⑧ 母集団とそのリストと対応が一対一でない場合のサムプリングについて

遠藤 健児
丸山 文行

一般に調査単位の集団から確率的にその一部を抽出するためには，何らかの形でその集団の構成を抽象したモデルとしてのリストが必要である。　最も典型的な場合は各々の単位をそのまゝ写しとったリスト，即ち調査単位との対応が一対一となっているリストが利用出来るときである。

しかし実際問題に於てはこの様な場合は稀であつて，多くは全集団をいくつかの単位から成る群に分割した時の，この群についてのリストを利用する。　これは調査単位を多対一に写すリストと考えることが出来る。　又一方に於ては，この様な意味で調査単位を一対多に写すリストが利用出来る場合もある。　例えば，一戸の家屋について或る測定値の調査を行うときに，戸別のリストはないが，それらの家屋に居住している世帯のリストが利用出来る場合である。　空家を除外すればこの時の対応は多対多であり，更に一つの世帯で二戸以上の家屋に居住するものがないと仮定すれば，この時の対応は一対多である。

此処では先ず対応が一対多である場合について考える。　この場合にはその様なリストに依つて或る世帯が抽出されたとき，その世帯が居住する家屋が標本とされるのであるが，この様な標本

について得た測定から次の様にして不偏推定を構成することが出來る。

今調査対象たる家屋の集団をΠとすれば，Πの各単位にはそれぞれν個の世帯で占められることからνと云う標識（$\nu=1,2,3,\cdots$）が與えられる。　標識νを持つ家屋から成る群をΠ_ν，その大さをN_νとすれば，世帯のリストで表現される家屋の集団Π^*は

$$\Pi_1\ ;\ \Pi_2,\Pi_2\ ;\ \Pi_3,\Pi_3,\Pi_3\ ;\ \cdots\cdots\ ;\ \overbrace{\Pi_\nu,\cdots\Pi_\nu}^{\nu\text{個}}\ ;\ \cdots$$

なる$1+2+3+\cdots\cdots$個の群の合併であり，この大さは

$$N'=N_1+2N_2+3N_3+\cdots\cdots=\sum_\nu \nu N_\nu$$

である。　Π_νに属する家屋についての或る性質Xの測度を

$$X_{\nu 1},\ X_{\nu 2},\ \cdots\cdots,\ X_{\nu N_\nu}$$

とすれば，Xの総計並に平均は夫々

$$X=\sum_\nu\sum_{j=1}^{N_\nu}X_{\nu j}\ ,\qquad \overline{X}=\sum_\nu\sum_{j=1}^{N_\nu}X_{\nu j}\Big/\sum_\nu N_\nu=X/N$$

である。　此処でNはΠの大さ，即ちΠに含まれる家屋の総数である。

扨て始めに述べた方法に依つてこのリストを使つて単純に等確率的に抽出されたn個の標本に関するn個の測定値を，それがどのΠ_νに属するかに依つて分類したものを

$$O_\nu:\ x_{\nu 1},\ x_{\nu 2},\ \cdots\cdots,\ x_{\nu n_\nu}\quad(\nu=1,2,\cdots)$$

とし，O_νの大さをn_ν，その合計を$S(x_\nu)$，その平均を$\overline{x}_\nu=S(x_\nu)/n_\nu$で表わす。　このとき上述の$x_{\nu j}$のうちには，全一の家屋から求められた等しい測定値も別なものとして含まれているわけである。　この時，例えば単なる標本平均

$$\overline{x}^*=\{S(x_1)+S(x_2)+\cdots\cdots\}/n$$

をとつたのでは，これはΠ^*の平均，即ち

$$\overline{X}^* = \sum_\nu \nu \sum_{j=1}^{N_\nu} X_{\nu j} \Big/ \sum_\nu \nu N_\nu$$

