

## ON MANGAT'S IMPROVED RANDOMIZED RESPONSE STRATEGY

H. P. Singh, N. Mathur

## 1. INTRODUCTION

To acquire the reliable data for estimating  $\pi$ , the proportion of population possessing a sensitive attribute A say, Warner (1965) was first to develop a procedure called the randomized response technique (RRT). Subsequently, several modifications of RRT model have been suggested by various authors, for instance, see Chaudhuri and Mukerjee (1988) and Hedayat and Sinha (1991). Mangat and Singh (1990) suggested an alternative randomized response procedure which uses two randomized response devices making the interview procedure a little cumbersome. This motivated Mangat (1994) to suggest a simple RRT. According to him, for estimating  $\pi$ , a simple random sample of  $n$  people is selected with replacement from the population. Each of  $n$  respondents is instructed to say, 'yes' if he or she has the attribute A. If he or she does not have attribute A, the respondent is required to use the Warner randomization device consisting of two statements:

- i) 'I belong to attribute A' and
- ii) 'I do not have attribute A',

represented with probabilities  $p$  and  $(1-p)$  respectively. Then he or she is to report 'yes' or 'no' according to the outcome of this randomization device and the actual status that he or she has with respect to attribute A. The procedure protects the privacy of the respondent.

Assuming that the reporting is completely truthful,  $\theta$ , the probability of 'yes' answer is given by

$$\theta = \pi + (1-p)(1-\pi) \quad (1)$$

Mangat (1994) suggested the maximum likelihood estimator (mle) of  $\pi$  as

$$\hat{\pi} = \frac{(\hat{\theta} - 1 + p)}{p}, \quad (2)$$

where  $\hat{\theta} = \frac{n_1}{n}$  is the observed proportion of 'yes' answers obtained from the  $n$  sampled people and  $n_1$  is the number of 'yes' answers obtained from  $n$  respondents.

The estimator  $\hat{\pi}$  has been shown to be unbiased and has the variance

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)}{np} \quad (3)$$

It is to be noted that situations may arise when the investigator is to use the same randomization device for more than one character, e.g. in multiple characteristic surveys. In such situations, it may happen that the value of  $\pi$  (to be estimated) is very small and  $p$  in the RR device is large (*i.e.* near to 'unity'). For example, let us visualize a situation where an investigator using a RR device with  $p = 0.85$ , desires to estimate the proportion of faculty members who are involved in illegal activity (e.g. drug usage, criminal activities, etc.) of an academic institution. In such a case  $\pi$  is expected to be quite small and hence a very small value of  $\theta$ . For such cases,  $n_1$  may assume zero value for not so large values of  $n$  and thus the estimate so obtained may depend entirely on  $p$  which is not desirable. The estimator  $\hat{\pi}$  will then depend entirely on  $p$ , which is also not desirable. The frequency of  $\hat{\pi}$  taking inadmissible values outside  $[0,1]$  is also increased in such cases. To overcome such problems Mangat and Singh (1991, 1995) advocated the use of an inverse binomial randomized response (IBRR) procedure.

The objective of the present paper is to suggest an alternative to the Mangat (1994) randomized response sampling procedure for estimating the proportion of human population having sensitive attribute using IBRR procedure and developed theoretical details. Two upper bounds of the variance of suggested estimator are also given.

## 2. INVERSE BINOMIAL RANDOMIZED RESPONSE TECHNIQUE

In this method, the sample size  $n$  is not fixed in advance. Instead, sampling is continued until a predetermined number  $m$  of respondents reporting 'yes' answer are drawn. Thus  $n$  is a random variable taking possible values  $m, m+1, m+2, \dots, \infty$  and follows a negative binomial distribution, namely

$$P(n) = {}^{n-1}C_{m-1} \theta^m (1-\theta)^{n-m} \quad (4)$$

It is well known that an unbiased estimator of  $\theta$  is

$$\hat{\theta}_u = \frac{(m-1)}{(n-1)} \quad (5)$$

Thus the unbiased estimator of  $\pi$  is given by

$$\hat{\pi}_u = \frac{(\hat{\theta}_u - 1 + p)}{p} \tag{6}$$

The variance of the estimator  $\hat{\pi}_u$  is defined by

$$V(\hat{\pi}_u) = E(\hat{\pi}_u^2) - \pi^2 = \frac{1}{p^2} [E(\hat{\theta}_u^2) - \theta^2] = \frac{V(\hat{\theta}_u)}{p^2} \tag{7}$$

