統計数理　第36巻　第1号（1988）

有限母集団における順位情報を用いる母平均の推定

弘前大学理学部　高　橋　宏　一*
二ッ矢　昌　夫

（1988年3月　受付）

1. 序

連続分布から，大きさ n の標本を独立に n 組抽出する．第 i 組の標本を $X_{i1}, X_{i2}, \dots, X_{in}$，その順序統計量を $X_{i(1)}, X_{i(2)}, \dots, X_{i(n)}$ とする．$\bar{Y}_{[n]}=\frac{1}{n}\sum_{i=1}^{n} X_{i(i)}$ と定義する．一方，大きさ n の標本の標本平均を $\bar{X}_n$ とする．勿論，$\bar{X}_n$ は母平均 μ の不偏推定量であり，分散は σ^2/n（σ^2 は母分散）である．McIntyre (1952), Takahasi and Wakimoto (1968) は，$\bar{Y}_{[n]}$ も μ の不偏推定量であり，$\bar{Y}_{[n]}$ の $\bar{X}_n$ に対する相対効率 $e_{[n]}=(\bar{X}_n$ の分散$/\bar{Y}_{[n]}$ の分散）は常に1より大きく（$n\geqq 2$ とする），$(n+1)/2$ 以下であることを示した．一様分布に対しては $e_{[n]}=(n+1)/2$ であり，$e_{[2]}$ についてみると，正規分布で1.47，指数分布で1.33，両側指数分布で1.39などである．

上記の推定量 $\bar{Y}_{[n]}$ が使用されるのは，母集団から抽出された大きさ n の標本の順位は一見しただけでわかるが，個々の要素の測定は困難であるといった状況のときである．たとえば，顕微鏡の視野にある細胞の長さのミクロメーターによる測定において長さの実測はある程度の時間を必要とするが，近くにいる2，3個の細胞の長さの順位は実測しなくとも一目でわかるというような場合である．順位付けが容易であるという状況を標本の抽出法にとり入れて，より有効に母平均を推定しようとする方法（the method of ranked set sampling）は，上記の2つの論文のほかに，Takahasi (1969, 1970), Dell and Clutter (1972), 柳川，白旗（1974), Yanagawa and Shirahata (1976), Stokes (1977), Yanagawa and Chen (1980) などにおいて，いろいろな面から検討されている．しかしながら，いずれにおいても順位付けのなされた要素からなる集合同士には独立性を仮定しているようである．

この論文では，有限母集団からの非復元抽出標本における順位の情報を利用した母平均の推定を問題にする．元に戻って連続分布の場合の $\bar{Y}_{[2]}$ を考えてみよう．通常の大きさ2の標本の標本平均 $\bar{X}_2$ に対する相対効率 $e_{[2]}$ は，前述のように，最大のとき（一様分布）で1.5であり，また，多くの分布でこれに近い値を示す（Takahasi and Wakimoto (1968) の Table 1 参照）．ところで，目的が有限母集団の母平均の推定であり，非復元抽出標本が使用されるときには，標本に含まれる要素間の相関の故に事情が異なってくることが想像される．たとえば，図1の正方形内の点の総数（1000個ある）を16個の小正方形を非復元抽出し，そこでの総数を4倍して推定してみよう．これを20回繰り返した結果が図2の下側である．一方，左右に隣り合う小正

