Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/060316

New Randomized Response Procedure for Finding Optimal Solution Using Branch and Bound Method

*Tanveer Ahmad Tarray**, *Muzafar Rasool Bhat and Kaiser J. Giri*

Islamic University of Science and Technology, Awantipora, Pulwama, Kashmir - India, 192122

Received: 12 Jun. 2017, Revised: 21 Sep. 2017, Accepted: 14 Oct. 2017
Published online: 1 Nov. 2017

Abstract: The crux of this paper is to consider a randomized response model using stratified random sampling based on Singh and Gorey (2017). In this paper the problem of optimal allocation in stratified random sampling where randomized response technique is used in presence of non response. The problem is formulated as a Nonlinear Programming Problem (NLPP) and is solved using Branch and Bound method. Also the results are formulated through LINGO.

Keywords: Randomized response technique, Optimum allocation, Stratified random sampling, Dichotomous population, Sensitive attribute, Branch and Bound method.

1 Introduction

The most serious problem in studying certain social problems that are sensitive in nature (e.g.drunk driving, use of marijuana, tax evasion, illicit drug use, induced abortion, shop lifting, child abuse, family disturbances, cheating in exams, HIV/AIDS and sexual behaviour induced abortion, etc.) is lack of reliable measure of their incidence or prevalence. Thus to obtain trustworthy data on such confidential matters, especially the sensitive ones, instead of open surveys alternative procedures are required. Such an alternative procedure known as randomized response technique (RRT) was first introduced by Warner (1965). It provides the opportunity of reducing response biases due to dishonest answers to sensitive questions. As a result, the technique assures a considerable degree of privacy protection in many contexts. Warner (1965) himself pointed out how one may get a biased estimate in an open survey when a population consists of individuals bearing a stigmatizing character A or its complement , which may or may not also be stigmatizing. Theoretical details for this model were given by Greenberg et al. (1969). This technique has generated much interest in the statistical literature since the publication of Warners (1965) randomized response (RR) model. Subsequently, several other workers have proposed different RR strategies for instance, see the review oriented references like Fox and Tracy (1986) and Tarray (2016). Some times in survey sampling certain amount of information is known about the elements of the population to be studied. For instance, information may be available on the geographical location of the area, e.g. if it is an inner city, a suburban or a rural area. Census information will provide a wealth of other information about the area, for instance, its population at the previous census, its rate of population change, the proportion of its population employed in manufacturing, or the proportion of its population with different origins. Supplementary information of this type can be used either at the design stage to improve the sample design, or at the analysis stage to improve the sample estimators, or both the essence of stratification is the classification of population in to sub-population or strata, based on some supplementary information and then the selection of separate samples from each of the strata. The benefits of stratification derive from the fact that the sample sizes in the strata are controlled by the sampler, rather than being randomly determined by the sampling process after the strata sample sizes are made proportional to the strata population sizes.

For the sake of completeness and convenience to the readers, we have given the descriptions of Singh and Gorey (2017) model.

The randomized response R_i device consists of a deck having three types of cards in Singh and Gorey (2017) model . In stratum i, p_{1i} proportions of cards carry the statement I belong to the sensitive category A, p_{2i} $(p_{1i} \neq p_{2i})$the proportion

* Corresponding author e-mail: tanveerstat@gmail.com

© 2017 NSP
Natural Sciences Publishing Cor.

