Jour. Ind. Soc. Ag. Statistics
Vol. LX, No. 3 (1988), pp. 218-222

ON NARAIN'S NECESSARY CONDITION IN SAMPLING

PADAM SINGH
Institute for Research in Medical Statistics (Delhi Chapter), New Delhi
(Received : February, 1988)

SUMMARY

Narain [1] gave a necessary condition for without replacement sampling to have smaller variance than with replacement sampling. In this paper that necessary condition has been disproved.

Keywords : Inclusion probabilities; Loading principle minor Negative semi-definite; Positive definite; Quadratic form; Sampling design.

Introduction

When a sample of n units is selected by unequal probabilities with replacement, an unbiased estimator of population total Y, with usual notations, is given by

$$\hat{Y}_{wr} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}. \tag{1.1}$$

The variance of (1.1) is given by

$$V(\hat{Y}_{wr}) = \frac{1}{n}\left(\sum_{i=1}^{N} \frac{y_i^2}{p_i} - Y^2 \right). \tag{1.2}$$

Also, when sampling is done by varying probabilities without replacement the estimator given by

$$\hat{Y}_{wtr} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \tag{1.3}$$

is unbiased for Y with its variance given by

$$V(\hat{Y}_{wtr}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} + \sum_{i \neq j}^{N} \frac{y_i \, y_j}{\pi_i \, \pi_j} \pi_{ij} - Y^2. \tag{1.4}$$

Narain [1] obtained a necessary condition for $V(\hat{Y}_{wtr})$ to be smaller then $V(\hat{Y}_{wr})$ for the situations when $\pi_i = np_i$, irrespective of y's as

$$\pi_{ij} \leqslant \frac{2\,(n-1)}{n} \pi_i \, \pi_j \qquad \text{for all } i, j \tag{1.5}$$

2. Comparison of $V(Y_{wr})$ and $V(Y_{wtr})$

The difference $V(\hat{Y}_{wr}) - V(\hat{Y}_{wtr})$ for the case $\pi_i = np_i$, takes the form

$$D = \frac{n-1}{n} \left[\sum_{i=1}^{N} \gamma_i^2 + \sum_{i \neq j}^{N} \gamma_i \, \gamma_j \left(1 - \frac{n}{n-1} \frac{\pi_{ij}}{\pi_i \, \pi_j} \right) \right]. \tag{2.1}$$

Clearly, the term in the bracket is the quadratic form

$$y A y' = [\gamma_1, \gamma_2 \ldots \gamma_N] \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots \gamma_{1N} \\ \lambda_{21} & \lambda_{22} & \ldots \lambda_{2N} \\ \vdots & & \\ \lambda_{N1} & \lambda_{N2} & \ldots \lambda_{NN} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_N \end{bmatrix} \tag{2.2}$$

$$\text{with} \quad \lambda_{ii} = 1 \quad \text{and} \quad \lambda_{ij} = 1 - \frac{n}{n-1} \frac{\pi_{ij}}{\pi_i \, \pi_j} \tag{2.3}$$

The core of Narain's proof lies in the following argument : Since (2.2) is a quadratic form, the necessary and sufficient condition for sampling without replacement to have smaller variance than sampling with replacement irrespective of y's is that all leading principle minors of A must be non-negative. The second order principle minors give

$$1 - \lambda_{ij}^2 \geqslant 0, \tag{2.4}$$

or $$\lambda_{ij} \leqslant 2\,\frac{(n-1)}{n}\,\pi_i\,\pi_j, \qquad (2.5)$$

which is the necessary condition according to Narain.

One important aspect in the proof has been overlooked. A quadratic form takes positive values not only in the case of the positive definite form but also many times in the indefinite form. Therefore the necessary condition for without replacement sampling to have smaller variance than with replacement sampling is that the quadratic form should not be negative semidefinite always. The condition for this is that the leading principle minors should not be negative and positive alternatively. This implies $\lambda_{ii} \geqslant 0$ which is always true. Thus, there is no minimum requirement for without replacement sampling to be better than with replacement sampling.

3. Illustration

Consider the expmple of 9 units with x and y values as given below:

Sl. No.	%	%
1	10	10
2	2	4
3	9	9
4	3	6
5	8	8
6	4	4
7	7	7
8	5	0
9	6	6

Further consider the following varying probability sampling design for sample size 2 for the above population

Sl. No. of Sample	*Units in the Sample*	*Probability of Selection*
1	1.2	1/90
2	1.3	11/135
3	1.4	1/45
4	1.5	19/270
5	1.6	1/30
6	1.7	8/135
7	1.8	2/45
8	1.9	13/270
9	2.3	1/540
10	2.4	1/60
11	2.5	1/270
12	2.6	2/135
13	2.7	1/180
14	2.8	7/540
15	2.9	1/135
16	3.4	1/90
17	3.5	13/180
18	3.6	1/45
19	3.7	11/180
20	3.8	1/90
21	3.9	1/20
22	4.5	1/540
23	4.6	7/270
24	4.7	1/270
25	4.8	13/540
26	4.9	1/180
27	5.6	1/90
28	5.7	17/270
29	5.8	1/45
30	5.9	7/135
31	6.7	1/540
32	6.8	19/540
33	6.9	1/270
34	7.8	1/90
35	7.9	29/540
36	8.9	1/540

It is easy to see that for above sampling design the inclusion probabilities of units are exactly proportional to their respective sizes. Also for the above sampling design :

$$\pi_2 = 2/27$$
$$\pi_4 = 1/9$$

and $$\pi_{24} = 1/60$$

Thus $\pi_2 \pi_4 - \pi_{24}$ is negative violating Narain's necessary condition. The variance of the estimate of population total based on without replacement πPS sample of size 2 and with replacement *PPS* sample of size 2 are as under

$$V(\hat{Y}_{wtr}) = 236$$

$$V(\hat{Y}_{wr}) = 270$$

From this example it has been verified that without replacement sampling provides more efficient estimate than with replacement sampling even when the Narain's necessary condition is violated.

REFERENCE

[1] Narain, R. D, (1951) : On sampling without replacement with varying probabilities, *Jour. Ind. Soc. of Ag. Stat*. 3.