* 昭和62年4月から63年3月まで統計数理研究所客員教官．

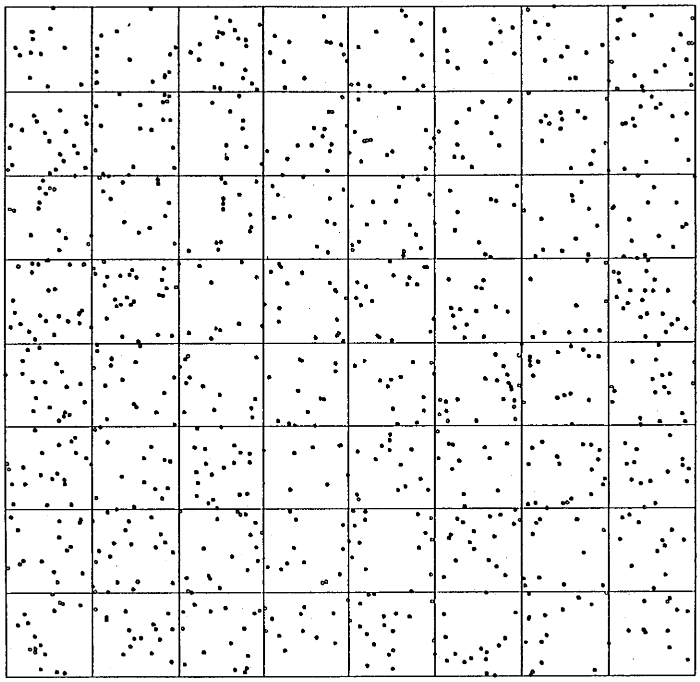

図1.　有限母集団の例

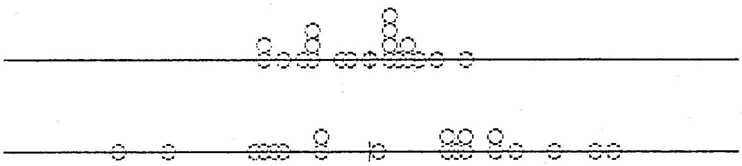

図2.　標本平均との比較（図1が母集団）

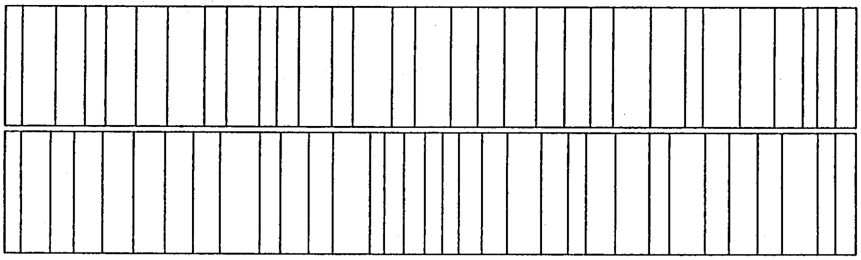

図3.　有限母集団の例

方形を組にして，全体を32組に分割しておき，そこから16組を非復元抽出し，最初の8組では個数の多い小正方形を，残りの8組では個数の少ない小正方形を選んで，合計16個の小正方形の総数を4倍して推定するということを20回繰り返した結果が図2の上側である．偏差平方の平均を求めてみると，前者で3124.6，後者で551.6であり，前者は後者の5.7倍になっている．同じことを図3の小長方形の平均幅の推定について行った結果が図4である．偏差平方の平均の比は2.1になっている．注意したいことは，図1，3は有限母集団であり，その平均値な

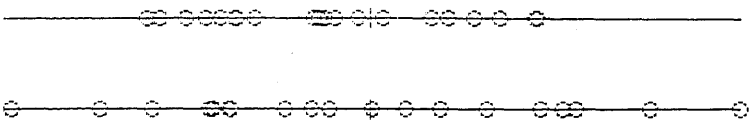

図 4.　標本平均との比較（図 3 が母集団）

り総計値の推定が目的であるということである（何らかの母集団からの実現値とみなし，背後の母集団の平均値を推定しようというのではない）.

この簡単な 2 つの実験例からでも，有限母集団の非復元抽出標本における順位情報の利用は，通常の連続分布の場合にくらべて，その効果が大きいことが予想される.

第 2 章で問題を定式化し，相対効率の表現を求める．第 3 章では種々の有限母集団モデルに対する相対効率を計算する.

2.　有限母集団の母平均推定に対する順位情報の利用

大きさ N の有限母集団の第 i 要素のもつ特性 X の値を x_i とする．$x_1, x_2, \dots, x_N$ を大きさの順に並べかえたものを $x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(N)}$ とする．この母集団から大きさ $n^2 r$ の単純無作為標本（非復元）をとる．これを $X_{ij}^{(k)}$ $(i, j=1, 2, \dots, n;\ k=1, 2, \dots, r)$ で表す．いいかえると，大きさ n の単純無作為標本を非復元で nr 回抽出する．各 i, k に対して，$X_{i1}^{(k)}, X_{i2}^{(k)}, \dots, X_{in}^{(k)}$ の順序統計量を

(2.1)
$$X_{i(1)}^{(k)} \leqq X_{i(2)}^{(k)} \leqq \cdots \leqq X_{i(n)}^{(k)}$$

とおく．この記号を用いて

(2.2)
$$Y_{[n]}^{(r)} = \frac{1}{nr} \sum_{k=1}^{r} \sum_{i=1}^{n} X_{i(i)}^{(k)}$$

と定義する．母平均を μ，母分散を σ^2 とする．この $Y_{[n]}^{(r)}$ が μ の不偏推定量であることは容易に示される(後述)．ところで，この $Y_{[n]}^{(r)}$ は大きさ $n^2 r$ の標本に基づいているが，(2.2) からわかるように，測定値を実際に用いているのは nr 個の要素に対してであって，残りの $n(n-1)r$ 個の要素は (2.1) において順位を定める際に用いられているだけである．第 1 章で述べたように，我々は一見して順位がわかるような状況を考えているので，μ の不偏推定量としての $Y_{[n]}^{(r)}$ の分散を，大きさ nr の単純無作為標本の標本平均 $\bar{X}_{nr}$ の分散

(2.3)
$$\frac{N-nr}{N-1} \cdot \frac{\sigma^2}{nr}$$

と比較することにする．相対効率 $e_{[n]}^{(r)}$ を

(2.4)
$$e_{[n]}^{(r)} = \frac{\mathrm{Var}\, \bar{X}_{nr}}{\mathrm{Var}\, Y_{[n]}^{(r)}}$$

で定義する.

大きさ n の単純無作為標本 $X_1, \dots, X_n$ の順序統計量を $X_{(1)} \leqq X_{(2)} \leqq \cdots \leqq X_{(n)}$ とし，

$$\mu_{n:i}=\mathrm{E}X_{(i)} \tag{2.5}$$

とおく．$\sum_{i=1}^{n} X_i=\sum_{i=1}^{n} X_{(i)}$ から直ちに

$$\mu=\frac{1}{n}\sum_{i=1}^{n}\mu_{n:i} \tag{2.6}$$

を得る．この関係から

$$\mathrm{E}\,Y_{[n]}^{(r)}=\frac{1}{nr}\sum_{k=1}^{r}\sum_{i=1}^{n}\mu_{n:i}=\mu \tag{2.7}$$

