STATISTICA, anno LXXIII, n. 4, 2013

AN ALTERNATIVE RANDOMIZED RESPONSE MODEL USING TWO DECK OF CARDS: A REJOINDER

Raghunath Arnab

Department of Statistics, University of Botswana, Botswana; and School of Statistics & Actuarial Sciences, University of Kwa-Zulu Natal, RSA

Sarjinder Singh

Department of Mathematics, Texas A&M University-Kingsville, USA

1. INTRODUCTION

While collecting information, directly from respondents, relating to sensitive issues such as induced abortion, drug addiction, duration of suffering from Aids and so on, the respondents very often report untrue values or even refuse to respond. Warner (1965) introduced an ingenious technique known as randomized response technique (RR) for estimating π, the proportion of population possessing certain stigmatized character A (say) by protecting the privacy of respondents and preventing the unacceptable rate of non-response. Warner's (1965) technique was modified by Horvitz et al. (1967), Greenberg et al. (1969), Raghavrao (1978), Kim (1978), Franklin (1989), Mangat and Singh (1990), Kuk (1990), Mangat and Singh (1991, 1992), Singh (1993), Kervliet (1994), Singh *et al.* (1994), Mahajan et al. (1994), Bhargava and Singh (2000), Singh et al. (2000), Singh and Mathur (2002, 2004), Arnab (2004), Espejo and Singh (2004), Sidhu and Bansal (2005/06), Pal and Sonali (2005/06), Javed and Grewal (2005/06), Zhimin (2005/06), Ryu et al. (2005/06), Zaizai (2005/06) and Singh (2010) among other researchers for improving greater co-operation and efficiency. One could also refer to Giordano and Perri (2011).

Recently Odumade and Singh (2009) proposed two decks of cards for estimating the population proportion π for RR surveys under SRSWR sampling. They showed on the basis of empirical investigation their model fares better than Warner (1965) and Mangat and Singh (1990) models if the value of $\pi \to 0$ or $\pi \to 1$. Arnab et al. (2012) extended Odumade and Singh (2009) RR technique for complex survey designs and wider classes of estimators while Abdelfatah et al. (2011) modified Odumade and Singh (2009) RR procedure under SRWR sampling method by using unrelated question. Abdelfatah et al. (2011) proved that the proposed method provides more efficient method of estimation π. In this paper we have shown that Abdelfatah et al. (2011) RR strategy (combination of estimator and sampling design) cannot provide efficient estimator than Warner (1965) model. For more clarity we describe Odumade and Singh (2009) and Abdelfatah et al. (2011) RR as follows:

1.1. *Odumade and Singh (2009)*

Odumade and Singh (2009) selected a sample of size n by SRSWR method. Each of the selected respondents in the sample is asked to select two cards, one card from the Deck-I and the other from the Deck-II. Each of the decks consists of two types of cards as in Warner (1965) model. The proportion of cards written "I belong to the sensitive group A" in the Deck-I and Deck-II are P and T respectively. The respondents are asked to report his/her response as (X, Y) where X indicates response from the card selected from the Deck-1 while Y indicates response from the card selected from the Deck-II. For example if a respondent selects a card written "I belong to the sensitive group A" from the Deck-I and selects the other card written "I do not belong to the sensitive group A" from the Deck-II, then he/she will supply with a response "Yes, No" if he/she belong to the sensitive group A. On the other hand if the respondent do not belongs to the group A, he/she will supply "No, Yes" as his/her response. Let out of the n responses n_{11}, n_{10}, n_{01} and n_{00} denote respectively the frequencies of the responses (Yes, Yes), (Yes, No), (No, Yes) and (No, No). An unbiased estimator for the population proportion π was obtained by Odumade and Singh (2009) as

$$\hat{\pi}_{os} = \frac{1}{2} + \frac{(P+T-1)(n_{11}-n_{00})(P-T)(n_{10}-n_{00})}{2n[(P+T-1)^2+(P-T)^2]} \tag{1}$$

