Extension of Mangat Randomized Response Model

Zawar Hussain
Department of Statistics, Quaid-i-Azam University 45320, Islamabad 44000, Pakistan
E-mail: zhlangah@yahoo.com

Salman Arif Cheema
Department of Statistics, Virginia Tech (VA), Blacksburg 26040, USA.

Sidra Zafar
Faculty of Mathematical Sciences, Fatima Jinnah Women University, Rawalpindi, Pakistan

Abstract

In this study, a Bayesian estimation of population proportion of a stigmatized attribute has been considered and a Bayes estimator is developed when the information from the survey respondents is collected using the randomized response technique (RRT) by Mangat and Singh (1990). Dominance picture of Bayes estimator has been portrayed for a wide range of values of population proportion by using simple Beta distribution as Prior information. It is observed Bayes estimator performed better than the Maximum Likelihood Estimator (MLE), in particular, for small and moderate samples.

Key Words: Sensitive attribute, Bayesian estimation, Simple random sampling and Mean squared error.

1. Introduction

Untruthful responses might be received from respondents in a direct questioning approach regarding sensitive attributes. For many reasons, it might be necessary to acquire information about incidence of stigmatized attribute(s) in the population. Warner (1965) was the first researcher who proposed a method of survey to collect information regarding stigmatized/ sensitive attributes by ensuring privacy of the respondent. Now a large number of improvements and variants of Warner's randomized response model (RRM) have been suggested in several studies, for example, Greenberg et al. (1969), Chaudhuri and Mukerjee (1988), Mangat and Singh (1990), Mangat et al. (1995), Tracy and Mangat (1996), Mahmood et al. (1998), Bhargava and Singh (2000), Singh et al. (2003), Christofides (2003), Kim and Warde (2004), Zaizai (2005-2006), Hussain et al. (2007), Perri (2008), Diana and Perri (2009, 2010) Hussain and Shabbir(2011), Hussain et al. (2011a, b), Abid et al. (2011),

However, in situations where prior information is available, Bayesian approach of estimation can be adopted in order to estimate the unknown parameter by combining sample and prior information. Winkler and Franklin (1979), Pitz (1980), Spurrier and Padgett (1980), O'Hagan (1987), Oh (1994), Migon and Tachibana (1997), Unnikrishnan and Kunte (1999), Bar-Lev and Bobovich (2003), Barabesi and Marcheselli (2006), and Kim et al. (2006) are some of the researchers who made efforts in Bayesian analysis of randomized response models. Using the Mangat and Singh (1990) RRM and applying the Bayesian estimation we intend to suggest an estimator of population proportion. Before moving to the formal development of Bayes estimator we present the Mangat Singh (1990) RRMs in the following section. The development of Bayes estimator is presented in Section 2 and a comparative study is presented in Section 3.

2. Mangat and Singh RRM

The basic rationale in Mangat and Singh (1990) RRM is to establish a random relationship between the sensitive questions and individual's responses. In this model, two randomization devices R_1 and R_2 are used. The device R_1 consists of the two statements: (i) "do you belong to sensitive group?", and (ii) " go to R_2", presented with probabilities T and $(1-T)$ respectively. The randomization device R_2 consists of the two statements: (i) "do you belong to sensitive group?",and (ii) "do you not belong to sensitive group?", presented with probabilities P and $(1-P)$. A respondent is selected and the response is recorded as *yes* if the respondents actual status matches with the selected question and *no* else wise.

For a particular respondent, the probability for a *yes* answer is given by

$$P_r(yes) = \phi = \pi\left\{2(T + P - PT) - 1\right\} + (1-P)(1-T), \quad (1)$$

The MLE of π is given by

$$\hat{\pi}_{ML} = \frac{\hat{\phi} - (1-P)(1-T)}{\left\{2(T+P-PT)-1\right\}}, \tag{2}$$

where, $\hat{\phi} = \frac{n'}{n}$ and n' is the number of *yes* responses in the sample of n respondents. Moreover,

$$\text{var}(\hat{\pi}_{ML}) = \frac{\pi(1-\pi)}{n} + \frac{(1-T)(1-P)\left\{1-(1-T)(1-P)\right\}}{n\left\{2P-1+2T(1-P)\right\}^2}. \tag{3}$$