がわかる．問題は $\mathrm{Var}\ Y_{[n]}^{(r)}$ を求めることである．

$$\begin{aligned}\mathrm{Var}\ Y_{[n]}^{(r)}&=\mathrm{E}\left(\frac{1}{nr}\sum_{k=1}^{r}\sum_{i=1}^{n}X_{i(i)}^{(k)}-\mu\right)^2\\&=\mathrm{E}\left(\frac{1}{nr}\sum_{k=1}^{r}\sum_{i=1}^{n}(X_{i(i)}^{(k)}-\mu_{n:i})\right)^2\\&=\frac{1}{n^2r^2}\sum_{k=1}^{r}\sum_{i=1}^{n}\sum_{h=1}^{r}\sum_{j=1}^{n}\mathrm{Cov}(X_{i(i)}^{(k)},\ X_{j(j)}^{(h)})\\&=\frac{1}{n^2r^2}\left\{\sum_{k=1}^{r}\sum_{i=1}^{n}\mathrm{Cov}(X_{i(i)}^{(k)},\ X_{i(i)}^{(k)})\right.\\&\qquad\left.+\sum_{k\neq h}^{r}\sum^{r}\sum_{i=1}^{n}\mathrm{Cov}(X_{i(i)}^{(k)},\ X_{i(i)}^{(h)})+\sum_{k=1}^{r}\sum_{h=1}^{r}\sum_{i\neq j}^{n}\sum^{n}\mathrm{Cov}(X_{i(i)}^{(k)},\ X_{i(i)}^{(h)})\right\}\\&=\frac{1}{n^2r^2}\left\{\sum_{i=1}^{n}\tilde{\alpha}_{n:i,i}+(r-1)\sum_{i=1}^{n}\tilde{\gamma}_{n:i,i}+r\sum_{i\neq j}^{n}\sum^{n}\tilde{\gamma}_{n:i,j}\right\}\\&=\frac{1}{n^2r^2}\left\{\sum_{i=1}^{n}\tilde{\alpha}_{n:i,i}-\sum_{i=1}^{n}\tilde{\gamma}_{n:i,i}+r\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\gamma}_{n:i,j}\right\},\end{aligned}\tag{2.8}$$

ただし，$\tilde{\alpha}_{n:i,j}$ は大きさ n の標本の i 番目と j 番目の順序統計量の共分散，$\tilde{\gamma}_{n:i,j}$ は大きさ $2n$ の標本 $(X_1,\ldots,X_n,\ Y_1,\ldots,Y_n)$ の前半 $X_1,\ldots,X_n$ の順序統計量を $X_{(1)},\ldots,X_{(n)}$, 後半 $Y_1,\ldots,Y_n$ の順序統計量を $Y_{(1)},\ldots,Y_{(n)}$ とするとき，$X_{(i)}$ と $Y_{(j)}$ の共分散である．ところで

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\gamma}_{n:i,j}&=\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{(i)}-\mu_{n:i})(Y_{(j)}-\mu_{n:j})\right)\\&=n^2\,\mathrm{Cov}(\bar{X},\ \bar{Y})=-\frac{n^2\sigma^2}{N-1},\end{aligned}\tag{2.9}$$

$$\begin{aligned}\sum_{i=1}^{n}\tilde{\alpha}_{n:i,i}&=\sum_{i=1}^{n}(\alpha_{n:i,i}-\mu_{n:i}^2)=n(\sigma^2+\mu^2)-\sum_{i=1}^{n}\mu_{n:i}^2\\&=n\sigma^2-\sum_{i=1}^{n}(\mu_{n:i}-\mu)^2\end{aligned}\tag{2.10}$$

である．ただし，$\alpha_{n:i,j}=\mathrm{E}X_{(i)}X_{(j)}$ である．(2.9), (2.10) を (2.8) に代入することによって

$$\begin{aligned}\mathrm{Var}\ Y_{[n]}^{(r)}&=\frac{N-nr}{N-1}\cdot\frac{\sigma^2}{nr}-\frac{1}{r}\left\{\frac{\sigma^2}{(N-1)n}+\frac{1}{n^2}\sum_{i=1}^{n}(\mu_{n:i}-\mu)^2+\frac{1}{n^2}\sum_{i=1}^{n}\tilde{\gamma}_{n:i,i}\right\}\\&=\mathrm{Var}\ \bar{X}_{nr}-\frac{1}{r}\left\{\frac{\sigma^2}{(N-1)n}+\frac{1}{n^2}\sum_{i=1}^{n}(\mu_{n:i}-\mu)^2+\frac{1}{n^2}\sum_{i=1}^{n}\tilde{\gamma}_{n:i,i}\right\},\end{aligned}\tag{2.11}$$

を得る．(2.11)式の｛　｝の中の量は r に依存しない．この量を g とおこう．すなわち

$$g=\frac{\sigma^2}{(N-1)n}+\frac{1}{n^2}\sum_{i=1}^{n}(\mu_{n:i}-\mu)^2+\frac{1}{n^2}\sum_{i=1}^{n}\tilde{\gamma}_{n:i,i}$$

である．(2.9)を σ^2 について解いたものを g 中の σ^2 に代入して整頓すると

$$g=\frac{1}{n^2}\left\{\left(1-\frac{1}{n}\right)\sum_{i=1}^{n}\gamma_{n:i,i}-\frac{1}{n}\sum_{i\neq j}^{n}\sum^{n}\gamma_{n:i,j}\right\} \tag{2.12}$$