The variance of $\hat{\pi}_{os}$ and an unbiased estimator of the variance of $\hat{\pi}_{os}$ were obtained respectively as

$$V(\hat{\pi}_{os}) = \frac{(P+T-1)^2\{PT+(1-P)(1-T)\}+(P-T)^2\{T(1-P)+P(1-T)\}}{4n[(P+T-1)^2+(P-T)^2]^2} - \frac{(2\pi-1)^2}{4n} - (2\hat{\pi}-1)^2] \tag{2}$$

and

$$\hat{V}(\hat{\pi}_{os}) = \frac{1}{4(n-1)}[\frac{(P+T-1)^2\{PT+(1-P)(1-T)\}+(P-T)^2\{T(1-P)+P(1-T)\}}{[(P+T-1)^2+(P-T)^2]^2} \tag{3}$$

1.2. Abdelfatah (2011)

Under this RR technique each respondent is asked to draw two cards; one from the "Deck(I)" and another from "Deck(II)". Deck (I) comprises two types of cards as in Warner (1965) viz. "I belong to the sensitive group A" with proportion W and " I do not belong to the sensitive group A" with proportion $1-W$. The respondent should answer truthfully "Yes" or "No".

The deck two comprises also two types of cards written "YES" with proportion Q and "NO" with proportion 1-Q. Regardless of his/her actual status, the respondent have to answer "YES" if he /she receives card written "YES". Alternatively, if the respondent receives the card written "NO" the respondent should answer "NO" as his or her response.

Deck (I)		Deck (II)
I $\in A$ with proportion W		"YES" with proportion Q
I $\in A^c$ with proportion W		"NO" with proportion $1-Q$

Let the responses of the selected sample of n units by SRSWR method be classified as follows:

Response from Deck (I)	Response from Deck (II)		Total
	Yes	No	
Yes	n_{11}	n_{10}	$n_{1\bullet}$
No	n_{01}	n_{00}	$n_{0\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 0}$	n

Abdelfatah et al. (2011) derived the following results:

(i) An unbiased estimator of the population π is

$$\hat{\pi}_f = \frac{1}{2} + \frac{Q(n_{11}/n - n_{\bullet 1}/n) + (1-Q)(n_{1\bullet}/n - n_{\bullet\bullet}/n)}{2(2W-1)[Q^2 + (1-Q)^2]}, \quad W \neq 0.5$$

(ii) The variance of $\hat{\pi}_f$ is

$$V(\hat{\pi}_f) = \frac{Q^3 + (1-Q)^3}{4n(2W-1)^2[Q^2 + (1-Q)^2]^2} - \frac{(2\pi - 1)^2}{4n}, \quad W \neq 0.5$$

(iii) An unbiased estimator of the variance of $\hat{\pi}_f$ is

$$\hat{V}(\hat{\pi}_f) = \frac{1}{4(n-1)}\left[\frac{Q^3 + (1-Q)^3}{(2W-1)^2[Q^2 + (1-Q)^2]^2} - (2\hat{\pi}_f - 1)^2\right]$$

2. PROPOSED IMPROVED ESTIMATOR

Suppose that from a finite population $U = (U_1, ..., U_N)$ of N identifiable units a sample s of size n is selected by SRSWR method. Then each of the selected respondent in the sample s is asked to perform randomizes device suggested by Abdelfatah et al. (2011). The respondents will provide one of the answers (Yes, Yes), (Yes, No), (No, Yes) and (No, No) as his/her randomize response. Let us denote

$$x_{11}(i) = \begin{cases} 1 \text{ } \textit{if the answer from the ith unit is (Yes, Yes)} \\ 0 \qquad \textit{otherwise} \end{cases} \tag{4}$$

$$x_{10}(i) = \begin{cases} 1 \text{ } \textit{if the answer from the ith unit is (Yes, No)} \\ 0 \qquad \textit{otherwise} \end{cases} \tag{5}$$

$$x_{01}(i) = \begin{cases} 1 \text{ } \textit{if the answer from the ith unit is (No, Yes)} \\ 0 \qquad \textit{otherwise} \end{cases} \tag{6}$$

$$x_{00}(i)=\begin{cases}1 \text{ if the answer from the ith unit is (No, No)} \\ 0 \qquad\qquad\qquad\qquad\qquad \text{otherwise}\end{cases} \tag{7}$$