の推定値となるに過ぎない。 Π の母平均 $\overline{X}$ の不偏推定値としては

$$\bar{x} = \frac{N'}{N} \cdot \frac{1}{n} \sum_\nu \frac{S(x_\nu)}{\nu}$$

をとればよい。 今ある $n_1, n_2, ----$ ($\sum_\nu n_\nu = n$) を固定したときの平均をとることを E' であらわせば

$$(1) \qquad E'(\bar{x}) = \frac{N'}{N} \cdot \frac{1}{n} \sum_\nu \frac{n_\nu}{\nu} E'(\bar{x}_\nu)$$

$$= \frac{N'}{N} \cdot \frac{1}{n} \sum_\nu \frac{n_\nu}{\nu} \overline{X}_\nu \qquad \left(\overline{X}_\nu = \frac{1}{N_\nu} \sum_{j=1}^{N_\nu} X_{\nu j} \right).$$

処で或る n_ν の組

$$(n_1, \ n_2, ----, n_\nu, ----) \qquad (\sum_\nu n_\nu = n)$$

の分布は，有限の場合の多項分布でこの母集団は大さが夫々

$$N_1, \ 2N_2, ----, \nu N_\nu, ---- \qquad (\sum \nu N_\nu = N')$$

なる有限個の部分群から成るから

$$E(n_\nu) = n \cdot \frac{\nu N_\nu}{N'}. \qquad (\nu = 1, 2, ----)$$

である。（此処で $\nu N_\nu / N'$ は最初の標本が Π_ν からとられる確率となつている。） 従つて

$$(2) \qquad \begin{aligned} E(\bar{x}) &= \frac{N'}{N} \cdot \frac{1}{n} \sum_\nu \frac{\overline{X}_\nu}{\nu} E(n_\nu) \\ &= \frac{N'}{N} \cdot \frac{1}{n} \sum_\nu \frac{\overline{X}_\nu}{\nu} \cdot n \cdot \frac{\nu N_\nu}{N'} \\ &= \frac{1}{N} \sum_\nu N_\nu \overline{X}_\nu = \overline{X} \end{aligned}$$

となり，$\bar{x}$ は母平均に対する不偏推定値であることが分る。

次に推定値 $\bar{x}$ の標準誤差を求めよう。

ある一組の（$n_1, n_2, ----$）を固定した時の分布に於る分散を，

D'^2 で表わせば，$x = \sum_{\nu} S(x_\nu)/\nu$ の分散は

$$(3) \qquad D^2(x) = E\{D'^2(x)\} + D^2\{E'(x)\}$$

となる。処で

$$E'(x) = \sum_{\nu} n_\nu \overline{X}_\nu / \nu \quad ,$$

従って

$$x - E'(x) = \sum_{\nu} \frac{n_\nu}{\nu} (\overline{x}_\nu - \overline{X}_\nu)$$

となる。故に

$$D'^2(x) = \sum_{\mu}\sum_{\nu} \frac{n_\mu n_\nu}{\mu\nu} E'(\overline{x}_\mu - \overline{X}_\mu)(\overline{x}_\nu - \overline{X}_\nu)$$

$$= \sum_{\nu} \frac{n_\nu^2}{\nu^2} E'(\overline{x}_\nu - \overline{X}_\nu)^2$$

$$= \sum_{\nu} \frac{n_\nu^2}{\nu^2} D'^2(\overline{x}_\nu)$$

処で $D'^2(\overline{x}_\nu) = \dfrac{\nu N_\nu - n_\nu}{\nu N_\nu - 1} \cdot \dfrac{\nu}{n_\nu} \sigma_\nu^2$ ，$\sigma_\nu^2 = \dfrac{1}{N_\nu} \sum_j (x_{\nu j} - \overline{X}_\nu)^2$

であるが，$E\left(\dfrac{n_\nu}{\nu N_\nu}\right) \doteq (\nu N_\nu \cdot n/N') \div (\nu N_\nu) = n/N'$ であるから抽出比 n/N が十分小さいならば

$$D'^2(\overline{x}_\nu) = \sigma_\nu^2 / n_\nu$$

と考えることが出来る。従って

$$D'^2(x) = \sum_{\nu} \frac{n_\nu}{\nu^2} \sigma_\nu^2$$

故に

$$(4) \qquad E\{D'^2(x)\} = \sum_{\nu} \frac{\sigma_\nu^2}{\nu^2} E(n_\nu) = \frac{n}{N'} \sum_{\nu} N_\nu \sigma_\nu^2 / \nu$$

