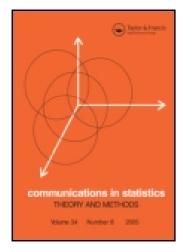
This article was downloaded by: [The University of Manchester Library]

On: 12 October 2014, At: 09:13 Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer

House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information: http://www.tandfonline.com/loi/lsta20

Note on umvsu-estimation under randomized response model

Parimal Mukhopadhyay ^a

^a Indian Statistical Institute , 203 B.T.Road, Calcutta, 700035

Published online: 27 Jun 2007.

To cite this article: Parimal Mukhopadhyay (1997) Note on umvsu-estimation under randomized response model, Communications in Statistics - Theory and Methods, 26:10, 2415-2420, DOI: 10.1080/03610929708832056

To link to this article: http://dx.doi.org/10.1080/03610929708832056

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

A NOTE ON UMVU-ESTIMATION UNDER RANDOMIZED RESPONSE MODEL

PARIMAL MUKHOPADHYAY Indian Statistical Institute 203 B.T.Road Calcutta 700035

Key words and Phrases Ordered design; uniformly minimum variance unbiased estimation; symmetric population function

ABSTRACT

Considering a class of randomized response trials for eliciting sensitive information from a sample survey and a class of ordered sampling designs, a uniformly minimum variance unbiased estimator of population variance (of the sensitive character) has been obtained. This note indicates that a theorem (theorem 3.9) of Cassel, Sarndal and Wretman (1977) and the results in the present note can be extended to estimation of any symmetric function of population values in the field of direct response surveys and randomized response surveys respectively.

1 Introduction and Preliminaries

Let \mathcal{P} be a finite population of units labelled $1, \ldots, i, \ldots, N$. Associated with i is a real quantity y_i , value of a character 'y' of interest $(i = 1, \ldots, N)$. We assume that y is a sensitive character (eg. containing some social

2415

Copyright © 1997 by Marcel Dekker, Inc.

stigma) about which the respondents may not answer truthfully to a direct question. In such circumstances generally some randomized response trials (RRT) are performed by virtue of which the respondent answers a value z_i when his true value is y_i . Such trials are performed independently from person to person. We are interested in estimating the population variance

$$V(\underline{y}) = a_1 \sum_{i=1}^{N} y_i^2 - a_2 \sum_{i \neq i'=1}^{N} y_i y_i'$$
 (1.1)

where

$$a_1=rac{1}{N}, \qquad a_2=rac{1}{N(N-1)}$$

by survey sampling, for which a sample s is selected with probability

$$p(s)\left(p(s) \geq 0, \sum_{s \in \mathcal{S}} p(s) = 1, \mathcal{S} = \{s\}\right).$$

$$\pi_k = \sum_{s \ni k} p(s), \pi_{kk'} = \sum_{s \ni (k,k')} p(s)$$

denoting the first-order and second-order inclusion probabilities respectively. We assume that the sampling design used in selecting s is an ordered design such that

- (i) each s is a sequence of n distinct labels and has nominal size n.
- (ii) all n! sequences s that are permutations of the same set of labels have equal probability of being selected.

(iii)
$$\pi_{ij} \geq 0 \forall i \neq j = 1, \ldots, N$$
.

Let $S=(K_1,\ldots,K_n)$ be a random vector taking values $s=(k_1,\ldots,k_n)$ where K_i is a random variable denoting the label of the units selected at the ith draw in S. Data obtained from the survey are $\mathcal{D}=((K_1,z_{K1}),\ldots,(K_n,z_{Kn}))$ available through two types of randomisation, one due to sampling design and the other due to RRT. An estimator e=(e,z) is a function defined SXR_N such that for a given s its value depends on $z(=(z_1,\ldots,z_N))$ only through $i\in s$. Let E_p , E_R denote expectation operator with respect to sampling design p and RR respectively. An estimator t is unbiased for V(y) if

$$E_p E_R(t(s,z)) = V(y) \forall y \in R_N$$
 (1.2)

