Jour. Ind. Soc. Ag. Statistics
44 (2), 1992 : 149–154

An Alternative Estimator for Randomized Response Technique

Sarjinder Singh and Ravindra Singh
Punjab Agricultural University, Ludhiana-141004
(Received : April, 1990)

Summary

A slightly biased alternative estimator, $\hat{\pi}_1$ of population proportion π is proposed. $\hat{\pi}_1$ is shown to have a smaller mean squared error than the Warner's [9] estimator, $\hat{\pi}$. Since $\hat{\pi}_1$ involves the use of an unknown population parameter λ, it has, therefore, little practical utility. Using an estimated value of λ in $\hat{\pi}_1$, another estimator, $\hat{\pi}_2$ has been proposed for use in practice. The estimator $\hat{\pi}_2$ is also shown to have the same asymptotic mean squared error as $\hat{\pi}_1$. A numerical investigation has also been undertaken to study the behaviour of relative efficiency of the estimator $\hat{\pi}_1$ with respect to the estimator $\hat{\pi}$.

Key words : Randomized response technique, simple random sampling, Warner's method, relative efficiency.

Introduction

Consider a population having some sensitive characteristic, say A. Warner [9] introduced a randomized response technique for eliciting sensitive information and thus estimating the proportion PI of the population having the characteristic A. The technique consists of using a device with outcomes A and not A with known probabilities p and $\bar{p}$ = (1–p) respectively. The respondent observes the device's outcome which remains unknown to the experimenter in order to protect the respondent's privacy. The respondent answers 'yes' if he has the characteristic shown by the device's outcome and 'no' otherwise. Hence the probability θ of a yes response is

$$\theta = p\pi + \bar{p}\,(1-\pi)$$

If n_1 is the total number of yes answers out of n responses then Warner proposed.

$$\hat{\pi} = \frac{\hat{\theta} - \bar{p}}{2p - 1}, \quad p \neq \frac{1}{2}$$

as an estimator of π, where $\hat{\theta} = \frac{n_1}{n}$. The variance of $\hat{\pi}$ is given by

$$V(\hat{\pi}) = \frac{\theta(1-\theta)}{n(2p-1)^2} \tag{1.1}$$

Singh [5] has pointed out that $\hat{\theta}$ and $\hat{\pi}$ are not the Maximum Likelihood Estimators (MLE) of θ and π respectively.

Fligner *et al* [2] have compared two randomized response methods taking into account the protection afforded to the respondent. In addition, they pointed out that the estimators, which previous authors have claimed to be the maximum likelihood estimators of the population proportion with the sensitive characteristic, are in fact not the maximum likelihood estimators. Singh [6] has shown that Takahasi and Sakesegawa [8] model can be used to estimate π, but the model needs modification in order to obtain the M.L.E. Singh [7] has also examined the admissibility aspect of certain estimators of π. Certain other recent developments on randomized response techniques are due to Mangat and Singh [4], Kuk [3] and Arnab [1].

2. *An Alternative Estimator*

In this section we propose an alternative estimator for π. It is

$$\hat{\pi}_1 = \frac{\lambda\hat{\theta} - \bar{p}}{2p - 1} \tag{2.1}$$

where constant λ is chosen to minimize the mean square error of $\hat{\pi}_1$. This mean square error (MSE) is given by

$$MSE(\hat{\pi}_1) = \frac{\lambda^2 \frac{\theta(1-\theta)}{n+\theta^2} + \theta^2 - 2\lambda\theta^2}{(2p-1)^2} \tag{2.2}$$

The MSE in (2.2) is minimized for

$$\lambda = \frac{n\theta}{1 + (n-1)\theta} \tag{2.3}$$

Hence the resultant minimum MSE of the estimator $\hat{\pi}_1$ is obtained as

$$\text{Min. MSE } (\hat{\pi}_1) = \frac{n\theta}{1 + (n-1)\theta} V(\hat{\pi}) \tag{2.4}$$

From (2.4) it is clear that MSE is always less than the variance of Warner's [9] estimator. The relative efficiency (RE) of $\hat{\pi}_1$ with respect to $\hat{\pi}$ is given by

$$RE = \frac{1 + (n-1)\theta}{n\theta} \tag{2.5}$$

3. *Alternative Estimator with Estimated value of* λ.

Since $\lambda = \dfrac{n\theta}{1+(n-1)\theta}$ is not known in practice, it is, therefore, advisable to replace λ with its estimated value

$$\hat{\lambda} = \frac{n\hat{\theta}}{1+(n-1)\hat{\theta}} \tag{3.1}$$

in the estimator $\hat{\pi}_1$. This yields us another estimator which can be used in practical situations. Thus we have

$$\hat{\pi}_2 = \frac{\hat{\lambda}\hat{\theta} - \bar{p}}{2p-1} \tag{3.2}$$

where $\hat{\lambda}$ is as given in (3.1). Now we prove the following theorem.

Theorem 3.1 To the order $O(n^{-2})$ the estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ are equally efficient.

