Austral. J. Statist., **27** (2), 1985, 135–137

DOMAIN ESTIMATION IN FINITE POPULATIONS[1]

ARIJIT CHAUDHURI AND RAHUL MUKERJEE
Indian Statistical Institute, Calcutta

Summary

This paper suggests unbiased estimators (UE's) for the size, mean and total of a domain, with specific features, in a given finite population on the basis of simple random sampling without replacement (SRSWOR) continued till a preassigned number of domain members is observed.

Key words: Domain parameters; finite population; inverse sampling; unbiased estimator.

1. Introduction

Consider a finite population of known size N having an unknown number, M, of units of a given category forming a domain A. Denoting the variate values for the M domain members by $X_1, \ldots, X_M$, suppose UE's are required for M, $T = \sum_1^M X_i$, $\mu = T/M$, the domain size, total and mean respectively. Haldane (1945) and Sampford (1962) considered inverse sampling with replacement to estimate $F = M/N$ and Rao (1975) gave biased ratio estimators for μ based on SRSWOR with a fixed number of draws. This paper gives UE's for the above three domain parameters based on inverse SRSWOR sampling and notes that SRSWOR sampling with a fixed number of draws yields a UE for μ only under severely restrictive conditions.

2. Unbiased estimation with inverse SRSWOR

For inverse SRSWOR let u denote the random number of draws required to realize a preassigned number, m, of domain members. Denote the parametric space of M by $\mathcal{M} = \{r, r+1, \ldots, N\}$ and, to avoid trivialities, assume that $1 < m \le r\ (\le N)$ which holds in most practical situations. Then the probability distribution of u is given by

$$P(u = n) = \frac{\binom{M}{m-1}\binom{N-M}{n-m}}{\binom{N}{n-1}} \cdot \frac{M-m+1}{N-n+1} = g_{Mn} \quad (m \le n \le N - M + m) \tag{2.1}$$

[1] Manuscript received February 15, 1984; revised February 18, 1985

Theorem 2.1. *Every parametric function $f(M)$ admits a unique UE involving u.*

Proof. For $h(u)$ to be a UE of $f(M)$ it is necessary and sufficient that for any M, $r \le M \le N$,

$$f(M) = \sum_{n=m}^{N-M+m} h(n) g_{Mn},$$

which may be written $\mathbf{f} = \mathbf{G}\mathbf{h}$, where $\mathbf{f} = [f(r), \ldots, f(N)]'$, $\mathbf{h} = [h(m), \ldots, h(N-r+m)]'$ and

$$\mathbf{G} = \begin{bmatrix} g_{rm} & g_{r,m+1} & \cdots & g_{r,N-r+m-1} & g_{r,N-r+m} \\ g_{r+1,m} & g_{r+1,m+1} & \cdots & g_{r+1,N-r+m-1} & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ g_{Nm} & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Since $\mathbf{G}$ is clearly nonsingular, $\mathbf{h} = \mathbf{G}^{-1}\mathbf{f}$ is the unique UE of $f(M)$ among functions of u.

Corollary 2.1. *The unique UE of M based on u is $\hat{M}(u) = N(m-1)/(u-1) = \hat{M}$.*

Formulae for $E(\hat{M}^2)$ and $V(\hat{M})$ are available by (2.1), where, as usual, E and V stand for expectation and variance operators. Let $S^2 = (M-1)^{-1} \sum_1^M (X_i - \mu)^2$, $q(u)$ and $l(u)$ be UE's, by Theorem 2.1, for M^{-1} and M^2 respectively, $\sum'$ denote summation over the A-units in the sample, $\bar{x} = m^{-1} \sum' X_i$, $Z = m^{-1} \sum' X_i^2$ and $s^2 = (m-1)^{-1} \sum' (X_i - \bar{x})^2$. The following results are obtained by conditioning on u.

(i) A UE for μ is $\bar{x}$ with $V(\bar{x}) = S^2(1/m - 1/M)$,

(ii) A UE for T is $\hat{T} = \hat{M}\bar{x}$ with

$$V(\hat{T}) = S^2(1/m - 1/M)E(\hat{M}^2) + \mu^2 V(\hat{M}),$$

(iii) $v(\bar{x}) = s^2(m^{-1} - q(u))$ is a UE for $V(\bar{x})$ and

(iv) $v(\hat{T}) = \hat{T}^2 - [l(u)(Z - s^2) + \hat{M}s^2]$ is a UE for $V(\hat{T})$.

3. Estimation from SRSWOR with a fixed number of draws

Let SRSWOR be based on a fixed number, d, of draws and c denote the number of A-units in a sample so drawn. If $d \ge N - r + 1$, then $P(c = 0) = 0$, $\forall M \in \mathcal{M}$ and $\hat{\mu} = c^{-1} \sum' X_i$ is a UE of μ. On the other hand, unless $d \ge N - r + 1$ it is possible to get a sample containing no members of the domain so that μ cannot be unbiasedly estimated. Thus μ admits a UE if and only if $d \ge N - r + 1$. The mathematical details in this regard can be worked out (vide Chaudhuri & Mukerjee (1983)) but are omitted here to save space.

Since $(N-r+1)$ is large in practice especially when r is small, SRSWOR with a fixed number of draws is clearly less convenient than inverse SRSWOR in unbiasedly estimating μ. For the former scheme, as with the latter, it is not difficult to find and use UE's for M and T.

Acknowledgement

The authors are thankful to a referee for his highly constructive suggestions which helped an improved draft over an earlier one.

References

CHAUDHURI, A. & MUKERJEE, R. (1983). Unbiased estimation of domain parameters in sampling without replacement. Tech. Rep. 42/83, Indian Stastical Institute, Calcutta.

HALDANE, J. B. S. (1945). On a method of estimating frequencies. *Biometrika* **33**, 222–225.

RAO, J. N. K. (1975). Analytical studies of simple survey data. *Survey Methodology* **1**, 1–76.

SAMPFORD, M. R. (1962). Methods of cluster sampling with and without repacement for clusters of unequal sizes. *Biometrika* **49**, 27–40.