3. *Bayesian Estimation of π using Mangat and Singh (1990) RRM*

For the Bayesian estimation of population proportion for this randomized response model, we assume Beta distribution as the prior distribution for population proportion (π) with parameters a and b, which is as follows:

$$f(\pi) = \frac{1}{\beta(a,b)} \pi^{a-1} (1-\pi)^{b-1}, 0 < \pi < 1, a, b > 0. \tag{4}$$

Let $X = \sum_{i=1}^{n} x_i$,where X represents the total number of *yes* responses in a sample of size n drawn from the population with simple random sampling with replacement. Here $x_i = 1$ with probability ϕ and $x_i = 0$ with probability $1-\phi$, where ϕ is as defined in (1). Then the conditional distribution of X given π is:

$$\begin{aligned} f_{X/\pi}(X/\pi) &= \frac{n!}{x!(n-x)!} \phi^x (1-\phi)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \left[(2T+2P-2PT-1)+PY-P-T+1\right]^x \\ &\times \left[1-\left\{\pi(2T+2P-2PT-1)+PY-P-T+1\right\}\right]^{n-x} \end{aligned}$$

Letting $f = \frac{(T-1)(P-1)}{2(T+P-TP)-1}$ and $h = \frac{3(1-TP)-2\left(T^2+P^2-2PT^2+T^2P^2\right)-T-P}{2(T+P-TP)-1}$, we have

$$f_{X/\pi}(X/\pi) = \frac{n!}{x!(n-x)!} \left[(2T+2P-2TP-1)\right]^n (\pi+f)^x (1-\pi+h)^{n-x}$$

or

$$f_{X/\pi}(X/\pi) = \frac{n!}{x!(n-x)!\left[(2T+2P-2TP-1)\right]^{-n}} \sum_{i=0}^{x} \sum_{j=0}^{n-x} \frac{(n-x)!}{j!(n-x-j)!} \frac{x!}{i!(x-i)!} f^{x-i} h^{n-x-j} \pi^i (1-\pi)^j, \tag{5}$$

where, $x = 0, 1, 2, \ldots.., n$.

The joint distribution of X and π is given by:

$$f(X,\pi) = \frac{1}{\beta(a,b)} \pi^{a-1} (1-\pi)^{b-1} \frac{n!}{x!(n-x)!} \left[(2T+2P-2TP-1)\right]^n \sum_{i=0}^{x} \sum_{j=0}^{n-x} \frac{(n-x)!}{j!(n-x-j)!} f^{x-i} h^{n-x-j} \pi^i (1-\pi)^j$$

Or

$$f\left(X,\pi\right)=\frac{1}{\beta\left(a,b\right)}\frac{n!}{x!\left(n-x\right)!}\left[\left(2T+2P-2TP-1\right)\right]^{n}$$
$$\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\pi^{a+i-1}\left(1-\pi\right)^{j+b-1}. \qquad (6)$$

Now the marginal distribution of X can be obtained by integrating the joint distribution of X and π. Thus the marginal distribution of X is given by:

$$f\left(X\right)=\frac{1}{\beta\left(a,b\right)}\frac{n!}{x!\left(n-x\right)!}\left[\left(2T+2P-2TP-1\right)\right]^{n}$$
$$\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\beta\left(a+i,j+b\right). \qquad (8)$$

Now, the posterior distribution of π given X is defined as:

$$f_{\pi/X}\left(\pi/X\right)=\frac{f\left(\pi,X\right)}{f\left(X\right)}. \qquad (9)$$

Thus the posterior distribution of π given X may be obtained as:

$$f_{\pi/X}\left(\pi/X\right)=\frac{\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\pi^{a+i-1}\left(1-\pi\right)^{j+b-1}}{\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\beta\left(a+i,b+j\right)}. \qquad (10)$$

The Bayes estimator, under the squared error loss function, is given by

$$\hat{\pi}_{Bayes}=\frac{\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\beta\left(a+i+1,b+j\right)}{\sum_{i=0}^{x}\sum_{j=0}^{n-x}\frac{\left(n-x\right)!}{j!\left(n-x-j\right)!}\frac{x!}{i!\left(x-i\right)!}f^{x-i}h^{n-x-j}\beta\left(a+i,b+j\right)}. \qquad (11)$$

