An ameliorated two-stage randomized response model for estimating a rare stigmatized characteristic using Poisson distribution

Garib Nath Singh , Surbhi Suman *

Dept. of Applied mathematics, Indian Institute of Technology (Indian School of Mines)

Dhanbad, Jharkhand, India 826004.

Corresponding author: surbhi.iitism@yahoo.com

Abstract

The present work sheds light on the estimation procedure of a mean number of persons in the population bearing a rare sensitive characteristic using the Poisson distribution. A modified two-stage randomized response model for the rare sensitive characteristic is used to acquire the truthful response. Subsequently unbiased estimators are proposed for two situations when the information on another supplementary rare non-sensitive characteristic is known as well as unknown. The variances of the proposed estimators and their estimates are derived. Empirical studies are executed to show the dominance of the proposed estimators over some contemporary estimators.

Keywords: Poisson distribution; randomized response; rare sensitive attribute; unrelated attribute; variance.

AMS Subject Classification: 62D05

1. Introduction

The problem of non-response occurs due to several factors in sample surveys. The sensitive (stigmatized) nature of a characteristic under study is one of the most likely reasons for inviting a non-response or misleading response in the survey data. Such data reduce the size of the desired sample and produce the bias estimates. The problem of non-response due to stigmatized nature of characteristics was addressed by Warner (1965), who introduced a randomized response technique for collecting the responses from interviewees selected in the sample. The randomized response technique was further improved by Greenberg *et al.* (1969), Mangat *et al.* (1992), Singh *et al.* (1994), Singh *et al.* (2003), Chaudhuri *et al.* (2016), Tarray & Singh (2017) among others. Mangat & Singh (1990) and Mangat (1992) introduced two-stage related and unrelated randomized response techniques which substantially improved the performance of resultant estimators over the Warner (1965) estimator.

Land *et al.* (2012) suggested an estimation procedure for the mean number of persons in the population bearing a rare sensitive characteristic. A large sample is required to be drawn from the population for estimating the parameter of a rare sensitive characteristic. Such situations also validate the use of the Poisson probability distribution in developing a suitable estimation procedure. Motivated with these arguments, Singh and Tarray (2014; 2017), Singh *et al.* (2018) suggested randomized response models and estimation procedures for the similar problems.

In follow up of the previous works, the present study introduces a modified two-stage unrelated randomized response model and estimation procedures for mean number of persons in the population who possess a rare sensitive attribute. The proposed model under the Poisson approximation is an improved version of Mangat (1992) and Singh *et al.* (1994) models, and the resultant estimation procedures were more accurate than Land *et al.* (2012), Singh & Tarray (2014; 2017) estimators. The properties of the suggested estimation procedures have been examined for the cases of known and unknown unrelated rare non-sensitive attribute.

2. The proposed estimation procedure when the proportion of an unrelated rare non-sensitive attribute in the population is known

Consider a finite population Ω of size N, in which some of the individuals possess a rare sensitive attribute A. Let π_a and π_b be the true proportions of the rare sensitive attribute A and unrelated rare non-sensitive attribute B in the population, respectively. To estimate the mean number of persons who possess the rare sensitive attribute in the population, a large sample of size n is drawn using simple random sampling with replacement scheme (SRSWR) such that for small π_a and π_b(i.e. $\pi_a \to 0$ and $\pi_b \to 0$), we have and for the large sample size n.

When the proportion π_b of an unrelated rare non-sensitive attribute is known, each individual selected in the sample was given two randomization devices (R_1 ,R_2) and requested to report his/ her response as per the following outcomes of the devices:

The first stage randomization device R_1 consists of the following statements:

(i) "I possess the rare sensitive attribute A" with probability U, and

(ii) "Go to the randomization device R_2" with probability $1-U$.

The second stage randomization device R_2 consists of the following statements:

(i) "I possess the rare sensitive attribute A" with probability P_1.

(ii) "I possess the rare non-sensitive attribute B " with probability P_2 .

(iii) Draw one more card with probability P_3 where $P_1+P_2+P_3=1$, . If the statement (iii) is selected by the respondent, then it is required to repeat the process without replacing the card. In the second draw, if the statement (iii) reappeared, then the respondent has to report "No".

Using the above randomization devices, the probability of obtaining answer "yes" from the respondent is given by

Using the above randomization devices, the probability of obtaining answer "yes" from the respondent is given by

$$\zeta_0 = U\pi_a + (1-U)\left(P_1\pi_a + P_2\pi_b\right)\left(1+P_3\frac{k}{k-1}\right) \qquad (1)$$

where k is the total number of cards in the randomization device R_2. Since the attributes A and B under study are rare in the population, therefore, we assume, for large and small ζ_0 i.e. $\zeta_0 \to 0$, we have $n\zeta_0 = \lambda_0 > 0$.

