Jour. Ind. Soc. Ag. Statistics
56(2), 2003 : 199-206

Modified Optional Randomized Response Sampling Techniques

Housila P. Singh and Nidhi Mathur
Vikram University, Ujjain-456010, M.P.
(Received : January, 2003)

SUMMARY

In this paper, two randomized response techniques are proposed and have shown that these techniques are superior to Singh [12] model.

Key words : Optional randomized response technique, Percent relative efficiency (PRE), Sensitive characters.

1. Introduction

The randomized response model was first envisaged by Warner [18] as a procedure to reduce response distortion of threatening or personal questions. According to him, each respondent included in the sample using the simple random sampling with replacement (SRSWR) scheme is provided with a suitable randomization device consisting of two statements of the form: (i) "I belong to sensitive group A" and (ii) "I do not belong to sensitive group A" represented with probabilities p and (1 – p), respectively. The respondents give answers "yes" or "no" according to their randomly selected statement and to his actual status with respect to the attribute, without revealing the statement chosen. If m persons in the sample (including repetition) answered "yes", then the Warner's estimator

$$\hat{\pi}_1 = \frac{(\hat{\theta} - 1 + p)}{(2p - 1)}, \; p \neq \frac{1}{2} \tag{1.1}$$

is unbiased for π, the proportion of population belonging to the sensitive group A, and its variance is given by

$$V(\hat{\pi}_1) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \tag{1.2}$$

where $\hat{\theta} = \frac{m}{n}$, the proportion of 'yes' answers obtained from the n respondents.

Subsequently, various authors have modified and proposed alternative randomized response procedures applicable to different situations for instance see, Raghvarao [10], Fox and Tracy [5], Chaudhuri and Mukerjee ([3], [4]), Hedayat and Sinha [6], Sheers [11], Bellhouse [1], Mangat and Singh [8], Singh and Singh [13], Singh [14], Mangat [7], Singh *et al.* ([15], [16], [17]) and Chang and Huang [2].

Singh [12] has developed two randomized response procedures designated as RRT1 and RRT2, which are described below

RRT1 : In this method, each interviewee in a with replacement simple random sample of size n respondents is provided with one randomized response device. It consists of the statement, "I belong to sensitive group A" with known probability p, exactly the same probability as used by Warner [18] and statement 'yes' with probability (1 – p). The interviewee is instructed to use the device and report 'yes' or 'no' for the outcome of the sensitive statement according to his actual status. Otherwise, he is simply to report the 'yes' statement observed on the randomized response device. The whole procedure is completed by the respondent, unobserved by the interviewer. Then θ_1, the probability of 'yes' answer in the population is

$$\theta_1 = p\pi + (1 - p) \tag{1.3}$$

An unbiased estimator of π, considered by Singh [12] is

$$\hat{\pi}_1 = \frac{\{\hat{\theta} - (1 - p)\}}{p} \tag{1.4}$$

with the variance

$$V(\hat{\pi}_1) = \frac{\pi(1 - \pi)}{n} + \frac{(1 - p)(1 - \pi)}{np} \tag{1.5}$$

RRT2 : This procedure is exactly like RRT1 except for a change in probabilities on the randomized reponse device, i.e. the probabilities for the 'sensitive' statement and 'yes' statement have been interchanged. Then θ_2, the probability of 'yes' answer in the population is

$$\theta_2 = (1 - p)\pi + p \tag{1.6}$$

Singh [12] suggested an unbiased estimator of π as

$$\hat{\pi}_2 = \frac{(\hat{\theta} - p)}{(1 - p)} \tag{1.7}$$

with the variance

$$V(\hat{\pi}_2) = \frac{\pi(1 - \pi)}{n} + \frac{p(1 - \pi)}{n(1 - p)} \tag{1.8}$$

In the present paper, motivated by Mangat and Singh [9] an attempt has been made to suggest a modification of Singh's [12] model. The situations under which the suggested strategy is superior to that due to Singh [12] have been identified through a numerical study.