Noting from Best (1974, equation 2) that

$$E(\hat{\theta}_u^2) = (m-1)(1-\theta) \left[ \sum_{r=2}^{m-1} \left(\frac{\theta}{1-\theta}\right)^r \frac{(-1)^r}{(m-r)} - (-1)^m \left(\frac{\theta}{1-\theta}\right)^m \log_e \theta \right] \tag{8}$$

Using (8) in (7) we get the variance of  $\hat{\pi}_u$  as

$$V(\hat{\pi}_u) = \frac{1}{p^2} \left[ (m-1)(1-\theta) \left\{ \sum_{r=2}^{m-1} \frac{(-1)^r}{(m-r)} \left(\frac{\theta}{1-\theta}\right)^r \right\} - (-1)^m \left(\frac{\theta}{1-\theta}\right)^m \log_e \theta \right] - \theta^2 \tag{9}$$

*Remark 2.1.* The predetermined number  $m$  must be greater than two (*i.e.*  $m > 2$ ) so that the variance  $V(\hat{\pi}_u)$  exists.

An unbiased estimator of the variance  $V(\hat{\theta}_u)$  has been given by Sukhatme *et al.* (1984) as

$$\hat{V}(\hat{\theta}_u) = \frac{\hat{\theta}_u(1-\hat{\theta}_u)}{(n-2)} \tag{10}$$

Replacing  $V(\hat{\theta}_u)$  by  $\hat{V}(\hat{\theta}_u)$  in (7), we get an unbiased estimator of  $V(\hat{\pi}_u)$  as

$$\hat{V}(\hat{\pi}_u) = \frac{\hat{\theta}_u(1-\hat{\theta}_u)}{p^2(n-2)} \tag{11}$$

We note from (9) that the variance expression is intractable as a function of  $\theta$  or of  $m$ . Therefore, it is desired to obtain simple upper bounds of the variance  $V(\hat{\pi}_u)$ . Sahai (1983) reported an upper bound of the variance  $V(\hat{\theta}_u)$  as

$$V_1(\hat{\theta}_u) = \frac{\theta}{6m} \left[ (A^2 - 12m\theta B)^{1/2} - A \right], \tag{12}$$

where

$$A = \left[ \left\{ m^2 + (3\theta - 1)m - 3\theta(1 - \theta) \right\} - \frac{6(1 - \theta)^2}{(m + 1)} \right],$$

$$B = \left\{ \frac{(m - 1)(1 - \theta)}{(m + 1)} - (m + 2) \right\} (1 - \theta).$$

Substitution of  $V_1(\hat{\theta}_u)$  in place of  $V(\hat{\theta}_u)$  in (7), we get the upper bound of  $V(\hat{\pi}_u)$  as

$$V_1(\hat{\pi}_u) = \frac{\theta}{6mp^2} \left[ (A^2 - 12m\theta B)^{1/2} - A \right], \quad (13)$$

where  $A$  and  $B$  are same as defined in (12).

Further, the upper bound of the variance  $\hat{\theta}_u$  due to Pathak and Sathe (1984) is given by

$$V_2(\hat{\theta}_u) = \frac{\theta^2(1 - \theta)}{m} \left[ 1 + \frac{2(1 - \theta)}{(m - 2)} - \frac{12\theta(1 - \theta)}{(m - 2) \left\{ (m + 3\theta - 2) + \left[ (m + 5\theta - 4)^2 + 16\theta(1 - \theta) \right]^{1/2} \right\}} \right] \quad (14)$$

Replacing  $V(\hat{\theta}_u)$  in (7) by  $V_2(\hat{\theta}_u)$ , we get another upper bound of  $V(\hat{\pi}_u)$ , as

$$V_2(\hat{\pi}_u) = \frac{\theta^2(1 - \theta)}{mp^2} \left[ 1 + \frac{2(1 - \theta)}{(m - 2)} - \frac{12\theta(1 - \theta)}{(m - 2) \left\{ (m + 3\theta - 2) + \left[ (m + 5\theta - 4)^2 + 16\theta(1 - \theta) \right]^{1/2} \right\}} \right] \quad (15)$$

The values of the variance  $V(\hat{\pi}_u)$  and its upper bounds  $V_j(\hat{\pi}_u)$ ;  $j=1,2$ ; have been computed for different values of  $\pi$ ,  $p$ ,  $m$ ,  $\theta$  and displayed in table 1.