を得る．ただし，$\gamma_{n:i,j}=\mathrm{E}X_{(i)}Y_{(j)}$ である．

ここで $n=2$ としてみる．$\gamma_{2:i,j}$ は大きさ2の標本を2組非復元抽出したときの第1組の i 番目の順序統計量 $X_{(i)}$ と第2組の j 番目の順序統計量 $Y_{(j)}$ の積の期待値であるから，これを大きさ4の非復元抽出標本の順序統計量を用いて表現することが考えられる．大きさ4の標本を $Z_1,\ldots,Z_4$, 順序統計量を $Z_{(1)}\leqq\cdots\leqq Z_{(4)}$ とする．このとき，たとえば

$$\begin{aligned}\gamma_{2:1,1}&=\mathrm{E}X_{(1)}Y_{(1)}=\mathrm{E}(\min\{Z_1,\ Z_2\}\cdot\min\{Z_3,\ Z_4\})\\&=\mathrm{E}[\mathrm{E}(\min\{Z_1,\ Z_2\}\cdot\min\{Z_3,\ Z_4\}\mid Z_{(1)},\ldots,Z_{(4)})]\\&=\mathrm{E}\left(\frac{2}{3}Z_{(1)}Z_{(2)}+\frac{1}{3}Z_{(1)}Z_{(3)}\right)\\&=\frac{2}{3}\alpha_{4:1,2}+\frac{1}{3}\alpha_{4:1,3}\end{aligned} \tag{2.13}$$

となることは，$Z_{(1)},\ldots,Z_{(4)}$ の順列を書き上げることによって容易にわかる．同じようにして

$$\gamma_{2:2,2}=\frac{2}{3}\alpha_{4:3,4}+\frac{1}{3}\alpha_{4:2,4} \tag{2.14}$$

を得る．これらを(2.12)に代入して

$$g=\frac{1}{12}(\alpha_{4:1,2}+\alpha_{4:3,4}-\alpha_{4:2,3}-\alpha_{4:1,4}) \tag{2.15}$$

を得る．$\alpha_{4:i,j}=\mathrm{E}(Z_{(i)}Z_{(j)})$ を代入し整頓すると

$$g=\frac{1}{12}\mathrm{E}[(Z_{(3)}-Z_{(1)})(Z_{(4)}-Z_{(2)})] \tag{2.16}$$

となる．母集団の要素の特性値の集合 $\{x_1,\ldots,x_N\}$ が相異なる値を少なくとも3つ含んでいれば(2.16)の右辺は明らかに正である．また，集合 $\{x_1,\ldots,x_N\}$ が2種類の値しか含まない場合でも，それぞれの値をもつ要素が複数個ずつあるならば(2.16)の右辺が正なることも明らかである．また，標本抽出は非復元を仮定しているので，(2.16)が意味をもつためには少なくとも $N\geqq 4$ が必要である．一般的にいうと，非復元抽出が可能なためには $N\geqq n^2r$ が必要である．以上，$n=2$ の場合をまとめて次の定理を得る．

定理1. $r\geqq 1,\ N\geqq 4r$ とする．大きさ N の有限母集団の要素の特性値の集合 $\{x_1,\ldots,x_N\}$ は少なくとも相異なる値を3つ含むか，2つしか含まない場合はそれぞれの値をもつ要素が複数個ずつあるものとする．この母集団からの大きさ4の単純無作為標本（非復元）の順序統計量を $Z_{(1)}\leqq\cdots\leqq Z_{(4)}$ とし，

$$g=\frac{1}{12}\mathrm{E}[(Z_{(3)}-Z_{(1)})(Z_{(4)}-Z_{(2)})]$$

とおく．このとき

（i）

$$g>0, \tag{2.17}$$

(ii)

(2.18) $$\mathrm{Var}\ Y_{[2]}^{(r)}=\mathrm{Var}\ \bar{X}_{2r}-\frac{g}{r}<\mathrm{Var}\ \bar{X}_{2r}$$

(iii)

(2.19) $$e_{[2]}^{(r)}=\left(1-\frac{N-1}{N-2r}\cdot\frac{2g}{\sigma^2}\right)^{-1},$$

(iv) $\tau_{[2]}^{(r)}=\dfrac{\mathrm{Var}\ \bar{X}_{2r}-\mathrm{Var}\ Y_{[2]}^{(r)}}{\mathrm{Var}\ \bar{X}_{2r}}$ とおけば

(2.20) $$\tau_{[2]}^{(r)}=\frac{N-1}{N-2r}\cdot\frac{2g}{\sigma^2}$$

である．ただし，$\bar{X}_{2r}$ は大きさ $2r$ の単純無作為標本の標本平均である．

定理1の母集団に関する制限は，実際的には，ほとんど考慮する必要がない程弱いものである．定理1の主要な結果である $\mathrm{Var}\ Y_{[2]}^{(r)}<\mathrm{Var}\ \bar{X}_{2r}$ に対応して，一般の n についても $\mathrm{Var}\ Y_{[n]}^{(r)}<\mathrm{Var}\ \bar{X}_{nr}$ が同じような方法で導かれることが予想されるが，現在のところ(2.16)に対応する簡単な表現は求まっていない．

定理1を応用するために，ここで $\alpha_{4:i,j}$ を求めておく．ただし，ここからは計算の都合のため母集団の各要素はすべて相異なる値をもつものと仮定する．大きさの順に並べかえたものを $x_{(1)}<x_{(2)}<\cdots<x_{(N)}$ とする．$(Z_{(i)},\ Z_{(j)})$ の同時分布は