The Mean Squared Errors (MSEs) of both the classical (MLE) and Bayesian estimators are defined for a fixed value of π and are written as

$$MSE\left(\hat{\pi}_{ML}\right)=E\left(\hat{\pi}_{ML}-\pi\right)^{2}=\sum_{x=0}^{n}\left(\hat{\pi}_{ML}-\pi\right)^{2}\phi^{x}\left(1-\phi\right)^{n-x} \qquad (12)$$

And

$$MSE\left(\hat{\pi}_{Bayes}\right)=E\left(\hat{\pi}_{Bayes}-\pi\right)^{2}=\sum_{x=0}^{n}\left(\hat{\pi}_{Bayes}-\pi\right)^{2}\phi^{x}\left(1-\phi\right)^{n-x}. \qquad (13)$$

4. Comparison

For different values of design parameters, using (12) and (13) we have calculated the MSEs of the both the estimators under consideration. For this purpose R is used to write the codes. The results are displayed graphically in the Figures 1-4(see Appendix). To save the space we have presented a limited number of graphs but we have observed that for all values of the design parameters P and T and population proportion π the Bayes estimator outperforms the MLE over the complete range of π. We assumed hyperparameters $a=1$ and $b=2$ so that prior mean is $\frac{1}{3}$. It is observed that Bayes estimator is bounded in the interval $\left[0,1\right]$ even in the extreme realization of *yes* responses. For these extremes the MLE may lie outside this interval which is not admissible. Thus, whenever it is easier to collect data through Mangat and Singh RRM Bayesian estimation may be used to have more reliable estimates.

References

1. Abid, M. Cheema, S. A. and Hussain, Z. (2011). An Efficiency Comparison of Certain Randomized Response Strategies. *International Journal of Business and Social Sciences*, 2(7), 188-198.
2. Barabesi, L.; Marcheselli, M. (2006) A practical implementation and Bayesian estimation in Franklin's randomized response procedure. *Communication in Statistics- Computation and Simulation.* 35, 563-573.
3. Bar-Lev, S. K. Bobovich, E. and . Boukai, B. (2003) A common conjugate prior structure for several randomized response models. *Test*, 12 (1), 101-113
4. Bhargava, M. and Singh, R. (2000). A modified randomization device for Warner's model, *Statistica*, 60, 315-321.
5. Chaudhuri, A. and Mukerjee, R. (1988). Randomized Response: Theory and Techniques. Marcel Dekker, New York.
6. Christofides, T. C. (2003). A generalized randomized response technique. *Metrika* 57, 195-200.
7. Diana, G. and Perri, P. F. (2009). Estimating a sensitive proportion through randomized response procedures based on auxiliary information, *Statistical Papers*, **50**, 661-672.
8. Diana, G. and Perri, P. F. (2010). New scrambled response models for estimating the mean of a sensitive character, *Journal of Applied Statistics*, 37(11), 1875-1890.
9. Greenberg, B. G., Abul-Ela Abdel-Latif, A., Simmons, W. R. and Horvitz, D. G. (1969). The unrelated question RR model: Theoretical framework, *Journal of the American Statistical Association*, 64, 52-539.
10. Hussain, Z., Shabbir, J. and Gupta S. (2007). An alternative to Ryu et al. randomized response model, *Journal of Statistics & Management Systems*, 10(4), 511-517.
11. Hussain, Z., Shabbir, J. (2011). An Estimation of Sensitive Proportion Utilizing Higher Order Moments of Auxiliary Variable. *International Journal of Business and Social Sciences*, 2(2), 121-125.
12. Hussain, Z., Hamraz, M. and Shabbir, J. (2011a). An improved Quantitative Randomized Response Model. *International Journal of Business and Social Sciences*, 2(4), 200-205.
13. Hussain, Z., Anjum, S. and Shabbir, J. (2011b). Improved logit estimation through Mangat randomized response model. *International Journal of Business and Social Sciences*, 2(5), 179-188.
14. Kim, J. M., Warde, D. W. (2004). A stratified Warner's randomized response model. *J. Statist. Plann. Inference*, Vol. 120/1-2 155-165.
15. Kim, J. M., Tebbs, J. M., An. S. W. (2006). Extensions of Mangat's randomized response model. *J. Statist. Plann. Inference*, 36(4) 1554-1567.
16. Mahmood, M. Singh, S. and Horn, S. (1998). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques, *Biometrical Journal*, 40, 237-242.
17. Mangat, N. S., Singh, R. (1990). An alternative randomized response procedure, *Biometrika*, **77**, 439-442.
18. Mangat, N. S., Singh, S. and Singh, R. (1995). A note on the inverse binomial randomized response procedure, *Journal of the Indian Society of Agricultural Statistics*, 47, 21-25.
19. Migon, H., Tachibana, V. (1997). Bayesian approximations in randomized response models. *Comput. Statist. Data Anal*, 24 401-409.
20. O'Hagan, A. (1987). Bayes linear estimators for randomized response models. *J. Amer. Statist. Assoc*, 82 580-585.
21. Oh, M. (1994). Bayesian analysis of randomized response models: a Gibbs sampling approach. *J. Korean. Statist*. Soc,23 463-482.
22. Pitz. G. (1980). Bayesian analysis of randomized response models. *J. Psychological Bull*, 87 209-212.
23. Perri, P.F. (2008). Modified randomized devices for Simmons' model, *Model Assisted Statistics and Applications*, 3, 233-239.
24. Spurrier, J., Padgett, W. (1980). The application of Bayesian techniques in randomized response. *Sociological Methodol*, 11 533-544.
25. Singh, S., Horn, S., Singh, R. and Mangat, N.S. (2003). On the use of modified randomization device for estimating the prevalence of a sensitive attribute, *Statistics in Transition*, 6(4), 515-522.
26. Tracy, D. and Mangat, N. (1996). Some development in randomized response sampling during the last decade-a follow up of review by Chaudhuri and Mukerjee, *Journal of Applied Statistical* Sciences, **4**, 533-544.