Proof. Let $\varepsilon = \dfrac{\hat{\theta}}{\theta} - 1$, so that

$$E(\varepsilon) = 0 \quad \text{and} \quad E(\varepsilon^2) = \frac{1-\theta}{n\theta}$$

Then (3.2) may be rewritten as

$$\hat{\pi}_2 = \frac{\frac{\theta}{1+n^{-1}(\theta^{-1}-1)}\left[1+\varepsilon-\frac{n^{-1}}{1+n^{-1}(\theta^{-1}-1)}\left\{\frac{1}{\theta}(-\varepsilon+\varepsilon^2+\ldots)\right\}-\bar{p}\right]}{2p-1} \tag{3.3}$$

and the mean square of $\hat{\pi}_2$ is given by

$$MSE(\hat{\pi}_2) = E(\hat{\pi}_2-\pi)^2$$

$$= \frac{1}{(2p-1)^2} E\left[\frac{n^{-2}\theta^2(\theta^{-1}-1)}{\{1+n^{-1}(\theta^{-1}-1)\}^2} + \frac{\theta^2\theta^2}{\{1+n^{-1}(\theta^{-1}-1)\}^2}\right] + O(n^{-2}) \tag{3.4}$$

If the terms of order $O(n^{-2})$ are neglected, then $MSE(\hat{\pi}_2)$ can be approximated by the first expression which simplifies to $MSE(\hat{\pi}_1)$

Asymptotically the neglected terms tend to zero and the two estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ become equally efficient.

A numerical investigation is undertaken to study the behaviour of the relative efficiency of $\hat{\pi}_1$ in relation to the Warner's estimator $\hat{\pi}$

4. *Numerical Illustration*

Let us study the relative efficiency of the proposed estimator $\hat{\pi}_1$ with respect to the estimator $\hat{\pi}$ for different values of π (close to zero) and p (close to one). Relative efficiency figures are shown in the Table.

It is found that the largest gains are obtained for small sample sizes. Frequently in practice, however, when observations are expensive, such sample sizes may be all that are available. Moreover, as $p \to 1$ and $\pi \to 0$ simultaneously, then there is rapid increase in the relative efficiency for the samples of small sizes. For example when $p = 0.99$ and $\pi = 0.03$, the percent relative efficiency figures corresponding to n = 5, 10 and 50 are 587.61, 343.81 and 148.76 respectively.

ACKNOWLEDGEMENT

The authors are grateful to the referee for his valuable suggestions which helped in bringing the paper to its present form.

Table. Percent relative efficiency of $\hat{\pi}_1$ over $\hat{\pi}$ for differnet values of n, p and π

p	Sample size (n)				
	5	10	50	100	500
$\pi = 0.6$					
0.9	114.48	107.24	101.45	100.72	100.14
0.8	115.71	107.85	101.57	100.78	100.16
0.7	117.03	108.52	101.70	100.85	100.17
0.6	118.46	109.23	101.85	100.92	100.18
$\pi = 0.3$					
0.9	138.83	119.41	103.88	101.94	100.39
0.8	132.63	116.32	103.26	101.63	100.33
0.7	127.26	113.81	102.76	101.38	100.28
0.6	123.48	111.74	102.35	101.17	100.23
$\pi = 0.1$					
0.9	119.11	145.56	109.11	104.56	100.91
0.8	156.92	128.41	105.69	102.85	100.57
0.7	138.82	119.41	103.88	101.94	100.39
0.6	127.62	113.81	102.76	101.38	100.28
$\pi = 0.0138$					
0.9	260.115	180.057	116.011	108.005	101.601
0.8	176.020	138.012	107.603	103.801	100.760
0.7	145.450	122.725	104.545	102.273	100.455
0.6	129.657	114.828	102.965	101.482	100.296

REFERENCES

[1] Arnab, R. 1990. On comutativity of design and model expectations in randomized response surveys. *Commun. Statist. Theor. Meth.*, **19**(10), 3751-3757.

[2] Fligner, M.A., Policello II, G.B. and Singh, J., 1977. A comparison of two randomized response survey methods with consmeration for the level of respondent protection. *Commun. Statist.-Theor. Meth.*, **A6**(15), 1511-1524.

[3] Kuk, Anthony, Y.C., 1990. Asking sensitive questions indirectly. *Biometrika*, **77**, 2. 436-438.

[4] Mangat, N.S. and Singh, R., 1990. An alternative randomized. response technique. *Biometrika*, **77**, 2. 439-442.

[5] Singh, J. 1976. A note on the randomized response technique proceedings of the social statistical section, *American Statistical Assoc.*, 772.

[6] Singh, J., 1978. A note on maximum likelihood estimation from randomized response models. Proceedings of the Social Statistical Section, *American Statistical Assoc.* 282-283.

[7] Singh, J., 1981. A note on estimation from randomized response surveys. *Jour. Ind. Soc. Ag. Stat.* **33**(2). 89-93.

[8] Takahasi, K. and Sakesegawa, H., 1977. A randomized response technique without making use of any randomizing device. *Ann. Inst. Statist. Math.* **29**, Part AW, 1-8.

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