Consider Erikson's (1973) RR model as follows. When the *i*th individual (with true value y_i) is selected in the sample he chooses a quantity at radom from the set $(a_{1i}, \ldots, a_{M,i})$ and a quantity b_i at random from the set $(b_{1i}, \ldots, b_{Q,i})$ and responds

$$Z_i = a_i y_i + b_i \tag{1.3}$$

It can be seen that

$$E_R(r_i) = y_i, E_R(R_i^2) = y_i^2, E_R(R_{ij}) = y_i y_j$$
 (1.4)

where

$$\begin{split} r_{i} &= \frac{z_{i} - \bar{a}_{i}}{\bar{b}_{i}}, \\ R_{i}^{2} &= \frac{z_{i}^{2} - \bar{b}_{i} - 2\bar{a}_{i}\bar{b}_{i}r_{i}}{\bar{a}_{i}}. \\ R_{ij} &= \frac{z_{i}z_{j} - \bar{a}_{i}\bar{b}_{j}r_{i} - \bar{a}_{j}\bar{b}_{i}r_{j} - \bar{b}_{i}\bar{b}_{j}}{\bar{a}_{i}\bar{a}_{j}}. \\ \bar{a}_{i} &= \frac{1}{M_{i}} \sum_{j=1}^{M_{i}} a_{ji}. \\ \bar{b}_{i} &= \frac{1}{Q_{i}} \sum_{j=1}^{Q_{i}} b_{ji}, \\ \bar{a}_{i} &= \frac{1}{M_{i}} \sum_{j=1}^{M_{i}} a_{ji}^{2}, \\ \bar{b}_{i} &= \frac{1}{Q_{i}} \sum_{j=1}^{Q_{i}} b_{ji}^{2}. \end{split}$$

Let $e = (e_1, \ldots, e_N) \in R_N$ be a vector of known constants. For a given sample $s = \{i_1, \ldots, i_n\}$, define the quantities:

$$\frac{n(R_{ik}^2 - e_{ik}^2)}{N\pi_{ik}} = T_{ik}^2, \quad (k = 1, \dots, n)$$

$$\frac{n(n-1)(R_{i_k i_e} - e_{i_k i_e})}{N(N-1)\pi_{i_k i_e}} = T_{i_k i_e} \ (k \neq e = 1, \ldots, n)$$

and the set of unlabelled data

$$\underset{\sim}{T_s} = (T_{i1}^2, \dots, T_{i_n}^2, T_{i_1 i_2}, \dots, T_{i_{n-1} i_n} \mid \text{labels dropped})$$

Clearly, the members of T_s may not be distinct. We shall consider in this note the class of estimators that depend on D only through T_s . We call this class of estimators as Δ .

2 An UMVU-Estimator

Lemma1. Let p be any fixed effective sampling design. If $\tilde{a}_i (\neq 0)$, $\tilde{a}_i (\neq 0)$, \tilde{b}_i , \tilde{b}_i are known constants and $g(T_s)$ is a real valued statistic such that.

$$E_p E_R \{g(T_s)\} = 0$$

for every RR-model (1.3) and every $y \in R_N$, then $g(T_s) = 0$ identically.

Proof. The lemma follows from lemma 3.7 of Cassel, Sarudal and Wretman (1977)(whose proof is similar to a proof in Das(1962)) and theorem 2.1 of Adhikari, Chaudhuri and Vijayan(1984)(which can be easily extended to Eriksson's model (1.3)).

Theorem 1. Under the RR-models satisfying (1.3), for any given ordered design satisfying conditions (i) -(iii) and for any constant vector $e \in R_N$, the generalised difference estimator of population variance,

$$t'_{GD}(\underline{e}) = \frac{1}{N} \sum_{i \in s} T_i^2 - \frac{1}{n(n-1)} \sum_{i \neq j \in s} T_{ij} + V(\underline{e})$$
$$= w(T_s) + V(\underline{e})$$
(2.1)

is UMV in the class of all unbiased estimators of V(y) in Δ .

