ON NON-NEGATIVE VARIANCE ESTIMATION IN IKEDA-MIDZUNO-SEN STRATEGY

By

ARIJIT CHAUDHURI

Indian Statistical Institute, Calcutta

(Received : December, 1979)

SUMMARY

Several unbiased estimators for the variance of the unbiased ratio estimator based on the scheme of sampling with probabilities proportional to aggregate size (PPAS) measures are considered along with simple and verifiable sufficient conditions for their uniform non-negativity. The study is extended to cover the well-known Horvitz/Thompson estimator.

INTRODUCTION

The problem of finding unifiormly non-negative estimators for variances of certain unbiased estimators of finite population totals is well-known. Many results are also available. A few relevant references inter alia are Ajgaonkar [1], Rao [16], [17], Lanke [8], Vijayan [19], Chaudhuri [4], [5], Rao/Vijayan [15], Rao [14], Bandyopadhyay/Chattopadhyay/Kundu [2] and Chaudhuri/Arnab [6]. In particular, we consider estimation of variance of (1) the ratio estimator based on the famous Ikeda/Midzuno/Sen's [9], [18] PPAS sampling scheme and (2) the Horvitz/Thompson [7] estimator (HTE, say). Here we suggest several variance estimators, supplementing the known ones, and indicate 'sufficient conditions' for their uniform non-negativity, which may be readily checked and verified in practice in any given survey situation as they are in terms of x-values which are at hand. The 'non-negativity' conditions are not 'identical' for the various alternative estimators mentioned. So, a policy to follow in practice may be to 'check' them one

Running Title : Non-negative variance estimation.

by one for the x-values given in a paritcular situation and 'use' the ones that satisfy the sufficiency conditions because we are sure about their uniform [in y-values] non-negativity. The other estimators for which the 'sufficient' (but not known if 'necessary' as well) non-negativity conditions are not met may still be non-negative of course but we cannot verify if they are really so. So they need not be discarded but it may be safe to use the former ones, if available. If more than one pass the test of non-negativiy then a choice among them may be guided by usual considerations. For instance we may examine their 'stabilities' considering their variances in general (*i.e.* not for the given set of x-values alone). Also one may apply Ajgaonkar's [1] criterion in choosing among all the available unbiased variance-estimators in terms of their probabilities of assuming negative values (in order to include those that pass the sufficient test for non-negativity). But our modest objective in this paper is only to suggest verifiable tests, given a set of size-measures, to check if certain variance estimators can be uniformly non-negative (and usable, if so). In section 3 we present some numerical data just to illustrate how even in simple situations some estimators may 'pass' while others may 'fail' the respective 'sufficiency tests' for uniform non-negatvity.

2. Notations, the estimators and the non-negativity conditions.

Let x_i's (>0 and known), y_i's (real, unknown) be variate-values for the units of a finite population U of size N. A typical sample (throughout having n units, all distinct) will be denoted as s with $p\ (s)$ as ites selection-probability for a design p to be suitably chosen. Let $Y=\Sigma Y_i$ (to be estimated), $X=\Sigma x_i$, the population totals, $Y_s=\Sigma' x_i$ (the sample totals, $\Sigma'\equiv$sum over sampled units),

$$y_i=x_i/X\,,\ P_s=\Sigma'\ P_i,\ M_r=\binom{N-r}{n-r}\ r=0,1,2;$$

$$a_{ij}=(y_i/p_i-y_j/p_j)^2\ p_ip_j.$$

Then the ratio estimator is $t=X\ y_s/x_s$; for the Ikeda/Midzuno/Sen (IMS, say) scheme, $p(s)=p_s/M_1$. Let $p\ (s\mid i)$=probability of choosing s given that unit i is chosen on the firts draw, $p\ (s/ij)$ = probability of choosing s given that i, j are the units chosen on the first two draws ($i\neq j$). Then the well-known

Murthy [10] estimator is $t_M = \Sigma' y_i \, p(s/i)/p(s)$ with a variance, (vide Rao [14]).

$$V(t_M) = \sum_{i>j}\sum a_{ij} \left[1 - \Sigma'' \; p(s/i)\, p(s/j) / p(s) \right]$$

