MMUN. STATIST.-THEORY METH., 20(7), 2291-2297 (1991)

ON OPTIMALITY OF THE HORVITZ-THOMPSON ESTIMATOR UNDER A MARKOV PROCESS MODEL

Arun Kumar Adhikary[1]
Department Of Mathematics
University Of Nairobi
P.O. Box 30197
Nairobi, Kenya.

Key Words and Phrases: Finite population ; Horvitz-Thompson estimator ; Markov process model ; Varying probability sampling.

ABSTRACT

or any varying probability sampling design the rvitz-Thompson (1952) estimator is shown to be optimal thin the class of all unbiased estimators of a finite ulation total under a Markov process model.

leave from Indian Statistical Institute, Calcutta.

1. INTRODUCTION

Consider a finite population U of size N and let y_{it} denote the value of a variate under study for the ith unit observed at time t ($i = 1,\ldots,N$). Our problem is to estimate the population total $Y_t = \sum_{i=1}^{N} y_{it}$ on the basis of a sample s drawn from the population with a probability p(s). Assuming that the variate values depend on time t we consider a Markov process model described below which is very useful in many practical situations ,e.g.in market research studies where the analysis of sales figures in successive weeks or months is an important problem. Considering the class of all unbiased estimators for Y_t, the Horvitz-Thompson (1952) estimator is shown to be optimal within it, in the sense of having minimum expected design variance under the same model .

2. THE RESULTS

Let C denote the class of all unbiased estimators of the form

$$e_t = e(s,\underline{y}_t) = e(s,y_{it} \mid i \in s) \qquad (2.1)$$

such that e_t depends on only those y_{it}'s which are in the sample s and

$$\sum_{S} e(s,\underline{y}_t)\, p(s) = Y_t. \qquad (2.2)$$

For each unit i we assume the following Markov process model

$$y_{it} = \alpha_i\, y_{i,t-1} + Z_{it} \quad , \quad i=1,\ldots,N \qquad (2.3)$$

where α_i ($|\alpha_i| < 1$) is a constant and $\{Z_{it}\}$ is a *purely* random process with mean zero and variance σ_i^2. We also assume that under the above model y_{it} and y_{jt} are independently distributed for $i \neq j$.

Using the backward shift operator B defined as

$B^j y_{it} = y_{i,t-j}$ for all j, the model (2.3) can be written as

$$(1-\alpha_i B)\, y_{it} = Z_{it} \tag{2.4}$$

so that we have

$$\begin{aligned} y_{it} &= \frac{1}{1-\alpha_i B} Z_{it} \\ &= [\, 1 + \alpha_i B + \alpha_i^2 B^2 + \ldots]\, Z_{it} \\ &= Z_{it} + \alpha_i Z_{i,t-1} + \alpha_i^2 Z_{i,t-2} + \ldots \\ &= \sum_{k=0}^{\infty} \alpha_i^k Z_{i,t-k} \end{aligned} \tag{2.5}$$

which is known as a moving average process of infinite order.

Writing $E_m(V_m)$ as the operator for expectation (variance) with respect to the above model, we have

$$E_m(y_{it}) = 0, \qquad i = 1,\ldots,N \tag{2.6}$$

and

$$\begin{aligned} V_m(y_{it}) &= V_m\left[\sum_{k=0}^{\infty} \alpha_i^k Z_{i,t-k}\right] \\ &= \sum_{k=0}^{\infty} \alpha_i^{2k} \sigma_i^2 \\ &= \frac{\sigma_i^2}{(1-\alpha_i^2)}, \qquad i = 1,\ldots,N. \end{aligned} \tag{2.7}$$

Writing E_p (V_p) as the operator for expectation (variance) with respect to the sampling design p and writing Π_i as the inclusion-probabilities (assumed positive) for the units we have the following theorem.

Theorem 2.1

For any sampling design p and for any estimatior e_t in the class C, we have

$$E_m V_p(e_t) \geq \sum_{i=1}^{N} \frac{\sigma_i^2}{(1-\alpha_i^2)} \left(\frac{1}{\Pi_i} - 1\right). \tag{2.8}$$

Proof. Writing the Horvitz-Thompson (1952) estimator as

$$e_t^* = e^*(s, \underline{y}_t) = \sum_{i \in s} \frac{y_{it}}{\Pi_i}$$

and following Godambe and Joshi (1965) we may write any other estimator e_t of the class C as

$$e_t = e(s, \underline{y}_t) = e^*(s, \underline{y}_t) + h(s, \underline{y}_t) = e_t^* + h_t$$

where

$$0 = E_p h(s, \underline{y}_t) = \sum_{s} h(s, \underline{y}_t)\ p(s)$$

implying

$$\sum_{s \ni i} h(s, \underline{y}_t)\ p(s) = - \sum_{s \not\ni i} h(s, \underline{y}_t)\ p(s) \tag{2.9}$$

Then,

$$E_p V_m(e_t) = E_p V_m(e_t^*) + E_p V_m(h_t) + 2E_p C_m(e_t^*, h_t)$$

(C_m denoting cavariance with respect to the model).