of cards carry the statement I do not belong to category A and p_{3i} proportion of cards in the deck are left blank so that $p_{1i}+p_{2i}+p_{3i}=1$. In case the blank card is drawn by the respondent, he/she will report no whatever be his actual status with respect to the sensitive character. The rest of the procedure is same as suggested Warner (1965) [see, Hong et al.(1994) and Kim and Warde (2004)]. A respondent belonging to the sample in different strata will perform different randomization devices, each having different preassigned probabilities. Let n_i denote the number of units in the sample from stratum i and n denote the total number of units in sample from all stratum so that $n=\sum_{i=1}^{k} n_i$. Under the assumption that these "Yes" or "No" reports are made truthfully and $Pi(\neq 0.5)$ is set by the researcher, the probability of a "Yes" answer in a stratum i for this procedure is

$$\theta_{1i}=P_{1i}\pi_{si}+P_{2i}(1-\pi_{si}),\, for(i=1,2,...,k)$$

where θ_{1i} is the proportion of Yes answers in a stratum i ,π_{si} is the proportion of respondents with the sensitive trait in a stratum i.
The unbiased estimate of π_{si} is shown to be

$$\hat{\pi}_{si}=\frac{(\hat{\theta}_{1i}-P_{2i})}{P_{1i}-P_{2i}}$$

Since each $\hat{\theta_{1i}}$ is a binomial distribution $B(n_i,\theta_{1i})$ and the selections in different strata are made independently, the maximum likelihood estimate of is easily shown to be

$$\hat{\pi}_{tm}=\sum_{i=1}^{k} w_i\hat{\pi}_{si}=\sum_{i=1}^{k} w_i\{\frac{(\hat{\theta}_{1i}-P_{2i})}{P_{1i}-P_{2i}}\}$$

The variance of the unbiased estimator $\hat{\pi}_{tm}$ is

$$V(\hat{\pi}_{tm})=\sum_{i=1}^{k}\frac{w_i^2}{n_i}\{\pi_{si}(1-\pi_{si})+\frac{\pi_{si}P_{3i}}{(P_{1i}-P_{2i})}+\frac{P_{2i}(1-P_{2i})}{(P_{1i}-P_{2i})^2}\} \tag{1}$$

In this paper we have developed the problem of optimal allocation in stratified sampling where randomized response technique is used in presence of non response and is formulated as a non response programming problem . The formulated problem is solved using Branch and Bound method and the results are obtained through LINGO.

2 Problem Formulation

In the proposed models, the population is partitioned into strata, and a sample is selected by simple random sampling with replacement (SRSWR) in each stratum. Hong et al. (1994) suggested a stratified RR technique that applied the same randomization device to every stratum. Stratified random sampling is generally obtained by dividing the population into two over lapping groups called strata and selecting a simple random sample from each stratum. An RR technique using a stratified random sampling gives the group characteristics related to each stratum estimator. Also, stratified sample protect a researcher from the possibility of obtaining a poor sample. Under Hong et al. (1994) Proportional sampling assumption, it may be easy to derive the variance of the proposed estimator; however, it may cause a high cost because of the difficulty in obtaining a proportional sample from some stratum. To rectify this problem, Kim and Warde (2004) present a stratified randomized response technique using an optimal allocation which is more efficient than a stratified randomized response technique using a proportional allocation.
To get the full benefit from stratification, we assume that the number of units in each stratum is known. Let n_i denote the number of units in the sample from stratum i and n denote the total number of units in sample from all strata so that $n=\sum_{i=1}^{k} n_i$. Under the assumption that these "Yes" or "No" reports are made truthfully and P_i is set by the researcher. The problem of optimum allocation involves determining the sample size say $n_1,n_2,...,n_i$ that minimize the total variance $V(\hat{\pi}_{tm})$ subject to sampling cost. The sampling cost function is of the form $\sum_{i=1}^{k} c_i n_i$, the cost is proportional to the size of the sample within any stratum. But when we move from stratum to stratum, the cost per unit i.e. c_i may vary. Under RRT model the interviewer have to approach the population units selected in the sample to get the answers from the each

© 2017 NSP
Natural Sciences Publishing Cor.