Here the variable under study y is an indicator variable and y_i the value of y obtained from the ith unit can be written as

$$y_i=\begin{cases}1 \text{ if the ith unit} \in A \\ 0 \text{ if the ith unit} \notin A\end{cases} \tag{8}$$

For the Abdelfatah RR technique described above, we have

Probability of getting "yes" answer from card type I = $Wy_i+(1-W)(1-y_i)$
Probability of getting "no" answer from card type I = $(1-W)y_i+W(1-y_i)$
Probability of getting "yes" answer from card type II = Q
Probability of getting "no" answer from card type II = $1-Q$

Hence,

θ_{11} = Probability of getting "yes, yes" answer = $\{Wy_i+(1-W)(1-y_i)\}Q$
θ_{10} = Probability of getting "yes, no" answer = $\{Wy_i+(1-W)(1-y_i)\}(1-Q)$
θ_{01} = Probability of getting "no, yes" answer = $\{(1-W)y_i+W(1-y_i)\}Q$
θ_{00} = Probability of getting "no, no" answer = $\{(1-W)y_i+W(1-y_i)\}(1-Q)$

The likelihood function of getting x_{11} "yes, yes"; x_{10} "yes, no" answer ; x_{01} "no, yes" and x_{00} "No, No" answers is

$$\begin{aligned}L=&[\{Wy_i+(1-W)(1-y_i)\}Q]^{x_{11}}\times[\{Wy_i+(1-W)(1-y_i)\}(1-Q)]^{x_{10}}\\ &\times[\{(1-W)y_i+W(1-y_i)\}Q]^{x_{01}}\times[\{(1-W)y_i+W(1-y_i)\}(1-Q)]^{x_{00}}\end{aligned} \tag{9}$$

Note that

x_{11}, x_{10} , x_{01} and x_{00} are binary and subject to $x_{11}+x_{10}+x_{01}+x_{00}=1$.

Differentiating the likelihood with respect to y_i we get

$$\begin{aligned}\frac{\partial L}{\partial y_i}=&\frac{x_{11}(2W-1)}{\{Wy_i+(1-W)(1-y_i)\}}+\frac{x_{10}(2W-1)}{\{Wy_i+(1-W)(1-y_i)\}}\\ &-\frac{x_{01}(2W-1)}{\{(1-W)y_i+W(1-y_i)\}}-\frac{x_{00}(2W-1)}{\{(1-W)y_i+W(1-y_i)\}}\end{aligned}$$

$$= (2W-1)\left[\frac{x_{1\bullet}}{Wy_i + (1-W)(1-y_i)} - \frac{x_{0\bullet}}{(1-W)y_i + W(1-y_i)}\right] \tag{10}$$

where $x_{1\bullet} = x_{11} + x_{10}$ and $x_{0\bullet} = x_{01} + x_{02}$

$\dfrac{\partial L}{\partial y_i} = 0$ implies

$$x_{1\bullet}\{(1-W)y_i + W(1-y_i)\} - x_{0\bullet}\{Wy_i + (1-W)(1-y_i)\} = 0 \tag{11}$$

The equation (11) yields

$$(1-2W)y_i + W(x_{1\bullet} + x_{2\bullet}) - x_{2\bullet} = 0$$

i.e. MLE of y_i is

$$\hat{y}_i = \frac{W - x_{0\bullet}}{(2W-1)} \tag{12}$$

Denoting E_R and V_R as expectation and variance over the RR technique we have

THEOREM 1.

(i) $E_R(\hat{y}_i) = \dfrac{W - E_R(x_{0\bullet})}{(2W-1)}$

(ii) $V_R(\hat{y}_i) = \dfrac{W(1-W)}{(2W-1)^2} = \varphi$

PROOF.

