Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2012-60

Estimating the Variance of the Forced Quantitative Randomized Response model by Gjestvang and Singh (2007)

Andreas Quatember

February 2012

Abstract:
Gjestvang and Singh (2007) presented a “forced randomized response model” for the estimation of a population mean of a sensitive variable, which was found to perform better than the model of Bar-Lev et al. (2004). Odumade and Singh (2008) extended the discussion of the statistical properties of the proposed estimator to general probability sampling designs. In this paper, this generalization is based on the estimation of the true value of the study variable for each sampling unit. These individual estimators build the basis for two different estimators of the variance of the generalized Gjestvang and Singh model. For the model’s practical applicability, the variance estimation is as important as its generalization to all probability sampling methods. A simulation study compares the performance of the proposed “simple” and the “bootstrap estimator” for different situations leading to recommendations for its practical use.
Key words: Randomized Response Technique; Statistical Disclosure Control; Complex Sampling Design; General Probability Sampling; Variance Estimation; Bootstrap

1 Introduction

Nonresponse and untruthful answering are very common in survey sampling. This is true especially when the variables asked are sensitive. For such situations, Warner (1965) presented the pioneering work in the field of randomized response questioning designs. Since then, various such techniques with different randomization devices have been proposed for all types of variables (cf. for a review for instance, Tracy and Mangat 1996; or a standardization of a group of strategies in Quatember 2009). All these strategies use a questioning design, which does not enable the data collector to identify the randomly selected question or instruction on which the respondent has given the answer. This shall reduce the individuals' fear of answering on a sensitive question. However, the strategy does still allow the estimation of the parameter under study because the probability mechanism of the randomization device is known.

Warner (1971) was the first to note that these techniques are also applicable in the field of statistical disclosure control as methods of masking confidential micro-data sets to allow their release for public use (cf. ibd., p.887). Such micro-data sets may contain variables with sensitive information on an individual. For the randomized response techniques to be applied in this field, either the survey units already perform the randomization of their answers at the survey's design stage or the statistical agency applies the probability mechanism of the technique on the micro-data file after the data collection before its release (cf. for instance, Gouweleeuw et al. 1998, or van den Hout and van der Heijden 2002).

For the estimation of a mean value μ_x of a sensitive quantitative variable x, Eichhorn and Hayre (1983) suggested that the answer given by the respondent k should not be the true value x_k but $y_k = z_k \cdot x_k$. Herein, variable z is the "scrambling" variable, of which, the distribution is known. Its expectation and standard deviation are denoted by μ_z and σ_z, respectively.

Bar-Lev et al. (2004) added a second possibility to this scrambled response. The answer y_k of survey unit k is

$$y_k = \begin{cases} x_k & \text{with probability } p_1, \\ z_k \cdot x_k & \text{with probability } p_2 = 1 - p_1 \end{cases}$$

($0 \leq p_1 \leq 1$). By rotating a spinner, drawing cards, throwing the dice, or using some other randomization instrument (for the use of house numbers or dates of birth see Diekmann 2011), the respondent then is instructed to report his or her true value of the variable under study with probability p_1 or to answer according to the Eichhorn and Hayre model with the remaining probability.

In the paper of Gjestvang and Singh (2007), this randomization device is modified again by adding a third possibility to the Bar-Lev et al. process. The randomized response of the survey unit k is

$$y_k = \begin{cases} x_k & \text{with probability } p_1, \\ z_k \cdot x_k & \text{with probability } p_2, \\ F & \text{with probability } p_3 \end{cases}$$

($p_1 + p_2 + p_3 = 1$; $0 \leq p_i \leq 1, i = 1, 2, 3$). Therein, survey unit k randomly has to answer either truthfully, or in the manner suggested by Eichhorn and Hayre, or with a fixed value F predetermined by the agency.

In section 2 of this paper, the extension of the discussion of the statistical properties (i.e. the estimator and the variance of the estimator of the mean value μ_x) of the Gjestvang and Singh model by Odumade and Singh (2008) to general probability sampling designs is presented. This extension, being of greatest importance for the practical application of the procedure, is derived herein using unbiased estimators $\widehat{x}_k$ calculated for this reason for each x_k obtained in the sample s ($k \in s$). Based on these estimated values, in the subsequent section two estimators of the estimator's variance are proposed. For the practical application, this estimation problem is considered to be as important as the generalization of the model to any probability sample. In the concluding section 4, a simulation study compares the quality of these proposed estimators in different situations leading to recommendations for its practical use.