stratum. In each stratum the interviewer have to travel from unit to contract them, this involves additional cost to the overhead cost. Also, we define

$$c^0 = C - C^0$$

The linear cost function is

$$C = C^0 + \sum_{i=1}^{k} c_i n_i$$

, where C^0 is the over head cost,c_i is the per unit cost of measurement in i^{th} stratum, C is the available fixed budget for the survey. Equation (1) can be rewritten as

$$V(\hat{\pi}_{tm}) = \sum_{i=1}^{k} \{\frac{w_i^2}{n_i}\} A_i$$

where

$$A_i = \pi_{si}(1-\pi_{si}) + \frac{\pi_{si}P_{3i}}{(P_{1i}-P_{2i})} + \frac{P_{2i}(1-P_{2i})}{(P_{1i}-P_{2i})^2} \tag{2}$$

The problem of optimum allocation can be formulated as a non linear programming problem (NLPP) for fixed cost as

$$Minimize V(\hat{\pi}_{tm}) = \sum_{i=1}^{k} \{\frac{w_i^2}{n_i}\} A_i$$

$$subject to \sum_{i=1}^{k} c_i n_i \leq c^0$$

$$2 \leq n_i \leq N_i$$

$$and n_i integers, i = 1, 2, ..., k. \tag{3}$$

The above NLPP can be solved using non linear integer programming technique. We can now apply Branch and Bound method to determine the optimal sample size in presence of non response. This method consists of two strategies , alternatively followed till the desired solution is obtained. One strategy consists in Branch a problem in to two sub problems and the other in solving each of the two sub problems to obtain the minimum or suitable lower bound of the objective function.

3 Solution Procedure

Let us now determine the solution of problems (3) by ignoring upper and lower bounds and integer requirements. The Lagragian function may be

$$\varphi = \sum_{i=1}^{k} \{\frac{w_i^2}{n_i}\} A_i + \lambda(\sum_{i=1}^{k} c_i n_i - c^0) \tag{4}$$

Differentiating (4) with respect to n_i and equate to zero, we get

$$\frac{\bar{V}\varphi}{Vn_i} = 0 \Rightarrow n_i = \frac{w_i\sqrt{A_i}}{\sqrt{c_i}\sqrt{\lambda}} \tag{5}$$

Again differentiating (4) with respect to λ in equation to zero, we get

$$\frac{\bar{V}\varphi}{V\lambda} = 0 \Rightarrow c^0 = \sum_{i=1}^{k} c_i n_i \tag{6}$$

Solving (5) and (6), we have

$$\sqrt{\lambda} = \sum_{i=1}^{k} c_i \frac{w_i\sqrt{A_i}}{\sqrt{c_i}} \tag{7}$$

© 2017 NSP
Natural Sciences Publishing Cor.

Substituting (7) in (5), we have

$$n_i = \frac{w_i\sqrt{A_i}}{\sum\limits_{i=1}^{k} c_i \frac{w_i\sqrt{A_i}}{c^0\sqrt{c_i}}\sqrt{c_i}} \Rightarrow \frac{c^0 w_i \frac{\sqrt{A_i}}{\sqrt{c_i}}}{(\sum\limits_{i=1}^{k} w_i\sqrt{A_i})\sqrt{c_i}} \tag{8}$$

The Branch and Bound method will require the solution of sub problems in which some of the n_i are fixed. Suppose that at r^{th}node, the fixed values of n_i are for $i\varepsilon 1r$. Then the required Lagrangian function is

$$\varphi = \sum_{i=1\varepsilon r}^{k} \{\frac{w_i^2}{n_i}\} A_i + \lambda(\sum_{i=1\varepsilon r}^{k} c_i n_i - c^0) \tag{9}$$

Further, differentiating (9) with respect to n_i and equating to zero, we have

$$n_i = \frac{w_i\sqrt{A_i}}{\sqrt{\lambda}\sqrt{c_i}} \tag{10}$$

At r^{th} node,

$$\sum_{i\varepsilon 1r}^{k} c_i n_i = c^0 - \sum_{i\varepsilon 1r}^{k} c_i n_i \tag{11}$$

$$\Rightarrow \lambda = \frac{c^0 - \sum\limits_{i\varepsilon 1r}^{k} c_i n_i}{\sum\limits_{i\varepsilon 1r}^{k} \sqrt{c_i} w_i \sqrt{A_i}} \tag{12}$$

