*Jour. Ind. Soc. Ag. Statistics*
*55(2), 2002 : 189-196*

# On Alternative Estimators in Randomized Response Technique

Housila P. Singh and Nidhi Mathur
*Vikram University, Ujjain-456010, M.P.*
(Received : May, 2001)

SUMMARY

This paper proposes an alternative estimator $\hat{\pi}_h$ of population proportion $\pi$ for the unrelated question randomized response technique and analyses its properties. Numerical illustrations are also given.

*Key words :* Sensitive characteristic, Population proportion, Unrelated question randomized response technique, Mean squared error.

## *1. Introduction*

The randomized response (RR) procedure to procure trustworthy data on a sensitive character by protecting privacy of the respondent was first introduced by Warner [5]. W.R. Simmons perceived that the belief of the respondent furnished by RR procedure might be further enhanced if one of the two questions is referred to a non-stigmatized attribute which is unrelated to the sensitive attribute. This led, Horvitz *et al.* [2] to develop an unrelated question RR model (U-model). For this model (see Greenberg *et al.* [1]), the expression of probability of a 'yes' reply is described below.

When $\pi_y$ is known, each respondent in a sample of n individuals, selected by simple random sampling with replacement (SRSWR) procedure, is provided with a random device, say $R_1$ consisting

(i) "I am a member of sensitive group A" and

(ii) "I am a member of non-sensitive group Y"

represented with probabilities p and (1 – p) respectively. The respondent selects randomly one of these two statements unobserved by the interviewer and reports "yes" if he possesses the characteristic indicated by the chosen statement and "no" otherwise. If $\pi$ is the proportion of sensitive group in the population, the probability of "yes" reply will be

$$\lambda = p\pi + (1-p)\pi_y \qquad (1.1)$$

The usual estimator of $\pi$ suggested by Greenberg *et al.* [1] is given by

$$\hat{\pi} = \frac{\left(\hat{\lambda} - (1-p)\pi_y\right)}{p} \tag{1.2}$$

where $\hat{\lambda} = \frac{n_1}{n}$, $n_1$ is the number of 'yes' reply by the individuals in the sample of size n. The estimator $\hat{\pi}$ is unbiased and its variance is given by

$$V(\hat{\pi}) = \frac{\lambda(1-\lambda)}{np^2} \tag{1.3}$$

The minimum value of variance $V(\hat{\pi})$ can be obtained by selecting p near to 1 and $\pi_y$ near to 0 or 1 according as $\pi < 0.5$ or $\pi > 0.5$. If $\pi = 0.5$, the $\left|\pi_y - 0.5\right|$ could be maximum on either side.

Mangat *et al.* [3] suggested a generalized estimator for $\pi$ as

$$\hat{\pi}_s = \frac{\left[\alpha\hat{\lambda} - \pi_y(1-p)\right]}{p} \tag{1.4}$$

where $\alpha$ is suitably chosen constant such that $MSE(\hat{\pi}_s)$ is minimum. The minimum MSE of $\hat{\pi}_s$ is given by

$$\min . MSE(\hat{\pi}_s) = \frac{\lambda^2(1-\lambda)}{np^2\left\{\lambda + \frac{(1-\lambda)}{n}\right\}} \tag{1.5}$$

for the optimum value of $\alpha$

$$\alpha = \frac{n\lambda}{\{1 + (n-1)\lambda\}} \tag{1.6}$$

Following Searls [4], one may define a class of estimators for $\pi$ as

$$\hat{\pi}_h = \frac{\delta\left(\hat{\lambda} - \bar{p}\pi_y\right)}{p} = \delta\hat{\pi} \tag{1.7}$$

where $\delta$ is a suitably chosen scalar such that MSE of $\hat{\pi}_h$ is minimum and $\bar{p} = (1-p)$.