2. The Suggested Procedure

In this technique, the RR device and method for sampling the respondents remain same as in Singh [12]. However, it differs in the sense that the respondent is free to give answer in terms of 'yes' and 'no' either by using RR device or without using it, without revealing to the interviewer which mode has been followed for giving answer.

If T is the probability that a respondent gives answer without using RR device then assuming completely truthful reporting, the probability of 'yes' answer with RRT1 and RRT2 are respectively given by

$$\theta_1^* = \pi T + (1-T)\theta_1 \tag{2.1}$$

and

$$\theta_2^* = \pi T + (1-T)\theta_2 \tag{2.2}$$

For these procedures, we consider the following estimator of π

$$\hat{\pi}^* = n_1/n \tag{2.3}$$

Since n_1 is distributed as a (i) binomial variate $B(n, \theta_1^*)$ with RRT1, and (ii) binomial variate $B(n, \theta_2^*)$ with RRT2, we therefore have the theorems below

Theorems 2.1. The estimator $\hat{\pi}^*$ is biased and the expression for bias under RRT1 is given by

$$B_1(\hat{\pi}^*) = (1-T)(1-p)(1-\pi) \tag{2.4}$$

Theorems 2.2. The estimator $\hat{\pi}^*$ is biased and the expression for bias under RRT2 is given by

$$B_2(\hat{\pi}^*) = (1-T)p(1-\pi) \tag{2.5}$$

Further, the mean square error of $\hat{\pi}^*$ under RRT1 and RRT2 are respectively given by

$$MSE_1(\hat{\pi}^*) = \frac{1}{n}[\pi(1-\pi)\{p+T(1-p)\}^2 + (1-T)(1-p)(1-\pi)\{1-(1-p)(1-T)\}] + [(1-\pi)(1-T)(1-p)]^2 \tag{2.6}$$

and

$$MSE_2(\hat{\pi}^*) = \frac{1}{n}[\pi(1-\pi)\{T+(1-T)(1-p)\}^2 + p(1-\pi)(1-T)\{1-p(1-T)\}] + [p(1-T)(1-p)]^2 \tag{2.7}$$

In order to identify the situations where the proposed procedure fares well taking into account the precision, an empirical study has been carried out.

The percent relative efficiency of the estimator $\hat{\pi}^*$ under RRT1 with respect to $\hat{\pi}_1$ is given by

$$\text{PRE}(\hat{\pi}^*, \hat{\pi}_1) = \frac{\theta_1(1-\theta_1)}{p^2[\theta_1^*(1-\theta_1^*) + n\{(1-\pi)(1-p)(1-T)\}^2]} \times 100 \quad (2.8)$$

and the percent relative efficiency of the proposed estimator $\hat{\pi}^*$ under RRT2 with respect to $\hat{\pi}_2$ is given by

$$\text{PRE}(\hat{\pi}^*, \hat{\pi}_2) = \frac{\theta_2(1-\theta_2)}{(1-p)^2[\theta_2^*(1-\theta_2^*) + n\{(1-\pi)(1-p)(1-T)\}^2]} \times 100 \quad (2.9)$$

We have computed the PRE's for different values of p, π, T and displayed in Table 2.1 and Table 2.2. Keeping (2.4) and (2.5) in view, we have taken the value of p close to 'unity' (i.e. $p > 1/2$) for computing PRE $(\hat{\pi}^*, \hat{\pi}_1)$ while for computing $\text{PRE}(\hat{\pi}^*, \hat{\pi}_2)$ the value of p is taken close to zero (i.e. less than 1/2).

3. Concluding Remarks

The percent relative efficiency of the suggested estimator $\hat{\pi}^*$ with respect to $\hat{\pi}_1$, is shown in Table 2.1. It is observed from the table that

(i) At $T = 0.10$, $n = 2$ and $p \in [0.80, 0.90]$, the PRE increases as π increases. For (n, p) = (10, 0.80), the gain in efficiency is obtained when $\pi \in [0.40, 0.90]$ and for (n, p) = (10, 0.90), it is obtained when $\pi \in [0.30, 0.90]$. Further we note that for (n, p) = (20, 0.80), the gain in efficiency is observed when $\pi \in [0.60, 0.90]$ and for (n, p) = (20, 0.90), it is seen when $\pi \in [0.40, 0.90]$. When sample size is very large (i.e. $n \geq 100$), the gain in efficiency is observed only for large values of π.