Let x_1, x_2,......, x_n be a random sample from the the Poisson distribution with parameter λ_0 where

$$\lambda_0 = U\lambda_a + (1-U)\left(P_1\lambda_a + P_2\lambda_b\right)\left(1+P_3\frac{k}{k-1}\right). \qquad (2)$$

The likelihood function of the random sample of n observations is written as

$$L = \prod_{i=1}^{n}\frac{e^{-\lambda_0}\lambda_0^{x_i}}{x_i!}. \qquad (3)$$

Taking the logarithm on equation (3), substituting the value of λ_0 from the equation (2), and maximizing with respect to the parameter λ_0, the maximum-likelihood estimator of the mean number of persons in the population possessing the sensitive characteristic A has been derived as

$$\hat{\lambda}_a = \frac{1}{U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)}\left[\frac{1}{n}\sum_{i=1}^{n}x_i - (1-U)P_2\left(1+P_3\frac{k}{k-1}\right)\lambda_b\right]. \qquad (4)$$

2.1 Properties of the proposed estimator

The properties of the proposed estimator $\hat{\lambda}_a$ are summarized in the following theorems:

Theorem 2.1. The suggested estimator $\hat{\lambda}_a$ is unbiased for the parameter .

Proof. Since $x_i \sim P(\lambda_0) \Rightarrow E(x_i)=\lambda_0$, , using this result, it may be easily proven that , $E(\hat{\lambda}_a)=\lambda_a$.

Theorem 2.2. The variance of the proposed estimator $\hat{\lambda}_a$ is given by

$$V(\hat{\lambda}_a) = \frac{\lambda_a}{n\left[U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)\right]} + \frac{(1-U)P_2\left(1+P_3\frac{k}{k-1}\right)\lambda_b}{n\left[U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)\right]^2} \qquad (5)$$

Proof. The variance of the proposed estimator $\hat{\lambda}_a$ is derived as

$$V(\hat{\lambda}_a) = \frac{\frac{1}{n^2}\sum_{i=1}^{n}V(x_i)}{\left[U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)\right]^2}.$$

This is because $x_i \sim P(\lambda_0) \Rightarrow V(x_i)=\lambda_0$.

Utilizing this result, then substituting the value of λ_0, and performing some algebraic simplification, we obtain the expression for the variance of the estimator $\hat{\lambda}_a$ as given in equation (5).

Theorem 2.3. The unbiased estimate of the variance $V(\hat{\lambda}_a)$ is given by

$$\hat{V}(\hat{\lambda}_a) = \frac{\sum_{i=1}^{n}x_i}{n^2\left[U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)\right]^2}. \qquad (6)$$

Proof. It may be seen that

$$E[\hat{V}(\hat{\lambda}_a)] = \frac{\sum_{i=1}^{n}\lambda_0}{n^2\left[U+(1-U)P_1\left(1+P_3\frac{k}{k-1}\right)\right]^2}.$$

Putting the value of λ_0 and performing some algebraic simplification, we have $E[\hat{V}(\hat{\lambda}_a)] = V(\hat{\lambda}_a)$.

3. The proposed estimation procedure when the proportion of an unrelated rare non-sensitive attribute in the population is unknown

When the true proportion π_b of an unrelated rare non-sensitive attribute B in the population is unknown, each respondent selected in the sample has provided two sets of randomization devices (R_{11}, R_{12}) and (R_{21} ,R_{22}) where each set of the randomization device consists of the similar statements with different probabilities as described in Section 2. Initially, respondents were provided the first set of randomization devices (R_{11}, R_{12}) for their use in two-stages. The randomization device R_{11} used in the first stage consists of the following statements:

(i) "I possess the rare sensitive attribute A " with probability U_1 , and

(ii) "Go to the randomization device " with probability $1-U_1$.

The randomization device R_{12} to be used in second stage consists of the following statements:

(i) "I possess the rare sensitive attribute A " with probability P_1.

(ii) "I possess the rare non-sensitive attribute B " with probability P_2 .

(iii) Draw one more card with probability P_3 .

If the statement (iii) is selected by the respondent, then it is required to repeat the process without replacing the card. On the second draw, if the statement (iii) reappeared, the respondent must report "No". Again, the respondents were provided with the second set of randomization devices (R_{21} ,R_{22}) for their use in two-stages. The randomization device R_{21} to be used in first stage consists of the following statements:

(i) "I possess the rare sensitive attribute "with probability U_2 , and

(ii) "Go to the randomization device R_{22} " with probability $1-U_2$.

The randomization device R_{22} to be used in second stage consists of the following statements:

(i) "I possess the rare sensitive attribute A " with probability Q_1 .

(ii) "I possess the rare non- sensitive attribute B" with probability Q_2 .

