*Jour. Ind. Soc. Ag. Statistics*
*47(1), 1995 : 21–25*

# A Note on the Inverse Binomial Randomized Response Procedure

N.S. Mangat and Ravindra Singh
*Punjab Agricultural University, Ludhiana-141 004*
(Received : February, 1992)

SUMMARY

In certain situations, the Warner's [4] randomized response procedure may result in zero number of "yes" answers. The estimator of (the proportion of respondents in the population possessing the sensitive attribute), in such cases depends entirely on the probability of the statement : "I belong to the sensitive group" in the randomization device. Such an estimator is clearly not desirable. Mangat and Singh [2] proposed an alternative procedure where such a situation does not arise. The present paper extends the results of Mangat and Singh's procedure to the case where the sampled respondents may not report truthfully.

*Key Words* : Inverse sampling; Randomized response technique; Estimation of proportion; Sensitive characterstics.

## *Introduction*

The randomized response (RR) procedure for collecting trustworthy data on sensitive characters was first introduced by warner [4]. Assuming truthful reporting by the respondents, he considered the following estimator of $\pi$ (proportion of the population possessing the sensitive attribute) :

$$\hat{\pi} = [(n'/n) - 1 + p]/(2p - 1), \quad p \neq 0.5, \tag{1.1}$$

where n′ is the number of persons who report "yes" answer in an equal probability with replacement sample of size n and p is the probability in the randomization device to point to the sensitive attribute.

Mangat and Singh [2] have pointed out that when the investigator is to use same randomization device for more than one character e.g. in multiple characteristic surveys, it may happen that the value of $\pi$ to be estimated and p in the RR device are on the opposite extremes of 0.5. For example let us visualize a situation where an investigator, using a RR device with p =0.8, wishes to estimate the proportion of married faculty members of an Indian university who are unfaithful to their spouses. In such a case $\pi$ is expected to be quite small and p and $\pi$ will lie on opposite sides of 0.5 In such cases, the probability of "yes" answer turns out to be very small and n′ may assume

zero value for not so large values of n. The estimator $\hat{\pi}$ will then depend entirely on p, which is not desirable. The frequency of $\hat{\pi}$ taking inadmissible values outside [0, 1] is also increased in such cases. To avoid these difficulties, Mangat and Singh [2] have suggested the use of an inverse binomial RR (IBRR) procedure.

While suggesting IBRR procedure, Mangat and Singh have assumed that the reporting is truthful. But the situation when reporting is not truthful has not been considerd by them. The objective of the present paper is to consider this aspect of the problem and develop theoretical details for this particular situation. It will help the investigator in having an idea about the magnitude of bias and the effect on the efficiency of the estimator proposed by them as the probability of reporting truth changes.

## 2. *Proposed Estimator*

In the IBRR procedure of Mangat and Singh [2], the sample size n is not fixed in advance. Instead, sampling is continued until a predetermined number m of individuals reporting "yes answer are selected. In this case, the probability of "yes" answer $\theta'$ is same as given by Greenberg *et. al.* [1] where

$$\theta' = \pi pT + \pi (1-p)(1-T) + (1-\pi)(1-p). \tag{2.1}$$

The probability of "no" answer thus becomes

$$\alpha' = 1 - \theta', \tag{2.2}$$

where T denotes the probability that respondents in sensitive category report truth. As the RR procedure ensures the protection of privacy, the value of T is expected to be close to 1.

As T is unknown, we consider the estimator proposed by Mangat and Singh [2] for estimating in this case also. Let this estimator be now denoted by $\hat{\pi}_1$. Then

$$\hat{\pi}_1 = [\hat{\theta} - 1 + p]/[2p - 1], \quad p \neq 0.5, \tag{2.3}$$

where

$$\hat{\theta} = (m-1)/(n-1).$$

As m is a predetermined number of "yes" answers fixed by the investigator, $\hat{\theta}$ never attains zero value. Therefore, the estimator $\hat{\pi}_1$ does not depend on p alone and thus tends to take values in the admissible range [0, 1] more frequently.

As $\hat{\theta}$ follows inverse binomial distribution with parameters m and $\theta'$, we have the following theorem the proof of which is obvious.

*Theorem 2.1* : The estimator $\hat{\pi}_1$ is biased for population proportion $\pi$ and the expression for bias is given by

$$B(\hat{\pi}_1) = \pi (T - 1). \tag{2.4}$$

It is interesting to note that the expression obtained for $B(\hat{\pi}_1)$ is same as reported by Greenberg *et al.* [1] for the binomial RR model. As T is expected close to 1 the bias does not seem to be serious.

In order to study the estimator $\hat{\pi}_1$ in detail, we need its mean square error (MSE). This is obtained in the theorem below :

*Theorem 2.2* : The MSE of the estimator $\hat{\pi}_1$ is given by

$$MSE(\hat{\pi}_1) = \frac{\alpha'(m-1)\{\sum_{(r=2)}^{m-1} \frac{(-\theta'/\alpha')^r}{(m-r)} - (-\theta'/\alpha')^m \log_e \theta'\} - \theta'^2}{(2p-1)^2} + [\pi(T-1)]^2. \tag{2.5}$$

The theorem can be easily proved by using (2.1), (2.2), (2.4) and the value os $E(\hat{\theta})^2$ obtained by replacing $\theta$ and $\alpha$ by $\theta'$ and $\alpha'$ in (2.3) of Mangat and Singh [2].