2 Estimating the Survey Units' True Values of the Sensitive Variable

Odumade and Singh (2008) generalized the discussion of the estimator and the variance of the model by Gjestvang and Singh (2007) to general probability sampling. In the following, these statistical properties of the generalized model are presented in a different way. This will serve as the basis for two proposed estimators of the model's variance in the next session.

In the Gjestvang and Singh model, the expectation of y_k with respect to the randomization R is given by

$$\begin{aligned} E_R(y_k) &= p_1 \cdot x_k + p_2 \cdot x_k \cdot \mu_z + p_3 \cdot F \\ &= x_k \cdot b + a \end{aligned}$$

with $a \equiv p_3 \cdot F$ and $b \equiv p_1 + p_2 \cdot \mu_z$. Hence, the term

$$\widehat{x}_k = \frac{y_k - a}{b} \tag{1}$$

($b \neq 0$) is unbiased for the true value x_k.

The extension of the theory of the Gjestvang and Singh model to a sample s drawn from a finite population U with N elements by a general probability sampling design P can make use of these "substitutes" for the unknown x_k's. Assuming full cooperation of the survey units because of the higher privacy protection offered by the questioning design, the following theorems apply (compare the results in Odumade and Singh 2008, p.246):

Theorem 1: Applying the randomization device by Gjestvang and Singh (2007) and using (1), for a probability sampling design P, the population mean μ_x of study variable x is unbiasedly estimated by

$$\widehat{\mu}^P = \frac{1}{N} \cdot \sum\nolimits_s \widehat{x}_k \cdot d_k. \tag{2}$$

In (2), the sample weights d_k are the reciprocals of the first order sample inclusion probabilities of the n sample units k ($k = 1, 2, ..., n$).

Proof The subsequent calculation proofs that $\widehat{\mu}^P$ is without bias:

$$\begin{aligned} E(\widehat{\mu}^P) &= \frac{1}{N} \cdot E_P\left(E_R\left(\sum\nolimits_s \widehat{x}_k \cdot d_k | s\right)\right) = \frac{1}{N} \cdot E_P\left(\sum\nolimits_s x_k \cdot d_k\right) \\ &= \frac{1}{N} \cdot \sum\nolimits_U x_k = \mu_x \end{aligned}$$

Herein, E_P and E_R denote the expectations over two random processes, the probability sampling design P and the randomization of responses R, respectively.

Theorem 2: For a probability sampling design P, the variance of $\widehat{\mu}^P$ is given by

$$\begin{aligned} V(\widehat{\mu}^P) &= \frac{1}{N^2} \cdot V_P\left(\sum\nolimits_s x_k \cdot d_k\right) + \\ &\quad + \frac{1}{N^2 b^2} \cdot \Big[(p_1 + (\sigma_z^2 + \mu_z^2) \cdot p_2 - b^2) \cdot \sum\nolimits_U x_k^2 \cdot d_k - \\ &\quad -2 \cdot a \cdot b \cdot \sum\nolimits_U x_k \cdot d_k + a \cdot (F - a) \cdot \sum\nolimits_U d_k\Big]. \end{aligned} \tag{3}$$

The first summand of variance (3) refers to the variance of the Horvitz-Thompson estimator for the total of variable x for a given probability sampling design P, when the question on variable x is asked directly. Then, the second summand in (3) can be seen as the price to be paid in terms of accuracy for the privacy protection offered by the questioning design.
Proof Let V_P and V_R denote the variances over P and R. The variance of $\widehat{\mu}^P$ is given by

$$V(\widehat{\mu}^P) = V_P(E_R(\widehat{\mu}^P|s)) + E_P(V_R(\widehat{\mu}^P|s)). \tag{4}$$

The first of the two summands on the right-hand side of (4) yields

$$V_P(E_R(\widehat{\mu}^P|s)) = \frac{1}{N^2} \cdot V_P\left(\sum\nolimits_s x_k \cdot d_k\right). \tag{5}$$