(iii) Draw one more card with probability Q_3, where $(P_1+P_2+P_3=1)$ and $(Q_1+Q_2+Q_3=1)$. If the statement (iii) is selected by the respondent, the process will remain as in the previous case. The probabilities of getting a yes-answer from the respondent using the above randomization response devices are

$$\zeta_1 = U_1\pi_a + (1-U_1)\left(P_1\pi_a + P_2\pi_b\right)\left(1+P_3\frac{k}{k-1}\right)$$

and

$$\zeta_2 = U_2\pi_a + (1-U_2)\left(Q_1\pi_a + Q_2\pi_b\right)\left(1+Q_3\frac{k}{k-1}\right).$$

For large n, as $\zeta_1 \to 0$ and $\zeta_2 \to 0$, we have $n\zeta_1 = \lambda_a^* > 0$ and $n\zeta_2 = \lambda_b^* > 0$. Let

$x_{11}, x_{12}, ..., x_{1n}$ and $x_{21}, x_{22}, ..., x_{2n}$

be the random samples from poisson distribution with aprameter λ_a^* and λ_b^* respectively. Proceeding in the similar way as described in Section 2, we have

$$\frac{1}{n}\sum_{i=1}^{n} x_{1i} = U_1\hat{\lambda}_a + (1-U_1)\left(P_1\hat{\lambda}_a + P_2\hat{\lambda}_b\right)\left(1+P_3\frac{k}{k-1}\right) \tag{7}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_{2i} = U_2\hat{\lambda}_a + (1-U_2)\left(Q_1\hat{\lambda}_a + Q_2\hat{\lambda}_b\right)\left(1+Q_3\frac{k}{k-1}\right). \tag{8}$$

Solving equations (7) and (8), the estimators of the mean number of persons in the population who possess a rare sensitive attribute A and a non-sensitive attribute B , respectively, are derived as

$$\hat{\lambda}_{au} = \frac{1}{n\Delta_1}\begin{bmatrix} Q_2(1-U_2)\left(1+Q_3\frac{k}{k-1}\right)\sum_{i=1}^{n} x_{1i} \\ -(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\sum_{i=1}^{n} x_{2i} \end{bmatrix} \tag{9}$$

and

$$\hat{\lambda}_{bu} = \frac{1}{n\Delta_2}\begin{bmatrix} U_2+(1-U_2)Q_1\left(1+Q_3\frac{k}{k-1}\right)\sum_{i=1}^{n} x_{1i} \\ -(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\sum_{i=1}^{n} x_{2i} \end{bmatrix} \tag{10}$$

where, $\Delta_1 = \beta_2\alpha_1 - \beta_1\alpha_2$, $\Delta_2 = \beta_1\alpha_2 - \beta_2\alpha_1$,

$$\alpha_1 = U_1 + (1-U_1)P\left(1+P_3\frac{k}{k-1}\right),$$

$$\alpha_2 = (1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right),$$

$$\beta_1 = U_2 + (1-U_2)Q_1\left(1+Q_3\frac{k}{k-1}\right),$$

and $$\beta_2 = Q_2(1-U_2)\left(1+Q_3\frac{k}{k-1}\right).$$

3.1 Properties of the proposed estimators $\hat{\lambda}_{au}$ and $\hat{\lambda}_{bu}$

The properties of estimators $\hat{\lambda}_{au}$ and $\hat{\lambda}_{bu}$ are given in the following theorems:

Theorem 3.1. The proposed estimators $\hat{\lambda}_{au}$ and $\hat{\lambda}_{bu}$ are unbiased for parameters λ_a and λ_b , respectively.

Proof. This property is the consequence of the results $E(x_{1i}) = \lambda_a^*$ and $E(x_{2i}) = \lambda_b^*$.

Theorem 3.2. The variances of the proposed estimators $\hat{\lambda}_{au}$ and $\hat{\lambda}_{bu}$ are

$$V(\hat{\lambda}_{au}) = \frac{1}{n\Delta_1^2}\begin{bmatrix}(\beta_2^2\alpha_1 + \alpha_2^2\beta_1 - 2\alpha_1\alpha_2\beta_1\beta_2)\lambda_a + \\ (\alpha_2\beta_2^2 + \alpha_2^2\beta_2 - 2\alpha_2^2\beta_2^2))\lambda_b\end{bmatrix} \tag{11}$$

and

$$V(\hat{\lambda}_{bu}) = \frac{1}{n\Delta_2^2}\begin{bmatrix}(\beta_1^2\alpha_1 + \alpha_1^2\beta_1 - 2\alpha_1^2\beta_1^2)\lambda_a + \\ (\alpha_2\beta_1^2 + \alpha_1^2\beta_2 - 2\alpha_1\alpha_2\beta_1\beta_2))\lambda_b\end{bmatrix} \tag{12}$$

Proof. The variance of the proposed estimator $\hat{\lambda}_{au}$ is derived as

$$V(\hat{\lambda}_{au}) = \frac{1}{n^2\Delta_1^2}\begin{bmatrix}\left\{Q_2(1-U_2)\left(1+Q_3\frac{k}{k-1}\right)\right\}^2\sum_{i=1}^{n}V(x_{1i}) \\ +\left\{(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\right\}^2\sum_{i=1}^{n}V(x_{2i}) \\ -P_2Q_2(1-U_2)(1-U_1)\left(1+Q_3\frac{k}{k-1}\right) \\ \left(1+P_3\frac{k}{k-1}\right)\sum_{i=1}^{n}Cov(x_{1i},x_{2i})\end{bmatrix}$$

$$= \frac{1}{n\Delta_1^2}\begin{bmatrix}\left\{Q_2(1-U_2)\left(1+Q_3\frac{k}{k-1}\right)\right\}^2\lambda_a^* \\ +\left\{(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\right\}^2\lambda_b^* \\ -P_2Q_2(1-U_2)(1-U_1) \\ \left(1+Q_3\frac{k}{k-1}\right)\left(1+P_3\frac{k}{k-1}\right)\lambda_{ab}^*\end{bmatrix} \tag{13}$$

