Journal of Modern Applied Statistical Methods

Volume 11 | Issue 1 Article 9

5-1-2012

Estimation of Multinomial Proportions using Higher Order Moments of Scrambling Variables in Randomized Response Sampling

Cheng C. Chen
Texas A&M University

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

Recommended Citation
Chen, Cheng C. and Singh, Sarjinder (2012) "Estimation of Multinomial Proportions using Higher Order Moments of Scrambling Variables in Randomized Response Sampling," *Journal of Modern Applied Statistical Methods*: Vol. 11 : Iss. 1 , Article 9.
DOI: 10.22237/jmasm/1335845280

Estimation of Multinomial Proportions using Higher Order Moments of Scrambling Variables in Randomized Response Sampling

Cover Page Footnote

The authors are thankful to the Editor Professor Sawilowsky and to Dr. Julie M. Smith to bring the original manuscript in the present form. The authors are also thankful to Ms. Rebecca West at TAMUK for editing the original version of the manuscript.

Journal of Modern Applied Statistical Methods
May 2012, Vol. 11, No. 1, 106-122

1538 – 9472/12/$95.00

Estimation of Multinomial Proportions using Higher Order Moments of Scrambling Variables in Randomized Response Sampling

Cheng C. Chen Sarjinder Singh
Texas A&M University,
Kingsville, TX

An extension to estimating multinomial proportions of potentially sensitive attributes in survey sampling is proposed using higher order moments of scrambling variables at the estimation stage to produce unbiased estimators. The variance and covariance expressions are derived and the relative efficiency of the proposed estimators based on scrambling variables is investigated.

Key words: Estimation of sensitive multinomial proportions, randomized response sampling, respondents protection.

Introduction

The problem of estimating the proportion of potentially sensitive attributes in survey sampling has been very well addressed in the literature following the pioneering work of Warner (1965), and the use of randomized response sampling in social, medical and environmental sciences has been well documented (Waltz, et al., 2004; Blank, 2008). Singh and Chen (2009) introduced the use of higher order moments of scrambling variables to improve single proportion estimates without affecting respondent cooperation in survey research.

The problem of estimating trinomial proportions has been very useful, especially during election periods in the United States of America. Voters in the US can be divided into three mutually exclusive groups: Democrat, Republican and Other. At this time, expressing preference for one of these three groups does not pose a threat to an individual's privacy; however, the competition for the presidential position is becoming more difficult and there may be a time when voters will not feel safe disclosing their preferences to vote in the US. A new method is developed here that could be useful in such circumstances to organizations that conduct surveys about the prediction of future president for the US in the forthcoming elections. It is assumed that the partition of voters will remain trinomial because it may not be easy to establish a new competitive party as strong as those that are currently functioning; however, the proposed model can be extended to the case of a multinomial distribution if required.

This same argument can be extended to other applications if a population can be captured completely within three mutually exclusive groups. As noted by Singh, Kim and Grewal (2008), a sensitive question in one survey could be non-sensitive in another survey depending on the situation, particularly when there are three categories that are feasible when an answer is: exactly known, exactly unknown and not sure; hence leading to the problem of trinomial proportions estimation.

In a careful examination of the literature in randomized response sampling (Tracy & Mangat, 1996), not much attention has been paid to estimate sensitive multinomial proportions. Abul-Ela, et al. (1967) extended Warner's (1965) design to the multichotomous case when

Cheng Chen is an Associate Professor in the Department of Mathematics. Email him at: cheng.chen@tamuk.edu. Sarjinder Singh is an Associate Professor in the Department of Mathematics. Email him at: sarjinder@yahoo.com.

a population can be considered to be divided into t disjoint classes C_j with unknown proportions π_j ($j = 1,2,...,t$, $0 < \pi_j < 1$, $\Sigma \pi_j = 1$). It is assumed that at least one of the classes carries a stigma and at least one carries no stigma. They suggested drawing s $(= t-1)$ independent simple random with replacement samples sized n_i ($i = 1,2,...,s$, $\Sigma n_i = n$), and then employing a randomized response device to each of the samples. Abul-Ela, et al. (1967) examined the extent of bias and the mean square error of estimators for $t = 3$.

Bourke and Dalenious (1973, 1974) proposed a Latin square measurement design to extend Warner's model to the multinomial case; their design uses t different possible responses and requires only one sample. A respondent is asked to select one of a t-type cards using a random device. Each of the t-mutually exclusive classes is described on each card, except that the order of the description is permuted from card to card and the permutation for t-cards forms a Latin square.

The respondent reads the cards selected and reports only the position of the card (i.e., $t = 1,\ 2,\ \ldots,\ (t-1)$ or t) of the statement describing the class to which he/she belongs. The unrelated question design was also extended by Bourke (1974) to estimate the proportion of a population in each of t mutually exclusive classes of which $(t-1)$ are sensitive. One sample is needed if the distribution of the unrelated character is known. The design uses a deck of cards, each of which contains a number of statements. The arrangement of the statements is a part of the design.