Let I_k indicate the sample inclusion of survey unit k ($k = 1, 2, ..., N$). Because the conditional covariance C_R of the two substitutes $\widehat{x}_k$ and $\widehat{x}_l$ ($k \neq l$) over the randomization R is $C_R(\widehat{x}_k, \widehat{x}_l|s) = 0$ and $E_P(I_k^2) = E_P(I_k) = 1/d_k$, the variance of $\widehat{\mu}^P$ over the randomization R conditioned on sample s yields:

$$V_R(\widehat{\mu}^P|s) = \frac{1}{N^2} \cdot V_R\left(\sum\nolimits_U I_k \cdot \widehat{x}_k \cdot d_k | s\right) = \frac{1}{N^2} \cdot \sum\nolimits_U I_k^2 \cdot d_k^2 \cdot V_R(\widehat{x}_k).$$

Hence, for the second summand on the right-hand side of (4)

$$E_P(V_R(\widehat{\mu}^P|s)) = \frac{1}{N^2} \cdot \sum\nolimits_U V_R(\widehat{x}_k) \cdot d_k$$

applies, where

$$V_R(\widehat{x}_k) = \frac{1}{b^2} \cdot V_R(y_k).$$

Moreover, $V_R(y_k) = E_R(y_k^2) - E_R^2(y_k)$. The first expectation over R yields

$$\begin{aligned} E_R(y_k^2) &= x_k^2 \cdot p_1 + E_R(x_k^2 \cdot z_k^2) \cdot p_2 + F^2 \cdot p_3 \\ &= x_k^2 \cdot (p_1 + (\sigma_z^2 + \mu_z^2) \cdot p_2) + F \cdot a. \end{aligned}$$

Furthermore, $E_R(y_k) = x_k \cdot b + a$. Altogether, the second summand on the right-hand side of equation (4) results in

$$E_P(V_R(\widehat{\mu}^P|s)) = \frac{1}{N^2b^2} \cdot \Big[(p_1 + (\sigma_z^2 + \mu_z^2) \cdot p_2 - b^2) \cdot \sum\nolimits_U x_k^2 \cdot d_k - \\ -2 \cdot a \cdot b \cdot \sum\nolimits_U x_k \cdot d_k + a \cdot (F - a) \cdot \sum\nolimits_U d_k\Big]. \quad (6)$$

Summing up (5) and (6) proves Theorem 2.

For the sampling design P, the design probabilities p_1 to p_3, and the distribution of z all given, in our presentation, the variance minimizing value of F results in

$$F = \frac{b \cdot \sum_U x_k \cdot d_k}{(1 - p_3) \cdot \sum_U d_k}. \quad (7)$$

In (7), the sum $\sum_U d_k$ is known. Applying the Gjestvang and Singh model as method of statistical disclosure control after the data collection, the sum $\sum_U x_k \cdot d_k$ can be estimated by the sum $\sum_s x_k \cdot d_k^2$. In other cases, observations from past surveys may help to estimate the sum needed in the enumerator of (7).

3 The Variance Estimators Proposed

With the estimators $\widehat{x}_k$ for the true values x_k presented in section 2, two differing solutions for the problem of estimating variance (3) can be proposed ($k \in s$).

For a closed form estimator of variance $V(\widehat{\mu}^P)$, on the right hand side of formula (3) three terms have to be estimated from the data obtained in the sample. The first one, $V_P(\sum_s x_k \cdot d_k)$, refers to the variance of the Horvitz-Thompson estimator for the total of variable x. It is estimated by the usual unbiased estimator $\widehat{V}_P(\sum_s x_k \cdot d_k)$ (cf. for instance, Särndal et al. 1992, p.43). Using the estimators $\widehat{x}_k$ from (1), the term $\sum_U x_k \cdot d_k$ depending on both random processes, P and R, is estimated unbiasedly by $\sum_s \widehat{x}_k \cdot d_k^2$ because $E_P[E_R(\sum_s \widehat{x}_k \cdot d_k^2|s)] = \sum_U x_k \cdot d_k$ applies. But, the third term to be estimated in (3), $\sum_U x_k^2 \cdot d_k$, cannot be estimated unbiasedly by $\sum_s \widehat{x}_k^2 \cdot d_k^2$, because $E_R(\widehat{x}_k^2) = V_R(\widehat{x}_k) + x_k^2 \geq x_k^2$ when there is a true randomization ($p_1 < 1$). This means that the "simple estimator"