where

$$\lambda_a^* = v(x_{1i}) = \begin{bmatrix}\left\{U_1 + (1-U_1)P_1\left(1+P_3\frac{k}{k-1}\right)\right\}\lambda_a \\ +\left\{(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\right\}\lambda_b\end{bmatrix}$$

$$\lambda_b^* = v(x_{2i}) = \left\{U_2 + (1-U_2)Q_1\left(1+Q_3\frac{k}{k-1}\right)\right\}\lambda_a + \left\{(1-U_2)Q_2\left(1+Q_3\frac{k}{k-1}\right)\right\}\lambda_b$$

$$\lambda_{ab}^* = Cov(x_{1i}, x_{2i})$$

$$= \left\{U_1 + (1-U_1)P\left(1+P_3\frac{k}{k-1}\right)\right\}$$

$$\left\{U_2 + (1-U_2)Q_1\left(1+Q_3\frac{k}{k-1}\right)\right\}\lambda_a.$$

$$+\left\{(1-U_1)P_2\left(1+P_3\frac{k}{k-1}\right)\right\}$$

$$\left\{Q_2(1-U_2)\left(1+Q_3\frac{k}{k-1}\right)\right\}\lambda_b$$

Putting the values of λ_a^*, λ_b^* and λ_{ab}^* in equation (13), and after some algebraic simplifications, we get the expression for the variance of the estimator $\hat{\lambda}_{au}$ as given in equation (11). We may derive the expression for variance of $\hat{\lambda}_{bu}$ as given in equation (12). in the same way.

Lemma 3.1. The unbiased estimates of the variances $\hat{\lambda}_{au}$ and $\hat{\lambda}_{bu}$ are given by

$$\hat{V}(\hat{\lambda}_{au}) = \frac{1}{n\Delta_1^2}\begin{bmatrix}(\beta_2^2\alpha_1 + \alpha_2^2\beta_1 - 2\alpha_1\alpha_2\beta_1\beta_2)\hat{\lambda}_a \\ +(\alpha_2\beta_2^2 + \alpha_2^2\beta_2 - 2\alpha_2^2\beta_2^2))\hat{\lambda}_b\end{bmatrix} \tag{14}$$

and

$$\hat{V}(\hat{\lambda}_{bu}) = \frac{1}{n\Delta_2^2}\left[(\beta_1^2\alpha_1 + \alpha_1^2\beta_1 - 2\alpha_1^2\beta_1^2)\hat{\lambda}_a + (\alpha_2\beta_1^2 + \alpha_1^2\beta_2 - 2\alpha_1\alpha_2\beta_1\beta_2))\hat{\lambda}_b\right] \tag{15}$$

4. Confidentiality protection

To perform the randomized response interviews on a rare sensitive attribute, we are generally concerned about the statistical properties of the suggested estimation procedures, while at the same time, in such cases, the privacy (confidentiality) protection of respondents is also an equally important issue. It is important to note that previous studies regarding rare sensitive characteristics do not truly achieve a significant degree of privacy protection/confidentiality for the respondents. This section, an attempt has been made to quantify this for the proposed randomized response model of a rare sensitive attribute. Leysieffer & Warner (1976) suggested a measure of privacy protection for the respondents when the survey involves the sensitive attribute. Following the work of Leysieffer & Warner (1976), a measure of respondent privacy protection for the proposed randomized response model was suggested as

$$g(Y/A) = \frac{P(Y/A)}{P(Y/A')}$$

$$= \frac{\left[\begin{array}{l} U+(1-U)P_1\left(1+P_3\dfrac{k}{k-1}\right) \\ +(1-U)P_2\left(1+P_3\dfrac{k}{k-1}\right)\lambda_b \end{array}\right]}{(1-U)P_2\left(1+P_3\dfrac{k}{k-1}\right)\lambda_b} \tag{16}$$

and

$$g(N/A') = \frac{P(N/A')}{P(N/A)} =$$

$$\frac{1-(1-U)P_2\left(1+P_3\dfrac{k}{k-1}\right)\lambda_b}{(1-U)\left\{1-P_1\left(1+P_3\dfrac{k}{k-1}\right)-P_2\left(1+P_3\dfrac{k}{k-1}\right)\lambda_b\right\}}. \tag{17}$$

Remark 4.1 It has been found that the above probability statements are admissible when the unrelated non-sensitive attribute is highly rare in the population, specifically $0 \le \lambda_b \le 1$.

5. Empirical comparisons

Generally, empirical comparison between two randomized response strategies is performed based on the variances or efficiencies of the resultant estimators. The researchers did not pay attention to the degree of privacy protection offered to the interviewees when a characteristic under study was related to the rare sensitive attribute. In this section, an attempt has been made for a comparison in terms of variance as well as privacy protection through empirical studies.