After simplification , we get formula for r^{th} node as

$$n_i = \frac{(c^0 - \sum\limits_{i\varepsilon 1r}^{k} c_i n_i)\frac{\sqrt{A_i}w_i}{\sqrt{c_i}}}{\sum\limits_{i\varepsilon 1r}^{k} \frac{\sqrt{A_i}w_i}{\sqrt{c_i}}} \tag{13}$$

where $1r$ is the set of indices which have been fixed at the r^{th} node.

4 Numerical illustration

To judge the performance of the proposed a numerical example is presented to illustrate the formulation of the problem. Assuming that C (available budget) = 4500 units including c^0 and c^0 =500 units (overhead cost). Therefore c^0 =4500-

Table 1: The stratified population with $P_1 = 0.4, P_2 = 0.3$ and $P_3 = 0.3$ is given as

Stratum i	N_i	w_i	π_{si}	c_i
1	400	0.7	0.08	15
2	800	0.3	0.03	20

500=4000 units. Also we assume that 400 and 800 are stratum sizes respectively as given in above table for i = 1,2 ,N = 400+800 = 1200. The values of A_i and $A_i w_i^2$ are calculated as given in table below.

Substituting the above calculated values of the parameters into (3) non linear programming problem NLPP, we have

$$\begin{aligned} Minimize V(\hat{\pi}_{tm}) &= \frac{10.444}{n_1} + \frac{1.893}{n_2} \\ subject to 15n_1 + 20n_2 &\leq 4000 \\ 2 \leq n_1 &\leq 400 \\ 2 \leq n_2 &\leq 800 \end{aligned}$$

© 2017 NSP
Natural Sciences Publishing Cor.

