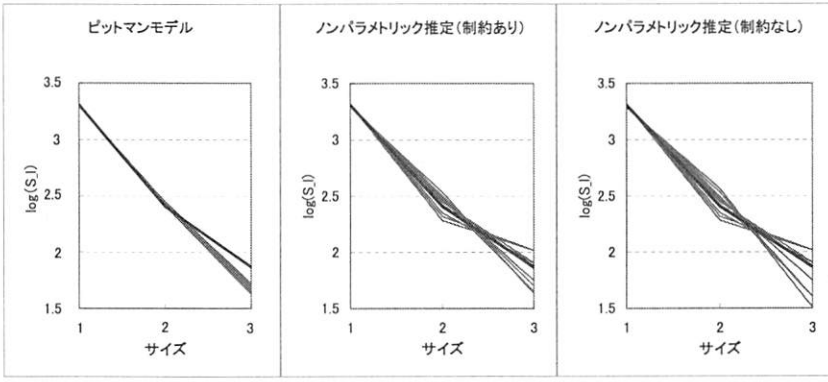


図3：母集団寸法指標の10回の推定結果

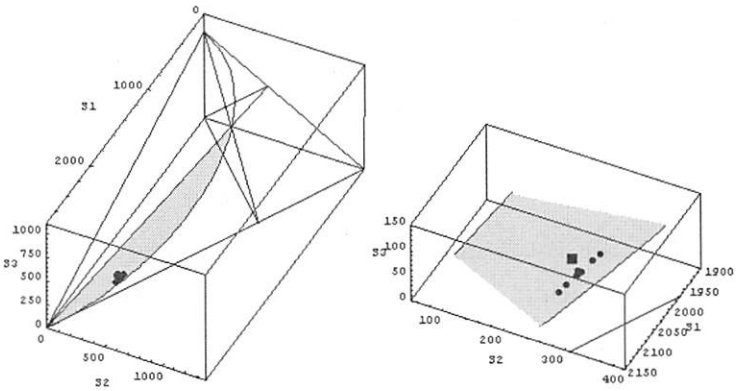


図4：ピットマンモデルでの推定結果

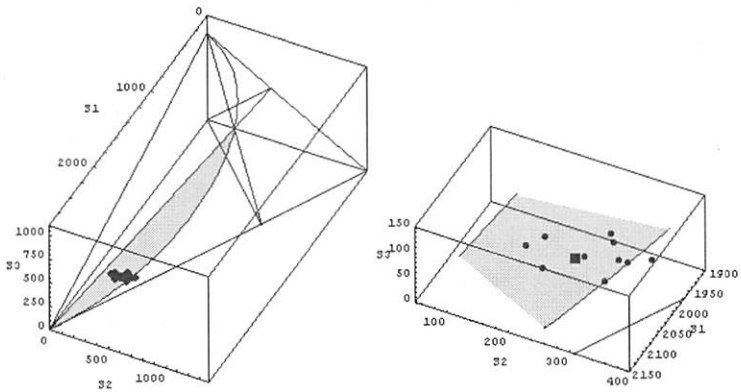


図5：ノンパラメトリック法（制約あり）での推定結果

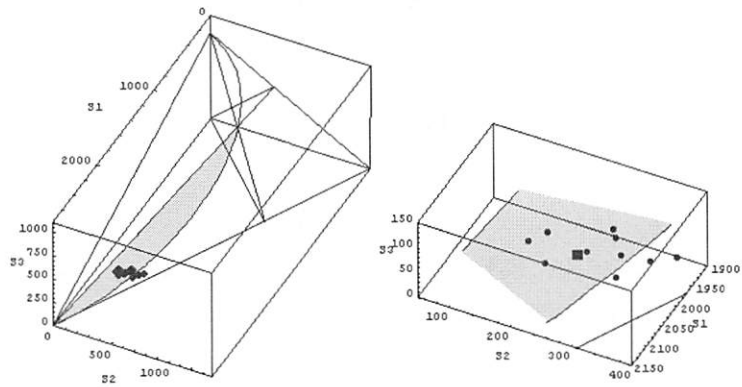


図6：ノンパラメトリック法（制約なし）での推定結果

4　多重寸法指標の導入

この節では所属学会問題に多重寸法指標を導入する。これにより所属する学会を区別した議論が可能となり，より自然なサンプリングを行うこともできる。

多重寸法指標については渋谷（2007），渋谷，佐井（2007）において説明されており，EPSF の多重指標確率分割への拡張も行われている。

3学会 G_1，G_2，G_3 について，学会 G_i に所属する人数を N_i とする。また，学会 G_i に所属することを $g_i = 1$ で，所属しないことを $g_i = 0$ で表し，所属パターン (g_1, g_2, g_3) に対する人数を $S_{g_1g_2g_3}$ と定義する。例えば S_{101} は G_1，G_3 にのみ所属する人数である。すべての所属パターンに対する人数を $\boldsymbol{S} = \{S_{100}, S_{010}, \ldots, S_{111}\}$ と表し，母集団多重寸法指標と呼ぶことにする。

学会 G_i から，他の学会とは独立に大きさ n_i の標本を非復元単純無作為抽出する。抽出率を $\lambda_i = n_i / N_i$ とする。また，母集団と同様に標本での所属パターンを (g_1', g_2', g_3') と表記し，標本多重寸法指標を $\boldsymbol{s} = \{s_{100}, s_{010}, \ldots, s_{111}\}$ と定義する。

ここで，標本多重寸法指標が得られたときの母集団多重寸法指標の尤度関数は，多変量超幾何分布を用いて

$$\mathrm{L}(\boldsymbol{S}\,|\,\boldsymbol{s}) = \frac{1}{\prod_{i=1}^{3} {}_{N_i}C_{n_i}} \sum_{C_2} \prod_{g_1,g_2,g_3} \left\{ \frac{S_{g_1g_2g_3}!}{\prod_{g_1',g_2',g_3'} k_{g_1g_2g_3,g_1'g_2'g_3'}!} \cdot \prod_{g_1',g_2',g_3'} \left(\prod_{i=1}^{3} {}_{g_i}C_{g_i'} \right)^{k_{g_1g_2g_3,g_1'g_2'g_3'}} \right\} \tag{5}$$