5.1 Comparison in terms of percent relative efficiency

To perform the empirical comparison, the percent relative efficiencies of the proposed estimators $\hat{\lambda}_a$ and $\hat{\lambda}_{au}$ are obtained for two different cases when the proportion of an unrelated rare attribute *B* (i) is known and (ii) is unknown. The percent relative efficiencies of the estimator $\hat{\lambda}_a$ with respect to $(\hat{\lambda}_1)_{t1}$ and $(\hat{\lambda}_1)_{t2}$ are defined as

$$E_{11} = \frac{V[(\hat{\lambda}_1)_{t1}]}{V[\hat{\lambda}_a]} \times 100 \text{ and } E_{12} = \frac{V[(\hat{\lambda}_1)_{t2}]}{V[\hat{\lambda}_a]} \times 100.$$

The percent relative efficiencies of the estimator $\hat{\lambda}_{au}$ with respect to $(\hat{\lambda}_{1u})_{t1}$ and $(\hat{\lambda}_{1u})_{t2}$ are defined as

$$E_{21} = \frac{V[(\hat{\lambda}_{1u})_{t1}]}{V[\hat{\lambda}_{au}]} \times 100 \text{ and } E_{22} = \frac{V[(\hat{\lambda}_{1u})_{t2}]}{V[\hat{\lambda}_{au}]} \times 100.$$

where $\left\{(\hat{\lambda}_1)_{t1}, (\hat{\lambda}_1)_{t2}\right\}$ and $\left\{(\hat{\lambda}_{1u})_{t1}, (\hat{\lambda}_{1u})_{t2}\right\}$ are the estimators based on the randomized response models proposed by Singh and Tarray (2014, 2017) for the cases (i) and (ii), respectively.

For fixed values of $k=100$ and $n=100$, the percent relative efficiencies E_{11}, E_{12}, E_{21} and E_{22} are calculated for different choices of

$\lambda_a = (0.5, 0.1, 1.5)$, $\lambda_b = (0.5, 0.1, 1.5)$,

(P_1, P_2, P_3) and (Q_1, Q_2, Q_3), where

$$P_3 = (1-P-P_2) \text{ and } Q_3 = (1-Q_1-Q_2)$$

Let the range of U be from 0.3 to 0.9 with a step of 0.2, and P_1 ranges from 0.4 to 0.8 with a step of 0.1. We also set $U_1=(0.5,0.7)$ and $U_2=(0.5,0.3)$.

The percent relative efficiencies E_{11}, E_{12}, E_{21} and E_{22} are calculated for all combinations of parametric choices but are only shown with respect to the most recent model suggested by Singh and Tarray (2017) (Tables 1 and 2). The variation in the percent relative efficiencies with respect to k can be observed for almost all combinations of parametric choices and in Figures 1 and 2 for a few parametric choices.

5.2 Comparison in terms of privacy protection

The measures of privacy protection are obtained for the $\hat{\lambda}_a$ with respect to $(\hat{\lambda}_1)_{t1}$ and $(\hat{\lambda}_1)_{t2}$ (Figures 3 and 4) for some parametric choices of (P_1, P_2, P_3).

The values of λ_b and U are selected in the range [0.1,1] and [0.3,0.6], respectively. L, L_1, L_2 and are the measures of privacy protection for respondents proposed, Singh and Tarray (2017) and Singh and Tarray (2014) models respectively, which are calculated as per the discussion in Section 4.

6. Interpretation of the results

(i) Data from the tables show that the calculated percent relative efficiencies exceed 100, which indicates that the proposed model and estimation procedures perform better than that of Singh and Tarray (2014; 2017). Since the model discussed by Singh and Tarray (2017) is better than the model discussed by Land *et al.* (2012), this research model is also better than Land's *et al.* (2012).

(ii) From the results, substantial gain is observed for the smaller values of λ_a and larger values of λ_b.

(iii) From Table 1, it is observed that the values of percent relative efficiencies increas as the values of U decrease.

(iv) From Table 2, it is visible that the values of

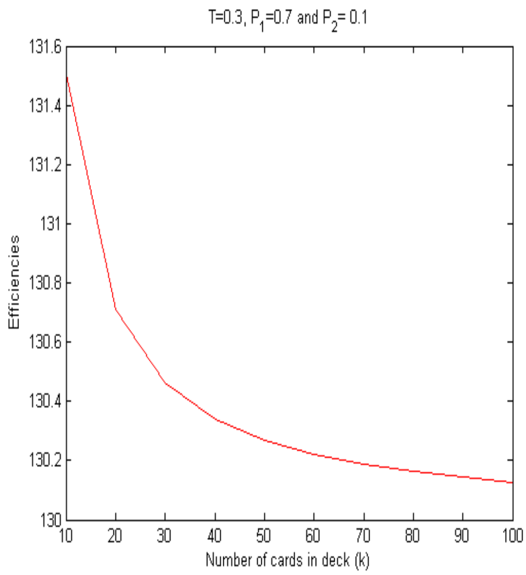

Fig. 1. Percent relative efficiencies with respect to k for T=0.3, P_1=0.7 and P_2=0.1.