Table 2: Calculated values of A_i and $A_i w_i^2$

Stratum (i)	A_i	$A_i w_i^2$
1	21.314	10.444
2	21.038	1.893

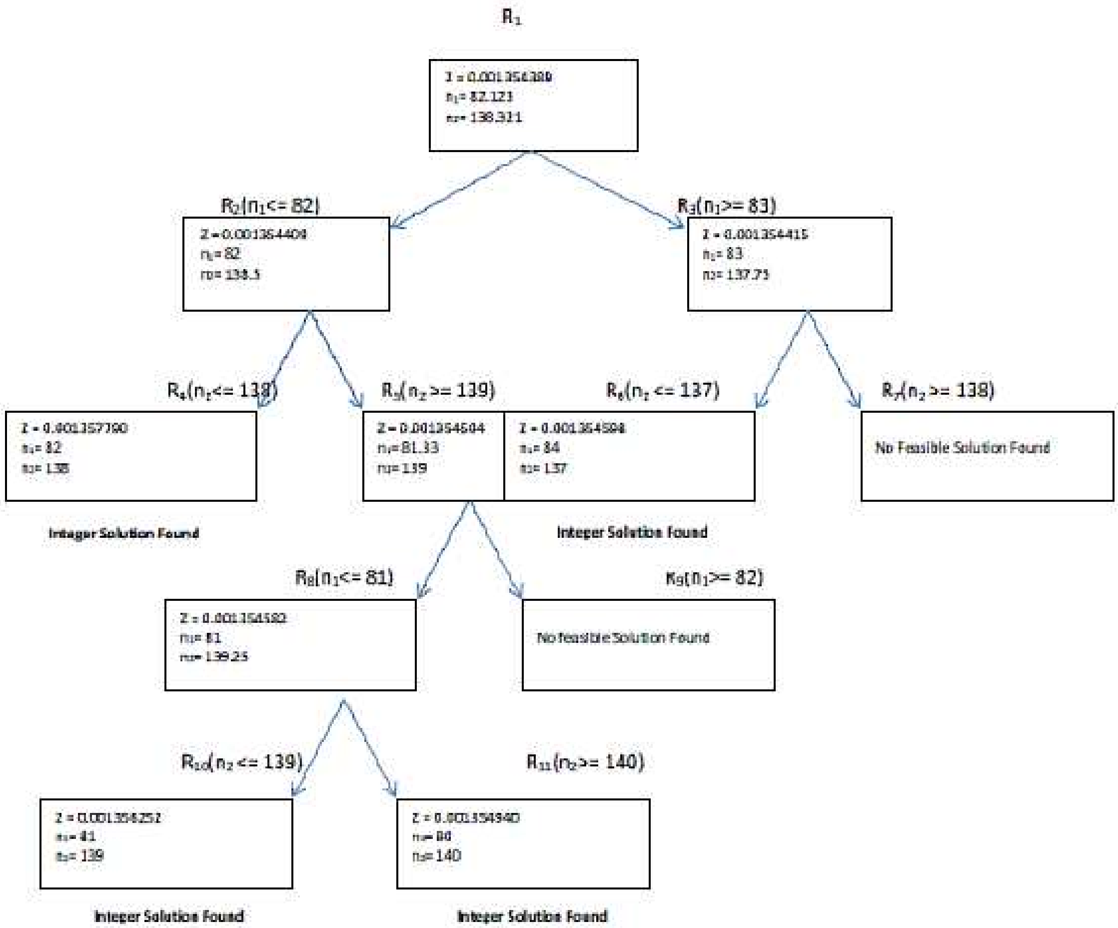

Fig. 1: various nodes for NLPP.

Using the above minimization problem , we get optimal solution as $n_1 = 82.123$, $n_2 = 138.321$ and optimal value is Minimize $V(\hat{\pi}_{tm}) = 0.001398345$. Since $n_1$and n_2 are required to be the integers, we branch problem R_1 into two sub problems R_2 and R_3 by introducing the constraints $n_1 \leq 82$ and $n_1 \geq 83$ respectively indicated by the value n_1=82.123 which lies between 82 and 83. This process of replacing a problem by two sub problems is called branching. The solution of these two sub problems can be obtained using LINGO software as shown in figure (1). Since these two sub problems have optimal solutions in which the variables n_2 is non-integral, none of the sub problems has been fathomed. So both problems R_2 and R_3 are further branched into sub problems $R_4, R_5, R_6 and R_7$ with additional constraints as $n_2 \leq 138, n_2 \geq 139, n_2 \leq 137 and n_2 \geq 138$ respectively .Problems R_4 stand fathomed as the optimal solution in each case is integral in n_1 and n_2. Problem R_5 has been further branched into sub problems R_8 and R_9 with additional constraints as $n_1 \leq 81, n_1 \geq 82$; respectively. Problem R_3 is not fathomed and is further branched into two sub problems, R_6 and R_7 by imposing the additional constraints $n_2 \leq 137, n_2 \geq 138$ respectively, which suggests that R_6 is fathomed and R_7 has no feasible solution. R_9 stands fathomed as the optimal solution in each case of n_1 and n_2 but problem R_8 is not fathomed and is

© 2017 NSP
Natural Sciences Publishing Cor.

required to further branching into two sub problems R_{10} and R_{11} by imposing the additional constraints $n_2 \leq 139, n_2 \geq 140$ respectively which are suggested by the non-integral value $n_2 = 139.25$. Problem R_{11} and problem R_{10} is fathomed with integer value. Now, all the terminal nodes are fathomed. The feasible fathomed node with the current best lower bound is node R_6. Hence the solution is treated as optimal. The optimal value is $n_1 = 156$ and $n_2 = 183$ and optimal solution is to Minimize $V(\hat{\pi}_{tm})$=0.001354598. The total cost under this allocation is 4000 units. It may be noted that the optimal integer values are same as obtained by rounding the n_i to the nearest integer. Let us suppose Minimize $V(\hat{\pi}_{tm})$=Z, the various nodes for the NLPP (3) utilizing table 1 and table 2, are presented below in figure (1).