と表される。ただし，$k_{g_1g_2g_3,g_1'g_2'g_3'}$ は，母集団において所属パターン (g_1, g_2, g_3) の人のうち，標本において所属パターン (g_1', g_2', g_3') と観測される人数で，C_2 は図7に示す $\boldsymbol{S}$ から $\boldsymbol{s}$ が生成されるような $k_{g_1g_2g_3,g_1'g_2'g_3'}$ についてのすべての組み合わせを表す。積の (g_1, g_2, g_3) は $(0,0,0)$ を含まず，(g_1', g_2', g_3') は $(0,0,0)$ を含む。なお，(5)式の第2行の積は1となるが，6

節における設定の拡張のために省略せずに示している。

図８に，通常の寸法指標と多重寸法指標について，サイズの可能な変化パターンを矢印で示す。

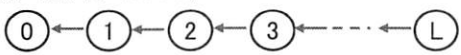
(5)式を，前節と同様にポアソン分布の確率関数の積で近似する，その議論は佐井（2002，2003）と同様である。

まず，学会 G_i の各レコードを他のレコードとは独立に確率 λ_i でサン

観測された標本多重寸法指標 ＼ 母集団多重寸法指標の候補	S_{000}	S_{100}	S_{010}	S_{001}	S_{110}	S_{011}	S_{101}	S_{111}
s_{000}	$k_{000,000}$	$k_{100,000}$	$k_{010,000}$	$k_{001,000}$	$k_{110,000}$	$k_{011,000}$	$k_{101,000}$	$k_{111,000}$
s_{100}		$k_{100,100}$			$k_{110,100}$		$k_{101,100}$	$k_{111,100}$
s_{010}			$k_{010,010}$		$k_{110,010}$	$k_{011,010}$		$k_{111,010}$
s_{001}				$k_{001,001}$		$k_{011,001}$	$k_{101,001}$	$k_{111,001}$
s_{110}					$k_{110,110}$			$k_{111,110}$
s_{011}						$k_{011,011}$		$k_{111,011}$
s_{101}							$k_{101,101}$	$k_{111,101}$
s_{111}								$k_{111,111}$