(writing Σ'' for sum over samples containing i and j $(i \neq j)$), the latter having a non-negative unbised estimator [vide Pathak/ Shukla [11]] $v_M = \Sigma t''' [p(s)\, p(s/ij) - p(s/i)\, p(s/j)]\, a_{ij}/p^2(s)$, where Σ''' denotes sum over i, j in s with $i > j$. Noting that for the IMS scheme, $p(s/i) = 1/M_1$, $p(s/ij) = 1/M_2$ and hence

$$t_M = t, \; V(t) = \sum_{i>j}\sum a_{ij} \left(1 - 1/M_1 \sum{}'' \, 1/p_s\right),$$

and also that the inclusion-probability for the pair of units i, j for this scheme is

$$I'_{ij} = \Sigma'' \, p(s) = ab + ac\,(p_i + p_j),$$

where

$$a = (n-1)/(N-1), \; b = (n-2)/(N-2)$$

$$c = (N-n)/(N-2),$$

we have two unbiased estimators for $V(t)$ as,

$$v_1 = \Sigma''' a_{ij} \,(N-1)/(n-1)\,(p_s - (n-1))\,(N-1))/p^2_s$$

$$= \Sigma''' \frac{1}{p_s} \left(\frac{1}{a} - \frac{1}{p_s} \right)$$

and $$v_2 = \Sigma''' a_{ij} \,(1 - 1/M_1 \;\; \Sigma''' \, 1/p_s)/I'_{ij} \,,$$

These coincide with what Rao/Vijayan [15] in a different way proposed earlier : their common non-negativity condition, as they also noted, is

$$\min p_s \geqslant a. \qquad \ldots (2.1)$$

Choosings two fixed (not depending on s, y_i's and x_i's) constants α, β in [0, 1] and recalling Raj's (1954) and Ajgaonkar's [1] works the following additional unbiased estimators for $V(t)$ are available, viz.

$$v_3 = 1/M_2 p(s)\, \Sigma''' \, a_{ij} \,(1 - 1/M_1 \Sigma'' \, 1/p_s)$$

$$v_4 = \Sigma''' a_{ij} \,[\, 1/I'_{ij} - 1/M_2 p_s \;\; \Sigma'' \, 1/p_s \,]$$

$$v_5 = \Sigma''' a_{ij} \,[\, 1/M_2\, p(s) - 1/M_1 \, I'_{ij} \, \Sigma \, 1/p_s]$$

$$v_6 = t^2 - M_1/p_s \,[\, 1/M_1 \, \Sigma' \;\; y_i^2 + 2/M_2 \, \Sigma''' \, y_i \, y_j]$$

$$v_7 = t^2 - [\, \Sigma' \, y_i^2 I'_i + 2 \, \Sigma''' \, y_i \, y_j]/I'_{ij}],$$

where $I'_i = a + c.p_i \equiv$ incleusion-probabiliting of i ($= 1, \ldots, N$),

$$v_8 = \Sigma''' \, a_{ij} \left[(\alpha/I'_{ij} + M_1/M_2 \, (1-\alpha)/p_s) - (\beta/M_2 \, M_1/p_s + (1-\beta)/I'_{ij})/M_1 \, \Sigma'' \, 1/p_s \right].$$

We will later write, for simplicity, [vide Table 3.2, section 3]

$$A_{ij}(s) = \alpha/I'_{ij} + M_1/M_2 \, (1-\alpha)/p_s$$

and $$B_{ij}(s) = [\beta/M_2 \, M_1/p_s + (1-\beta)/I_{ij}]/M_1 \, \Sigma'' \, 1/p_s].$$

Then, sufficient conditions for non-negativity respectively for v_i ($i = 3, 4, 5$) are

$$\Sigma'' \, 1/p_s \leqslant M_1 \; \forall i, j \qquad \ldots(2.2)$$

$$p_s \geqslant 1/M_1 M_2 \, (\Sigma'' \, p_s) \Sigma'' \, 1/p_s) \forall s \ni i, j \; (i \neq j) \qquad \ldots(2.3)$$

and $$p_s \leqslant (N\text{-}1)/\,(n\text{-}1) \, (\Sigma'' \, ps)/\,(\Sigma'' \, 1/p_s) \; \forall \; s \ni i.j, \; \forall \; i.j. \; (i \neq j) \qquad \ldots(2.4)$$

Remark I. If (2.1) holds then does (2.2) and hence the latter is a weaker condition rendering v_3 preferable to v_1 and v_2 more frequently.