Proof. It follows by virtue of the assumptions (i)-(iii) on the class of sampling designs, that the random variables $(T_{i1}^2, \ldots, T_{i_n}^2)$ have exchangeable joint distribution with common expectation $E_p E_R(T_{i_k}^2) = \sum_{i=1}^N (y_i^2 - e_i^2)/N$. Similarly for all permutations of (i_1, \ldots, i_n) , the random variables $(T_{i_1 i_2}, \ldots, T_{i_{n-1} i_n})$ have the same joint distribution with common expectation $E_p E_R(T_{i_k i_e}) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N (y_i y_j - e_i e_j)$. Hence $E_p E_R(T_{i_k}^2 - T_{i_k i_l}) = V(y) - V(e)$. Again, it follows from lemma 3.6 of Cassel, Sarndal and Wretman(1977)[with a slight modification] that a MVUE of V(y) - V(e) in the

UMVU-ESTIMATION 2419

class of all estimators in Δ must be symmetric in T_s . Now $w(T_s)$ is a symmetric function in T_s and is unbiased for V(y) - V(e). Lemma 1 states that under randomisation due to sampling design (satisfying(i)-(iii)) and RR-models satisfying (1), distribution of T_s is complete. Hence $w(T_s)$ is the unique unbiased estimator of V(y) - V(e). Consequently $t'_{GD}(e)$ is the UMVUE of V(y) in the class of all unbiased estimators in Δ .

Corollary 1. Under the RR-models satisfying (1.3), for any given ordred design satisfying (i)-(iii), the Horvitz-Thompson estimator $\sum_{i \in s} \frac{R_i^2}{N\pi_i} - \sum_{i \neq j \in s} \frac{R_{ij}}{N(N-1)\pi_{ij}}$ is UMVUE for $V(\underline{y})$ in the class of all unbiased estimators that depend on $\mathcal D$ only through $T_s^0 = \underline{T}_s$ with $\underline{e} = \underline{e}$.

Remarks. Clearly, result on UMV in theorem 3.9 of Cassel and others (1977) can be extended to estimation of any symmetric function of population values, say, population moments, population cumulants, etc. in direct-response surveys. Again results in theorem 1 (which is extension of theorem 1 in Tracy and Mukhopadhyay (1994)) can be extended to any symmetric function of population values in randomized response surveys satisfying (1.3).

Note 2.1. We note that the conditions (i)-(iii) on sampling designs is not required for lemma 3.7 of Cassel, Sarndal and Wretman (1977). Thus theorem 2.1 of Adhikari, Chaudhuri and Vijayan (1984) can be modified as follows.

Lemma 2.2. Let p be any given FES(n) design. Then for the class of RR-models (1.3) $E_pE_R\{g(Z_s)\}=0 \forall \underline{y}\in R_N\Rightarrow g(Z_s)=0 \forall s: p(s)\geq 0$ and $\forall \underline{y}\in R_N$, where $g(z_s)$ is a function that depends on $\mathcal D$ only through the sets of unlabelled responses Z_s .

BIBLIOGRAPHY

Adhikari, A.K., Chaudhuri, A., and Vijayan, K. (1984): Optimum sampling strategies for randomized response trials. Int. Statist. Rev., 52, 115-129.

Cassel, C.M., Sarndal, C.E. and Wretman, J.H. (1977): Foundation of Inference in Survey Sampling, John Wiley, New York.

2420 MUKHOPADHYAY

Das, A.C.(1962): On MVU estimates of parameters of a finite population based on varying probability sampling. Cal. Statist. Assoc.Bull. 11,39-48.

- Eriksson, S.A.(1973); A new model for randomized response. Int. Statist Rev., 41, 101-103.
- Tracy, D.S. and Mukhopadhyay, P. (1994): On UMVU-estimation under randomized response models. Statistics, 25, 173-175.

Received August, 1995; Revised April, 1997.