$$\widehat{V}_{simple} = \frac{1}{N^2} \cdot \widehat{V}_P\left(\sum\nolimits_s x_k \cdot d_k\right) + \\ + \frac{1}{N^2b^2} \cdot \Big[(p_1 + (\sigma_z^2 + \mu_z^2) \cdot p_2 - b^2) \cdot \sum\nolimits_s \widehat{x}_k^2 \cdot d_k^2 - \\ -2 \cdot a \cdot b \cdot \sum\nolimits_s \widehat{x}_k \cdot d_k^2 + a \cdot (F - a) \cdot \sum\nolimits_U d_k\Big]. \quad (8)$$

overestimates the true variance of $\widehat{\mu}^P$ if $V_R(\widehat{x}_k) > 0$. The overestimation of $V(\widehat{\mu}^P)$ increases with increasing $V_R(\widehat{x}_k)$ (see also section 4).

A second possibility for the estimation of the variance applies the bootstrap method. The bootstrap method was originally designed by Efron (1979) to estimate the sampling distribution of a statistic on the basis of data observed under i.i.d. conditions. For this purpose, the empirical distribution of the variable under study can be regarded as a set-valued Maximum-Likelihood-estimator of its unknown true probability distribution. Adapting the bootstrap method to complex sampling, the finite population U with N elements plays the role of the unknown probability distribution, for which a set-valued estimator U^* has to be generated from the sample data according to the Maximum-Likelihood-principle (cf. Chao and Lo 1994). Its basic idea is to replicate each observed x_k d_k-times ($k \in s$) (cf. Gross 1980, or for complex samples, Sitter 1992). Then, resamples are drawn from the "bootstrap population" U^* by the same sampling method P by which the original sample was drawn from the true population U. These resamples build the basis for the estimation of the sampling distribution of the statistic under study. To implement data imputation for missing values into the bootstrap procedure, Shao and Sitter (1996) proposed re-imputing in each bootstrap sample in exactly the same way as in the original sample.

In our case, we generate a bootstrap population U^* replicating the estimators $\widehat{x}_k$ instead of the unknown true values x_k. The sample weights d_k of the sampling design P are the replication factors applied on the $\widehat{x}_k$'s of the original sample. In the next step, b resamples are drawn from U^* applying the same sampling method P as in the original sample. In each of the b bootstrap samples, the randomization process R used in the original sample is applied ("re-randomization") and the estimator $\widehat{\mu}^P_{(.)}$ is calculated yielding $\widehat{\mu}^P_{(1)}, \widehat{\mu}^P_{(2)}, ..., \widehat{\mu}^P_{(b)}$. The variance of $\widehat{\mu}^P$ can be estimated by the "bootstrap estimator"

$$\widehat{V}_{boot} = \frac{1}{b-1} \cdot \sum\nolimits_{i=1}^{b} (\widehat{\mu}^P_{(i)} - \overline{\widehat{\mu}^P_{(.)}})^2 \tag{9}$$

with

$$\overline{\widehat{\mu}^P_{(.)}} = \frac{1}{b} \cdot \sum\nolimits_{i=1}^{b} \widehat{\mu}^P_{(i)}.$$

The estimator $\widehat{V}_{boot}$ overestimates the true variance $V(\widehat{\mu}^P)$ as a rule, because $V(\widehat{x}_k) > V(x_k)$ applies, if there is a real randomization process used in the Gjestvang and Singh model. The bias increases with increasing $V(\widehat{x}_k)$. In the next section, a simulation study compares the two proposed estimators with respect to efficiency.

4 A Simulation Study

A simulation study was carried out to investigate the performance of the two variance estimators, $\widehat{Y}_{simple}$ and $\widehat{Y}_{boot}$, proposed in section 3. For this purpose, a population U of $N = 1,000$ elements was generated from a normal distribution with a mean of 1,000 and a standard deviation of 200. The population parameters of these $1,000$ values for variable x were $\mu_x = 995.739$ and $\sigma_x = 197.657$.