図７：条件 C_2（母集団サイズから標本サイズへの移動パターン）

通常の寸法指標の場合（最大サイズ：L）



多重寸法指標の場合（最大サイズ：111）

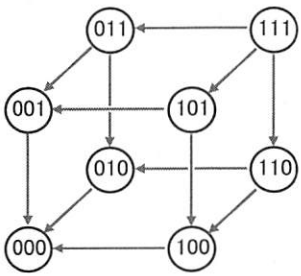


図８：母集団サイズから標本サイズへの移動の可能性

プリングする場合の尤度関数を考える。このような抽出法は Särndal et al. (1992) ではベルヌーイ抽出と呼ばれており，ここでもこの用語を用いる。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，標本の大きさを n_i，母集団の大きさを N_i にそれぞれ固定すると(5)式に対応する尤度関数は

$$L_B(\boldsymbol{S}\,|\,\boldsymbol{s}) = \prod_{i=1}^{3} \lambda_i^{n_i}(1-\lambda_i)^{N_i-n_i}$$

$$\cdot \sum_{C_2} \prod_{g_1,g_2,g_3} \left\{ \frac{S_{g_1g_2g_3}!}{\prod_{g_1',g_2',g_3'} k_{g_1g_2g_3,g_1'g_2'g_3'}!} \cdot \prod_{g_1',g_2',g_3'} \left(\prod_{i=1}^{3} {}_{g_i}C_{g_i'} \right)^{k_{g_1g_2g_3,g_1'g_2'g_3'}} \right\} \tag{6}$$

と書け，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単純無作為抽出の尤度関数の定数倍となる。従って，標本寸法指標が単純無作為抽出で得られたときの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，学会 G_i の大きさが N_i 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での各所属パターン (g_1, g_2, g_3) について，すべての $k_{g_1g_2g_3,g_1'g_2'g_3'}$ の組 $\boldsymbol{k}_{g_1,g_2,g_3}$ が，他の所属パターンとは独立に多項分布に従い，その確率関数は

$$f(\boldsymbol{k}_{g_1,g_2,g_3}) = \frac{S_{g_1g_2g_3}!}{\prod_{g_1',g_2',g_3'} k_{g_1g_2g_3,g_1'g_2'g_3'}!} \prod_{g_1',g_2',g_3'} \left\{ \prod_{i=1}^{3} {}_{g_i}C_{g_i'} \lambda_i^{g_i'}(1-\lambda_i)^{g_i-g_i'} \right\}^{k_{g_1g_2g_3,g_1'g_2'g_3'}} \tag{7}$$

である。ここですべての抽出率 λ_i が十分小さいものと仮定する。その場合，$\boldsymbol{k}_{g_1,g_2,g_3}$ のうち $k_{g_1g_2g_3,000}$ を除くと，その確率は小さく，互いの共分散も小さい。そこで $k_{g_1g_2g_3,g_1'g_2'g_3'}$ をそれぞれ独立なポアソン分布で

$$f\left(k_{g_1g_2g_3,g_1'g_2'g_3'}\right)$$
$$= e^{-\left\{S_{g_1g_2g_3}\cdot\prod_{i=1}^{3}{}_{g_i}C_{g_i'}\lambda_i^{g_i'}(1-\lambda_i)^{g_i-g_i'}\right\}}$$
$$\cdot\left\{S_{g_1g_2g_3}\cdot\prod_{i=1}^{3}{}_{g_i}C_{g_i'}\lambda_i^{g_i'}(1-\lambda_i)^{g_i-g_i'}\right\}^{k_{g_1g_2g_3,g_1'g_2'g_3'}}\Big/k_{g_1g_2g_3,g_1'g_2'g_3'}!\qquad(8)$$

