5-1-2014

Stochastic Randomized Response Model for a Quantitative Sensitive Random Variable

Sarjinder Singh
Texas A&M University-Kingsville, sarjinder@yahoo.com

Stephen A. Sedory
Texas A&M University-Kingsville, stephen.sedory@tamak.edu

Stochastic Randomized Response Model for a Quantitative Sensitive Random Variable

Sarjinder Singh
Texas A&M University-Kingsville
Kingsville, TX

Stephen A. Sedory
Texas A&M University-Kingsville
Kingsville, TX

A new stochastic randomized response model is introduced that is useful for estimating the population mean of a sensitive quantitative variable. The proposed stochastic randomized response model is an extension of the stochastic randomized response model from a qualitative sensitive variable to a quantitative variable found in Singh (2002). The stochastic nature of a randomized response device helps increase a respondent's cooperation while collecting information on sensitive variables in a society. The Bar-Lev, Bobovitch, and Boukai (2004) model is shown to be a special case of the proposed model.

Keywords: Sensitive variable; estimation of population mean, stochastic randomized response device

Introduction

The collection of data through personal interview surveys on sensitive issues, such as induced abortion, drug abuse and family income, is a serious issue. For example, some questions are sensitive:

- By how much did you underreport your income on your 2009 tax return?
- Are you a Baath Party Member?
- How many abortions have you had?
- How many children have you molested?
- Do you use illegal drugs?

Randomized response techniques are one way to encourage people to answer truthfully. Warner (1965) considered the case where the respondents in a

---

Dr. Singh is an Associate Professor in the Department of Mathematics. Email him at: sarjinder@yahoo.com. Dr. Sedory is a Professor in the Department of Mathematics. Email him at stephen.sedory@tamuk.edu.

population can be divided into two mutually exclusive groups: one group with stigmatizing or otherwise sensitive characteristic *A*, and the other group without it. For estimating π, the proportion of respondents in the population belonging to the sensitive group *A*, a simple random sample of *n* respondents is selected with replacement from the population. For collecting information on the sensitive characteristic, Warner (1965) made use of a randomization device. One such device could be a deck of cards. On each card is written one of the following two statements: "I belong to group A", or "I do not belong to group A." The statements occur with relative frequencies p_0 and $(1-p_0)$ respectively in the deck of cards. Each respondent in the sample is asked to select a card at random from the well-shuffled deck. Without showing the card to the interviewer, the interviewee answers the question, "Is the statement true for you?" The number of people, n_1, who answer yes is binomially distributed with parameters $p_0\pi + (1-p_0)(1-\pi)$ and *n*. The maximum likelihood estimator of π exists for $p_0 \neq 0.5$ is given by

$$\hat{\pi}_w = \frac{(n_1/n)-(1-p_0)}{2p_0-1}.$$
(1)

The estimator is unbiased with variance:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p_0(1-p_0)}{n(2p_0-1)^2}.$$
(2)

In a randomized response procedure, the cooperation of respondents depends on the confidentiality of their responses – the greater the confidentiality, the greater the cooperation from the respondents. Conversely, if the magnitude of response confidentiality is increased, the efficiency of the estimator of population proportion π is adversely affected. It is necessary, therefore, to strike a balance between response confidentiality and estimator efficiency. Several researchers have tried to modify data collection procedures to increase the confidentiality of responses. Horvitz, et al. (1967) felt that by providing the respondent with the opportunity of replying to one of two questions in which one question is completely innocuous and unrelated to the sensitive attribute, the sense of confidentiality among the respondents could possibly be increased. The theoretical framework for their approach was developed by Greenberg, et al. (1969).

Singh (2002) considered another procedure that may result in a greater sense of response confidentiality among the sampled individuals. The procedure can be used in surveys where respondents selected in the sample assemble at common place for the conduction of the survey. This could be a situation of collecting data from a small town, community or organization. The procedure invokes K decks of cards (called a stochastic randomization device) with different proportions of cards carrying the statement, "I belong to group A." After explaining to the respondents how the randomization device provides confidentiality to their responses, the investigator asks one of the assembled respondents to randomly select a deck of cards from the box containing K decks of cards. The deck is then used to collect information on the sensitive attribute from the respondents. Every sampled respondent draws one card from the selected deck of cards and reads the statement on it. In the proposed procedure every respondent is provided with two identical slips of paper with yes or no printed on them. According to his status in relation to the statement printed on the card drawn, each respondent is requested to put one of the two slips of paper into an empty box. After the survey is completed, the number of yes answers is counted from the box and the proportion, p^*, for the deck used in the survey is noted. Random selection of one randomization device from several such devices may help in increasing the sense of confidentiality among the respondents. The choice of values of p for preparing K decks of cards for the survey is important in this procedure. These K values of p could either be purposively selected by an investigator, or they could be taken as a random sample from a known discrete or continuous density function. Let this density function be denoted by $f(p)$. The value of p corresponding to the deck used in the survey will be selected from this random sample of p-values with equal probabilities. Thus, the value of p^* used in the survey is a random variable with $f(p)$ as its probability density function. When $f(p)$ is a one-point distribution, the proposed procedure reduces to Warner (1965). Singh (2002) assumes let n_1 persons in the sample answered yes and $(n-n_1)$ answer no. Because the probability of a yes answer for a particular choice of p^* is given by

$$\theta = p^*\pi + (1-\pi)(1-p^*).$$
(3)