(i) $E_R(\hat{y}_i) = \dfrac{W - E_R(x_{0\bullet})}{(2W-1)}$

$$= \frac{W - (\theta_{01} + \theta_{00})}{2W-1}$$

$$= y_i$$

(ii) $V_R(\hat{y}_i) = \dfrac{V_R(x_{0\bullet})}{(2W-1)^2}$

$$= V_R(x_{01}) + V_R(x_{00}) + 2C_R(x_{01}, x_{00}) \tag{13}$$

Now

$$V_R(x_{01}) = \theta_{01} - \theta_{01}^2$$

$$= \{(1-W)y_i + W(1-y_i)\}Q - [\{(1-W)y_i + W(1-y_i)\}Q]^2$$

$$= \{(1-W)y_i + W(1-y_i)\}Q - \{(1-W)^2 y_i + W^2(1-y_i)\}Q^2$$

(Since y_i can take values 0 or 1 with positive probabilities)

$$= Q(1-Q)(1-2W)y_i + WQ(1-WQ) \tag{14}$$

$$V_R(x_{00}) = \theta_{00} - \theta_{00}^2$$

$$= Q(1-Q)(1-2W)y_i + W(1-Q)\{1-W(1-Q)\} \tag{15}$$

And

$$C_R(x_{01}, x_{00}) = -\theta_{01}\theta_{00}$$

$$= -\{(1-W)y_i + W(1-y_i)\}^2 Q(1-Q)$$

$$= -Q(1-Q)\{(1-2W)y_i + W^2\} \tag{16}$$

Finally substituting (14), (15) and (16) in (13) we get

$$R_R(\hat{y}_i) = \frac{V_R(x_{01}) + V_R(x_{00}) + 2C_R(x_{01}, x_{00})}{(2W-1)^2}$$

$$= \frac{WQ(1-WQ) + W(1-Q)\{1-W(1-Q)\} - 2Q(1-Q)W^2}{(2W-1)^2}$$

$$= \frac{W(1-W)}{(2W-1)^2} \tag{17}$$

REMARK 1. It is important the MLE of y_i is independent of Q that is the force response from the card type II has no role in estimating y_i.

THEOREM 2. For SRSWR

(i) $\hat{\pi} = \frac{1}{n}\sum_{i \in s} \hat{y}_i = \frac{W - \hat{p}_{01}}{(2W-1)}$ *is an unbiased estimator of* π

(ii) Variance of $\hat{\pi}$ *is*

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{W(1-W)}{n(2W-1)^2}$$

(iii) An unbiased estimator of is

$$\hat{V}(\hat{\pi}) = \frac{1}{n(n-1)} \sum_{i \in s} (\hat{y}_i - \hat{\pi})^2$$

where $\sum_{i \in s}$ *denote sum over units in s with repetition and* $\hat{p}_{01} = \sum_{i \in s} x_{0\bullet} / n$.

PROOF.

(i) $E(\hat{\pi}) = E_p\{E_R(\hat{\pi})\}$

(where E_p denotes expectation over sampling design p (SRSWR))

$$= E_p[\frac{1}{n} \sum_{i \in s} E_R(\hat{y}_i)]$$

$$= \frac{1}{N} \sum_{i \in U} y_i$$

$$= \pi$$

(ii) $V(\hat{\pi}) = E[V_R(\hat{\pi})] + V_p[E_R(\hat{\pi})]$

(where V_p is the variance with respect to the sampling design p)

$$= \frac{1}{n^2} \sum_{i \in s} V_R(\hat{y}_i) + V_p(\frac{1}{n} \sum_{i \in s} y_i)$$

$$= \frac{\phi}{n} + \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \text{ where } \bar{Y} = Y / N$$

$$= \frac{\pi(1-\pi)}{n} + \frac{W(1-W)}{n(2W-1)^2}$$

(iii) $E[\hat{V}(\hat{\pi})] = V(\hat{\pi})$

The following Theorem 3 shows that the proposed estimator $\hat{\pi}$ is always superior to the estimator $\hat{\pi}_f$ proposed by Abdelfatah et al. (2011).