(2.21)
$$\begin{aligned}
&\Pr\{Z_{(1)}=x_{(i)},\ Z_{(2)}=x_{(j)}\}=\binom{N-j}{2}\Big/\binom{N}{4}, && (1\leqq i<j\leqq N-2)\\
&\Pr\{Z_{(3)}=x_{(i)},\ Z_{(4)}=x_{(j)}\}=\binom{i-1}{2}\Big/\binom{N}{4}, && (3\leqq i<j\leqq N)\\
&\Pr\{Z_{(2)}=x_{(i)},\ Z_{(3)}=x_{(j)}\}=\binom{i-1}{1}\binom{N-j}{1}\Big/\binom{N}{4}, && (2\leqq i<j\leqq N-1)\\
&\Pr\{Z_{(1)}=x_{(i)},\ Z_{(4)}=x_{(j)}\}=\binom{j-1-i}{2}\Big/\binom{N}{4} && (1\leqq i<i+3\leqq j\leqq N)
\end{aligned}$$

となる．これより

(2.22)
$$\begin{aligned}
\alpha_{4:1,2}&=\frac{1}{\binom{N}{4}}\sum\sum_{1\leqq i<j\leqq N-2}\binom{N-j}{2}x_{(i)}x_{(j)},\\
\alpha_{4:3,4}&=\frac{1}{\binom{N}{4}}\sum\sum_{3\leqq i<j\leqq N}\binom{i-1}{2}x_{(i)}x_{(j)},\\
\alpha_{4:1,4}&=\frac{1}{\binom{N}{4}}\sum\sum_{1\leqq i<i+3\leqq j\leqq N}\binom{j-1-i}{2}x_{(i)}x_{(j)},\\
\alpha_{4:2,3}&=\frac{1}{\binom{N}{4}}\sum\sum_{2\leqq i<j\leqq N-1}\binom{i-1}{1}\binom{N-j}{1}x_{(i)}x_{(j)}
\end{aligned}$$

となる．これらを(2.15)に代入して

$$g=\frac{2}{N(N-1)(N-2)(N-3)} \cdot \sum\sum_{1\leqq i<j\leqq N}\left\{\binom{i-1}{2}+\binom{N-j}{2}-\binom{j-1-i}{2}-\binom{i-1}{1}\binom{N-j}{1}\right\}x_{(i)}x_{(j)}, \tag{2.23}$$

ただし，$\binom{a}{b}$ は a が負のときや，$a<b$ のときは0と約束する．

例1． 離散的一様分布 $x_{(i)}=i$ $(i=1, 2, \ldots, N)$ の場合，(2.23)を計算すると

$$g=\frac{1}{360}(N+1)(5N+4) \tag{2.24}$$

となる．$\sigma^2=\frac{(N-1)(N+1)}{12}$ であるから

$$\tau_{[2]}^{(r)}=\frac{5N+4}{15(N-2r)} \tag{2.25}$$

$$e_{[2]}^{(r)}=\frac{15N-30r}{10N-30r-4}=\frac{3N-6r}{2N-6r-\frac{4}{5}} \tag{2.26}$$

となる．$e_{[2]}^{(r)}$ は N について単調減少，r について単調増加であることは容易にわかる．r を固定したとき，N の範囲は $4r\leqq N<\infty$ であることに注意すれば

$$\frac{3}{2}<e_{[2]}^{(r)}\leqq\frac{15r}{5r-2}=3+\frac{6}{5r-2} \tag{2.27}$$

を得る．(2.27)の右辺は r について単調減少であるから，結局

$$\frac{3}{2}<e_{[2]}^{(r)}\leqq 5 \tag{2.28}$$

である．すなわち，離散的一様分布の有限母集団においては，相対効率は常に連続分布の場合の上限3/2より大きく，$r=1, N=4$ のときに最大値5をとる．r を固定し，$N\to\infty$ とすると，

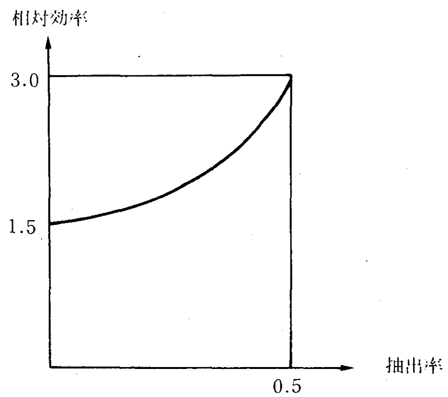

図5．相対効率

これは無限母集団に対応することになるが，このときに相対効率は下限の 3/2 になる．また $r=[N/4]$ ($N/4$ の整数部分を表す) として $N\to\infty$ にすれば相対効率は 3 に近づいていく．$f=\frac{2r}{N}$ (抽出率) を用いて相対効率を表せば

$$(2.29)\qquad e_{[2]}^{(r)}=(3-3f)\Big/\left(2-3f-\frac{4}{5N}\right)$$

であり，N が十分大きいときは

$$(2.30)\qquad \frac{3-3f}{2-3f}=\frac{3}{2}\left(1+\frac{1}{\frac{2}{f}-3}\right),\qquad \left(0<f\leqq\frac{1}{2}\right)$$