Singh (2002) considers the unbiased estimator of π as

$$\hat{\pi}_R = \frac{\hat{\theta} - (1-p^*)}{2p^* - 1},$$
(4)

382

where $\hat{\theta} = n_1/n$ is the proportion of yes answers in the sample, with variance

$$V(\hat{\pi}_R) = \frac{\pi(1-\pi)}{n} + \int_a^b \frac{p(1-p)}{n(2p-1)^2} f(p)dp \ldots a \le p \le b, \, p \ne 0.5 \qquad (5)$$

where $f(p)$ denotes the probability density function (p.d.f.) of p. Singh (2002) showed that the stochastic version of the Warner (1965) model remains more efficient than the pioneer Warner (1965) model. In the same article, Singh (2002) also considers stochastic version of the Kuk (1990) model and showed its benefits over the original Kuk (1990) model. Recent work on randomized response techniques is found in Abdelfatah, et al. (2013).

Quantitative Randomized Response Model

In the randomized response model due to Bar-Lev, Bobovitch, and Boukai (2004), hereafter the BBB model, the distribution of responses is given by:

$$Y_i = \begin{cases} X_i S & \text{with probability } (1-p) \\ X_i & \text{with probability } p \end{cases} \qquad (6)$$

In other words, each respondent is requested to rotate a spinner unobserved by the interviewer and if the spinner stops in the shaded area then he/she is requested to report the real response on the sensitive variable, for example X_i; and if the spinner stops in the non-shaded area then he/she is requested to report the scrambled response, for example X_iS, where S is any scrambling variable and its distribution is assumed to be known. In other words, $E(S) = \theta$ and $V(S) = \gamma^2$ are assumed to be known. Let p be the proportion of the shaded area of the spinner and $(1-p)$ be the non-shaded area of the spinner as shown in the Figure 1.
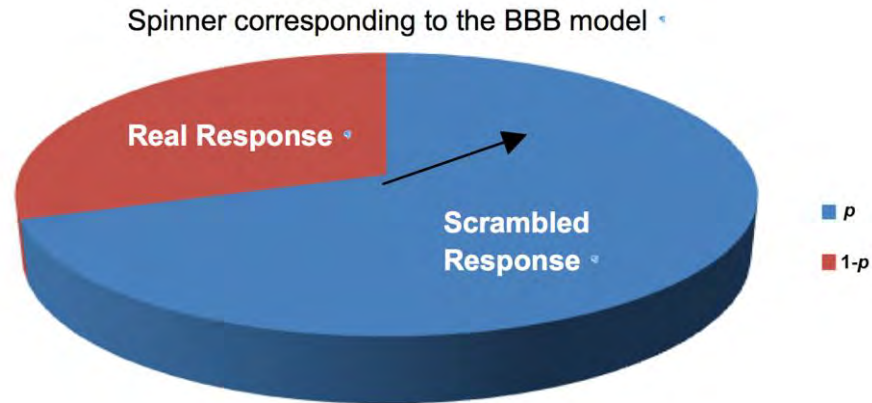
Figure 1. BBB Randomized Response Device.

An unbiased estimator of population mean μ_x is given by

$$\hat{\mu}_{X(BBB)} = \frac{1}{n\{(1-p)\theta + p\}} \sum_{i=1}^{n} Y_i \tag{7}$$

with the variance of the estimator $\hat{\mu}_{X(BBB)}$ given by

$$V\left[\hat{\mu}_{x(BBB)}\right] = \frac{1}{n}\left[\sigma_x^2 + \mu_x^2\left(1 + C_\gamma^2\right)C_s^2(p)\right] \tag{8}$$

where $C_s^2(p) = \dfrac{(1-p)\theta^2(1+C_\gamma^2) + p}{[(1-p)\theta + p]^2} - 1$, and $C_\gamma^2 = \gamma^2/\theta^2$.

Proposed Stochastic Quantitative Randomized Response Model

Let p^* be the stochastic proportion of cards in a deck bearing the statement, "Please report the real response X_i" and $(1-p^*)$ be the stochastic proportion of cards in the same deck bearing the statement, "Please report the scrambled

384

response $X_i S^*$, where S^* is also a stochastic scrambling variable. Let θ^* be the mean value, between pre-decided limits a and b, of a scrambling variable S^*. Again this procedure can be used in surveys where the respondents selected in the sample assemble at a common place to take a survey on a sensitive quantitative variable. This could be a situation of collecting data from a small town, community or organization, or a homogeneous stratum.