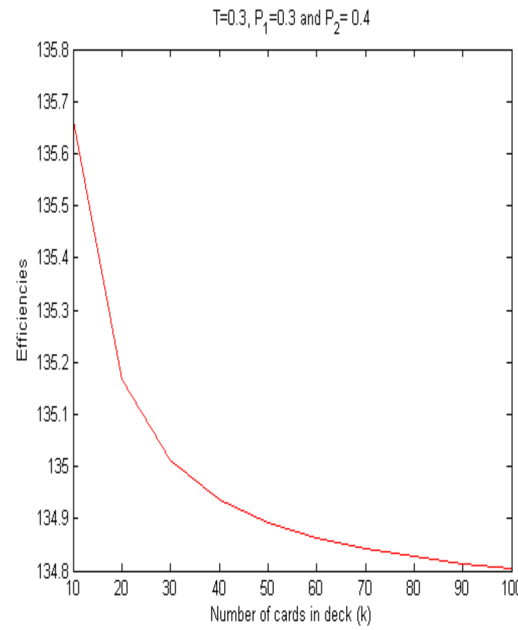

Fig. 2. Percent relative efficiencies with respect to k for T=0.3, P_1=0.4 and P_2=0.5.

Table 1. Percentage relative efficiencies of the estimator $\hat{\lambda}_a$ with respect to the estimator $\left(\hat{\lambda}_1\right)_{t2}$

		P_1	0.4		0.5		0.6		0.7		0.8
		P_2	0.5	0.1	0.4	0.2	0.2	0.1	0.2	0.1	0.1
		P_3	0.2	0.5	0.1	0.3	0.2	0.3	0.1	0.2	0.1
U	λ_a	λ_b									
0.3	0.5	0.5	110.69	187.12	111.89	144.16	128.41	147.56	113.65	130.13	114.31
		1	113.26	235.68	115.28	162.95	140.13	172.14	119.04	144.92	120.94
		1.5	114.71	274.38	117.37	176.47	148.84	192.72	123.16	157.54	126.67
	1	0.5	108.49	157.89	109.25	131.74	120.88	133.38	110.27	121.73	110.60
		1	110.69	187.12	111.89	144.16	128.41	147.56	113.65	130.13	114.31
		1.5	112.19	212.86	113.81	154.39	134.74	160.42	116.54	137.83	117.75
	1.5	0.5	107.49	147.24	108.15	126.98	118.05	128.33	109.01	118.75	109.30
		1	109.33	168.07	110.23	136.17	123.54	138.26	111.45	124.61	111.87
		1.5	110.69	187.12	111.89	144.16	128.41	147.56	113.65	130.13	114.31
0.5	0.5	0.5	106.34	149.43	107.34	126.79	117.99	129.72	108.99	119.62	109.68
		1	108.30	178.90	109.87	139.48	126.13	145.73	112.85	129.63	114.32
		1.5	109.56	204.05	111.62	149.45	132.67	159.82	116.02	138.53	118.51
	1	0.5	104.89	132.72	105.60	119.06	113.12	120.86	106.71	114.15	107.16
		1	106.34	149.43	107.34	126.79	117.99	129.72	108.99	119.62	109.68
		1.5	107.44	164.77	108.73	133.54	122.30	137.99	111.02	124.77	112.06
	1.5	0.5	104.30	126.81	104.92	116.22	111.36	117.77	105.90	112.25	106.29
		1	105.42	138.45	106.22	121.76	114.81	123.88	107.50	116.01	108.01
		1.5	106.34	149.43	107.34	126.79	117.99	129.72	108.99	119.62	109.68
0.7	0.5	0.5	103.26	124.58	103.88	113.99	109.70	115.83	105.00	110.82	105.51
		1	104.54	140.16	105.49	121.34	114.51	124.73	107.34	116.54	108.26
		1.5	105.49	154.37	106.76	127.68	118.70	132.96	109.41	121.87	110.83
	1	0.5	102.45	116.20	102.91	109.85	107.02	111.12	103.71	107.80	104.07
		1	103.26	124.58	103.88	113.99	109.70	115.83	105.00	110.82	105.51
		1.5	103.95	132.56	104.73	117.81	112.19	120.37	106.21	113.73	106.91
	1.5	0.5	102.15	113.31	102.55	108.39	106.08	109.50	103.26	106.77	103.58
		1	102.74	119.04	103.25	111.27	107.94	112.71	104.15	108.82	104.56
		1.5	103.26	124.58	103.88	113.99	109.70	115.83	105.00	110.82	105.51
0.9	0.5	0.5	100.95	106.99	101.16	104.13	102.94	104.74	101.56	103.33	101.75

	1	101.42	111.69	101.73	106.54	104.53	107.52	102.35	105.16	102.65
	1.5	101.84	116.24	102.26	108.82	106.05	110.23	103.10	106.95	103.54
1	0.5	100.70	104.59	100.85	102.88	102.11	103.33	101.15	102.40	101.29
	1	100.95	106.99	101.16	104.13	102.94	104.74	101.56	103.33	101.75
	1.5	101.19	109.36	101.45	105.35	103.74	106.14	101.96	104.25	102.20
1.5	0.5	100.61	103.78	100.75	102.46	101.83	102.85	101.01	102.09	101.13
	1	100.79	105.39	100.96	103.30	102.39	103.80	101.28	102.71	101.44
	1.5	100.95	106.99	101.16	104.13	102.94	104.74	101.56	103.33	101.75