によって近似できる．このグラフを図 5 に示す．

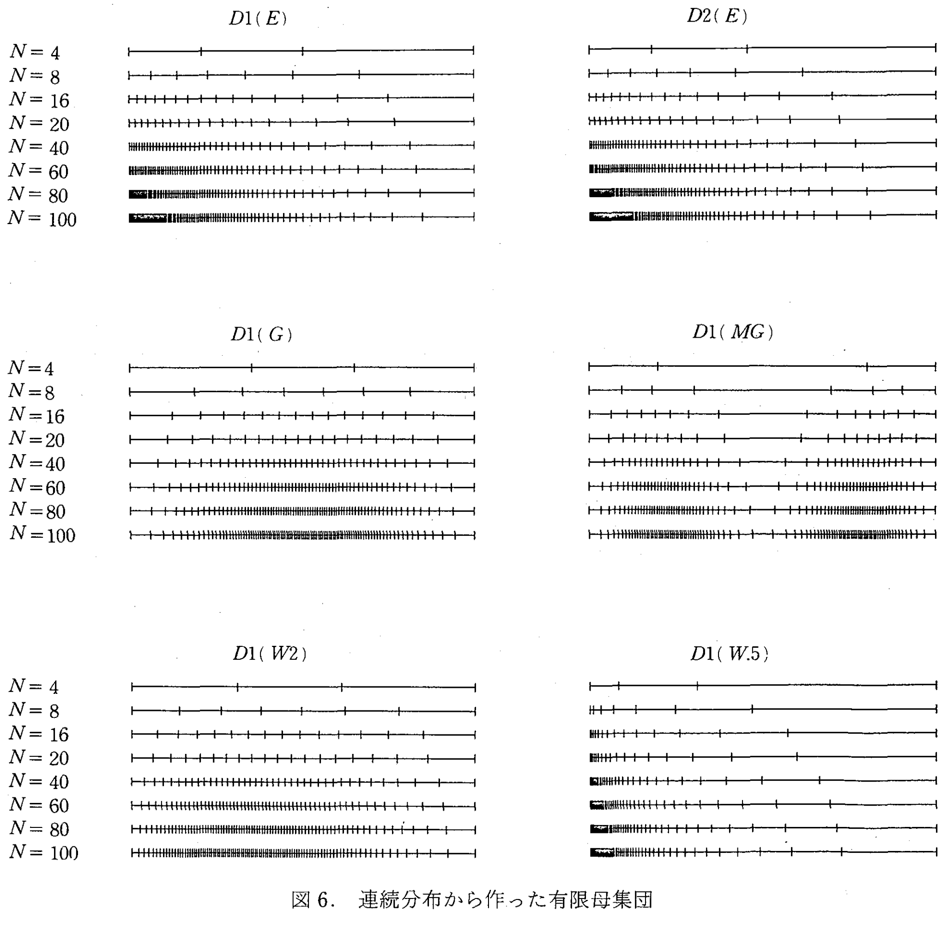

図 6. 連続分布から作った有限母集団

3. 種々の有限母集団における相対効率の数値例

$n=2$ の場合の相対効率 $e_{[2]}^{(r)}$ は, (2.19), (2.23) を用いることによって, 任意の有限母集団に対して数値的に求めることができる. この章の目的は相対効率が母集団のタイプによって, どのように変化するかを調べることであるが, 有限母集団を次のいずれかで作り出す:
ある連続分布の分布関数を $F(x)$ とする.

(i) $x_{(i)}=F\left(\dfrac{i}{N+1}\right),\qquad i=1, 2, \ldots, N,$

(ii) $x_{(i)}=F(x)$ からの大きさ N の標本の i 番目の順序統計量の期待値, $i=1, 2, \ldots, N,$

(iii) $F(x)$ からの大きさ N の標本の実現値.

連続分布としては, 一様分布 (U で表す), 正規分布 (G で表す), 指数分布 (E で表す), 型母数 2 及び 0.5 のワイブル分布 ($W2$, $W.5$ で表す), 型母数 2 及び 0.5 のガンマ分布 ($\Gamma 2$, $\Gamma .5$ で表す), $N(0, 1)$ と $N(5, 1)$ の混合比 1/2 の混合分布 (MG で表す) をとり上げる. $D1(F)$, $D2(F)$ $D3(F)$ でそれぞれ連続分布 F から (i), (ii), (iii) の方法で作った有限母集団を表すことにする.

表 1. R_N の値

母集団 \ N	4	8	16	20	40	60	80	100	200	400
$D1(U)$	0.533	0.419	0.373	0.365	0.349	0.344	0.341	0.339	0.336	0.335
$D1(G)$	0.517	0.409	0.364	0.356	0.339	0.333	0.330	0.328	0.324	0.321
$D1(E)$	0.468	0.363	0.316	0.307	0.286	0.278	0.273	0.270	0.263	0.258
$D2(E)$	0.435	0.339	0.298	0.289	0.272	0.266	0.263	0.260	0.256	0.253
$D1(MG)$	0.628	0.415	0.364	0.355	0.339	0.334	0.331	0.330	0.327	0.325
$D1(W2)$	0.515	0.406	0.361	0.352	0.335	0.329	0.326	0.324	0.320	0.317
$D1(W.5)$	0.308	0.239	0.202	0.194	0.172	0.163	0.157	0.152	0.141	0.133