In practice, it is suggested that, at the gathering place, there is a collection of K_1 decks of pink cards in a box. Every pink deck of cards consists of two types of cards bearing the two statements (a) and (b) with stochastic proportions p^* and $(1-p^*)$ respectively. In another box, there are K_2 green decks of cards and each green deck can produce stochastic scrambling variable with different mean values of θ^* in the range $a < \theta^* < b$. In the presence of all the respondents and the interviewer, a lottery method is used. A huge number of pink decks of cards are left in box I, and a huge number of green decks of cards are left in box II. One green deck is selected and another pink deck is selected by the lottery method. The values of p^* and θ^* remain unknown during and after the survey. Both decks are either returned back to the boxes or are destroyed without looking at the particular values of p^* and θ^* used in the survey. This ensures respondents cooperation and privacy. The decks selected by the lottery method are used in the entire survey. Also note that the values of a and b are assumed to be known to the interviewer and interviewees. Thus, in the proposed stochastic randomized response model the distribution of the responses is given by

$$Y_i^* = \begin{cases} X_i S^* & \text{with probability } (1-p^*) \\ X_i & \text{with probability } p^* \end{cases} \tag{9}$$

The following theorems result.

Theorem 1

An unbiased estimator of the population mean μ_x is given by

$$\hat{\mu}_{S(BBB)} = \frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}} \sum_{i=1}^{n} Y_i^* \tag{10}$$

where the values of p^* and θ^* remains unknown to both the interviewer and interviewees, unlike Warner (1965) model. These values p^* and θ^* are derived from the known joint density of p^* and θ^* to get an estimate from the observed responses.

Proof. Let E_1 denote the expected value over all possible samples and E_2 denote the expected value over the randomization device for given values of θ^* and p^*. Taking expected value on both sides of (10), results in

$$E\left[\hat{\mu}_{S(BBB)}\right] = E\left[\frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\sum_{i=1}^{n}Y_i^*\right] = E_1 E_2\left[\frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\sum_{i=1}^{n}Y_i^* \mid p^*,\theta^*\right]$$

$$= E_1\left[\frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\sum_{i=1}^{n}E_2\left\{Y_i^* \mid p^*,\theta^*\right\}\right] = E_1\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \mu_x,$$

thus proving the theorem.

Theorem 2

The variance of the unbiased estimator $\hat{\mu}_{S(BBB)}$ of the population mean μ_x is given by

$$V\left(\hat{\mu}_{S(BBB)}\right) = \frac{1}{n}\left[\sigma_x^2 + \mu_x^2\left(1+C_x^2\right)\int_a^b\int_0^1\left\{\frac{p(1-p)(1-\theta)^2 + (1-p)\gamma^2}{\left\{p+(1-p)\theta\right\}^2}\right\}f(p,\theta)\,dpd\theta\right] \quad (11)$$

where $\gamma^2 = V(S^*)$ is constant and known.

Proof. Let V_1 be the variance over all possible samples and V_2 denote the variance for the given values of the randomization device p^* and θ^*. By the definition of the variance,

$$V\left(\hat{\mu}_{S(BBB)}\right) = E_1 V_2\left(\hat{\mu}_{S(BBB)} \mid p^*, \theta^*\right) + V_1 E_2\left(\hat{\mu}_{S(BBB)} \mid p^*, \theta^*\right) \tag{12}$$

$$= E_1 V_2\left[\frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\sum_{i=1}^{n} Y_i^* \mid p^*, \theta^*\right] + V_1 E_2\left[\frac{1}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\sum_{i=1}^{n} Y_i^* \mid p^*, \theta^*\right]$$

$$= E_1\left[\frac{1}{n^2\left\{\left(1-p^*\right)\theta^* + p^*\right\}^2}\sum_{i=1}^{n} V_2\left(Y_i^*\right)\right] + V_1\left[\frac{\sum_{i=1}^{n} E_2\left(Y_i^*\right)}{n\left\{\left(1-p^*\right)\theta^* + p^*\right\}}\right]$$

Now

$$V_2\left(Y_i^* \mid p^*, \theta^*\right) = E_2\left(Y_i^{*2}\right) - \left(E_2\left(Y_i^*\right)\right)^2$$

$$= p^* X_i^2 + \left(1-p^*\right)X_i^2 E_2\left(S^2\right) - \left[p^* X_i + \left(1-p^*\right)X_i\theta\right]^2$$

$$= p^* X_i^2 + \left(1-p^*\right)X_i^2\left(\gamma^2 + \theta^2\right)$$

$$\quad - \left[p^{*2} X_i^2 + \left(1-p^*\right)^2 X_i^2\theta^{*2} + 2p^*\left(1-p^*\right)X_i^2\theta^*\right]$$

$$= p^*\left(1-p^*\right)X_i^2 + \left(1-p^*\right)X_i^2\gamma^2 + X_i^2\theta^{*2}p^*\left(1-p^*\right)$$

$$\quad - 2p^*\left(1-p^*\right)X_i^2\theta^*$$

$$= p^*\left(1-p^*\right)X_i^2\left[1+\theta^{*2}-2\theta^*\right] + \left(1-p^*\right)X_i^2\gamma^2$$

$$= p^*\left(1-p^*\right)X_i^2\left(1-\theta^*\right)^2 + \left(1-p^*\right)X_i^2\gamma^2$$

$$= \left[p^*\left(1-p^*\right)\left(1-\theta^*\right)^2 + \left(1-p^*\right)\gamma^2\right]X_i^2 \tag{13}$$