The minimum MSE of $\hat{\pi}_h$ for optimum value of $\delta$

$$\delta = \frac{(\lambda - \bar{p}\pi_y)^2}{(\lambda - \bar{p}\pi_y)^2 + \frac{\lambda(1-\lambda)}{n}} \tag{1.8}$$

is given by

$$\min.\mathrm{MSE}(\hat{\pi}_h)=\frac{(\lambda-\bar{p}\pi_y)^2\lambda(1-\lambda)}{p^2\left[n(\lambda-\bar{p}\pi_y)^2+\lambda(1-\lambda)\right]} \tag{1.9}$$

Thus from (1.3), (1.5) and (1.9) it can be easily proved that

$$\min.\mathrm{MSE}(\hat{\pi}_h)\le \min.\mathrm{MSE}(\hat{\pi}_s)\le V(\hat{\pi}) \tag{1.10}$$

which implies that the proposed estimator $\hat{\pi}_h$ is more efficient than Greenberg *et al.* [1] estimator $\hat{\pi}$ and Mangat *et al.* [3] estimator $\hat{\pi}_s$ at its optimum condition.

It is to be mentioned that estimators $\hat{\pi}_s$ and $\hat{\pi}_h$ cannot be used in practice as they are based on unknown parameters. This led authors to propose estimators based on estimated values of parameters and discuss their properties.

## 2. *Estimators Based on Estimated Optimum*

The estimated optimum value of $\delta$ is given by

$$\hat{\delta}_1=\frac{(\hat{\lambda}-\bar{p}\pi_y)^2}{\left((\hat{\lambda}-\bar{p}\pi_y)^2+\frac{\hat{\lambda}(1-\hat{\lambda})}{n}\right)} \tag{2.1}$$

Substitution of (2.1) in (1.7) yields an estimator of $\pi$ as

$$\hat{\pi}_h^{(1)}=\frac{(\hat{\lambda}-\bar{p}\pi_y)^3}{p\left\{(\hat{\lambda}-\bar{p}\pi_y)^2+\frac{\hat{\lambda}(1-\hat{\lambda})}{n}\right\}} \tag{2.2}$$

Another consistent estimator of $\delta$ is given by

$$\hat{\delta}_2=\frac{(\hat{\lambda}-\bar{p}\pi_y)^2}{p\left\{(\hat{\lambda}-\bar{p}\pi_y)^2+\frac{\hat{\lambda}(1-\hat{\lambda})}{1-n}\right\}} \tag{2.3}$$

Putting $\hat{\delta}_2$ in place of $\delta$ in (1.7), we get another estimator of $\pi$ as

$$\hat{\pi}_h^{(2)}=\frac{(\hat{\lambda}-\bar{p}\pi_y)^3}{p\left\{(\hat{\lambda}\ \bar{p}\pi_y)^2+\frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)}\right\}} \tag{2.4}$$

A more flexible form of the estimator $\pi$ is given by

$$\hat{\pi}_h^{(k)} = \frac{\left(\hat{\lambda} - \bar{p}\pi_y\right)^3}{p\left[\left(\hat{\lambda} - \bar{p}\pi_y\right)^2 + k\frac{\hat{\lambda}\left(1-\hat{\lambda}\right)}{n}\right]}; \left(k \geq 0\right) \tag{2.5}$$

where $k\ (\geq 0)$ is a constant. For $k = 1$, $\hat{\pi}_h^{(k)}$ reduces to $\hat{\pi}_h^{(1)}$ while it reduces to $\hat{\pi}_h^{(2)}$ for $k = \frac{n}{n-1}$.

The MSE of an estimator T can be computed from

$$\text{MSE}(T) = E(T-\pi)^2 = \sum_{n_1=0}^{n} (T-\pi)^2\ {}^nC_{n_1}\ \lambda^{n_1}\ (1-\lambda)^{n-n_1} \tag{2.6}$$

The percent relative efficiency of T with respect to $\hat{\pi}$ is given by

$$\text{PRE}(T, \hat{\pi}) = \frac{\lambda(1-\lambda)}{np^2}\left[\sum_{n_1=0}^{n} (T-\pi)^2\ {}^nC_{n_1}\ \lambda^{n_1}(1-\lambda)^{n-n_1}\right]^{-1} \times 100 \tag{2.7}$$

The percent relative efficiencies of different estimators $\left(\hat{\pi}_h^{(1)}, \hat{\pi}_h^{(2)} \text{ and } \hat{\pi}_h^{(k)}\right)$ with respect to $\hat{\pi}$ have been computed for different values of n, p, $\pi$, k and $\pi_y$ as shown in Tables 2.1 (a) and Table 2.1(b).