表 2. R_N の値

母集団 \ N	4	8	16	20	40	60	80	100	200	400
$D3(U)$	0.650	0.404	0.361	0.357	0.344	0.341	0.339	0.338	0.335	0.334
$D3(U)$	0.384	0.408	0.353	0.352	0.347	0.342	0.337	0.337	0.336	0.334
$D3(G)$	0.536	0.382	0.343	0.348	0.317	0.330	0.324	0.331	0.325	0.321
$D3(G)$	0.223	0.369	0.336	0.317	0.333	0.332	0.308	0.332	0.311	0.319
$D3(E)$	0.225	0.331	0.306	0.305	0.240	0.230	0.310	0.241	0.284	0.269
$D3(E)$	0.519	0.242	0.348	0.267	0.258	0.293	0.315	0.227	0.255	0.245
$D3(\Gamma 2)$	0.281	0.347	0.303	0.314	0.286	0.324	0.291	0.320	0.298	0.298
$D3(\Gamma 2)$	0.188	0.413	0.205	0.333	0.327	0.280	0.290	0.274	0.285	0.290
$D3(\Gamma .5)$	0.061	0.252	0.267	0.310	0.246	0.235	0.169	0.240	0.146	0.182
$D3(\Gamma .5)$	0.025	0.393	0.275	0.148	0.227	0.229	0.199	0.242	0.204	0.226

表 3. 相対効率（$N=64$）

母集団 \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$D1(U)$	1.53	1.56	1.59	1.63	1.67	1.71	1.76	1.82	1.89	1.96	2.06	2.17	2.32	2.50	2.74	3.08
$D1(G)$	1.51	1.54	1.56	1.60	1.63	1.67	1.72	1.77	1.83	1.91	1.99	2.10	2.23	2.39	2.60	2.89
$D1(MG)$	1.51	1.54	1.57	1.60	1.64	1.68	1.72	1.78	1.84	1.91	2.00	2.10	2.23	2.40	2.61	2.91
$D1(E)$	1.39	1.41	1.43	1.45	1.48	1.50	1.53	1.57	1.61	1.66	1.71	1.77	1.85	1.94	2.05	2.19
$D2(E)$	1.37	1.39	1.40	1.42	1.45	1.47	1.50	1.53	1.57	1.61	1.66	1.72	1.78	1.87	1.97	2.09
$D1(W2)$	1.50	1.53	1.55	1.59	1.62	1.66	1.71	1.76	1.82	1.89	1.97	2.07	2.19	2.35	2.55	2.83
$D1(W.5)$	1.20	1.20	1.21	1.22	1.23	1.24	1.25	1.27	1.28	1.30	1.32	1.34	1.36	1.39	1.43	1.46
$D3(\Gamma 2)$	1.44	1.46	1.48	1.51	1.54	1.57	1.61	1.65	1.70	1.75	1.82	1.90	1.99	2.11	2.25	2.44
$D3(\Gamma 2)$	1.46	1.48	1.51	1.54	1.57	1.60	1.64	1.69	1.74	1.80	1.87	1.96	2.06	2.19	2.35	2.57
$D3(\Gamma .5)$	1.26	1.27	1.28	1.29	1.31	1.32	1.34	1.36	1.38	1.41	1.43	1.47	1.50	1.55	1.60	1.66
$D3(\Gamma .5)$	1.18	1.18	1.19	1.20	1.21	1.22	1.23	1.24	1.25	1.27	1.28	1.30	1.32	1.35	1.37	1.41
Fig. 3の例	1.53	1.56	1.59	1.62	1.66	1.70	1.75	1.81	1.88	1.95	2.05	2.16	2.30	2.48	2.72	3.04

表 4. 相対効率（$n=3$）

母集団	N \ r	1	2	3	4	5	6	8	10	15	20	25	30	40	50	60	80	100
$D1(U)$	9	5.32																
	18	2.66	4.57															
	36	2.26	2.58	3.13	4.26													
	45	2.20	2.42	2.74	3.26	4.21												
	90	2.09	2.17	2.27	2.39	2.53	2.70	3.20	4.10									
	135	2.06	2.11	2.17	2.23	2.30	2.38	2.57	2.83	4.07								
	180	2.04	2.08	2.12	2.16	2.21	2.26	2.38	2.52	3.02	4.05							
	225	2.03	2.06	2.09	2.13	2.16	2.20	2.28	2.37	2.68	3.17	4.04						
	450	2.02	2.03	2.05	2.06	2.08	2.09	2.12	2.16	2.25	2.37	2.51	2.67	3.15	4.02			
	900	2.01	2.02	2.02	2.03	2.04	2.04	2.06	2.07	2.11	2.16	2.20	2.25	2.37	2.50	2.67	3.15	4.01
$D1(E)$	9	3.57																
	18	2.18	3.10															
	36	1.89	2.08	2.36	2.85													
	45	1.85	1.98	2.15	2.40	2.79												
	90	1.75	1.80	1.86	1.92	2.00	2.08	2.31	2.66									
	135	1.72	1.75	1.78	1.82	1.86	1.90	2.00	2.12	2.60								
	180	1.71	1.73	1.75	1.77	1.80	1.83	1.89	1.95	2.19	2.57							
	225	1.70	1.71	1.73	1.75	1.77	1.79	1.83	1.88	2.02	2.23	2.55						
	450	1.67	1.68	1.69	1.70	1.70	1.71	1.73	1.75	1.80	1.85	1.92	2.00	2.20	2.49			
	900	1.66	1.66	1.67	1.67	1.67	1.68	1.68	1.69	1.71	1.74	1.76	1.78	1.84	1.90	1.98	2.17	2.46

まず,(i)で作った有限母集団を図示してみると図6のようになっている((ii)は E についてのみ行っているので,これも図6に入れておく.$D1(U)$ は省略).