Thus, plugging (13) into (12):

$$V\left(\hat{\mu}_{S(BBB)}\right) = E_1\left[\frac{p^*\left(1-p^*\right)\left(1-\theta^*\right)^2 + \left(1-p^*\right)\gamma^2\sum_{i=1}^{n} X_i^2}{n^2\left\{p^* + \left(1-p^*\right)\theta^*\right\}^2}\right] + V_1\left[\frac{\sum_{i=1}^{n}\left\{p^* + \left(1-p^*\right)\theta^*\right\}X_i}{n\left(p^* + \left(1-p^*\right)\theta^*\right)}\right]$$

$$= E_1\left[\frac{\left[p^*(1-p^*)(1-\theta^*)^2 + (1-p^*)\gamma^2\right]}{n\left\{p^* + (1-p^*)\theta^*\right\}^2}\right]\left[\frac{1}{N}\sum_{I=1}^{N} X_I^2\right] + V_1\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= E_1\left[\left|\frac{\left[p^*(1-p^*)(1-\theta^*)^2+(1-p^*)\gamma^2\right]}{n\left\{p^*+(1-p^*)\theta^*\right\}^2}\right|\left(\sigma_x^2+\mu_x^2\right)+\frac{\sigma_x^2}{n}\right]$$

$$=\frac{1}{n}\left[\sigma_x^2+\mu_x^2\left(1+C_x^2\right)E_1\left\{\frac{p^*(1-p^*)(1-\theta^*)^2+(1-p^*)\gamma^2}{\left[p^*+(1-p^*)\theta^*\right]^2}\right\}\right]$$

$$=\frac{1}{n}\left[\sigma_x^2+\mu_x^2\left(1+C_x^2\right)\int_a^b\int_0^1\left\{\frac{p(1-p)(1-\theta)^2+(1-p)\gamma^2}{\left[p+(1-p)\theta\right]^2}\right\}f(p,\theta)\,dp\,d\theta\right]$$

which proves the theorem.

Theorem 3

A joint probability density function of p and θ is given by

$$f(p,\theta)=\frac{\left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1}}{(b-a)\left[\mathrm{B}(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)\mathrm{B}(\alpha,\beta+2)+(b+a)\mathrm{B}(\alpha+1,\beta+1)\right]} \quad (14)$$

$$0\le p\le 1, \quad a\le\theta\le b$$

Proof.   Consider

$$f(p,\theta)=k\left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1}, \quad 0<p<1, a<\theta<b \quad (15)$$

Therefore,

$$\int_a^b\int_0^1 f(p,\theta)\,dp\,d\theta=k\int_a^b\left[\int_0^1\left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1}\,dp\right]d\theta, \quad a\le\theta\le b$$

$$=k\int_a^b\left[\int_0^1\left\{p^2+(1-p)^2\theta^2+2\theta p(1-p)\right\}p^{\alpha-1}(1-p)^{\beta-1}\,dp\right]d\theta$$

$$=k\int_a^b\left[\int_0^1 p^{(\alpha+2)-1}(1-p)^{\beta-1}\,dp+\theta^2\int_0^1 p^{\alpha-1}(1-p)^{(\beta+2)-1}\,dp+2\theta\int_0^1 p^{(\alpha+1)-1}(1-p)^{(\beta+1)-1}\,dp\right]d\theta$$

$$= k \int_a^b \left[B(\alpha+2,\beta) + \theta^2 B(\alpha,\beta+2) + 2\theta B(\alpha+1,\beta+1) \right] d\theta$$

$$= k \left[B(\alpha+2,\beta) \int_a^b d\theta + B(\alpha,\beta+2) \int_a^b \theta^2 d\theta + 2B(\alpha+1,\beta+1) \int_a^b \theta d\theta \right]$$

which implies

$$k \left[(b-a)B(\alpha+2,\beta) + \frac{1}{3}(b^3-a^3)B(\alpha,\beta+2) + B(\alpha+1,\beta+1)(b^2-a^2) \right] = 1$$

or

$$k(b-a) \left[B(\alpha+2,\beta) + \frac{(a^2+ab+b^2)}{3}B(\alpha,\beta+2) + (b+a)B(\alpha+1,\beta+1) \right] = 1$$

or

$$k = \frac{1}{(b-a)\left[B(\alpha+2,\beta) + \frac{(a^2+ab+b^2)}{3}B(\alpha,\beta+2) + (b+a)B(\alpha+1,\beta+1) \right]} \tag{16}$$

Substituting (16) into (15), proves the theorem.