Since the estimation of the sampling variance of an estimator of the mean value is well-known given the sampling method P, the simulations were concentrated solely on the uncertainty added through the randomization process R. Therefore, a Gjestvang and Singh model with

$$y_k = \begin{cases} x_k & \text{with probability } p_1 = 0.8, \\ z_k \cdot x_k & \text{with probability } p_2 = 0.16 \text{ and } z = N(1, \sigma_z), \\ \mu_x & \text{with probability } p_3 = 0.04 \end{cases}$$

was applied 10,000-times directly to U. In practice, F can be fixed at its variance-minimizing value μ_x when the method is used as masking method for statistical disclosure

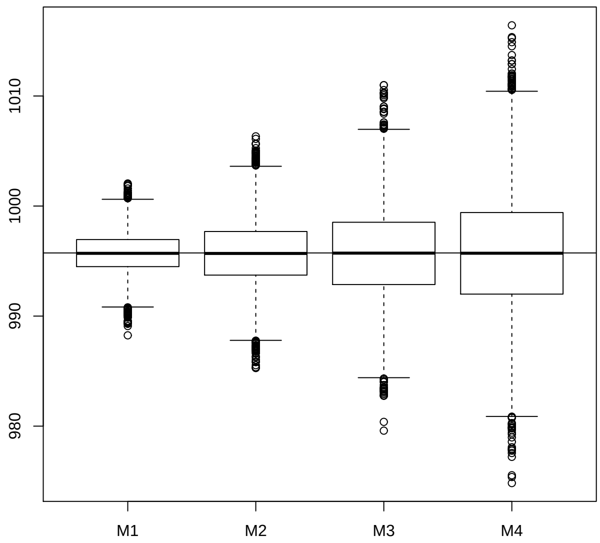

Figure 1: The 10,000 simulated estimators of μ_x for different σ_z (M1: $\sigma_z = 0.1$, M2: $\sigma_z = 0.2$ and so on)

control after the data collection. The distribution of the 10,000 simulated mean estimators calculated according to (3) is shown in Figure 1 for different σ_z's.

On the one hand, in each of the 10,000 simulations, after the population was "masked" with the described randomization process, the simple variance estimator $\widehat{V}_{simple}$ according to (8) and a usual confidence interval at the approximate confidence level $1 - \alpha = 0.95$ assuming approximate normality was calculated. Then, the rate of intervals covering the true mean value $\mu_x = 995.739$ was observed. On the other hand, in each of the 10,000 masked populations U^*, b bootstrap data sets were generated by "re-randomizing" in the same way as the original data set was randomized. These 10,000 data sets built the basis for the calculation of the 10,000 bootstrap variance estimators $\widehat{V}_{boot}$ (9) and 10,000 approximate confidence intervals. Also for these confidence intervals the coverage rate was calculated.

Table 1 shows the simulation results for the performance of the simple and bootstrap variance estimator (respectively standard deviation estimators) for differing values of σ_z. The standard deviation of the scrambling variable z, σ_z was used to vary $V(\widehat{x}_k)$. As indicated in section 3, the results show that both variance estimation methods proposed do overestimate in average the true standard deviation. This overestimation increases absolutely as well as relatively with increasing $V(\widehat{x}_k)$ (Figure 2 shows the box-plots for the simulated simple and bootstrap estimators of the theoretical standard deviation of $\widehat{\mu}^P$). Both methods of variance estimation perform quite well with respect to the coverage rate of approximate 95%-confidence intervals. But, the more σ_z increases the more does the coverage rate "overachieve" the targeted rate. The main difference in the quality of the two variance (respectively standard deviation) estimators is their differing inherent variability (see Figure 2 and columns $S_{\widehat{S}_1}$ and $S_{\widehat{S}_2}$ in Table 1). The simple estimators have

σ_z	Parameter $S(\widehat{\mu}^P)$	$\widehat{S}_1 \equiv \widehat{V}^{0.5}_{simple}$ $\overline{x}_{\widehat{S}_1}$	Coverage	$S_{\widehat{S}_1}$	$\widehat{S}_2 \equiv \widehat{S}_{boot}$ $\overline{x}_{\widehat{S}_2}$	Coverage	$S_{\widehat{S}_2}$
0.1	1.8481	1.8877	0.9545	0.0095	1.8857	0.9519	0.1338
0.2	2.9636	3.0346	0.9556	0.0160	3.0278	0.9537	0.2175
0.3	4.2105	4.3292	0.9562	0.0269	4.3225	0.9540	0.3109
0.4	5.5002	5.6876	0.9567	0.0428	5.6796	0.9542	0.4014