但し $\sigma_\nu^2 = \dfrac{1}{N_\nu} \sum_j (X_{\nu j} - \overline{X}_\nu)^2$

次に

$$E'(x) = \Sigma n_\nu \bar{X}_\nu / \nu \qquad \text{であるから}$$

$$D^2\{E'(x)\} = \sum_\mu \sum_\nu \frac{\bar{X}_\mu \bar{X}_\nu}{\mu\nu} E(n_\mu - E(n_\mu))(n_\nu - E(n_\nu))$$

此処で N, 従つて N' が十分大きいならば

$$E\{n_\mu - E(n_\mu)\}(n_\nu - E(n_\nu)) = \frac{n\nu N_\nu}{N'}\left(\delta_{\mu\nu} - \frac{\nu N_\nu}{N'}\right)$$

$$\therefore \quad D^2\{E'(x)\} = \frac{n}{N'} \sum_\nu \frac{N_\nu \bar{X}_\nu^2}{\nu} - n\left(\frac{1}{N'} \Sigma N_\nu \bar{X}_\nu\right)^2$$

$$(5) \qquad = \frac{n}{N'} \sum_\nu \frac{N_\nu \bar{X}_\nu^2}{\nu} - \frac{nN^2}{N'^2} \bar{X}^2$$

(4), (5) を (3) の右辺に代入すれば

$$D^2(x) = \frac{n}{N'} \sum_\nu \frac{N_\nu}{\nu}(\sigma_\nu^2 + \bar{X}_\nu^2) - \frac{nN^2}{N'^2} \bar{X}^2$$

故に

$$D^2(\bar{x}) = \frac{1}{n} \cdot \frac{N'}{N} \cdot \frac{1}{N} \sum_\nu \frac{N_\nu}{\nu}(\sigma_\nu^2 + \bar{X}_\nu^2) - \frac{1}{n}\bar{X}^2$$

処で

$$\sigma^2 = \frac{1}{N} \Sigma\Sigma (X_{\nu j} - \bar{X})^2$$

$$= \frac{1}{N} \Sigma N_\nu \sigma_\nu^2 + \frac{1}{N} \Sigma N_\nu (\bar{X}_\nu - \bar{X})^2$$

$$= \frac{1}{N} \Sigma N_\nu (\sigma_\nu^2 + \bar{X}_\nu^2) - \bar{X}^2$$

$$\text{or} \qquad \bar{X}^2 = \frac{1}{N} \Sigma N_\nu (\sigma_\nu^2 + \bar{X}_\nu^2) - \sigma^2$$

これを上に得た最后の結果の式の右辺に代入すれば

$$(6) \qquad D^2(\bar{x}) = \frac{\sigma^2}{n} + \frac{1}{n} \cdot \frac{1}{N} \sum_\nu N_\nu \left(\frac{N'}{N\nu} - 1\right)(\bar{X}_\nu^2 + \sigma_\nu^2)$$

但し $\sigma_\nu^2 = \frac{1}{N_\nu} \sum_j (x_{\nu j} - \bar{X}_\nu)^2$, $\sigma^2 = \frac{1}{N} \sum_\nu \sum_j (X_{\nu j} - X)^2$

この右辺の第二項は

$$\frac{1}{n}\cdot\frac{1}{N^2}\sum_{\mu}\sum_{\nu}\left(\frac{\mu}{\nu}-1\right)N_\mu N_\nu(\overline{X}_\nu^2+\sigma_\nu^2)$$