Theorem 4

Under the joint probability density function $f(p,\theta)$, the variance of the estimator $\hat{\mu}_{S(BBB)}$ is given by

$$V(\hat{\mu}_{S(BBB)}) = \frac{1}{n}\left[\sigma_x^2 + \frac{\mu_x^2(1+C_x^2)\left\{ B(\alpha+1,\beta+1)\left(1+\frac{1}{3}(a^2+ab+b^2)-(b+a)\right)+\gamma^2 B(\alpha,\beta+1)\right\}}{B(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)} \right] \tag{17}$$

Proof.

$$E_1\left[V(\hat{\mu}_x)\right] = \int_a^b \int_0^1 V(\hat{\mu}_x) f(p,\theta) dp d\theta$$

$$= \frac{1}{n(b-a)\left[B(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)\right]}$$

$$\times \int_a^b \int_0^1 \left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1}\left[\sigma_x^2+\mu_x^2(1+C_x^2)\left\{\frac{p(1-p)(1-\theta)^2+(1-p)\gamma^2}{\{p+(1-p)\theta\}^2}\right\}\right] dp d\theta$$

$$= \frac{1}{n(b-a)\left[B(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)\right]}$$

$$\times \left[\int_a^b \int_0^1 \sigma_x^2\left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1} dp d\theta\right.$$

$$\left.+\int_a^b \int_0^1 \mu_x^2(1+C_x^2)\left\{p(1-p)(1-\theta)^2+(1-p)\gamma^2\right\}p^{\alpha-1}(1-p)^{\beta-1} dp d\theta\right]$$

$$= \frac{1}{n(b-a)\left[B(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)\right]}$$

$$\times \left[\sigma_x^2\int_a^b \int_0^1 \left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1} dp d\theta\right.$$

$$\left.+\mu_x^2(1+C_x^2)\int_a^b \int_0^1 \left\{p(1-p)(1-\theta)^2+(1-p)\gamma^2\right\}p^{\alpha-1}(1-p)^{\beta-1} dp d\theta\right]$$

$$= \frac{\sigma_x^2 I_1 + \mu_x^2(1+C_x^2)I_2}{n(b-a)\left[B(\alpha+2,\beta)+\frac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)\right]} \qquad (18)$$

So that

$$I_1 = \int_a^b \int_0^1 \left[p+(1-p)\theta\right]^2 p^{\alpha-1}(1-p)^{\beta-1} dp d\theta$$

$$= \int_a^b \int_0^1 \left[p^2+(1-p)^2\theta^2+2\theta p(1-p)\right]p^{\alpha-1}(1-p)^{\beta-1} dp d\theta$$

$$= \int\limits_a^b \left[\int\limits_0^1 p^{(\alpha+2)-1}(1-p)^{\beta-1}dp + \theta^2 \int\limits_0^1 p^{\alpha-1}(1-p)^{(\beta+2)-1}dp + 2\theta \int\limits_0^1 p^{(\alpha+1)-1}(1-p)^{(\beta+1)-1}dp \right] d\theta$$

$$= \int\limits_a^b \left[\mathrm{B}(\alpha+2,\beta) + \theta^2\mathrm{B}(\alpha,\beta+2) + 2\theta\mathrm{B}(\alpha+1,\beta+1) \right] d\theta$$

$$= \mathrm{B}(\alpha+2,\beta)\int\limits_a^b d\theta + \mathrm{B}(\alpha,\beta+2)\int\limits_a^b \theta^2 d\theta + 2\mathrm{B}(\alpha+1,\beta+1)\int\limits_a^b \theta d\theta$$

$$= \mathrm{B}(\alpha+2,\beta)(b-a) + \mathrm{B}(\alpha,\beta+2)\left(\frac{b^3-a^3}{3}\right) + \mathrm{B}(\alpha+1,\beta+1)(b^2-a^2)$$

$$= (b-a)\left[\mathrm{B}(\alpha+2,\beta) + \frac{1}{3}(a^2+ab+b^2)\mathrm{B}(\alpha,\beta+2) + (b+a)\mathrm{B}(\alpha+1,\beta+1) \right]$$

and

$$I_2 = \int\limits_a^b \int\limits_0^1 \left[p(1-p)(1-\theta)^2 + (1-p)\gamma^2 \right] p^{\alpha-1}(1-p)^{\beta-1}dpd\theta$$

$$= \int\limits_a^b \int\limits_0^1 p^{(\alpha+1)-1}(1-p)^{(\beta+1)-1}(1-\theta)^2 dpd\theta + \gamma^2 \int\limits_a^b \int\limits_0^1 p^{\alpha-1}(1-p)^{(\beta+1)-1}dpd\theta$$

$$= \int\limits_a^b (1-\theta)^2 \left[\int\limits_0^1 p^{(\alpha+1)-1}(1-p)^{(\beta+1)-1}dp \right] d\theta + \gamma^2 \int\limits_a^b \left[\int\limits_0^1 p^{\alpha-1}(1-p)^{(\beta+1)-1}dp \right] d\theta$$

