数理解析研究所講究録
第 272 巻 1976 年 89-97

有限母集団における推定量の許容性について

静大　工学部　多賀保志

§1　序　論

有限母集団に関するサンプリングの理論は，数理統計学とかなりの距離をおいて発展してきたが，Basu（1958）は有限母集団からとられたサンプル（サンプル数一定として）についての十分統計量を与えた。これを契機としてTakeuchi（1966）らによりサンプリング理論の数理統計学的定式化が行われた。それによると，母集団の有限性が十分統計量の表現に特殊性を与え，それが完備十分統計量の不存在，ひいては一様最小分散（UMV）をもつ不偏推定量の不存在につながることが示された。一方，Godambe（1965），Hanurav（1966），Murthy（1969）らによって，有限母集団における推定量の許容性あるいは超許容性が研究された。さらにRamakrishnan（1973）は，サンプリング・デザインと推定量の対（ストラテジーと）に関す

る許容性を考え，HT（Horvitz－Thompson）型のストラテジーがすべての不偏なストラテジーのクラスにおいて許容的であることを示した（定理と証明の記述に条件が不足して不完全な点がみられる）。

ここでは上述の諸結果を要約し，若干の補足と展望を述べることにしよう。

§2　十分性と完備性

有限母集団からとられるサンプルの確率分布は，逐次サンプリングの場合もふくめて，つぎのように定式化される。大きさNの母集団の標識集合を$\{\theta_1, \cdots, \theta_N\}$とし，$\theta_i$に無関係な確率分布$P_1$にしたがって1つの要素をとり出し，その識別番号（母集団の各要素に対応する番号）をI_1，その観測量をX_1とする。つぎに，(I_1, X_1)の実現値(i_1, x_1)に依存してもよいがθには依存しないある確率分布P_2にしたがって才2の要素をとり出し，その識別番号と観測量をI_2, X_2と表す。

一般にk個の要素がすでにとられて，つぎに$(k+1)$番目の要素が$(i_1, \cdots, i_k, x_1, \cdots, x_k)$に依存してもよいが$\theta$には依存しない確率分布$P_{k+1}$にしたがってとり出されるとする。ただし$I_{k+1}=0$となった場合には，サンプリングはそこで中止され，$P_0\{n<1\}=1$とする。するとサンプル$(\mathbb{I}, \mathbb{X}, n)=(I_1, \cdots, I_n,$

$X_1, \cdots, X_n, n$）の確率分布 $p_\theta(\boldsymbol{i}, \boldsymbol{x}, n)$ は

$$(2.1)\quad p_\theta(\boldsymbol{i}, \boldsymbol{x}, n) = p_1(i_1) \prod_{k=1}^{n+1} p_k(i_k | i_1, \cdots, i_{k-1}, x_1, \cdots, x_{k-1}) \times \prod_{k=1}^{n} \chi_\theta(i_k, x_k)$$

と表わされる。ここで，$\theta = (\theta_1, \cdots, \theta_N)$，$\theta_i \in R^p$（$1 \leqq i \leqq N$），$\boldsymbol{i} = (i_1, \cdots, i_n)$，$\boldsymbol{x} = (x_1, \cdots, x_n)$，$i_{n+1} = 0$，

$$\chi_\theta(i_k, x_k) = \begin{cases} 1 & (x_k = \theta_{i_k} \text{ のとき}) \\ 0 & (x_k \neq \theta_{i_k} \text{ のとき}) \end{cases}$$

とし，観測は誤差なく行われるものとする。この確率分布族を $\mathcal{P}_1 = \{p_\theta(\boldsymbol{i}, \boldsymbol{x}, n), n = 1, 2, \cdots\}$ とすると，$\mathcal{P}_1$ に関する十分統計量（$\mathbb{J}$，$\mathbb{Y}$，m）が次の形で与えられる（Takeuchi）。

「$\mathbb{I} = (I_1, \cdots, I_n)$ の要素のうちで相異なるものを，$J_1 < J_2 < \cdots < J_m$ とし，$Y_k = X_{J_k}$（$1 \leqq k \leqq m$）とおくと，統計量（$\mathbb{J}$，$\mathbb{Y}$，m）は $\mathcal{P}_1$ に関して十分である。ただし $\mathbb{J} = (J_1, \cdots, J_m)$，$\mathbb{Y} = (Y_1, \cdots, Y_m)$ とする。」