From table 1, we observe that the variance bounds  $V_1(\hat{\pi}_u)$  and  $V_2(\hat{\pi}_u)$  are decrescent function of  $m$ . The value of  $m$  required for a given precision also depends on the choice of parameter  $p$ . If the value of  $p$  is close to '1' is sufficiently to ensure co-operation.

When  $\pi=0.01, p=0.75, \theta = 0.26$ , it is observed that for  $5 \leq m \leq 7$ , the upper bound  $V_2(\hat{\pi}_u)$  is nearer to the exact variance  $V(\hat{\pi}_u)$  followed by  $V_1(\hat{\pi}_u)$ . For  $8 \leq m \leq 15$ , the values of upper bound  $V_2(\hat{\pi}_u)$  coincides with the exact variance  $V(\hat{\pi}_u)$  followed by  $V_1(\hat{\pi}_u)$  while for  $m \geq 14$ , the values of upper bounds  $V_1(\hat{\pi}_u)$  and  $V_2(\hat{\pi}_u)$  are same and equal to the exact variance  $V(\hat{\pi}_u)$ .

When  $\pi=0.05, p=0.85, \theta = 0.19$ , it is shown that for  $m = 5, 6$ , the upper bound  $V_2(\hat{\pi}_u)$  is nearer to the exact variance  $V(\hat{\pi}_u)$  followed by  $V_1(\hat{\pi}_u)$ . For  $7 \leq m \leq 15$ , we see that  $V_2(\hat{\pi}_u) = V(\hat{\pi}_u)$  followed by  $V_1(\hat{\pi}_u)$ , while for  $m \geq 11$ ,  $V_1(\hat{\pi}_u) = V_2(\hat{\pi}_u) = V(\hat{\pi}_u)$ .

When  $\pi=0.1, p = 0.8, \theta = 0.28$ , we see that for  $5 \leq m \leq 7$ , the upper bound  $V_2(\hat{\pi}_u)$  gives the value closest to the exact variance  $V(\hat{\pi}_u)$  while for  $m \geq 8$ ,  $V_2(\hat{\pi}_u) = V(\hat{\pi}_u)$  followed by  $V_1(\hat{\pi}_u)$ .

Thus it is observed that these upper bounds are sufficiently accurate and may serve the objective, as  $m$  does not generally assume a very small value in practice. There is need to fix a higher value of  $m$  in order to have more accurate estimates. Between the two upper bounds  $V_1(\hat{\pi}_u)$  and  $V_2(\hat{\pi}_u)$ , the preference goes to  $V_2(\hat{\pi}_u)$  in practice, particularly when  $\theta$  is small.

TABLE 1

Values of the variance  $V(\hat{\pi}_u)$  and its bounds  $V_j(\hat{\pi}_u), j=1,2$ .