$$= \int\limits_a^b (1-\theta)^2 \mathrm{B}(\alpha+1,\beta+1)d\theta + \gamma^2 \int\limits_a^b \mathrm{B}(\alpha,\beta+1)d\theta$$

$$= \mathrm{B}(\alpha+1,\beta+1)\int\limits_a^b (1+\theta^2-2\theta)d\theta + \gamma^2\mathrm{B}(\alpha,\beta+1)(b-a)$$

$$= \mathrm{B}(\alpha+1,\beta+1)\left[(b-a) + \frac{1}{3}(b^3-a^3) - (b^2-a^2) \right] + \gamma^2\mathrm{B}(\alpha,\beta+1)(b-a)$$

$$= (b-a)\left[\mathrm{B}(\alpha+1,\beta+1)\left\{ 1 + \frac{1}{3}(a^2+ab+b^2) - (b+a) \right\} + \gamma^2\mathrm{B}(\alpha,\beta+1) \right]$$

Substituting the values of I_1 and I_2 into (18), proves theorem.

Simulation Study

A numerical study was performed to investigate the various choices of parameters where the proposed stochastic randomized response model is more efficient than the non-stochastic BBB model. The percent relative efficiency of the proposed stochastic randomized response model estimator $\hat{\pi}_{S(BBB)}$ with respect to the non-stochastic BBB model estimator $\hat{\pi}_{x(BBB)}$ is given by

$$
\begin{aligned}
RE &= \frac{V(\hat{\pi}_{x(BBB)})}{V(\hat{\pi}_{S(BBB)})} \times 100\% \\
&= \frac{1 + \dfrac{\dfrac{(1+C_x^2)}{C_x^2}\left\{ B(\alpha+1,\beta+1)\left(1+\dfrac{1}{3}(a^2+ab+b^2)-(b+a)\right)+\gamma^2 B(\alpha,\beta+1) \right\}}{B(\alpha+2,\beta)+\dfrac{1}{3}(a^2+ab+b^2)B(\alpha,\beta+2)+(b+a)B(\alpha+1,\beta+1)}}{1+\dfrac{(1+C_x^2)C_s^2(p)}{C_x^2}} \times 100\%
\end{aligned}
$$

Clearly the relative efficiency depends only on the value of P, θ, a, b, C_x, C_γ, α and β. Certain parameters were fixed as $P=0.7$, $\theta=15$, $a=5$, and $b=25$. Note that here $a=5$, and $b=25$ are not the lower limit and upper limit of the scrambling variable, but these are the limits for the mean values θ^* of various scrambling variables used in a survey. The value of C_x was changed from 0.1 to 0.9 with step of 0.2; the value of C_γ was also changed from 0.1 to 0.9 with a step of 0.2; the value of α was changed from 0.5 to 3.5 with a step of 1.5; and the value of β was changed between 0.5 to 5.0 with a step of 1.5. A box plot showing the magnitude of the RE is for each value of π between 0.1 to 0.9 with a step of 0.2 is given in Figure 2.
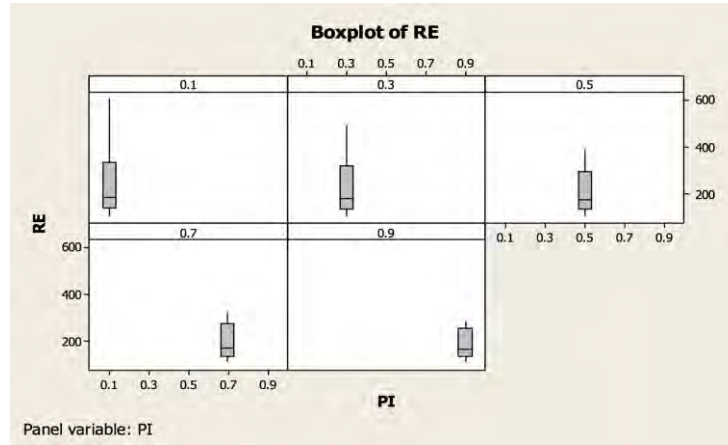
Figure 2. Relative efficiency plot.

For each combination of parameters, the percent relative efficiency of the proposed estimator was computed. The percent relative efficiency results so obtained are presented in Table 1.

Table 1. Percent relative efficiency of the proposed stochastic randomized response model.