Further, to obtain an approximate expression of MSE of $\hat{\pi}_h^{(k)}$, we write $\varepsilon = \left(\hat{\lambda} - \lambda\right)/\lambda$ such that $E(\varepsilon) = 0$ and $E\left(\varepsilon^2\right) = \frac{(1-\lambda)}{(n\lambda)}$

Expressing $\hat{\pi}_h^{(k)}$ in terms of $\varepsilon$, we have

$$\hat{\pi}_h^{(k)} = \pi\left(1 + \frac{\lambda}{p\lambda}\varepsilon\right)\left[1 + \frac{k\lambda(1-\lambda)}{np^2\pi^2}(1+\varepsilon)\left(1 - \frac{\lambda}{1-\lambda}\varepsilon\right)\left(1 + \frac{\lambda}{p\pi}\varepsilon\right)^{-2}\right]^{-1}$$

or

$$\left(\hat{\pi}_h^{(k)} - \pi\right) \cong \frac{\lambda}{p}\varepsilon - \frac{k\lambda(1-\lambda)}{np^2\pi}\left[1 + \frac{\lambda}{p\pi}\left\{\frac{p\pi(1-2\lambda)}{\lambda(1-\lambda)} - 1\right\}\varepsilon\right]$$

Squaring both sides of the above expression and then taking expectation, we get the MSE of $\hat{\pi}_h^{(k)}$ to terms of order $n^{-2}$ as

**Table 2.1(a).** Percent relative efficiency $\hat{\pi}_h^{(1)}, \hat{\pi}_h^{(2)}$ and $\hat{\pi}_h^{(k)}$ with respect to $\hat{\pi}$

| p ↓ | $\pi_y$ ↓ | n → / Estimator ↓ | $\pi = 0.05$, n = 5 | n = 10 | n = 20 |
|---|---|---|---|---|---|
| 0.6 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 195.62 | 162.57 | 146.00 |
| | | $\hat{\pi}_h^{(2)}$ | 211.63 | 166.99 | 147.69 |
| | | $\hat{\pi}_h^{(k=6)}$ | 330.26 | 232.54 | 193.63 |
| | | $\hat{\pi}_h^{(k=16)}$ | 358.87 | 239.32 | 191.81 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 165.12 | 151.90 | 167.65 |
| | | $\hat{\pi}_h^{(2)}$ | 174.78 | 155.84 | 170.12 |
| | | $\hat{\pi}_h^{(k=6)}$ | 249.18 | 217.34 | 254.38 |
| | | $\hat{\pi}_h^{(k=16)}$ | 271.57 | 228.13 | 263.24 |
| 0.7 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 200.51 | 165.66 | 138.33 |
| | | $\hat{\pi}_h^{(2)}$ | 218.20 | 170.20 | 139.77 |
| | | $\hat{\pi}_h^{(k=6)}$ | 347.11 | 232.58 | 174.06 |
| | | $\hat{\pi}_h^{(k=16)}$ | 373.13 | 234.69 | 167.80 |
| | 0.2 | $\hat{n}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 174.36 | 149.94 | 154.53 |
| | | $\hat{\pi}_h^{(2)}$ | 185.61 | 153.76 | 156.28 |
| | | $\hat{\pi}_h^{(k=6)}$ | 269.07 | 211.29 | 207.69 |
| | | $\hat{\pi}_h^{(k=16)}$ | 290.95 | 218.24 | 206.14 |
| 0.8 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 205.60 | 169.78 | 134.76 |
| | | $\hat{\pi}_h^{(2)}$ | 225.24 | 174.59 | 135.95 |
| | | $\hat{\pi}_h^{(k=6)}$ | 366.84 | 233.79 | 158.05 |
| | | $\hat{\pi}_h^{(k=16)}$ | 388.91 | 230.05 | 147.42 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 185.90 | 153.65 | 137.77 |
| | | $\hat{\pi}_h^{(2)}$ | 199.79 | 157.45 | 139.15 |
| | | $\hat{\pi}_h^{(k=6)}$ | 298.35 | 210.69 | 171.45 |
| | | $\hat{\pi}_h^{(k=16)}$ | 318.74 | 212.69 | 164.50 |