(ii) At T = 0.50, n = 2 and $p \in [0.80, 0.90]$, the PRE decreases as π increases. For (n, p) = (10, 0.80), the PRE decreases as π increases whereas it increases for (n, p) = (10, 0.90). It is further observed that for $(n, p) = (20, 0.80)$ the substantial gain in efficiency is obtained when $\pi \in [0.20, 0.90]$ and for $(n, p) = (20, 0.90)$ it is obtained for $\pi \in [0.05, 0.90]$ For (n, p) = (100, 0.80), the considerable gain in efficiency is observed when $\pi \in [0.80, 0.90]$, and for (n, p) = (100, 0.90), it is seen when $\pi \in [0.60, 0.90]$.

Table 2.1. Percent relative efficiency of $\hat{\pi}^*$ under RRT1 with respect to $\hat{\pi}_1$

T	π	n = 2		n = 10		n = 20		n = 100	
		p = 0.8	p = 0.9	p = 0.8	p = 0.9	p = 0.8	p = 0.9	p = 0.8	p = 0.9
0.1	0.05	123.57	116.16	61.35	80.45	37.65	58.12	9.20	18.05
	0.10	128.13	117.75	69.11	88.85	43.86	67.99	11.18	23.62
	0.20	134.76	119.44	83.14	99.61	56.22	82.50	15.66	34.74
	0.30	139.35	120.32	95.47	106.22	68.50	92.65	21.01	45.81
	0.40	142.72	120.87	106.39	110.69	80.71	100.15	27.54	56.84
	0.50	145.29	121.24	116.14	113.92	92.84	105.92	25.65	67.83
	0.60	147.32	121.51	124.88	116.35	104.91	110.50	46.02	78.77
	0.70	148.97	121.71	132.78	118.26	116.90	114.21	59.74	89.67
	0.80	150.33	121.87	139.94	119.79	128.82	117.29	78.74	100.52
	0.90	151.47	122.00	146.47	121.05	140.66	119.89	106.80	111.33
0.5	0.05	200.67	165.45	133.04	138.44	93.60	114.98	27.76	48.80
	0.10	185.19	148.41	134.10	131.74	99.72	115.52	32.68	58.20
	0.20	167.91	134.10	135.54	125.45	109.22	116.09	42.78	72.69
	0.30	158.50	127.77	136.48	122.45	116.28	116.38	53.24	83.35
	0.40	152.58	124.21	137.13	120.69	121.72	116.56	64.10	91.51
	0.50	148.51	121.92	137.61	119.53	126.05	116.68	75.38	97.97
	0.60	145.55	120.32	137.99	118.72	129.57	116.77	87.09	103.21
	0.70	143.29	119.15	138.28	118.11	132.50	116.83	99.27	107.54
	0.80	141.51	118.25	138.52	117.64	134.96	116.88	111.94	111.18
	0.90	140.07	117.54	138.72	117.26	137.07	116.93	125.14	114.28
0.9	0.05	438.72	272.63	420.05	269.17	398.83	264.96	284.04	235.53
	0.10	300.79	195.31	293.53	194.02	284.92	192.02	230.81	180.58
	0.20	211.94	150.97	209.42	150.50	206.35	149.92	184.67	145.43
	0.30	178.41	135.20	177.12	134.95	175.54	134.64	163.83	132.22
	0.40	160.80	127.12	160.04	126.97	159.10	126.78	151.95	125.29
	0.50	149.94	122.21	149.46	122.11	148.87	121.99	144.29	121.02
	0.60	142.58	118.91	142.27	118.84	141.89	118.76	138.93	118.13
	0.70	137.26	116.53	137.07	116.49	136.83	116.44	134.97	116.04
	0.80	133.24	114.75	133.13	114.72	132.99	114.69	131.92	114.46
	0.90	130.08	113.35	130.04	113.34	129.98	113.33	129.51	113.23