C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE
0.1	0.1	0.5	0.5	164.9	0.3	0.7	2.0	3.5	149.9	0.7	0.3	2.0	5.0	153.1
0.1	0.1	0.5	2.0	316.6	0.3	0.7	2.0	5.0	181.7	0.7	0.5	0.5	0.5	161.7
0.1	0.1	0.5	3.5	463.2	0.3	0.7	3.5	5.0	123.6	0.7	0.5	0.5	2.0	235.8
0.1	0.1	0.5	5.0	604.5	0.3	0.9	0.5	0.5	190.1	0.7	0.5	0.5	3.5	279.9
0.1	0.1	2.0	3.5	125.5	0.3	0.9	0.5	2.0	266.4	0.7	0.5	0.5	5.0	309.1
0.1	0.1	2.0	5.0	164.4	0.3	0.9	0.5	3.5	309.2	0.7	0.5	2.0	3.5	134.1
0.1	0.3	0.5	0.5	170.0	0.3	0.9	0.5	5.0	336.5	0.7	0.5	2.0	5.0	161.3
0.1	0.3	0.5	2.0	307.4	0.3	0.9	2.0	2.0	122.9	0.7	0.5	3.5	5.0	111.2
0.1	0.3	0.5	3.5	425.8	0.3	0.9	2.0	3.5	159.8	0.7	0.7	0.5	0.5	171.0
0.1	0.3	0.5	5.0	528.6	0.3	0.9	2.0	5.0	189.5	0.7	0.7	0.5	2.0	238.5
0.1	0.3	2.0	3.5	131.5	0.3	0.9	3.5	3.5	110.5	0.7	0.7	0.5	3.5	276.1
0.1	0.3	2.0	5.0	169.5	0.3	0.9	3.5	5.0	134.3	0.7	0.7	0.5	5.0	299.9
0.1	0.5	0.5	0.5	178.0	0.3	0.9	5.0	5.0	105.3	0.7	0.7	2.0	2.0	111.1
0.1	0.5	0.5	2.0	295.0	0.5	0.1	0.5	0.5	154.0	0.7	0.7	2.0	3.5	144.1

C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE
0.1	0.5	0.5	3.5	381.3	0.5	0.1	0.5	2.0	256.2	0.7	0.7	2.0	5.0	170.6
0.1	0.5	0.5	5.0	447.5	0.5	0.1	0.5	3.5	332.0	0.7	0.7	3.5	5.0	121.3
0.1	0.5	2.0	3.5	141.2	0.5	0.1	0.5	5.0	390.3	0.7	0.9	0.5	0.5	179.8
0.1	0.5	2.0	5.0	177.4	0.5	0.1	2.0	3.5	122.1	0.7	0.9	0.5	2.0	240.9
0.1	0.5	3.5	5.0	113.1	0.5	0.1	2.0	5.0	153.6	0.7	0.9	0.5	3.5	272.9
0.1	0.7	0.5	0.5	186.3	0.5	0.3	0.5	0.5	159.1	0.7	0.9	0.5	5.0	292.5
0.1	0.7	0.5	2.0	284.0	0.5	0.3	0.5	2.0	255.6	0.7	0.9	2.0	2.0	121.2
0.1	0.7	0.5	3.5	346.7	0.5	0.3	0.5	3.5	323.0	0.7	0.9	2.0	3.5	154.0
0.1	0.7	0.5	5.0	390.2	0.5	0.3	0.5	5.0	372.7	0.7	0.9	2.0	5.0	179.3
0.1	0.7	2.0	2.0	112.6	0.5	0.3	2.0	3.5	127.6	0.7	0.9	3.5	3.5	109.8
0.1	0.7	2.0	3.5	151.9	0.5	0.3	2.0	5.0	158.7	0.7	0.9	3.5	5.0	131.4
0.1	0.7	2.0	5.0	185.8	0.5	0.5	0.5	0.5	167.4	0.9	0.1	0.5	0.5	144.3
0.1	0.7	3.5	5.0	124.3	0.5	0.5	0.5	2.0	254.8	0.9	0.1	0.5	2.0	214.6
0.1	0.9	0.5	0.5	193.6	0.5	0.5	0.5	3.5	310.6	0.9	0.1	0.5	3.5	257.8
0.1	0.9	0.5	2.0	275.8	0.5	0.5	0.5	5.0	349.2	0.9	0.1	0.5	5.0	286.9
0.1	0.9	0.5	3.5	323.0	0.5	0.5	2.0	3.5	136.6	0.9	0.1	2.0	3.5	118.9
0.1	0.9	0.5	5.0	353.6	0.5	0.5	2.0	5.0	167.0	0.9	0.1	2.0	5.0	144.1
0.1	0.9	2.0	2.0	123.5	0.5	0.5	3.5	5.0	111.9	0.9	0.3	0.5	0.5	149.2
0.1	0.9	2.0	3.5	161.8	0.5	0.7	0.5	0.5	176.5	0.9	0.3	0.5	2.0	217.5
0.1	0.9	2.0	5.0	193.1	0.5	0.7	0.5	2.0	253.9	0.9	0.3	0.5	3.5	258.1
0.1	0.9	3.5	3.5	110.8	0.5	0.7	0.5	3.5	299.1	0.9	0.3	0.5	5.0	285.0
0.1	0.9	3.5	5.0	135.2	0.5	0.7	0.5	5.0	328.7	0.9	0.3	2.0	3.5	123.7
0.1	0.9	5.0	5.0	105.4	0.5	0.7	2.0	2.0	111.6	0.9	0.3	2.0	5.0	148.9
0.3	0.1	0.5	0.5	160.3	0.5	0.7	2.0	3.5	147.0	0.9	0.5	0.5	0.5	157.3
0.3	0.1	0.5	2.0	288.7	0.5	0.7	2.0	5.0	176.0	0.9	0.5	0.5	2.0	222.0
0.3	0.1	0.5	3.5	398.6	0.5	0.7	3.5	5.0	122.4	0.9	0.5	0.5	3.5	258.6
0.3	0.1	0.5	5.0	493.5	0.5	0.9	0.5	0.5	184.9	0.9	0.5	0.5	5.0	282.1
0.3	0.1	2.0	3.5	124.1	0.5	0.9	0.5	2.0	253.2	0.9	0.5	2.0	3.5	132.0
0.3	0.1	2.0	5.0	159.8	0.5	0.9	0.5	3.5	290.1	0.9	0.5	2.0	5.0	156.9
0.3	0.3	0.5	0.5	165.4	0.5	0.9	0.5	5.0	313.3	0.9	0.5	3.5	5.0	110.6
0.3	0.3	0.5	2.0	284.0	0.5	0.9	2.0	2.0	122.1	0.9	0.7	0.5	0.5	166.6
0.3	0.3	0.5	3.5	377.0	0.5	0.9	2.0	3.5	156.9	0.9	0.7	0.5	2.0	226.8
0.3	0.3	0.5	5.0	451.7	0.5	0.9	2.0	5.0	184.4	0.9	0.7	0.5	3.5	259.1
0.3	0.3	2.0	3.5	129.9	0.5	0.9	3.5	3.5	110.2	0.9	0.7	0.5	5.0	279.2
0.3	0.3	2.0	5.0	164.9	0.5	0.9	3.5	5.0	132.8	0.9	0.7	2.0	2.0	110.6