THEOREM 3.

$V(\hat{\pi}) \le V(\hat{\pi}_f)$, *equality holds if* $Q = 1/2$

PROOF.

Consider the difference

$$\frac{Q^3 + (1-Q)^3}{[Q^2 + (1-Q)^2]^2} - 1$$

$$= \frac{Q^3 + (1-Q)^3 - [Q^2 + (1-Q)^2]^2}{[Q^2 + (1-Q)^2]^2}$$

Now

$$Q^3 + (1-Q)^3 - [Q^2 + (1-Q)^2]^2$$

$$= Q^3 + (1-Q)^3 - \{Q^4 + 2Q^2(1-Q)^2 + (1-Q)^4\}$$

$$= Q(1-Q)\{4Q^2 + 1 - 4Q\}$$

$$= Q(1-Q)\{4Q^2 + 1 - 4Q\} \tag{18}$$

$$V(\hat{\pi}_f) = \frac{Q^3 + (1-Q)^3}{4n(2W-1)^2[Q^3 + (1-Q)^3]^2} - \frac{(2\pi - 1)^2}{4n}$$

$$\ge \frac{1}{4n(2W-1)^2} - \frac{(2\pi-1)^2}{4n}$$

(using (18))

$$= \frac{1}{4n(2W-1)^2} - \frac{1}{4n} + \frac{1}{4n} - \frac{(2\pi-1)^2}{4n}$$

$$= \frac{W(1-W)}{4n} + \frac{\pi(1-\pi)}{4n}$$

$$= V(\hat{\pi})$$

ACKNOWLEDGEMENT

The authors would like to thank the Editor, a referee and Giovanna Galatà Executive Editor for their valuable suggestions.

REFERENCES

S. ABDELFATAH, R. MAZLOUM, S. SINGH (2011). *An alternative randomized response model using two decks of card.* Statistica, LXXI (3), 381-390.

R. ARNAB (2004). *Optional randomized response techniques for complex survey designs*. "Biometrical Journal", 46, pp. 114-124.

R. ARNAB, S. SINGH, D. NORTH (2012), *Use of two decks of cards in randomized response techniques for complex survey designs*. Communications in Statistics-Theory and Methods, 41:16-17, 3198-3210.

M. BHARGAVA, R. SINGH (2000). *A modified randomization device for Warner's model*. Statistica, 60(2), 315–321.

L.A. FRANKLIN (1989). *A comparison of estimators for randomized response sampling with continuous distribution from dichotomous Populations.* Communications in Statistics, Theory- Methods, 18, pp. 489-505.

S. GIORDANO, P.F. PERRI (2012). *Efficiency comparison of unrelated question models based on same privacy protection degree.* Statistical Papers, 53, pp. 987–999.

C.R. GJESTVANG, S. SINGH (2006). *A new randomized response model.* Journal of the Royal Statistical Society, B, 68, pp. 523-530.

B.D. GREENBERG, A.L.A. ABUL-ELA, W.R. SIMMONS, D.G. HORVITZ (1969). *The unrelated question randomized response model. Theoretical framework.* Journal of American Statistical Association, 64, pp. 520-539.

D.G. HORVITZ , B.V. SHAH, W.R. SIMMONS (1967). *The unrelated question randomized response model.* Proceedings of Social Statistical section, American Statistical Association, pp. 65-72.

ZHIMIN, HONG (2005/06). *Estimation of mean in randomized response surveys when answers are incompletely truthful*. Model Assist. Stat. Appl. 1(4), pp. 221-230.

M. JAVED, I.S. GREWAL (2005/06). *On the relative efficiencies of randomized response devices with Greenberg unrelated question model*. Model Assist. Stat. Appl. 1(4), pp. 291-297.

J. KERKVLIET (1994). *Estimating a logit model with randomized data: the case of cocaine use.* Australian & Newzealand Journal of Statistics, pp. 36, 9-20.

J.I. KIM (1978). *Randomized response technique for surveying human populations.* Ph.D. Dissertation, Temple University, Philadelphia, USA.