Table 2.2. Percent relative efficiency of $\hat{\pi}^*$ under RRT2 with respect to $\hat{\pi}_2$

T	π	n = 2		n = 10		n = 20		n = 100	
		p = 0.8	p = 0.9	p = 0.8	p = 0.9	p = 0.8	p = 0.9	p = 0.8	p = 0.9
0.1	0.05	116.16	132.11	80.45	52.23	58.12	29.75	18.05	6.69
	0.10	117.75	138.51	88.85	58.31	67.99	33.83	23.62	7.76
	0.20	119.44	149.55	99.61	70.87	82.50	42.76	34.74	10.24
	0.30	120.32	158.72	106.22	84.01	92.65	52.89	45.81	13.34
	0.40	120.87	166.47	110.69	97.75	100.15	64.48	56.84	17.32
	0.50	121.24	173.10	113.92	112.14	105.92	77.86	67.83	22.60
	0.60	121.51	178.83	116.35	127.23	110.50	93.50	78.77	29.96
	0.70	121.71	183.85	118.26	143.07	114.21	112.01	89.67	40.93
	0.80	121.87	188.26	119.79	159.72	117.29	134.27	100.52	59.03
	0.90	122.00	192.19	121.05	177.24	119.89	161.53	111.33	94.51
0.5	0.05	165.45	231.89	138.44	126.82	114.98	80.96	48.80	20.80
	0.10	148.41	220.01	131.74	131.40	115.52	87.40	58.20	23.76
	0.20	134.10	204.08	125.45	139.06	116.09	99.46	72.69	30.34
	0.30	127.77	193.90	122.45	145.21	116.38	110.52	83.35	37.96
	0.40	124.21	186.83	120.69	150.24	116.56	120.70	91.51	46.90
	0.50	121.92	181.63	119.53	154.44	116.68	130.10	97.97	57.54
	0.60	120.32	177.65	118.72	158.00	116.77	138.81	103.21	70.40
	0.70	119.15	174.50	118.11	161.05	116.83	146.90	107.54	86.27
	0.80	118.25	171.95	117.64	163.70	116.88	154.44	111.18	106.32
	0.90	117.54	169.84	117.26	166.02	116.93	161.48	114.28	132.49
0.9	0.05	272.63	614.70	269.17	565.05	264.96	513.24	235.53	296.06
	0.10	195.31	423.50	194.02	402.60	192.43	397.20	180.58	258.86
	0.20	150.97	287.39	150.50	280.01	149.92	271.31	145.43	217.29
	0.30	135.20	233.05	134.95	229.35	134.64	224.89	132.22	194.61
	0.40	127.12	203.81	126.97	201.67	126.78	199.05	125.29	180.34
	0.50	122.21	185.55	122.11	184.22	121.99	182.59	121.02	170.52
	0.60	118.91	173.06	118.84	172.22	118.76	171.19	118.13	163.36
	0.70	116.53	163.97	116.49	163.46	116.44	162.82	116.04	157.90
	0.80	114.75	157.07	114.72	156.78	114.69	156.69	114.46	153.61
	0.90	133.35	151.65	113.34	151.52	113.33	151.37	113.23	150.14

(iii) At T = 0.90, the substantial gain in efficiency is observed for all values of (n, π, p) in decreasing order.

Finally, we conclude that the gain in efficiency increases as the value of T increases and decreases as sample size n increases. Moreover, we observe that at $p = 0.90$, the gain in efficiency is more in comparison to $p = 0.80$. We further note that for substantial gain in efficiency for all (n, p, π) it is advisable to choose T greater than 0.50 (i.e. $T > 0.50$).