C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE	C_x	C_Y	α	β	RE
0.3	0.5	0.5	0.5	173.5	0.5	0.9	5.0	5.0	105.1	0.9	0.7	2.0	3.5	141.8
0.3	0.5	0.5	2.0	277.4	0.7	0.1	0.5	0.5	148.5	0.9	0.7	2.0	5.0	166.2
0.3	0.5	0.5	3.5	349.3	0.7	0.1	0.5	2.0	231.4	0.9	0.7	3.5	5.0	120.3
0.3	0.5	0.5	5.0	402.0	0.7	0.1	0.5	3.5	286.4	0.9	0.9	0.5	0.5	175.6
0.3	0.5	2.0	3.5	139.3	0.7	0.1	0.5	5.0	325.4	0.9	0.9	0.5	2.0	231.2
0.3	0.5	2.0	5.0	173.1	0.7	0.1	2.0	3.5	120.3	0.9	0.9	0.5	3.5	259.5
0.3	0.5	3.5	5.0	112.6	0.7	0.1	2.0	5.0	148.2	0.9	0.9	0.5	5.0	276.7
0.3	0.7	0.5	0.5	182.3	0.7	0.3	0.5	0.5	153.5	0.9	0.9	2.0	2.0	120.4
0.3	0.7	0.5	2.0	271.2	0.7	0.3	0.5	2.0	233.1	0.9	0.9	2.0	3.5	151.6
0.3	0.7	0.5	3.5	326.0	0.7	0.3	0.5	3.5	283.7	0.9	0.9	2.0	5.0	175.2
0.3	0.7	0.5	5.0	363.1	0.7	0.3	0.5	5.0	318.7	0.9	0.9	3.5	3.5	109.5
0.3	0.7	2.0	2.0	112.2	0.7	0.3	2.0	3.5	125.4	0.9	0.9	3.5	5.0	130.2

Table 2 gives the descriptive statistics for the percent relative efficiency values for different values of π.

Table 2. Descriptive statistics of the relative efficiency values.

π	Mean	StDev	Minimum	Median	Maximum
0.1	244.9	131.9	105.4	185.8	604.5
0.3	228.6	109.5	105.3	181.7	493.5
0.5	210.0	87.5	105.1	176.0	390.3
0.7	198.2	71.0	109.8	170.8	325.4
0.9	188.0	61.3	109.5	166.4	286.9

Table 1 shows that overall the minimum RE value is 105.1% and maximum RE value is 604.5%. The average value the RE is 214.2% with a standard deviation of 97.06%. The median value of the percent relative efficiency is 175.61%. Thus, in conclusion, it is possible to make a stochastic randomization device which will remain more efficient than the BBB model and more cooperation could be expected from the respondents.

References

Abdelfatah, S., Mazloum, R, & Singh, S. (2013). Efficient use of a two-stage randomized response procedure. *Brazilian Journal of Probability and Statistics, 7*(4): 608-617.

Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika, 60*(3): 255-260.

Franklin, L. A. (1989). A comparison of estimators for randomized response sampling with continuous distributions from a dichotomous population. *Communications in Statistics - Theory and Methods, 18*(2): 489-505.

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

Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question randomized response model. *Proceedings of Social Statistics Section.* American Statistical Association*, pp. 65-72.

Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika, 77*(2): 436-438.

Singh, S. (2002). A new stochastic randomized response technique. *Metrika, 56*(2): 130-142.

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