Also. $$(\Sigma'' \, p_s) \, (\Sigma'' \, 1/p_s) \geqslant 1$$

and $$(\Sigma'' \, p_s) \, /(\Sigma'' \, 1/p_s) \leqslant 1 \; \forall \; i.j.$$

Still, (2.3), (2.4) look less stringent than (2.1). A sufficient non-negativity condition for v_8 is obvious but its relative stringency is hard to specify. We are unable to give easy non-negativity conditions for v_6 and v_7 rendering their use less decisive in practice.

Remark II. The above approach immediately extends to take care of HTE for any p with fixed n. Writing I_i, I_{ij} as inclusion-probabilities of first two orders for any such p, the HTE is

$$t_1 = \Sigma' \, y_i/I_i \text{ with } V(t_1) = \Sigma \, \Sigma b_{ij} \, (I_i \, I_j - I_{ij}),$$

where $$b_{ij} = (y_i/I_i - y_j/\,I_j)^2.$$

To the existing unbiased estimators for $V(t_1)$ we may add

$$v'_1 = \Sigma'' \, b_{ij} \, (I_i I_j/I_{ij} - I_{ij}/M_2 \, p \, (s)\,) \text{ and}$$

$$v'_2 = \Sigma''' \, b_{ij} \, (I_i I_j/p \, (s) \, M_2 - 1).$$

Their easily verifiable non-negativity conditions are respectively

$$p(s) > I_i{}^2{}_j / I_i I_j M_2 \ \forall\ i,j \text{ for } i,j \in s \qquad \ldots \qquad \ldots \qquad (2.5)$$

and $$p(s) < I_i I_j / M_2 \ \forall\ i,j \in s. \qquad \ldots \qquad \ldots \qquad (2.6)$$

So, in practice, for the chosen design one can ascertain if any of the estimators among the available alternatives is uniformly non-negative and if so may use safely any of them, if none is, then one should think of further alternatives. Any way one can use only any one which turns out non-negative (whether it satisfies the sufficiency condition or not) for the survey data at hand.

APPLICATION OF THE THEORY TO NUMBERICAL DATA.

From the x_i values as eye-estimates of the number of households [vide Table 9.1. p. 198, Raj [13]; also referred to by Ajgaonkar [1] and Horvitz/Thompson [7] in city blocks as reported by Raj [13] we take just 5 as given below. Thus we take a population of size 5. Also we takes samples of size 3 using (1) IMS scheme and (2) also Brewer's [3] scheme of choosing two units in first two draws with inclusion probabilities proportional to sizes (IPPS) and following up with another draw with an equal probability from the remaining units (to be called scheme B, say). Denoting the inclusion probabilities of the first two order for Brewer's [3] scheme in the first two draws by $\pi_i(2)$ $\pi_{ij}(2)$ and in the over-all sample of size 3 for scheme B by $\pi_i(3)$, $\pi_{ij}(3)$ we have

$$\pi_i(2) = 2p_i, \pi_{ij}(2) = \frac{\pi_i(2)\,\pi_j(2)}{2 + \sum \frac{\pi_i(2)}{1-\pi_i(2)}} \left(\frac{1}{1-\pi_i(2)} + \frac{1}{1-\pi_j(2)} \right)$$

and $$I_i = \pi_i(3) = \tfrac{1}{3} + \tfrac{2}{3}\pi_i(2)\,,\ I_{ij} = \pi_{ij}(3) = \tfrac{1}{3}\big(\pi_i(2) + \pi_j(2) + \pi_{ij}(2)\big).$$

We may also note that the selection-probability of a typical sample $s=(i,j,k)$, say, according to Scheme B is $p(s)=\frac{1}{3}(\pi_{ij}(2)+\pi_{ik}(2)+\pi_{jk}(2))$. Some numerical findings (in brief, to save space) are given below illustrating the application of the method of verifying sufficiency conditions of non-negativity of variance estimators. For $x_1=30$, $x_2=27$, $x_3=26$,

$x_4=21$, $x_5=19$. (the eye estimates) $p_1=.2439=\pi_1(2)/2$, $P_2=.2195=\pi_2(2)/2$, $P_3=.2114=\pi_3(2)/2$, $P_4=.1707=\pi_4(2)/2$, $p_5=.1545=\pi_5(2)/2$, the p_s values are: .6748, .6341, .6172, .6260, .6098. .6016, .5854, .5447, .5366, .5691 respectively for the samples $s=(123)$, (124), (125), (134), (135), (234), (235), (245), (345) and (145). Since $a=\frac{n-1}{N-1}=\frac{1}{2}$ and $p_s>1/2$ for every s, the condition (2.1) is satisfied implying non-negativity of v_1, v_2, v_3. For the others calculations are needed as in the table 3.1 below. For Scheme *B*, we find $I_1=.6585$ $I_2=:6260, I_3=.6152$, $I_4=.5609$. $I_5=.5393$.