Table 2.2 exhibits the percent relative efficiency of the suggested estimator $\hat{\pi}^*$ with respect to $\hat{\pi}_2$. From the table, we observe that

(i) At $T = 0.10$ and $n = 2$, the substantial gain in efficiency is achieved for all values of (π, p) in increasing order. For $(n, p) = (10, 0.10)$, the gain in efficiency is observed when $\pi \in [0.30, 0.90]$ whereas for $(n, p) = (10, 0.30)$, it is observed when $\pi \in [0.50, 0.90]$. We also note that for $(n, p) = (20, 0.10)$, the gain in efficiency is achieved when $\pi \in [0.40, 0.90]$ and for $(n, p) = (20, 0.30)$ it is seen when $\pi \in [0.70, 0.90]$. For large sample size $(n = 100)$, the gain in efficiency is observed only at $p = 0.10$ and $\pi \in [0.80, 0.90]$.

(ii) At $T = 0.50$, the gain in efficiency is obtained for all values of (n, p, π) except $(n = 20, \; p = 0.30, \pi \leq 0.20)$, $(n = 100, \; p = 0.10, \pi \leq 0.50)$ and $(n = 100, p = 0.30, \pi \leq 0.70)$.

(iii) At $T = 0.90$, the gain in efficiency is obtained for all values of (n, p, π) in decreasing form.

Finally, we infer that for larger gain in efficiency, the higher value of T is to be preferred. It is also seen that the PRE can be reasonably high for small sizes even if the value of T is low. Thus there is enough scope of choosing the value of T, so that the performance of the suggested estimator is better than Singh's [12] estimator for small sample sizes as well as moderately large sample sizes.

ACKNOWLEDGEMENTS

The authors are thankful to the referee for his constructive suggestions, which helped in bringing the paper in its final form.

REFERENCES

[1] Bellhouse, D.R. (1995). Estimation of correlation in randomized response. *Survey Methodology*, **21**, 13-19.

[2] Chang, H.J. and Huang, K.C. (2001). Estimation of proportion and sensitivity of a qualitative character. *Metrika*, **53**, 269-280.

[3] Chaudhuri, A. and Mukerjee, R. (1987). Randomized response techniques: A review. *Statistica Neerlandica*, **41**, 27-44.

[4] Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Responses: Theory and Techniques*. Marcel Dekker, New York.

[5] Fox, J. and Tracy, P. (1986). *Randomized Response: A Method for Sensitive Surveys*. Sage Publication, Beverly Hills.

[6] Hedyat, A.S. and Sinha, B.K. (1991). *Design and Inference in Finite Population Sampling*. Wiley, New York.

[7] Mangat, N.S. (1994). An improved randomized response strategy. *J. Roy. Statist. Soc.*, **B56**, 93-95.

[8] Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, **77 (2)**, 439-442.

[9] Mangat, N.S. and Singh, S. (1994). An optional randomized response sampling techniques. *Jour. Ind. Stat. Assoc.*, **32**, 71-75.

[10] Raghvarao, D. (1978). On an estimation problem in Warner's randomized response technique. *Biometrics*, **34**, 87-90.

[11] Sheers, N. (1992). A review of randomized response technique. *Measurement and Evaluation in Counselling and Development*, **25**, 27-41.

[12] Singh, S. (1993). An alternative to Warner's randomized response technique. *Statistica, anno* **LIII(1)**, 67-71.

[13] Singh, S. and Singh, R. (1993). Generalized Franklin's model for randomized response sampling. *Commu. Statist. -Theory and Methods*, **22(2)**, 741-755.

[14] Singh, S. (1994). Unrelated question randomized response sampling using continuous distributions. *Jour. Ind. Soc. Agril. Stat.*, **46 (3)**, 349-361.

[15] Singh, S., Singh, R., Mangat, N. S. and Tracy, D. S. (1994). An alternative device for randomized responses. *Statistica, anno* **LIV(2)**, 233-243.

[16] Singh, S., Horn, S. and Chowdhury, S. (1998). Estimation of stigmatized characteristics of a hidden gang in a finite population. *Austral. and New Zealand J. Statist.*, **40(3)**, 291-298.

[17] Singh, S., Singh, R. and Mangat, N.S. (2000). Some alternative strategies to Moor's model in randomized response sampling. *Jour. Statist. Plann. Inference*, **83**, 243-255.

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