とかける。 ここで例えば N_ν が 0 でないものについては，
$\overline{X}_1^2+\sigma_1^2 \fallingdotseq \overline{X}_2^2+\sigma_2^2 \fallingdotseq \cdots\cdots$ の場合を考えると

$$0 \leqq \sum N_\nu(\nu x-1)^2/\nu=(\sum\nu N_\nu)x^2-2(\sum N_\nu)x-\sum N_\nu/\nu$$

なることから

$$\sum_{\mu}\sum_{\nu}\left(\frac{\mu}{\nu}-1\right)N_\mu N_\nu=\left(\sum\frac{N_\nu}{\nu}\right)(\sum\nu N_\nu)-(\sum N_\nu)^2 \geqq 0$$

故にこのときには一対一のリストが利用出来る場合に比較して誤差が増大するものと見做し得る。 一般的には必ずしもこの様なことは結論されない。 Π が Π_1 と Π_2 とだけから成る様な場合，$\overline{X}_1^2+\sigma_1^2 \ll \overline{X}_2^2+\sigma_2^2$ ならば誤差は却って減少する。

次に二戸以上の家屋に居住する世帯が存在する場合を考えよう。 一つの世帯が μ 個の家屋に居住するときは，その家屋の標識 ν としては $1/\mu$ と云う値を與え，標識 $1/\mu$ を持つ家屋より成る群を $\Pi_{\frac{1}{\mu}}$ で表わす。 此処で，例えば二戸の家屋 A 及び A' に居住する世帯 B が抽出されたときには，調査を実際に行うときに A 又は A' のいづれか一方を等確率的に選定して測定値をとるものとする。この様にすれば始めに述べたことがそのまゝ適用出來て

$$\overline{x}=\frac{N'}{N}\cdot\frac{1}{n}\left(\cdots+2S(x_{\frac{1}{2}})+S(x_1)+\frac{1}{2}S(x_2)+\cdots\right)$$

が母平均 $\overline{X}$ の不偏推定値であつて，その分散は (6) と同様の式

$$D^2(\overline{x}) \doteqdot \frac{\sigma^2}{n}+\frac{1}{n}\cdot\frac{1}{N}\sum_{\nu=1}^{\infty}N_\nu\left(\frac{N'}{N\nu}-1\right)(\overline{X}_\nu^2+\sigma_\nu^2)$$

$$+\frac{1}{n}\frac{1}{N}\sum_{\nu=2}^{\infty}N_{\frac{1}{\nu}}\left(\frac{N'\nu}{N}-1\right)(\overline{X}_{\frac{1}{\nu}}^2+\sigma_{\frac{1}{\nu}}^2)$$

に依つて與えられる。 此処で添字が $\frac{1}{\nu}$ なる量は $\Pi_{\frac{1}{\nu}}$ に関する対応する量であり，又

$$N=\sum_{\nu=2}^{\infty} N_{\frac{1}{\nu}}+\sum_{\nu=1}^{\infty} N_{\nu}, \quad N'=\sum_{\nu=2}^{\infty} \frac{1}{\nu} N_{\frac{1}{\nu}}+\sum_{\nu=1}^{\infty} \nu N_{\nu} .$$

このときには $\Pi_{\frac{1}{\nu}}$ に属する一つの家屋が第一の標本にとられる確率は

$$\frac{1}{\nu} N_{\frac{1}{\nu}} \Big/ N'$$

であつて，Π^{*} に於ては $\Pi_{\frac{1}{\nu}}$ は $1/\nu$ にかぞえられていることになる。

（これは文部省科学研究費に依る研究の一部である。）

On the sampling problem where the correspondence between individuals of population and the available list of them is not one to one.

by K. Endō and F. Maruyama

Four cases may be pointed out about the correspondence between individuals of the population and their list, i.e.,

(1) one to one,

(2) many to one,

(3) one to many,

(4) many to many.

and if the simple random sampling method is desired, the list of type (1) will be indispensable, whereas the cluster sampling may be carried out by making use of certain lists of type (2).

In this paper the sampling method and the construction of an unbiased estimator are considered when the available lists are of type (3) or (4).