と近似すると，ポアソン分布の再生性から $k_{g_1g_2g_3,g_1'g_2'g_3'}$ の (g_1,g_2,g_3) に関する和である標本多重寸法指標 $s_{g_1'g_2'g_3'}$ 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。

$$f(s_{g_1'g_2'g_3'})=\frac{e^{-\mu_{g_1'g_2'g_3'}}\ \mu_{g_1'g_2'g_3'}{}^{s_{g_1'g_2'g_3'}}}{s_{g_1'g_2'g_3'}!}.\qquad(9)$$

ただし

$$\mu_{g_1'g_2'g_3'}=\sum_{g_i(\geq g_i')}\left\{S_{g_1g_2g_3}\cdot\prod_{i=1}^{3}{}_{g_i}C_{g_i'}\lambda_i^{g_i'}(1-\lambda_i)^{g_i-g_i'}\right\}\qquad(10)$$

である。したがって尤度関数(5)式は

$$\mathrm{L}(S\,|\,s)=c_2\cdot\prod_{g_1',g_2',g_3'}\frac{e^{-\mu_{g_1'g_2'g_3'}}\ \mu_{g_1'g_2'g_3'}{}^{s_{g_1'g_2'g_3'}}}{s_{g_1'g_2'g_3'}!}\qquad(11)$$

と近似される。積の (g_1',g_2',g_3') は $(0,0,0)$ を含まない。ただし，

$$c_2=\frac{1}{\prod_{i=1}^{3}{}_{N_i}C_{n_i}\lambda_i^{n_i}(1-\lambda_i)^{N_i-n_i}}\qquad(12)$$

は $S_{g_1g_2g_3}$ を含まない定数である。

5　母集団多重寸法指標のノンパラメトリック推定

3学会 G_1，G_2，G_3 の母集団の大きさが $N_1=376$，$N_2=996$，$N_3=1395$ となっており，図1から母集団多重寸法指標は

$$\begin{aligned} \boldsymbol{S} &= \{S_{100}, S_{010}, S_{001}, S_{110}, S_{011}, S_{101}, S_{111}\} \\ &= \{157, 788, 1090, 24, 110, 121, 74\} \end{aligned}$$

と書けるが，ここでは未知と考えて推定目標とする。

各学会の個人レコードから，それぞれ大きさ $n_1=188$，$n_2=498$，$n_3=697$ の標本を独立に非復元無作為抽出することにする。抽出率は $\lambda_1=\lambda_2=0.5$，$\lambda_3 \fallingdotseq 0.5$ である。

得られた標本から標本多重寸法指標を求めると，例えば

$$\begin{aligned} \boldsymbol{s} &= \{s_{100}, s_{010}, s_{001}, s_{110}, s_{011}, s_{101}, s_{111}\} \\ &= \{122, 438, 608, 16, 39, 45, 5\} \end{aligned}$$

となる。

母集団多重寸法指標に対する制約条件については，3節に倣って

(A) $S_{g_1g_2g_3} \geq 0$

(B) $\sum_{g_1,g_2,g_3}(g_1+g_2+g_3)\cdot S_{g_1g_2g_3} = N$

(C) $S_{g_1g_2g_3} \geq S_{\underline{g}_1\underline{g}_2\underline{g}_3}\ (g_1 \leq \underline{g}_1,\ g_2 \leq \underline{g}_2,\ g_3 \leq \underline{g}_3)$

(D) $2\cdot\log S_{011} \leq \log S_{010} + \log S_{111}$
$2\cdot\log S_{011} \leq \log S_{001} + \log S_{111}$
$2\cdot\log S_{101} \leq \log S_{001} + \log S_{111}$

とする。ただし(C)では $g_1g_2g_3=000$ のすべての場合と $S_{110} \geq S_{111}$ を除くことにする。

多重寸法指標は構造が1次元的でなく，しかも個票データのリスク評価の場合と異なり，(C)の単調減少などの条件は成り立たない場合も考えられる。ここでは未知である母集団多重寸法指標の形状を基に制約条件を考えた。

図9に独立な10個の標本からノンパラメトリック推定された母集団多重寸法指標の推定値を示す。図は左から，制約条件(A)～(D)を用いた場合，(A)～(C)を用いた場合，(B)のみを用いた場合の結果を示し，横軸はサイズを，縦軸は $S_{g_1g_2g_3}$ を表す。また各図において太線は実際の母集団多重寸法指標