(2.19)から,相対効率 $e_{[2]}^{(r)}$ の大小は g/σ^2 の大小と一致する.ところで,g/σ^2 は r にはよらないが有限母集団の大きさ N に関係している.そこで

$$R_N=\frac{2g}{\sigma^2} \tag{3.1}$$

とおいて,これを種々の有限母集団に対して求めてみる.その結果をまとめたのが表1である.

(iii)の方法で作られる有限母集団は,いわゆる超母集団からの確率標本であるから,勿論一意に定まらない.一つの連続分布から2回ずつ実験した結果を表2に示す.

表1,2を概観するとき,$N=4$ の場合を除くと,(イ)離散的一様分布で効率が最大である,(ロ)多くの母集団でこれに近い値を示している,(ハ)L字型分布から作り出した母集団では効率が小さい,ということがわかる.これらの傾向は連続分布の場合に類似しているといえる(Takahasi and Wakimoto (1968) の Table 1 参照).

$N=4$ の場合(勿論 $r=1$ である)は特殊な状況を示す.たとえば,1, 2, a, $a+1$ $(a>2)$ なる母集団を考えると $\frac{2g}{\sigma^2}=\frac{(a-1)^2}{3(a^2-2a+2)}$ となり,$e_{[2]}^{(1)}\to\infty$ $(a\to\infty)$ なることがわかる.したがって,$N=4$ の場合は離散的一様分布より効率の大きい母集団が存在する.しかも効率の上限は∞である.$N>4$ の場合については効率の上限は現在のところわからない.

すでに求めた g/σ^2 をもとに,$e_{[2]}^{(r)}$ を $N=64$ の場合に計算した結果を表3に示す.

4. あとがき

順位を補助情報として利用する母平均の推定法を,有限母集団からの非復元抽出の場合について考察してきた.順位を判定する要素の個数を一般にして議論を進めたかったが,途中からその個数を2に限定した.理由は数学的なことなので,近く,一般の場合も解決されるものと思われる.個数が3の場合について2つの例を表4に示しておく.やはり有限母集団の場合,相対効率が連続分布の場合にくらべて大きいことがこれからだけでもうかがわれる.

なお,二ツ矢は第3章の執筆と全体の数値計算を,高橋はそれ以外を担当した.

謝 辞

研究の機会を与えて下さった統計数理研究所 田口時夫 教授と論文の改良に役立った多くの有益なコメントを下さった査読者の方々に厚くお礼申し上げます.

参 考 文 献

Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, **28**, 545-553.

McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets, *Austral. J. Agric. Res.*, **3**, 385-390.

Stokes, S.L. (1972). Ranked set sampling with concomitant variables, *Comm. Statist. A—Theory Methods*, **6**, 1207-1211.

Takahasi, K. (1969). On the estimation of the population mean based on ordered samples from an equicorrelated multivariate distribution, *Ann. Inst. Statist. Math.*, **21**, 249-255.

Takahasi, K. (1970). Practical note on estimation of population means based on samples stratified by means of ordering, *Ann. Inst. Statist. Math.*, **22**, 421-428.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering, *Ann. Inst. Statist. Math.*, **20**, 1-31.

Yanagawa, T. and Chen, Shan-Huo (1980). The *MG*-procedure in ranked set sampling, *J. Statist. Plann. Inference*, **4**, 33-44.

柳川 堯, 白旗慎吾 (1974). 順位に関する情報を補助情報として利用する母集団平均の推定, 応用統計学, **4**, 55-63.

Yanagawa, T. and Shirahata, S. (1976). Ranked set sampling theory with selective probability matrix, *Austral. J. Statist.*, **18**, 45-52.

Ranked Set Sampling from a Finite Population

Koiti Takahasi and Masao Futatsuya
(Faculty of Science, Hirosaki University)

The theory of ranked set sampling to provide a more efficient estimator than the sample mean usually assumes that the ranked sets of elements are independently distributed. In this paper we consider the ranked set sampling without replacement from a finite population.

Suppose that a simple random sample of size $n^2 r$ is drawn from a finite population of size N with mean μ and variance σ^2 without replacement and that the sample is divided into nr groups of size n. Let $X_{i,k}$ be the i-th order statistic of the $(n(k-1)+i)$-th group for $i=1, \ldots, n$ and $k=1, \ldots, r$. We consider the unbiased estimator $Y_{[n]}^{(r)} = \frac{1}{nr} \cdot \sum_{k=1}^{r} \sum_{i=1}^{n} X_{i,k}$ of the mean μ, especially, in the case of $n=2$. The relative efficiency $e_{[2]}^{(r)}$ of $Y_{[2]}^{(r)}$ to the sample mean $\bar{X}_{2r}$ of a sample of size $2r$ obtained by simple random sampling without replacement is given by $\left(1-\frac{N-1}{N-2r} \cdot \frac{2g}{\sigma^2}\right)^{-1}$, where g is $\frac{1}{12} \mathrm{E}[(Z_{(4)}-Z_{(2)})(Z_{(3)}-Z_{(1)})]$ and $Z_{(1)}, \ldots, Z_{(4)}$ are the order statistics of a simple random sample of size 4 from the finite population without replacement.

The numerical examples of g/σ^2 and $e_{[2]}^{(r)}$ are given in Table 1, 2 and 3, respectively, for various kinds of finite populations.

Key words: Ranked set sampling, finite population, order statistics.