Hochberg (1975) outlined an alternative scheme for estimating the t group proportions of which, at most, $(t-2)$ are stigmatizing. The realizations for any sampled individuals constitute a two-stage scheme. The second stage is conditional on the random individual's response in the first stage. Drane (1976) used a forced yes stochastic model to estimate the proportion of more than one sensitive character. The use of supplemented block, balanced incomplete block and spring balance weighing designs were introduced by Raghavarao and Federer (1979); their models allow the surveyor to obtain answers to several sensitive questions. Mukhopadhyay (1980), Mukherjee (1981), Tamhane (1981), Bourke (1981, 1982, 1990), Silva (1983) and Christofides (2003) have also considered the estimation of multi-attribute parameters.

Guerriero and Sandri (2007) pointed out that the family of models proposed by Kuk (1990) is better than the Simmons' family (refer to Greenbeg *et al.* (1969)) in terms of efficiency and privacy protection. From an empirical standpoint, van der Heijden, et al. (2000) showsed that Kuk's procedure performs slightly better than the forced-response procedure and markedly better than face-to-face direct questioning and computer assisted self-interviewing.

They also noted that the recommendations and successful applications of Kuk's procedure have been reported in van den Hout and van der Heijden (2002), and these results should be even more marked for the model proposed by Christofides (2003). In addition, an adequate analysis of the efficiency and the respondents' protection is always necessary when proposing new randomized response models. Thus, following Guerriero and Sandri (2007), it is worthwhile to extend the Kuk (1990) and Franklin (1989) type models. Note that the the Mangat (1994), Mangat and Singh (1990), Gjestvang and Singh (2006) and Kuk (1990) models are special cases of the Franklin (1989) model. Additional work on randomized response sampling is available in Singh and Kim (2011), Diana and Perri (2009), Tan, et al. (2009) and Esponda and Guerrero (2009).

Proposed Randomized Response Technique

In the proposed randomized response device, if a person selected in the sample belongs to the first sensitive group A_1 then that person is requested to draw a random number S_1 from a density function $f_1(s)$ and report to the interviewer; if that person belongs to second sensitive group A_2 then that person is requested to draw a random number S_2 from a density function $f_2(s)$ and report to the interviewer; and if that person belongs to the third sensitive group A_3, then that person is requested to draw a random number S_3 from a density function $f_3(s)$ and report to the interviewer. The respondent is further requested not to disclose the mode of response.

Let Ω be the population under study; $\Omega = \bigcup_{k=1}^{3} A_k$ and the groups A_k are mutually exclusive. The choice of the three densities $f_1(s)$, $f_2(s)$ and $f_3(s)$ are comprised such that respondents should feel safe in reporting the random number drawn. In other words, to maintain the privacy of respondents from all the three groups, the mean values and the variances of the three densities should not deviate too greatly from each other. In particular, the densities $f_1(s)$, $f_2(s)$ and $f_3(s)$ could be normal, beta, gamma or some other distribution.

Let π_1, π_2 and π_3 represent the true proportions of persons belonging to groups A_1, A_2 and A_3 respectively such that $\pi_1 + \pi_2 + \pi_3 = 1$. Assume that E denotes the expected value over the proposed randomization response device, and let $\theta_1 = E(S_1)$, $\theta_2 = E(S_2)$, $\theta_3 = E(S_3)$, and $\gamma_{abc} = E[(S_1 - \theta_1)^a (S_2 - \theta_2)^b (S_3 - \theta_3)^c]$, where a, b and c are non-negative integers and are known moments of the three scrambling variables used in the proposed randomization device. Consider a simple random with replacement sample (SRSWR) of n respondents. Interestingly, it can be shown that, based on only single sample information, three unbiased estimates of the three different parameters can be proposed. The distribution of the responses will be as follows:

$$Z_i = \begin{cases} S_1 \text{ with probability } \pi_1 \\ S_2 \text{ with probability } \pi_2 \\ S_3 \text{ with probability } \pi_3 \end{cases} \tag{2.1}$$

If

$$E(Z_i) = \pi_1\theta_1 + \pi_2\theta_2 + (1 - \pi_1 - \pi_2)\theta_3 \tag{2.2}$$

then, following Singh and Chen (2009),

$$Z_i^2 = \begin{cases} S_1^2 \text{ with probability } \pi_1 \\ S_2^2 \text{ with probability } \pi_2 \\ S_3^2 \text{ with probability } \pi_3 \end{cases} \tag{2.3}$$

where $E(S_1^2) = \gamma_{200} + \theta_1^2$, $E(S_2^2) = \gamma_{020} + \theta_2^2$ and $E(S_3^2) = \gamma_{002} + \theta_3^2$.