Table 2. Percent relative efficiencies of the estimator $\hat{\lambda}_{au}$ with respect to the estimator $\left(\hat{\lambda}_{1u}\right)_{t2}$

			P_1	0.6	0.6	0.7	0.7	0.8	0.8
			P_2	0.2	0.2	0.15	0.15	0.1	0.1
U_1	U_2	λ_a	P_3	0.2	0.2	0.15	0.15	0.1	0.1
			Q_1	0.1	0.4	0.1	0.4	0.1	0.4
		λ_b	Q_2	0.45	0.3	0.45	0.3	0.45	0.3
			Q_3	0.45	0.3	0.45	0.3	0.45	0.3
			0.5	120.22	146.98	113.22	124.50	107.72	111.59
		0.5	1	128.55	165.77	119.06	135.10	111.32	116.94
			1.5	135.71	181.98	124.31	144.64	114.68	121.94
			0.5	115.52	136.41	110.05	118.74	105.83	108.78
	0.5	1	1	120.22	146.98	113.22	124.50	107.72	111.59
			1.5	124.55	156.74	116.22	129.94	109.55	114.31
			0.5	113.87	132.68	108.95	116.74	105.18	107.82
0.7		1.5	1	117.14	140.03	111.13	120.69	106.46	109.73
			1.5	120.22	146.98	113.22	124.50	107.72	111.59
			0.5	110.06	123.32	108.20	114.64	105.76	108.24
		0.5	1	114.41	132.88	111.91	121.09	108.47	112.08
			1.5	118.12	141.11	115.22	126.88	110.99	115.66
	0.3		0.5	107.60	117.94	106.18	111.14	104.33	106.22
		1	1	110.06	123.32	108.20	114.64	105.76	108.24
			1.5	112.33	128.29	110.11	117.95	107.13	110.19
			0.5	106.73	116.04	105.47	109.93	103.84	105.53
		1.5	1	108.44	119.78	106.86	112.33	104.81	106.90
			1.5	110.06	123.32	108.20	114.64	105.76	108.24
			0.5	161.97	388.75	133.94	193.98	116.95	132.36
		0.5	1	181.81	480.28	146.73	228.94	124.18	146.02
			1.5	197.15	551.64	157.39	258.20	130.63	158.23
			0.5	149.68	332.54	126.53	173.82	112.99	124.89
	0.5	1	1	161.97	388.75	133.94	193.98	116.95	132.36
			1.5	172.57	437.54	140.64	212.27	120.67	139.39
			0.5	145.14	311.86	123.88	166.63	111.62	122.30
		1.5	1	153.99	352.21	129.09	180.76	114.34	127.43
			1.5	161.97	388.75	133.94	193.98	116.95	132.36
0.5			0.5	119.48	178.17	115.46	139.92	110.45	118.80
		0.5	1	126.00	203.26	121.42	154.95	114.96	126.80
			1.5	130.96	222.73	126.34	167.50	118.96	133.95
			0.5	115.39	162.70	111.98	131.23	107.98	114.42
	0.3	1	1	119.48	178.17	115.46	139.92	110.45	118.80
			1.5	122.98	191.56	118.59	147.79	112.78	122.92
			0.5	113.86	156.99	110.73	128.13	107.12	112.90
		1.5	1	116.83	168.12	113.18	134.22	108.82	115.91
			1.5	119.48	178.17	115.46	139.92	110.45	118.80

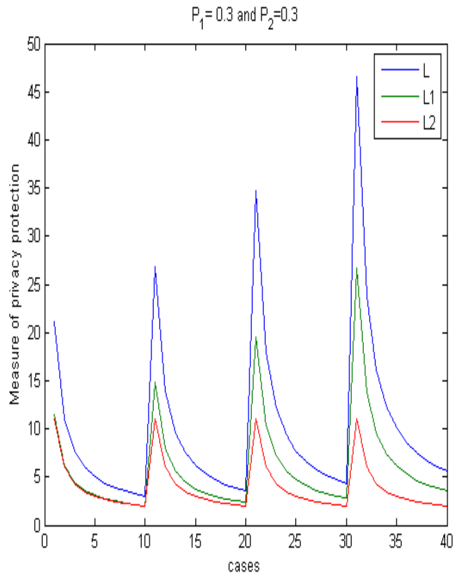

Fig. 3. Measures of privacy protection of different models

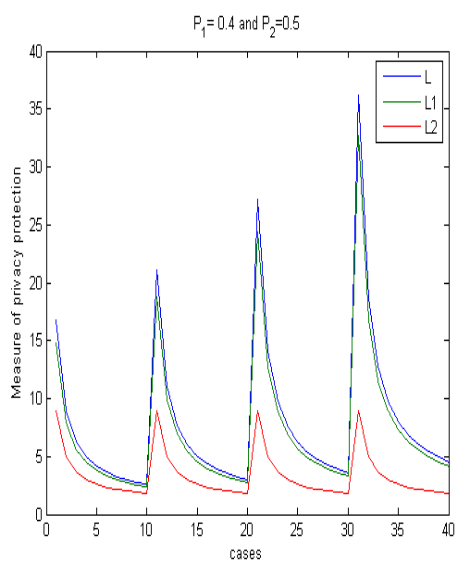

Fig. 4. Measures of privacy protection of different models

percent relative efficiencies increase and decrease for decreasing values of U_1 and U_2 , respectively. (v) From Figures 1 and 2, the values of percent relative efficiencies increase as the values of k decrease. (vi) Figures 3 and 4 clearly show that the level of privacy protection of the proposed randomized response model is better than the contemporary randomized response models of Singh and Tarray (2014; 2017).