を，細線はその推定値を示す。

この場合は抽出率が高いため，母集団多重寸法指標に対して強い制約を課さなくても比較的安定した推定がなされている。

ここで推定の安定性を見るために，図10に示すような人工的な所属デー

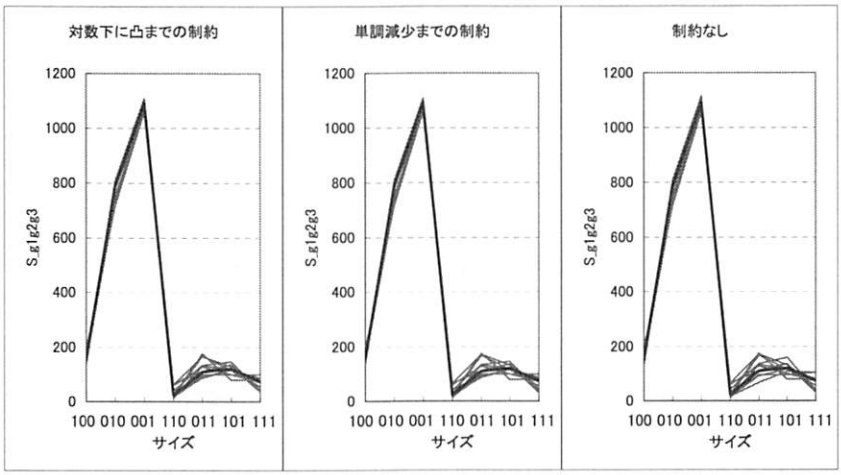


図９：母集団多重寸法指標の10回の推定結果

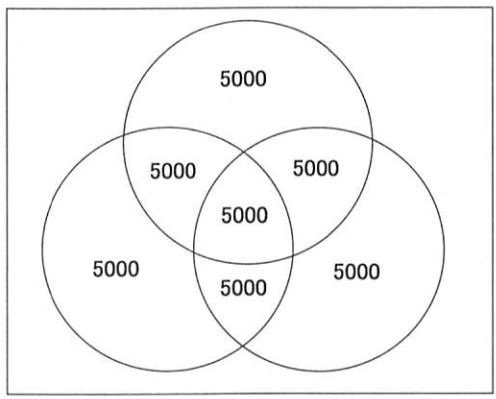


図10：人工的な所属データ

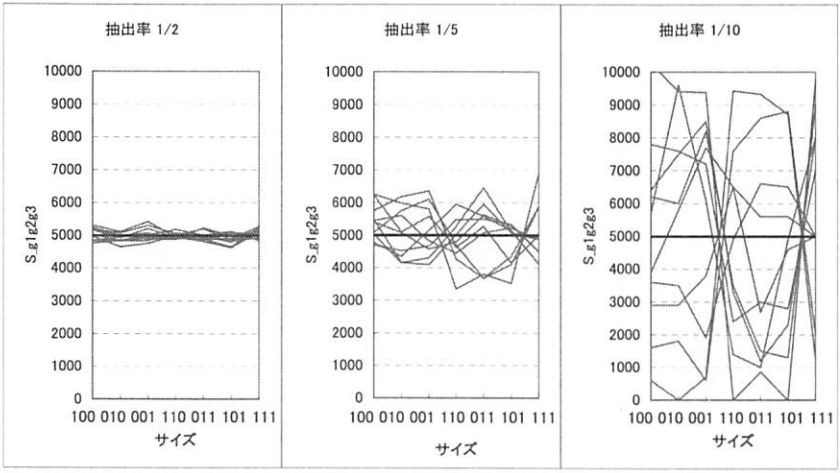


図11：10回の推定結果から見る推定の安定性

タを母集団として考える。サンプリング法や推定法は前述の方法と同様であるが，ここではすべての i に対して $\lambda_i = 0.5,\ 0.2,\ 0.1$ という３つの抽出率の場合について，推定の安定性を見る。なお制約条件としては(B)のみを課す。

結果を図11に示す。この程度の母集団の大きさの場合，ある程度抽出率が高くなければ推定は安定せず，母集団多重寸法指標に対する強めの制約条件が必要になることが分かる。