Table 1: The simple and bootstrap estimator of the standard deviation of $\widehat{\mu}^P$ for different σ_z ($\overline{x}_\bullet$, $S_\bullet$ ... the mean value and the standard deviation of the results of 10,000 simulations with $b = 100$)

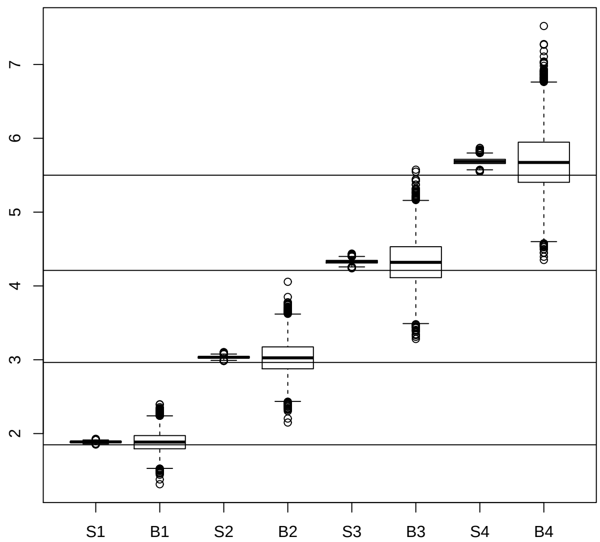

Figure 2: The simple and bootstrap estimators of the standard deviation for different σ_z (S1 and B1 for $\sigma_z = 0.1$, S2 and B2 for $\sigma_z = 0.2$ and so on)

	$\widehat{S}_2 \equiv \widehat{V}_{boot}^{0.5}$		
b	$\overline{x}_{\widehat{S}_2}$	Coverage	$S_{\widehat{S}_2}$
10	2.9496	0.9256	0.7034
50	3.0209	0.9492	0.3095
100	3.0278	0.9537	0.2175
500	3.0321	0.9545	0.0971
1000	3.0356	0.9548	0.0716

Table 2: The bootstrap estimator of the standard deviation of $\widehat{\mu}^P$ for different numbers of bootstrap samples b ($\overline{x}_{\widehat{S}_2}$, $S_{\widehat{S}_2}$ ... the mean value and the standard deviation of the results of 10,000 simulations with $\sigma_z = 0.2$)

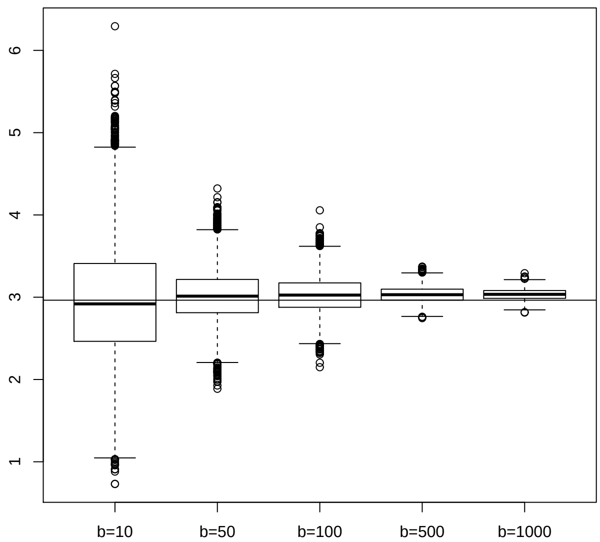

Figure 3: The bootstrap estimators of the standard deviation for different b

a much smaller inherent dispersion than the estimators based on the bootstrap.

The performance of the bootstrap estimator can be improved by raising the number b of bootstrap samples. For a constant $\sigma_z = 0.2$ (see the second row in Table 1) we additionally varied the number b of bootstrap samples. The results are presented in Table 2 and Figure 3, respectively. These results show, how the standard deviation of the bootstrap standard deviation estimator can be reduced by increasing the number b of bootstrap samples. The process is not too time-consuming. A single generation of 1,000 re-randomized bootstrap populations, as needed for the simulations, took only about 30 seconds.