27. Unnikrishnan, N., Kunte, S. (1999). Bayesian analysis for randomized response models. *Sankhya* ,Ser. B, 61 422-432.
28. Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.
29. Winkler,R., Franklin,L. (1979). Warner's randomized response model: A Bayesian approach. *J. Amer. Statist. Assoc*, 74 207-214.
30. Zaizai, Y. (2005, 2006). Ratio method of estimation of population proportion using randomized response technique, *Model Assisted Statistics and Applications*, 1, 125-130.

Appendices

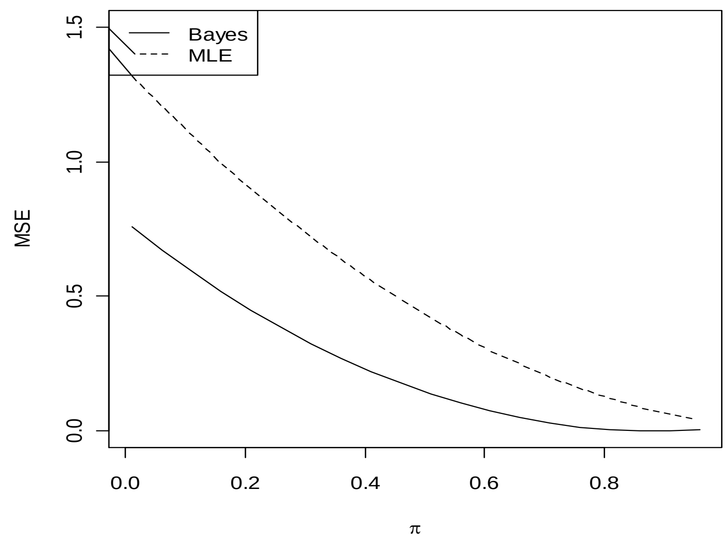

Figure 1: Graph of MSEs of $\hat{\pi}_{ML}$ and $\hat{\pi}_{Bayes}$ for $n = 25$, $P = 0.6$ and $T = 0.7$