6　個票データのリスク評価に対する多重寸法指標の適用可能性

官庁統計調査など，標本調査で得られる個票データを公開する際のリスク評価法として，個票データから計算される標本寸法指標を基に母集団寸法指標を推定することについて，これまでも議論が重ねられてきた。

個票データは，個体（個人，事業所など）ごとに調査項目の結果の値をレコードとして並べたものである。調査項目のうち，年齢や性別のように，個体を特定するために用いられる項目をキー変数と呼ぶが，すべてのキー変数の組み合わせによってセルが構成されていると考える。例えば年

齢と性別のみがキー変数の場合，「40歳男性」などというセルが構成される。そのセルに含まれる個体数が l のとき，そのセルをサイズ l と呼び，そのようなセルの数を標本では s_l と，母集団では S_l と表し，寸法指標と呼んでいる。個票データが標本調査で得られている場合，標本寸法指標が観測され，その情報に基づいて母集団寸法指標を推定するが，特に S_1 など小さいサイズの頻度の推定が重要である。

個票データは1回の調査分についてまとめられているのが一般的である。パネル調査などの継続調査において同一個体が数回の調査でサンプリングされている場合でも，数回分のレコードをリンクした個票データを作成すると，リスク評価の方法は通常の個票データの場合と同様となる。

ここで新たに次のようなリスク評価方法を提案する。

同じ項目に関する数回にわたる継続調査があり，個体はその都度，非復元単純無作為抽出されていると考える。この設定は多くの官庁統計調査に当てはまる。例えば1ヶ月に1回ずつ m ヶ月にわたって同じ調査が行われたとする。年齢と性別のみがキー変数の場合，各月の個票データから「40歳男性」のセルが1つずつ構成されるが，このように個票データ（標本）において各月のサイズが $\boldsymbol{t}' = (t_1', t_2', ..., t_m')$ であるセル数を $s_{t'}$，母集団において各月のサイズが $\boldsymbol{t} = (t_1, t_2, ..., t_m)$ であるセル数を S_t とし，すべてのサイズの組み合わせに対する組を4節のように，それぞれ $\boldsymbol{s}$ と $\boldsymbol{S}$ で表す。このとき $\boldsymbol{s}$ を基にした $\boldsymbol{S}$ のノンパラメトリック推定は，本報告で所属学会問題に対して提案したものと同様となるが，所属学会問題では多重寸法指標の添字が $g_i = 0, 1$ だったのに対して，$t_i = 0, 1, ..., L_i$ となる点が異なる。ただし L_i はセル内の個体数の最大値である。

4節と同様の定義の下で，非復元単純無作為抽出によって標本多重寸法指標 $\boldsymbol{s}$ が得られたときの母集団多重寸法指標の尤度関数は，

$$\mathrm{L}(\boldsymbol{S}\,|\,\boldsymbol{s}) = \frac{1}{\prod_{i=1}^{m} {}_{N_i}C_{n_i}} \sum_{C_2} \prod_{t} \left\{ \frac{S_t!}{\prod_{t'} k_{t,t'}!} \cdot \prod_{t'} \left(\prod_{i=1}^{m} {}_{t_i}C_{t_i'} \right)^{k_{t,t'}} \right\} \tag{13}$$

と書くことができ，ポアソン近似された尤度関数は

$$L(S|s) = c_3 \cdot \prod_{t'} \frac{e^{-\mu_{t'}} \mu_{t'}^{s_{t'}}}{s_{t'}!} \tag{14}$$

と書ける。ただし

$$c_3 = \frac{1}{\prod_{i=1}^{m} {}_{N_i}C_{n_i} \lambda_i^{n_i} (1-\lambda_i)^{N_i - n_i}}, \tag{15}$$

$$\mu_{t'} = \sum_{t(\geq t')} \left\{ S_t \cdot \prod_{i=1}^{m} {}_{t_i}C_{t_i'} \lambda_i^{t_i'} (1-\lambda_i)^{t_i - t_i'} \right\} \tag{16}$$