この定理および Rao－Blackwell の定理より，θ の関数 $g(\theta)$ の推定量としては，（$\mathbb{J}$，$\mathbb{Y}$，m）の関数のみを考えればよいこととなる（これよりサンプル数 n が一定ならば，復元サンプルよりも非復元サンプルをもちいた方が良い推定量がえられることがわかる）。さらに（$\mathbb{J}$，$\mathbb{Y}$，m）が完備性をもてば，それより一様最小分散をもつ $g(\theta)$ の不偏推定量がえられることになるが，一般にその完備性は成立しない。ただ

し，「mが一定，(2.1)の右辺の条件つき確率分布P_kが$x_1, \cdots,$ x_kに依存しないで($k=1, 2, \cdots$)，$\theta=(\theta_1, \cdots, \theta_N)$の要素の任意の置換に関して$P_\theta(i, x, n)$が不変であれば，$\mathbb{Y}$は完備となる」ことが示されている(Takeuchi)。

θに関して何らの事前情報もえられていない場合には，母集団における要素の番号をしめす$\mathbb{I}=(I_1, \cdots, I_n)$を推定量の構成に利用することは，一般に必ずしも有利とは考えられない。そこで$\mathbb{I}$を除いた$(\mathbb{X}, n)$の確率分布の族$\mathcal{P}_2=\{g_\theta(x, n)=\sum_{i} P_\theta(i, x, n),\ n=1, 2, \cdots\}$を考えると，それについての十分統計量$\mathbb{Z}$がつぎの形で与えられる。

「サンプリングは非復元で行われ，iとxの要素の任意の置換$\tau(i)$，$\tau(x)$に対して$P_\theta(\tau(i), \tau(x), n)=P_\theta(i, x, n)$が各$n$に対して成り立つ($P_\theta$が対称)とする。$\mathbb{X}=(X_1, \cdots, X_n)$の要素を辞書式順序に並びかえてえられる統計量を$\mathbb{Z}=(Z_1, \cdots, Z_n)$とすると，$\mathbb{Z}$は$\mathcal{P}_2$に関して十分である」ことが示される。さらに$\mathbb{Z}$の完備性についてはつぎのことがいえる。

「サンプリング方式が非復元かつ正則で，nが一定とすれば，$\mathbb{Z}$は完備となる」(これは$\mathbb{Z}$の関数が$\mathbb{X}$の要素についての対称関数$h(\mathbb{X})$とあらわされること，$h(\mathbb{X})$の期待値がすべてのθについて零となれば$h(x)\equiv 0$となることが帰納的に示されることよりいえる)。

以上2つの事柄をもちいると，「サンプリング方式が非復元でかつ正則で，nが一定，かつ $P_\theta(\mathfrak{i}, \mathfrak{x}, n)$ が対称であれば，$\mathbb{Z}$は$\mathcal{P}_2$に関して完備十分統計量となる」となることがいえる。これよりただちに，「サンプリング方式が非復元かつ正則，nが一定，$\mathcal{P}_2$ は対称なP_θより導かれる$\mathbb{X}$の確率分布g_θのすべてよりなる族とする。$\mathbb{X}$の要素について対称な$g(\theta)$の不偏推定量$T = t(\mathbb{X})$は一様最小分散をもち，それは$\mathcal{P}_2$に関して一意的である。」

§3　許容性

$g(\theta)$が推定可能すなわち$g(\theta)$の不偏推定量が存在する場合，$\theta = \theta^\circ$で分散0をもつ推定量T_0の存在がTakeuchi によって示された。たとえばθの各要素θ_iがスカラーで（$p=1$），$g(\theta)=\frac{1}{N}\sum_{i=1}^{N}\theta_i$（母平均）を大きさ$n$の単純ランダム・サンプル（$\mathbb{I}$, $\mathbb{X}$）で推定する場合，つぎの推定量T_0を考える：

$$(3.1)\qquad T_0 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \theta^\circ_{I_j}\right) + \frac{1}{N}\sum_{i=1}^{N}\theta^\circ_i,$$

$$\text{ここで } X_j = \theta_{I_j},\quad \theta^\circ = (\theta^\circ_1, \cdots, \theta^\circ_N)$$

するとT_0は$g(\theta)$の不偏推定量で，かつ$\theta = \theta^\circ$のとき$T_0 \equiv g(\theta_0)$であるから，$\theta = \theta^\circ$でT_0の分散$V(T_0) = 0$となる。$\theta^\circ \in R^N$は任意にとってよいから，一般に$g(\theta)$に対する一様最小分散不偏推定量は存在しないことがわかる。