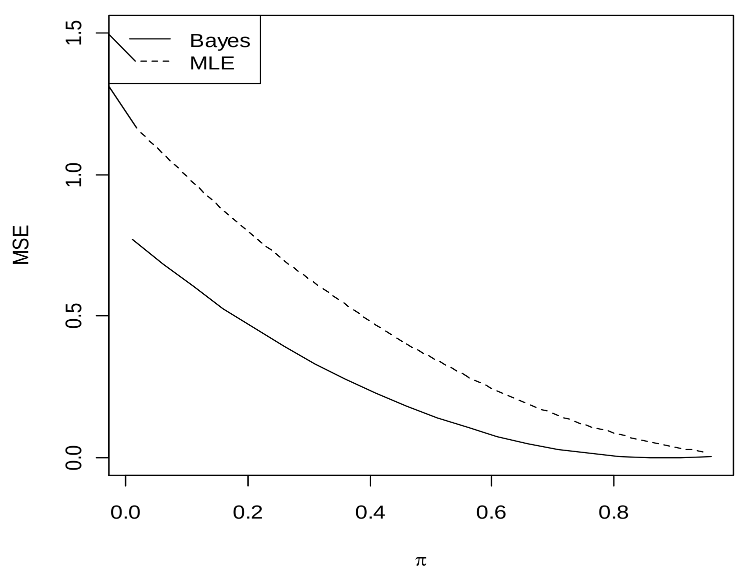

Figure 2: Graph of MSEs of $\hat{\pi}_{ML}$ and $\hat{\pi}_{Bayes}$ for $n = 25$, $P = 0.6$ and $T = 0.8$

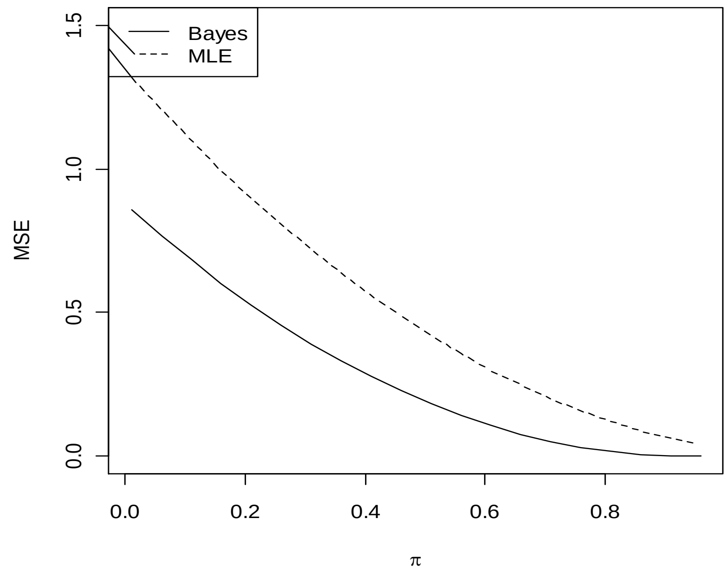

Figure 3: Graph of MSEs of $\hat{\pi}_{ML}$ and $\hat{\pi}_{Bayes}$ for $n = 50$, $P = 0.6$ and $T = 0.7$

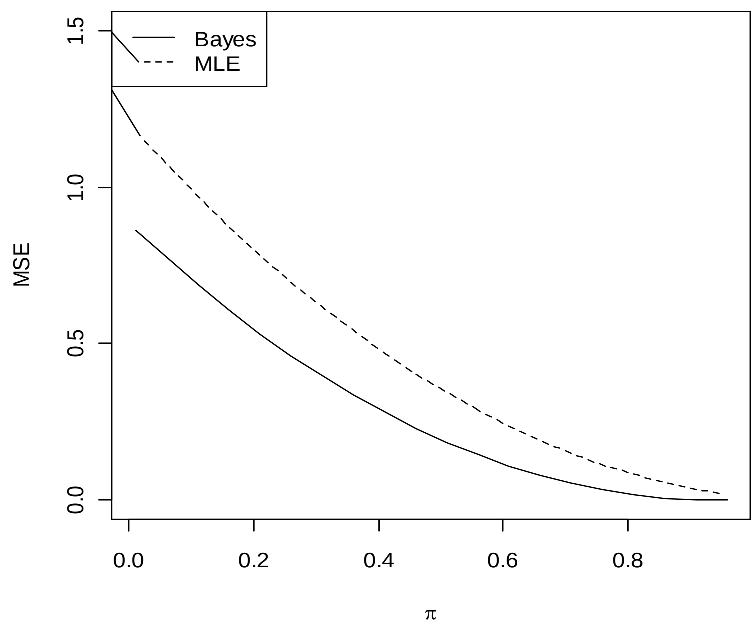

Figure 4: Graph of MSEs of $\hat{\pi}_{ML}$ and $\hat{\pi}_{Bayes}$ for $n = 50$, $P = 0.6$ and $T = 0.8$