m	$\pi = 0.01, p = 0.75, \theta = 0.26$			$\pi = 0.05, p = 0.85, \theta = 0.19$			$\pi = 0.1, p = 0.8, \theta = 0.28$		
	$V(\hat{\pi}_u)$	$V_1(\hat{\pi}_u)$	$V_2(\hat{\pi}_u)$	$V(\hat{\pi}_u)$	$V_1(\hat{\pi}_u)$	$V_2(\hat{\pi}_u)$	$V(\hat{\pi}_u)$	$V_1(\hat{\pi}_u)$	$V_2(\hat{\pi}_u)$
5	0.024124	0.024196	0.024157	0.011883	0.011926	0.011897	0.023994	0.066424	0.024027
6	0.019012	0.019040	0.019018	0.009281	0.009297	0.009284	0.018961	0.053429	0.018968
7	0.015664	0.015676	0.015665	0.007603	0.007610	0.007603	0.015651	0.044573	0.015653
8	0.013309	0.013315	0.013309	0.006434	0.006437	0.006434	0.013316	0.038180	0.013316
9	0.011565	0.011568	0.011565	0.005574	0.005576	0.005574	0.011582	0.033363	0.011582
10	0.010223	0.010225	0.010223	0.004916	0.004917	0.004916	0.010245	0.029610	0.010245
11	0.009158	0.009160	0.009158	0.004397	0.004397	0.004397	0.009184	0.026606	0.009184
12	0.008294	0.008295	0.008294	0.003976	0.003976	0.003976	0.008321	0.024150	0.008321
13	0.007578	0.007579	0.007578	0.003629	0.003629	0.003629	0.007606	0.022105	0.007606
14	0.006976	0.006976	0.006976	0.003337	0.003337	0.003337	0.007004	0.020377	0.007004
15	0.006462	0.006462	0.006462	0.003089	0.003089	0.003089	0.006490	0.018898	0.006490

## ACKNOWLEDGEMENTS

The authors are grateful to the referee for constructive suggestions which helped in bringing the manuscript to its present form.

## REFERENCES

- D. J. BEST (1974), *The variance of the inverse binomial estimator*, "Biometrika", 61, pp. 385-386.
- A. CHAUDHURI, R. MUKERJEE (1988), *Randomized Response: Theory and Techniques*, Marcel Dekker, N. Y..
- A. S. HEDAYAT, B. K. SINHA (1991), *Design and Inference in finite population sampling*, New York : Wiley.
- N. S. MANGAT, R. SINGH (1990), *An alternative randomized response procedure*, "Biometrika", 77, pp. 439-442.
- N. S. MANGAT, R. SINGH (1991), *Alternative approach to randomized response survey*, "Statistica", LI, 3, pp. 327-332.
- N. S. MANGAT, R. SINGH (1995), *A note on the inverse binomial randomized response procedure*, "Journal of Indian Society of Agricultural Statistics", 47, pp. 21-25.
- N. S. MANGAT (1994), *An improved randomized response strategy*, "Journal of Royal Statistical Society", B 56, pp. 93-95.
- P.K. PATHAK, Y.S. SATHE, (1984), *A new variance formula for unbiased estimation in inverse sampling*, "Sankhya", 46,B,3, pp. 301-305.
- P. V. SUKHATME, B. V. SUKHATME, S. SUKHATME, C. ASOK, (1984), *Sampling theory of surveys with applications*, 3<sup>rd</sup> edition, Iowa State University Press, U. S.A. and Indian Society of Agricultural Statistics, New Delhi (India).
- A. SAHAI (1983), *On a systematic sharpening of variance bounds of MVU estimator in inverse binomial sampling*, "Statistica", XLIII, 4, pp.621-624.
- Y.S. SATHE (1977), *Sharper variance bounds for unbiased estimation in inverse sampling*, "Biometrika", 64, pp. 425-426.
- S. L. WARNER (1965), *Randomized response: a survey technique for estimating evasive answer bias*, "Journal of the American Statistical Association", 60, pp.63-69.

## RIASSUNTO

*Sugli sviluppi della strategia di risposta randomizzata proposta da Mangat*

Il lavoro considera una soluzione alternativa alla procedura di campionamento proposta in Mangat (1994) per stimare la proporzione di popolazione in possesso di un carattere sensibile. Lo stimatore suggerito è corretto ed assume meno di frequente rispetto allo stimatore di Mangat (1994) valori implausibili esterni all'intervallo  $[0,1]$ . Per lo stimatore proposto sono anche dati 2 limiti superiori per la formula della varianza esatta.

SUMMARY

*On Mangat's improved randomized response strategy*

In this paper, an alternative to the Mangat (1994) randomized response sampling procedure for estimating the proportion of human population possessing a sensitive characteristic is suggested. The suggested estimator is unbiased and takes inadmissible values outside the range  $[0,1]$  less frequently as compared to the Mangat (1994) estimator. Two upper bounds of the exact variance formula of the suggested estimator are given.