| p ↓ | $\pi_y$ ↓ | n → / Estimator ↓ | $\pi = 0.05$ 5 | 10 | 20 |
|---|---|---|---|---|---|
| 0.9 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 210.86 | 174.82 | 135.14 |
| | | $\hat{\pi}_h^{(2)}$ | 232.77 | 180.11 | 136.15 |
| | | $\hat{\pi}_h^{(k=6)}$ | 390.33 | 236.70 | 145.95 |
| | | $\hat{\pi}_h^{(k=16)}$ | 406.60 | 225.70 | 130.58 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 199.84 | 163.89 | 130.17 |
| | | $\hat{\pi}_h^{(2)}$ | 217.91 | 168.21 | 131.19 |
| | | $\hat{\pi}_h^{(k=6)}$ | 343.52 | 218.75 | 147.34 |
| | | $\hat{\pi}_h^{(k=16)}$ | 360.29 | 212.94 | 135.35 |

**Table 2.1(b).** Precent relative efficiency $\hat{\pi}_h^{(1)}, \hat{\pi}_h^{(2)}$ and $\hat{\pi}_h^{(k)}$ with respect to $\hat{\pi}$

| p ↓ | $\pi_y$ ↓ | n → / Estimator ↓ | $\pi = 0.1$ 5 | 10 | 20 |
|---|---|---|---|---|---|
| 0.6 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 169.61 | 139.25 | 123.64 |
| | | $\hat{\pi}_h^{(2)}$ | 180.15 | 142.05 | 124.48 |
| | | $\hat{\pi}_h^{(k=6)}$ | 249.41 | 175.82 | 135.27 |
| | | $\hat{\pi}_h^{(k=16)}$ | 258.79 | 171.06 | 122.37 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 149.08 | 138.78 | 140.21 |
| | | $\hat{\pi}_h^{(2)}$ | 156.42 | 141.79 | 141.60 |
| | | $\hat{\pi}_h^{(k=6)}$ | 210.18 | 182.94 | 173.42 |
| | | $\hat{\pi}_h^{(k=16)}$ | 222.54 | 183.28 | 162.43 |
| 0.7 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 168.96 | 136.61 | 114.46 |
| | | $\hat{\pi}_h^{(2)}$ | 179.46 | 139.02 | 115.01 |
| | | $\hat{\pi}_h^{(k=6)}$ | 243.50 | 162.73 | 115.05 |
| | | $\hat{\pi}_h^{(k=16)}$ | 247.33 | 153.66 | 100.29 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 152.31 | 131.39 | 126.64 |
| | | $\hat{\pi}_h^{(2)}$ | 159.86 | 133.94 | 127.41 |
| | | $\hat{\pi}_h^{(k=6)}$ | 211.28 | 164.54 | 136.38 |
| | | $\hat{\pi}_h^{(k=16)}$ | 219.33 | 160.14 | 122.09 |