Godambe は母総計$g(\theta) = \sum_{i=1}^{N}\theta_i$に対する一様最小分

散不偏推定量の不存在を，別の立場から示した。したがってそれ以後，推定量の許容性についての研究が Godambe, Hanurav, Murthy らによって行われてきた。ここで $g(\theta)$ の不偏推定量 T^* が許容的であるとは，任意の θ に対して $V_\theta(T) \leqq V_\theta(T^*)$，かつある θ' に対して $V_{\theta'}(T) < V_{\theta'}(T^*)$ となるような不偏推定量 T が存在しないことをいう。

$g(\theta)=\sum_{i=1}^{N}\theta_i$（母総計）の線形不偏推定量としては，H－T 型推定量

$$(3.2)\qquad T^* = \sum_{j=1}^{n} X_j / \pi_{I_j} ,$$

ここで $X_j=\theta_{I_j}$，π_i は θ_i がサンプルにふくまれる才1次包含確率（$i=1,\cdots,N$）

がよく知られており，その分散は

$$(3.3)\qquad V_\theta(T^*) = \sum_{i=1}^{N}\pi_i\left(\frac{\theta_i}{\pi_i}\right)^2 + \sum_{i\neq j}^{N}\pi_{ij}\frac{\theta_i\theta_j}{\pi_i\pi_j} - \left(\sum_{i=1}^{N}\theta_i\right)^2$$

で与えられる（π_{ij} は θ_i と θ_j がともにサンプルにふくまれる才2次の包含確率）。この推定量 T^* は Midzuno により見出され，その分散の一般形（3.3）が Horwitz and Thompson により与えられた。また T^* の許容性は，Hanurav や Murthy らによって示され，さらに $g(\theta)=\sum_{i=1}^{N}c_i\theta_i$ なるパラメターについても，T^* を少し変形した線形不偏推定量の許容性（これを T^* の超許容性という）が示されている。

さらにRamakrishnan（1973）は，サンプリング・デザインp_θと推定量Tの対（p_θ, T）をストラテジーと名づけ，Tが$g(\theta)$の不偏推定量であるときストラテジー（p_θ, T）は不偏であると定義し，その許容性を考えた。それによると，「H-T型推定量T^*と任意のサンプリング・デザインp_θ°との対であるH-T型ストラテジー（p_θ°, T^*）は，平均サンプル数μをもつ不偏ストラテジーの族で許容的である」ことが示された。その証明として，任意の不偏なストラテジー（p_θ, T）があって，すべてのθに対して

$$(3.4)\qquad V(T \mid p_\theta) \leqq V(T^* \mid p_\theta^\circ)$$

が成り立つとすれば，p_θとp_θ°より導かれるすべての包含確率の2つの系が一致し，それより$T = T^*$（a.e.）を導くことができることが示される。ただし，この証明過程において，「推定量Tが$T = t(a_{I_1}X_1, \cdots, a_{I_n}X_n)$なる形で，かつ$\mathbb{I} = (I_1, \cdots, I_n)$の任意の置換に対して不変である」という条件を付加する必要がある（ただし$a_1, \cdots, a_N$は常数）。

サンプリング方式（サンプリング・デザインp_θ）を任意に設定しうることが'有限母集団におけるサンプリングの特色であるから，ストラテジー（p_θ, T）の許容性を扱った上の議論は興味深い。今後の課題としては，θに関する補助情報

が与えられたとき，それを利用してなるべく良いストラテジーをえらぶ規準と方法の開発が望まれる。

参 考 文 献

[1] Basu, D. (1958): On sampling with and without replacement; Sankhyā, 20, 287-294.

[2] Godambe, V. P. (1965): A review of the contributions towards a unified theory of sampling from finite populations; Rev. ISI, 33, 242-258.

[3] Hanurav, T. V. (1966): Some aspects of unified sampling theory; Sankhyā (A), 28, 175-204.

[4] Horwitz, D. G. and Thompson, D. J. (1952): A generalization of sampling without replacement from finite population; J. A. S. A., 47, 663-685.

[5] Miduno, H. (1950): An outline of the theory of sampling system; J. I. S. M., 1, 149-156.

[6] Murthy, M. N. and Singh, M. P. (1969): On the concept of best and admissible estimators in sampling theory; Sankhyā (A), 31, 343-354.

[7] Takeuchi, K. (1966): Some remarks on general theory for unbiased estimation of a real

parameter of a finite population ; Japanese J. of Math., 35, 73-84.