Now,

$$\begin{aligned} E_p C_m\ (e_t^*, h_t) &= E_p E_m \left[\ e^*(s.\underline{y}_t)\ h(s, \underline{y}_t)\ \right] \\ &= E_p E_m \left[\sum_{i \in s} \frac{y_{it}}{\Pi_i}\ h(s, \underline{y}_t)\ \right] \\ &= E_m \left[\ \sum_{s} \sum_{i \in s} \frac{y_{it}}{\Pi_i}\ h(s, \underline{y}_t)\ p(s)\ \right] \\ &= E_m \left[\sum_{i=1}^{N} \frac{y_{it}}{\Pi_i} \sum_{s \ni i} h(s, \underline{y}_t)\ p(s)\ \right] \end{aligned}$$

$$= E_m\left[\sum_{i=1}^{N}\frac{y_{it}}{\Pi_i}\left\{-\sum_{s\not\ni i} h(s,\underline{y}_t)\ p(s)\right\}\right]$$

by using (2.9)

$= 0$, by independence of y_{it} and y_{jt}'s $\forall\ i\neq j$.

So, $\quad E_pV_m(e_t) \geq E_pV_m(e_t^*)$. (2.10)

Following Godambe and Thompson (1977), we can write

$$E_mV_p(e_t) = E_pV_m(e_t) + E_p\Delta_m^2(e_t) - V_m(Y_t),$$

where $\Delta_m(e_t) = E_m(e_t) - E_m(Y_t)$,

so that $E_mV_p(e_t) \geq E_pV_m(e_t^*) - V_m(Y_t)$ (2.11)

by using (2.10)

$$= \sum_{i=1}^{N}\frac{\sigma_i^2}{(1-\alpha_i^2)}\left[\frac{1}{\Pi_i}-1\right].$$

Next we consider the following theorem of practical importance.

Theorem 2.2

For any sampling design p and for any estimator e_t in the class C, we have

$$E_mV_p(e_t) \geq E_mV_p(e_t^*).$$

Proof. From the Godambe-Thompson (1977) formula applied to e_t^*, we write

$$E_mV_p(e_t^*) = E_pV_m(e_t^*) + E_p\Delta_m^2(e_t^*) - V_m(Y_t),$$

and use the fact that $\Delta_m(e_t^*) = 0$ to obtain

$$E_mV_p(e_t^*) = E_pV_m(e_t^*) - V_m(Y_t)$$

$$\leq E_mV_p(e_t),\ \text{by (2.11)}.$$

Remark 1. We may note that under the present stationary time series model all the estimators have been standardized to have model expectation zero.

Remark 2 Postulating various super-population models the several results on optimality of the Horvitz-Thompson (1952) estimator are available in the literature viz. Godambe (1955) ,Godambe and Joshi (1965) and many others where both the Horvitz-Thompson (1952) estimator and its competitors are based on an inclusion probability proportional to size (IPPS) sampling design .But it is interesting to note that under the present stationary time series model neither the Horvitz-Thompson (1952) estimator nor its competitor must be based on an IPPS sampling design or even a fixed sample size design and we get the optimality of the Horvitz-Thompson (1952) estimator for any varying probability sampling design .Also we may note that no restrictions on the model parameters are needed to establish the optimality.

ACKNOWLEDGEMENT

The author is grateful to the referee for many helpful suggestions which substantially improved on an earlier version of the paper.

BIBLIOGRAPHY

Godambe,V.P.(1955).A unified theory of sampling from finite populations. *J.Roy.Statist.Soc.Ser.B.*,17,269-278.

Godambe,V.P.and Joshi,V.M.(1965).Admissibility and Bayes estimation in sampling from finite populations -I. *Ann.Math.Statist.*36,1707-1722.

Godambe,V.P.and Thompson,M.E.(1977).Robust near optimal estimation in survey practice. *Bull. Int. Statist. Inst.*47,129-146.

Horvitz,D.G.and Thompson,D.J.(1952).A generalization of sampling without replacement from finite universes. *J.Amer.Statist.Assoc.*47,663-685.

Received June 1990; Revised March 1991.

Recommended by P. S. R. S. Rao, University of Rochester, Rochester, NY.

Refereed by James H. Drew, GTE Labs, Waltham, MA.