The expression for $MSE(\hat{\pi}_1)$ seems to be difficult to calculate numerically for large m. We, therefore, obtain its upper bound by following Sathe [3] and Mangat and Singh [2]. The upper bound $MSE_1(\hat{\pi}_1)$ so obtained is given below:

$$MSE_1(\hat{\pi}_1) = \frac{2\theta'^2\alpha'}{\{(m - 2\alpha' + \sqrt{(m-2\alpha')^2 + 4\theta'\alpha'})\}(2p-1)^2} + \pi^2(T-1)^2.$$

Neglecting $4\theta'\alpha'$ in comparison to $(m - 2\alpha')^2$, we get a much simpler upper bound $MSE_2(\hat{\pi}_1)$ of $MSE(\hat{\pi}_1)$. It is given by

$$MSE_2(\hat{\pi}_1) = \theta'^2\alpha'/[(2p-1)^2(m-2\alpha')] + \pi^2(T-1)^2. \tag{2.6}$$

We now examine the behaviour of efficiency for the estimator $\hat{\pi}_1$ with respect to T. The relative efficiency of the estimator $\hat{\pi}_1$ when the reporting is truthful with respect to the situation when the respondents do not tell the truth is defined as

$$RE = V_2(\hat{\pi}_1)/MSE_2(\hat{\pi}_1),$$

where $V_2(\hat{\pi}_1)$ is the upper bound of variance of estimator $\hat{\pi}_1$ for truthful reporting case and is given by Mangat and Singh [2] as

$$V_2(\hat{\pi}_1) = \theta^2 \alpha/[(2p-1)^2 (m-2\alpha)]. \tag{2.7}$$

The values of $\theta$ and $\alpha$ can be obtained by putting T = 1 in (2.1) and (2.2), respectively. On using (2.6) and (2.7), the expression for RE becomes

$$RE = \frac{\theta^2 \alpha/[(2p-1)^2 (m-2\alpha)]}{[\theta'^2 \alpha'/\{(2p-1)^2 (m-2\alpha')\}] + [\pi (T-1)]^2}.$$

The behaviour of RE with respect to T has been examined numerically. The values of RE have been worked out for m = 30, 60 and 90 by using different values of T and $\pi$. The optimal value of p have been taken using Theorem 2.4 of Mangat and Singh [2]. The values of RE thus obtained are given in Table 1.

**Table 1.** RE (in percent) of less than completely truthful situation with respect to the situation of truthful reporting for estimator $\hat{\pi}_1$ for m = 30, 60 and 90.

| m | T | $\pi$ = .1 p = .8 | $\pi$ = .3 p = .8 | $\pi$ = .5 p = .8 | $\pi$ = .7 p = .2 | $\pi$ = .9 p = .2 |
|---|---|---|---|---|---|---|
| 30 | 1.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | .9 | 99.9 | 96.2 | 87.2 | 58.4 | 33.2 |
| | .8 | 9.3 | 77.8 | 58.6 | 28.2 | 12.0 |
| | .7 | 93.2 | 57.5 | 37.4 | 15.4 | 5.9 |
| | .6 | 84.7 | 41.6 | 24.6 | 9.5 | 3.5 |
| | .5 | 75.2 | 30.6 | 17.1 | 6.4 | 2.3 |
| 60 | 1.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | .9 | 99.6 | 87.2 | 73.4 | 43.3 | 21.0 |
| | .8 | 91.5 | 58.2 | 39.0 | 16.9 | 6.5 |
| | .7 | 78.7 | 36.8 | 21.7 | 8.5 | 3.0 |
| | .6 | 65.5 | 24.2 | 13.3 | 5.0 | 1.8 |
| | .5 | 53.4 | 16.7 | 8.9 | 3.3 | 1.1 |
| 90 | 1.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | .9 | 97.5 | 79.7 | 63.4 | 34.5 | 15.4 |
| | .8 | 84.8 | 46.5 | 29.1 | 12.1 | 4.5 |
| | .7 | 68.4 | 27.1 | 15.2 | 5.8 | 2.1 |
| | .6 | 53.5 | 17.0 | 9.1 | 3.4 | 1.2 |
| | .5 | 41.5 | 11.5 | 6.0 | 2.2 | 0.8 |

The results obtained show that the RE is equal to 1 or < 1 according as T = 1 or < 1 and the RE is an increasing function of T . It is also observed that for a given combination of p and $\pi$, the relative efficiency decreases with increasing value of m. The rate of decrease in relative efficiency is also seen generally to be a decreasing function of m.

The empirical investigation shows that deviation of T from 1 affects the efficiency of the estimator under consideration seriously like other RR estimators, particularly when $\pi$ is not small. It is, therefore, stressed that the investigator should make all out efforts to enhance the co-operation of the respondents. It can be done by explaining the method to the respondents thoroughly and convincing them that their privacy is not affected at all by the procedure being used.

## REFERENCES

[1] Greenberg, B.G., Abul-Ela, A.L.A, Simmons, W.R. and Horvitz, D.G., 1969. Thee unrelated question randomized response model : theoretical frame work. *J. Amer. Statist. Assoc.*, **64**, 520-39.

[2] Mangat, N.S. and Singh, R., 1991. an alternative approach to randomized response survey. *Statistica*, LI (3), 327-332.

[3] Sathe, Y.S., 1977. Sharper variance bounds for unbiased estimation in inverse sampling. *Biometrika*, **64**, 425-426.

[4] Warner, S.L., 1965. Randomized response : A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60**, 63-69.