TABLE 3.1

Showing data to check non-negativity of variance estimators

s	ij (es)	$1/I_{ij}$	$1/M_2P_s\ \Sigma''\ 1/p_s$	M_1/M_2P_s	$1/M_1I'_{ij}\ \Sigma''\ 1/p_s$
123	12	3.1140	2.3105		2.4275
	13	3.1404	2.3311	2.9638	2.4900
	23	8.3894	2.3960		2.7411
124	12	3.1140	2.4587		2.4275
	14	3.2801	2.5924	3.1540	2.6960
	24	3.3700	2.6678		2.8505
125	12	3.1140	2.5232		2.4275
	15	3.3393	3.70S7	2.2368	2.7914
	25	3.4325	2.7850		2.9534
134	13	3.1404	2.5128		2.4700
	14	3.2801	2.6259	3.1948	2.6960
	34	3.4010	2.7279		2-9040
135	13	3.1404	2.5797		2.4700
	15	3.3393	2.7416	3.2798	2.7914
	35	3.4646	2.8482		3.0094
I45	14	3.3801	2.8886		2.6960
	15	3.3393	2.9377	3.5144	2.7914
	45	3.6555	3.1962		3.3062
234	23	3.3895	2.6885		2.7411
	24	3.3700	2.8119	3.3244	2.8505
	34	3.4010	2.8386		2.9040
235	23	3.3894	2.7629		2.7411
	25	3.4325	2.9325	3.4164	2.9534
	35	3.4646	2.9675		3.0094
245	24	3.3700	3.1058		2.8505
	25	3.4325	3.1593	3.6718	2.9534
	45	3.6355	3.3392		3.3063
345	34	3.4010	3.1825		2.9040
	35	3.4646	32375.	3.7272	3.0094
	45	3.6355	3.3897		3.3063

From the above it follows that given the x_i 's as before, v_4, v_5 are non-negative uniformly in all y_i' s. Taking $\alpha=.4$ and $\beta=.2$, in v_8, the following table 3.2 shows its non-negativity. However this table also shows that (2.5) and (2.6) are not satisfied implying uncertainty about the signs of v_1', v_2' when based on Scheme *B*. But (2.5) and (2.6) are satisfied if the *HTE* is based on *IMS* scheme as revealed from Table 3.2.