A.Y.C. KUK (1990). *Asking sensitive question indirectly.* Biometrika 77, pp. 436-438.

P.K. MAHAJAN, J.P. GUPTA, R. SINGH (1994). *Determination of optimum strata boundaries for scrambled randomized response*. Statistica, 54(3), pp. 375–381

N.S. MANGAT, R. SINGH (1990). *An alternative randomized response procedure.* Biometrika 77, pp. 349-442.

N.S. MANGAT (1994). *An improved Randomized response strategy.* Journal of the Royal Statistical Society, B, 56, pp.93-95.

N. S. MANGAT (1991). *An optional randomized response sampling technique using nonstigmatized attribute*. Statistica, 51(4), pp. 595-602.

N. MANGAT, RAVINDRA SINGH (1992). *An alternative approach to randomized response survey*. Statistica, 51(3), pp. 327-332.

O. ODUMADE, S. SINGH (2009). *Efficient use of two decks of cards in randomized response sampling*. Commun. Statist.-Theory Meth., 38: pp. 439–446.

SANGHAMITRA PAL, S. CHAKRABORTY (2005/06). *Improvement upon Warner's model with an optional randomized response technique.* Model Assist. Stat. Appl, 1(4), pp. 299–304.

M. RUIZ ESPEJO, H.P. SINGH (2004). *Protection of privacy with objective prior distribution in randomized response*. Statistica, 63(4), pp. 697–701.

S.S. SIDHU, M.L. BANSAL (2008). *Estimator of population total using RAO, Hartley and Cochran's scheme using optional randomized response technique in multi-character surveys*. Model Assist. Stat. Appl, 3(3), pp. 259-267.

D. RAGHAVRAO (1978). *On estimation problem in Warner's randomized response Techniques.* Biometrics 34, pp. 87-90.

H.P. SINGH, N. MATHUR (2002). *On Mangat's improved randomized response strategy*. Statistica, 62(3), pp. 397–403.

JEA-BOK RYU, JONG-MIN KIM, TAE-YOUNG HEO, CHUN GUN PARK (2005/06). *On stratified randomized response sampling*. Model Assist. Stat. Appl. 1(4), pp. 31-36.

H.P. SINGH, N. MATHUR (2005), *Improved estimation of population proportion possessing sensitive attribute with unknown repeated trials in randomized response sampling*. Statistica, 64(3), pp. 537-544.

SUKHMINDER, SINGH (1993), *An alternative to Warner's randomized response technique*. Statistica, 53(1), pp. 67-71.

S SINGH (2010), *Proposed optimal orthogonal new additive model.* Statistica, 70(1), pp. 73-81.

S. SINGH, R. SINGH, N.S. MANGAT, D.S. TRACY (1994), " An alternative device for randomized responses." Statistica, 54(2), 233-243.

S. SINGH, N.S. MANGAT (2000), *Some alternative strategies to Moor's model in randomize response sampling.* Journal of Statistical Planning and Inference, 83, pp. 243-255.

S.L.WARNER (1965), *Randomize response: a survey technique for eliminating evasive answer bias.* Journal of American Statistical Association,60, pp. 63-69.

Y. ZAIZAI (2005/06). *Ratio method of estimation of population proportion using randomized response technique*. Model Assist. Stat. Appl., 1(2), pp. 125-130.

SUMMARY

An alternative randomized response model using two deck of cards: a rejoinder

The Randomized response (RR) technique with two decks of cards proposed by Odumade and Singh (2009) can always be made more efficient than the RR techniques proposed by Warner (1965), Mangat and Singh (1990), and Mangat (1994) by adjusting the proportion of cards in the decks. Abdelfatah et al. (2011) modified Odumade and Singh (2009) RR technique and claimed that their method can be more efficient than the Warner (1965) model. In this paper it is shown that such claim is not valid and the RR technique proposed by Abdelfatah et al. (2011) is in fact less efficient than the Warner (1965) technique at equal protection of respondents. Such finding are recently shown by Giordano and Perri (2011).