| | | | $\pi = 0.1$ | | |
|---|---|---|---|---|---|
| $p \downarrow$ | $\pi_y \downarrow$ | $n \rightarrow$ Estimator $\downarrow$ | 5 | 10 | 20 |
| 0.8 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 168.27 | 134.69 | 109.01 |
| | | $\hat{\pi}_h^{(2)}$ | 178.75 | 136.77 | 109.29 |
| | | $\hat{\pi}_h^{(k=6)}$ | 273.26 | 150.96 | 100.03 |
| | | $\hat{\pi}_h^{(k=16)}$ | 235.43 | 138.04 | 83.87 |
| | 0.2 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 156.46 | 128.22 | 112.83 |
| | | $\hat{\pi}_h^{(2)}$ | 164.55 | 130.24 | 113.26 |
| | | $\hat{\pi}_h^{(k=6)}$ | 213.67 | 149.72 | 109.88 |
| | | $\hat{\pi}_h^{(k=16)}$ | 216.25 | 140.78 | 94.47 |
| 0.9 | 0.1 | $\hat{\pi}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 167.57 | 133.35 | 106.43 |
| | | $\hat{\pi}_h^{(2)}$ | 178.03 | 135.18 | 106.50 |
| | | $\hat{\pi}_h^{(k=6)}$ | 230.90 | 140.65 | 88.98 |
| | | $\hat{\pi}_h^{(k=16)}$ | 223.46 | 124.33 | 71.53 |
| | 0.2 | $\hat{n}_h^{(1)} = \hat{\pi}_h^{(k=1)}$ | 161.36 | 128.94 | 105.24 |
| | | $\hat{\pi}_h^{(2)}$ | 170.40 | 130.62 | 105.38 |
| | | $\hat{\pi}_h^{(k=6)}$ | 217.89 | 138.80 | 91.93 |
| | | $\hat{\pi}_h^{(k=16)}$ | 213.60 | 125.00 | 75.37 |

$$MSE\left(\hat{\pi}_h^{(k)}\right) = V(\hat{\pi}) + k\frac{[V(\hat{\pi})]^2}{\pi^2}\left[k - 2\left\{\frac{p\pi(1-2\lambda)}{\lambda(1-\lambda)} - 1\right\}\right] \tag{2.8}$$

which is less than that of $\hat{\pi}$ if

either $$0 < k < 2\left\{\frac{p\pi(1-2\lambda)}{\lambda(1-\lambda)} - 1\right\}$$

or $$2\left\{\frac{p\pi(1-2\lambda)}{\lambda(1-\lambda)} - 1\right\} < k < 0$$

Tables 2.1 (a) and 2.1 (b) show that

(i) The gain in efficiency due to the proposed estimators over $\hat{\pi}$ are large for small values of $\pi$.

(ii) The larger gain in efficiency due to the estimator $\hat{\pi}_h^{(k)}$ can be obtained through increasing the value of k suitably.

(iii) The estimator $N_i$ is more efficient than $\hat{\pi}_h^{(1)}$.

(iv) The efficiencies of the estimators decrease as $\pi_y$ increases.

(v) For fixed $\pi$, p, $\pi_y$, the gain in efficiencies decreases as n increases.

(vi) The maximum gain in efficiency (406.60%) is seen at $\pi = 0.05$, $p = 0.9$, $\pi_y = 0.1$, $n = 5$ and $k = 16$.

Lastly, we conclude that the suggested estimators $\hat{\pi}_h^{(1)}, \hat{\pi}_h^{(2)}$ and $\hat{\pi}_h^{(k)}$ are preferable over $\hat{\pi}$ for small values of n, $\pi$ and $\pi_y$. In practice, small sample sizes are desirable when the survey procedure like RRT, is expensive.

ACKNOWLEDGEMENT

The authors are thankful to the referee for his valuable suggestions regarding the improvement of the paper.

REFERENCES

[1] Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response model. (Theoretical frame-work). *Jour. Amer. Statist. Assoc.*, **64**, 520-539.

[2] Horvitz, D.G., Shal, B.V. and Simmons, W.R. (1967). The unrelated question randomized response model. *Proc. Social Statistics Sections, Jour. Amer. Statist. Assoc.*, 65-72.

[3] Mangat, N.S., Singh, R., Singh, S. (1991). Alternative estimators in randomized response technique. *Aligarh Journal of Statistics*, **11**, 75-80.

[4] Searls, D.T. (1964). The utilisation of a known coefficient of variation in the estimation procedure. *Jour. Amer. Statist. Assoc.*, 1225-1226.

[5] Warner, S.L. (1965). Randomized response : A survey technique for eliminating evasive answer bias. *Jour. Amer. Statist. Assoc.*, 63-69.