TABLE 3.2

Showing data to check signs of variance estimators

s	$ij(\epsilon s)$	$A_{ij}(s)$	$B_{ij}(s)$	$\pi_{ij}(2)$	I_{ij}	$I^2_{ij}/I_i I_t$	$I_t I_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
123	12	3.0239	2.4256	.1472	.3580	.3109 (.5438)	.4122 (.7585)
	13	3.0344	2.4422	.1399	.3502	.3027 (.5383)	.4051 (.7534)
	23	3.1340	2.6722	.1201	.3273	.2782 (.4714)	.3851 (.7387)
124	12	3.1380	2.4338	.1472	.3580	.3109 (.5438)	.4122 (.7585)
	14	3.2044	2.6758	.1064	.3189	.2634 (.5106)	.3694 (.7281)
	24	3.2404	2.8140	.0911	.2905	.2403 (.4934)	.3511 (.7138)
125	12	3.1877	2.4467	.1472	.3580	.3109 (.5438)	.4122 (.7585)
	15	3.2778	2.7742	.0943	.2979	.2448 (.4996)	.3551 (7180)
	25	3.3151	2.9197	.0806	.2762	.2260 (.4722)	.3376 (.7040)
134	13	3.1730	24786	.1399	.3502	.3027 (.5383)	.4051 (.7534)
	14	3.2289	2.6820	.1064	.3119	.2634 (.5106)	.3694 (.7281)
	34	3.2773	2.8688	.0864	.2835	.2329 (.4877)	.3451 (.7091)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
135	13	3.2240	2.4920	.1399	.0502	.3027 (.5383)	.4051 (.7534)
	15	3.3036	2.7814	.0943	.2970	.2484 (.4996)	.3551 (.7180)
	35	3.3537	2.9773	.0765	.2694	.2188 (.4765)	.3318 (.6993)
1 5	14	3.4207	2.7345	.1064	.3119	.2634 (.5106)	.3694 (.7281)
	15	3.4444	2.8206	.0943	.2970	.2484 (.4996)	.3551 (.7180)
	45	3.5628	3.2843	.0576	.2360	.1812 (.4478)	.3025 (.6758)
234	23	3.3504	2.7306	.1201	.3273	.2782 (.4714)	.3851 (.7387)
	24	3.3426	2.8428.	.0911	.2905	.2403 (.4934)	.3511 (.7138)
	34	3.3550	2.8909	.0864	.2135	.2329 (4877)	.3451 (.7091)
235	23	3.4056	2.7554	.1201	.3273	.2782 (4714)	.3851 (.7387)
	25	3.4228	2.9506	.0806	.2762	.2260 (4822)	.3367 (.7040)
	35	3.4357	3.0010	.0765	.2694	.2188 (.4765)	.3318 (.6993)
245	24	3.5511	2.9016	.0911	.2905	.2403 (.4934)	.3511 (.7138)
	25	3.5761	2.9946	.0806	.2762	.2260 (.4822)	.3376 (.7040)
	45	3.6573	3.3129	.0576	.2360	.1812 (.4478)	.3025 (.9758)
345	34	3.5967	2.9597	.0864	.2835	.2329 (.4877)	.3451 (.7091)
	35	3.6222	3.0550	.0765	.2694	.2188 (.4765)	.3318 (.6993)
	45	3.6904	3.3230	.0576	.2360	.1812 (.4478)	.3025 (.6758)

In order to check (2.5) and (2.6) for *IMS* scheme we have to check if for the samples so drawn, the p_s values (as given earlier) respectively (1) exceed $2I'^2_{ij}/I'_i I'_j \; \forall s \; ij(\forall i \neq j)$ and (2) fall short of (2) $2I'_i I'_j \forall \; s \in i, j(\forall \; i \neq j)$. For the given data I'_i values are $I'_i = .6220$, $I'_2 = .6098$. $I_3 = .6057$, $I'_4 = .5854$ and $I'_5 = .5773$. The values of $2I'^2_{ij}/I'_i$ and $2I'_i I'_j$ are given for various ij's in Table 3.2 respectively in parentheses in columns (7) and (8).

ACKNOWLEDGEMENT

The author is grateful to the referee for his helpful comments on earlier drafts.

REFERENCES

[2] Ajgaonkar, S.G.P. (1967) : Unbiased estimators of the variance of the Narain, Horvitz and Thompson estimator. *Sankhya*, Ser A **29**, 55-60

[2] Bandyopadhyay, S., Chattopadhyay, A.K., and Kundu, S. (1977) : On estimation af population total. *Sankhyā*, Ser C, **39** 28-42.

[3] Brewer, K.R.W. (1963) : A model of systematic sampling with unequel probabitities. *Aust. J St.* **5**, 5-13.

[4] Chaudhuri, A. (1976) : A non-negativity criterion for a certain variance estimator. *Metrika*, **23**, 201-205.

[5] Cnaudhuri, (1981) : Non negative unbiased varirnce estimators. In *Current tapies in survey sampling* Ed. Krewski, D., Platck. M. and Rao. J.N.K., 317-328.

[6] Chaudhri, A and Arnab. R. (1981) : On non-negative variance estimation. *Metrika*, **28**, 1-12.

[7] Horvitz, D.G. and Thompson, D.J. (1952) : A generalization of sampling without replacement from a finite universe. *J. Amer, State. Assn.* **47**, 663-685.

[8] Lanké, J. (1974) : On non-negative variance estimators in survey sampling. *Sankhyā*, Ser. C, **36**. 33-42

[9] Midzuno, H. (1952) : On the sampling system with probability proportional to sum of sizes. *Ann. Math. Stat.* **3**, 99-107.

[10] Murthy, M.N. (1957) : Ordered and unordered estimators in sampling without replacement *Sankhya* **18**, 379-390.