である。

図12に，３回（１月，２月，３月）の継続調査における母集団と標本におけるセルのイメージを示す。

母集団多重寸法指標の推定によって，例えば３回の調査とも母集団においてサイズ１であるセル数 S_{111} の推定が可能となる。個々の個票データから推定される母集団でサイズ１であるセル数 S_1 だけでは，毎回の調査で同じセルがサイズ１となる傾向が強いか，あるいは毎回異なるセルがサイズ１となる傾向が強いかを判断することはできない。母集団多重寸法指

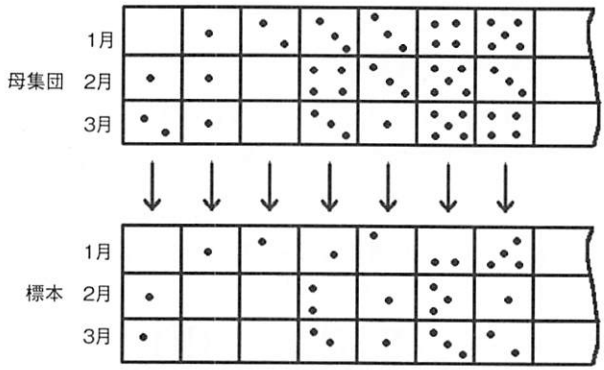


図12：継続調査におけるセルのイメージ

標を推定することにより，S_{111} のように危険性の高い個体数の推定が可能となり，個票データに秘匿措置を施す際にも，秘匿するターゲットを絞ることも可能となるであろう。

7　おわりに

本報告では，寸法指標を拡張した多重寸法指標に対するノンパラメトリック推定法を提案した。

いくつかの学会に重複して所属している会員がある場合，所属会員リストからサンプリングした標本の情報を基に母集団での所属状況を推定する，いわゆる所属学会問題に対して寸法指標を導入することにより，単純な設定の下での推定が可能となった。より自然な設定の下での推定を可能とするために多重寸法指標を導入し，計算時間上の問題を克服するためにポアソン近似を用いたノンパラメトリック推定法を提案した。

多重寸法指標を官庁統計などで得られる個票データのリスク評価に対して応用する点についても，その方法を述べるとともに重要性を指摘したが，今後は実際の個票データについて適用し，その有効性を議論することが必要となろう。

謝　辞

本論文で使用した所属学会データを提供していただいた統計数理研究所の馬場康維先生に感謝致します。また所属学会データに寸法指標を適用するアイディアを提供してくださった金沢大学の星野伸明先生に感謝致します。慶應義塾大学名誉教授の渋谷政昭先生からは，多重指標に関する数々の有益な助言をいただきました。ここに感謝致します。

本論文は科学研究費補助金（課題番号19300098，18200019，16203014），統計数理研究所共同利用研究プログラム（19－共研－2028）の研究成果に基づくものである。

参考文献

[1] Bethlehem, J. G., Keller, W. J. and Pannekoek, J. (1990). Disclosure control of microdata, *Journal of the American Statistical Association*, **85**, 38–45.
[2] Bureau of the Census (1993). 1990 Census of Population and Housing, Public Use Microdata Samples (microdata), Washington, D. C.
[3] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions, *Probability Theory and Related Fields*, **102**, 145–148.
[4] Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer, New York.
[5] 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，第37巻，第3号，61－79.
[6] 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，第51巻，第2号，183－198.
[7] 佐井至道（2005）．部分母集団の情報を用いた母集団寸法指標の推定とノンパラメトリック推定の改良，岡山商大論叢，第40巻，第3号，33－53.
[8] 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，第42巻，第1号，1－21.
[9] 佐井至道（2007a）．最大サイズが有限である母集団寸法指標の推定について，科学研究費補助金・統計数理研究所共同利用研究合同研究集会「官庁統計データの公開における諸問題の研究」資料，1－4.
[10] 佐井至道（2007b）．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.
[11] 渋谷政昭（2007）．多重指標の確率分布，科学研究費補助金・統計数理研究所共同利用研究合同研究集会「官庁統計データの公開における諸問題の研究」資料，1－18.
[12] 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.
[13] 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.
[14] 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.