5 Conclusion

A stratified randomized response method assists to solve the limitations of randomized response that is the loss of individual characteristics of the respondents. Formulating non linear programming problem (NLPP) of optimum allocation in stratified sampling with linear cost function in presence of non responses using Branch and Bound algorithm based on Singh and Gorey (2017) provides the optimum integer solution.

Acknowledgements

The authors are thankful to the Editor in- Chief and to the anonymous learned referees for there valuable suggestions regarding improvement of the paper.

References

[1] Chaudhuri A., Mukerjee R. (1988): Randomized Response: Theory and Techniques. Marcel- Dekker, New York, USA.
[2] Cochran W. G. (1977): Sampling Technique, 3rd Edition. New York: John Wiley and Sons, USA.
[3] Fox J. A. and Tracy P. E. (1986): Randomized Response: A method of Sensitive Surveys. Newbury Park, CA: SEGE Publications.
[4] Greenberg B., Abul- Ela A., Simmons W. R. and Horvitz D.G. (1969): The unreleased question randomized response: Theoretical framework. Jour. Amer. Statist. Assoc., 64,529-539.
[5] Hong K., Yum J. and Lee H. (1994): A stratified randomized response technique. Korean Jour. Appl. Statist., 7, 141-147.
[6] Kim J. and Warde W. (2004): A stratified Warner randomized response model. Jour. Statist. Plan. Infer., 120, 155-165.
[7] Singh H.P. and Gorey S.(2017):An Efficient Stratified Randomized Response Model. Jour. Statist. Theo. Prac., DOI: 10.1080/15598608.2017.1350607.
[8] Tarray T. A. (2016): Statistical Sample Survey Methods and Theory. Elite Publishers (onlinegatha.com), INDIA, ISBN: 978-93-86163-07-03.
[9] Warner S.L. (1965): Randomized response: A survey technique for eliminating evasive answer bias. Jour. Amer. Statist. Assoc., 60, 63-69.

Tanveer Ahmad Tarray Has received his Masters Degree in Statistics and B.Ed. from University of Kashmir India, M.Phil. and Ph.D. in Statistics from Vikram University Ujjain M.P. , India respectively. The Author is working as an Assistant Professor in the Department of Computer Science and Engineering, Islamic University of Science & Technology, Awantipora, J & K, India. His area of interests include - Survey Sampling, Applied Statistics, Operation Research and Probability Distributions. He has rich experience in teaching the graduate and postgraduate classes. He has published more than forty-five quality research papers at National and International Levels in reputed Journals. He is also the Recipient of Young Scientist Award from M.P. Council of Science and Technology India .

© 2017 NSP
Natural Sciences Publishing Cor.

Muzafar Rasool Bhat Has received his Masters Degree in Computer Applications from University of Kashmir, India and is perusing his Ph.D. from the same University. The Author is working as an Assistant Professor in the Department of Computer Science, Islamic University of Science & Technology, Awantipora, J & K, India. He has rich experience in teaching the graduate and postgraduate classes. The interest arrears of Author include Applied Mathematics, Signal Processing, Algorithms, Machine Learning and Information Security.

Kaiser J. Giri Has received his Masters Degree, M.Phil., Ph.D. in Computer Applications from University of Kashmir, India. The Author is working as an Assistant Professor in the Department of Computer Science, Islamic University of Science & Technology, Awantipora, J & K, India since 2006. He has rich experience in teaching the graduate and postgraduate classes. The interest arrears of Author include Image Processing, Signal Processing, and Information Security.

© 2017 NSP
Natural Sciences Publishing Cor.