7. Conclusions and recommendations

The proposed two-stage randomized response model and subsequent estimation procedures have shown exceptional performance in comparison to similar types of models and estimation procedures. It is worthwhile to mention that the suggested model and resultant estimators are much more efficient than the contemporary estimators, and at the same time, provide more safeguards in terms of privacy protection. The theoretical and empirical results also reveal that the proposed model is more adequate in terms of ascertaining truthful responses from respondents, and subsequent estimation procedures are more effective at estimating the mean number of persons in the population who possess a rare sensitive attribute.

ACKNOWLEDGMENTS

The authors are thankful to the Indian Institute of Technology Dhanbad (Indian School of Mines) for providing the financial support and infrastructural support need to carry out the present research. The researchers are also thankful to the reviewers for their valuable suggestions which improved the quality of the paper.

References

Chaudhuri, A., Christofides, T.C. & Rao, C.R. (2016). Handbook of statistics 34. Data gathering, analysis and protection of privacy through randomized response techniques: Qualitative and quantitative human traits.

Greenberg, B.G., Abul-Ela, A.A., Simmins, W.R. & Horvitz, D.G. (1969). The unrelated question randomized response model: Theoretical framework. Journal of the American Statistical Association, **64**(326): 520-539.

Land, M., Singh, S. & Sedory, S.A. (2012). Estimation of a rare sensitive attribute using Poisson distribution. Statistics, **46**(3): 351-60.

Leysieffer, F.W. & Warner, S.L. (1976). Respondent jeopardy and optimal designs in randomized response models. Journal of the American Statistical Association, **71**(355): 649-656.

Mangat, N.S., Singh, R. & Singh, S. (1992). An improved unrelated question randomized response strategy. Calcutta Statistical Association Bulletin, **42**(3-4): 277-282.

Mangat, N.S. & Singh, R. (1990). An alternative randomized procedure. Biometrika, **77**: 439- 442.

Mangat, N.S. (1992). Two-stage randomized response sampling procedure using unrelated question. Journal of the Indian Society of Agricultural Statistics, **44**(1): 82-87.

Singh, G.N., Singh, C., Suman, S. & Kumar, A. (2018). A two-stage unrelated randomized response model for estimating a rare sensitive attribute in probability proportional to size sampling using Poisson distribution. Communications in Statistics-Theory and Methods, **47**(15): 3744-3766.

Singh, H.P. & Tarray, T.A. (2014). A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. Statistics and Probability Letters, **90**(1): 42–45.

Singh, H.P. & Tarray, T.A. (2017). A two- stage Land *et al*.'s randomized response model for estimating a rare sensitive attribute using Poisson distribution. Communications in Statistics-Theory and Methods, **46**(1): 389-405.

Singh, S., Horn, S., Singh, R. & 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.

Singh, S., Singh, R., Mangat, N.S. & Tracy, D.S. (1994). An alternative device for randomized responses. Statistica, **54**(2): 233-243.

Tarray, T.A. & Singh, H.P. (2017). A simple way of improving the Bar–Lev, Bobovitch and Boukai randomized response model. Kuwait Journal of Science, **44**(4): 83-90.

Warner, S. L. (1965) Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, **60**(309): 63-69.

Submission : 30/10/2017
Revision : 06/08/2018
Acceptance : 14/08/2018

نموذج استجابة عشوائية ثنائي الطور ومُحَسن لتقدير صفة نادرة باستخدام توزيع بواسون

غاريث ناث سينغ، سوربي سومان
قسم الرياضيات التطبيقية، المعهد الهندي للتكنولوجيا (المدرسة الهندية للمناجم)
دانباد، جهارخاند، الهند

الملخص

العمل الحالي يلقي الضوء على إجراء تقدير لمتوسط عدد الأشخاص من بين السكان الذين يحملون صفة نادرة حساسة باستخدام توزيع بواسون. وقد تم استخدام نموذج استجابة عشوائية ثنائي الطور ومُعدل لتلك الصفة الحساسة النادرة للحصول على الإجابة الصحيحة. وفيما بعد، تم اقتراح مقدرات غير متحيزة عندما تكون المعلومات المتوفرة عن صفة أخرى غير حساسة معروفة أو عندما تكون المعلومات غير معروفة. وقد تم استنتاج متغيرات المقدرات المقترحة، كما تم عمل دراسات تجريبية لتوضيح مدى هيمنة المقدرات المقترحة على بعض المقدرات المعاصرة.