Anyhow, the simple variance estimators, $\widehat{V}_{simple}$, did perform better than the bootstrap based ones, $\widehat{V}_{boot}$. Yet, in cases where the distribution of the statistic $\widehat{\mu}^P$ is not approximately normal, for example in samples drawn from a highly skewed distribution of the variable of interest, x, the desired coverage level of the usual approximate confidence interval will not be valid. In such cases, the bootstrap procedure offers the "percentile method" as an alternative, because the bootstrap distribution estimates the sampling distribution of $\widehat{\mu}^P$ (cf. Efron 1981, ch.4). Therefore, instead of calculating a variance estimate, $\widehat{V}_{boot}$, for an approximate 95%-confidence interval assuming normality, such an interval can be obtained directly from the distribution of the $\widehat{\mu}^P_{(.)}$'s by taking its 2.5 and 97.5 percentile.

5 Summary

Gjestvang and Singh (2007) presented a standardization of different techniques of randomized response for the estimation of the mean value of a sensitive quantitative variable. Odumade and Singh (2008) extended the theory of the Gjestvang and Singh model to general probability sampling, which is very necessary for its applicability in practice. The subject of the current paper was the unresolved question of the variance estimation for the estimator of the mean value, which is another important question to be answered for the model to be applicable in practice.

Two methods are proposed, a "simple" and a "bootstrap estimator". Both estimators are biased with an increasing bias when the variance of the randomized responses of an individual is increased. But, as this variance is under control of the agency the amount of bias is under control. By and large, both estimators work fairly well with respect to the coverage rates of approximate confidence intervals. For approximately normal distributed mean estimators the simple estimator can be recommended as it is less dispersed compared to the bootstrap method. The bootstrap estimator has its merits when the distribution of the mean estimator is not approximately normal. Then, alternatively the bootstrap's "percentile method" can be applied to calculate valid confidence intervals.

References

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

Chao, M.-T., and Lo, S.-H. (1994). Maximum Likelihood Summary and the Bootstrap Method in Structured Finite Populations. *Statistica Sinica*, 4, 389–406.

Diekmann, A. (2011). Making Use of "Benford's Law" for the Randomized Response Technique. A Research Note. Paper presented at the conference of the European Survey Research Association, Lausanne, 19th to 22nd of July. Unpublished Manuscript, ETH Zurich.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jacknife. *Annals of Statistics*, 7, 1–26.

Eichhorn, B.H., and Hayre, L.S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, 7, 307–316.

Gjestvang, C.R., and Singh, S. (2007). Forced quantitative randomized response model: a new device. *Metrika*, 66, 243–257.

Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J., and de Wolf, P.-P. (1998). Post Randomisation for Statistical Disclosure Control: Theory and Implementation. *Journal of Official Statistics*, 14 (4), 463–478.

Gross, S. (1980). Median Estimation in Sample Surveys. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 181–184.

Odumade, O., and Singh, S. (2008). Generalized Forced Quantitative Randomized Response Model: A Unified Approach. *Journal of the Indian Society of Agricultural Statistics*, 62 (3), 244–252.

Quatember, A. (2009). A Standardization of Randomized Response Strategies. *Survey Methodology*, 35 (2), 143–152.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Shao, J., und Sitter, R.R. (1996). Bootstrap for Imputed Survey Data. *Journal of the American Statistical Association*, 91, No. 435, 1278–1288.

Sitter, R.R. (1992). Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics*, 20 (2), 135–154.

Tracy, D.S. and Mangat, N.S. (1996). Some Developments in Randomized Response Sampling during the last Decade – A Follow Up of Review by Chaudhuri and Mukerjee. *Journal of Applied Statistical Science*, 4 (2/3), 147–158.

van den Hout, A., and van der Heijden P.G.M. (2002). Randomized Response, Statistical Disclosure Control and Misclassification: a Review. *International Statistical Review*, 70 (2), 269–288.

Warner, S.L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60, 63–69.

Warner, S.L. (1971). The Linear Randomized Response Model. *Journal of the American Statistical Association*, 66, 884–888.