If

$$E(Z_i^2) = \pi_1(\gamma_{200} + \theta_1^2) + \pi_2(\gamma_{020} + \theta_2^2) + (1 - \pi_1 - \pi_2)(\gamma_{002} + \theta_3^2) \tag{2.4}$$

and defining

$$\Delta = (\theta_1 - \theta_3)\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\} - (\theta_2 - \theta_3)\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\}, \tag{2.5}$$

several theorems and lemmas may be put forth (see Appendix A).

Empirical Comparisons

It is possible to use the Warner (1965) model three times to estimate the three non-overlapping parameters π_k, $k = 1,2,3$. Each respondent selected in the sample could be provided with three randomization devices, for example, R_k, $k = 1,2,3$. The randomization R_k bears two types of statements, are you a member of group A_k?, and are you a member of group A_k^c? with probabilities P_k and $(1 - P_k)$, respectively. Based on a sample of n respondents, if n_k reports yes related to the k^{th} group, then the unbiased estimator of π_k is

$$\hat{\pi}_{k(w)} = \frac{n_k/n - (1 - P_k)}{2P_k - 1}, \quad P_k \neq 0.5 \tag{3.1}$$

with variance

$$V(\hat{\pi}_{k(w)}) = \frac{\pi_k(1-\pi_k)}{n} + \frac{P_k(1-P_k)}{n(2P_k-1)^2}\,. \qquad (3.2)$$

The relative efficiency of the proposed estimator $\hat{\pi}_k$ (as defined in the Appendix) with respect to the corresponding estimator $\hat{\pi}_{k(w)}$ (Warner, 1965) is:

$$RE(k) = \frac{V(\hat{\pi}_{k(w)})}{V(\hat{\pi}_k)} \times 100\%, \qquad (3.3)$$
$$k = 1, 2, 3.$$

Results

Choosing $P_k = 0.7$, $k = 1,2,3$, based on Warner (1965) is a reasonable and practical choice for the model, considering the problem of estimation of π_k with their respective estimators $\hat{\pi}_{k(w)}$ for $k = 1,2,3$. A privacy protection criterion is suggested, that is:

$$\lambda_{Z_{k,i}} = \frac{f(Z_k \mid k \in A_i)}{f(Z_k \mid k \notin A_i)} \qquad (3.4)$$

and refers to the privacy protection with respect to response Z_k for a respondent k being a member of A_i. For these measures $0 \le \lambda_{Z_{k,i}} < \infty$ applies with $\lambda_{Z_{k,1}} = 1$ indicating data privacy protection for unit k being a member of group A_i. This means that the value Z_k contains absolutely no information on the variable of interest; the more the λ-measure differs from unity the more information on the variable under study is contained in the response, meaning the less the privacy protection. The maximum $\lambda_{Z_{k,i}} = \infty$ (or 0) describes a situation where membership or the non-membership of A_i may be concluded based on the answer Z_k directly. A respondent would answer untruthfully or not answer at all in such a case.

Bearing in mind the proposed privacy protection criterion in (3.4), choice of the known parameters of the scrambling variables was: $\theta_1 = 57$, $\theta_2 = 62$, $\theta_3 = 60$, $\gamma_{200} = 0.5$, $\gamma_{020} = 3.5$, and $\gamma_{002} = 4.5$ in the proposed model with three scrambling variables. Based on the three sigma empirical rule, most of the values of the scrambling variables S_1, S_2 and S_3 could be any real numbers in the ranges: (54.87, 59.12); (56.38, 67.61), and (53.63, 66.36) respectively, but those values are not 100% bounded to these intervals. Due to an overlap between the three intervals, it is difficult to guess the status of the respondents based on their reported responses. Using the four sigma rule the scrambling variables S_1, S_2 and S_3 could, respectively, be any real numbers in the ranges: (54.17, 59.82); (54.51, 69.48), and (51.51, 68.48), and the six sigma empirical rule can be considered in a similar manner.

To study the effect of known higher order moments, such as skewness and kurtosis, of the scrambling variables on the relative efficiencies $RE(k)$ of the proposed estimators we studied different values of $\gamma_3 = \gamma_{300} = \gamma_{030} = \gamma_{003}$ as –2, 0, 3, 5, 10 and 20; and the values of the $\gamma_4 = \gamma_{400} = \gamma_{040} = \gamma_{004}$ as 2, 3, 5 and 10. The minimum relative efficiency of 103% was retained by assuming that a minimum 3% gain is enough if one methodology could gain over the other without affecting the respondents' cooperation (see Table 1).

Note that, while estimating rare attributes in two groups such as $\pi_1 = 0.1$, and $\pi_2 = 0.1$, then $\pi_3 = 0.8$ and based on 3 observations, the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 615.3%, 451.1%, 713.4% for $\gamma_{300} = \gamma_{030} = \gamma_{003} = -2$ and $\gamma_{400} = \gamma_{040} = \gamma_{004} = 10$. Keeping the same value of $\gamma_{400} = \gamma_{040} = \gamma_{004} = 10$, for $\gamma_{300} = \gamma_{030} = \gamma_{003} = 0$, the $RE(1)$, and $RE(3)$ values become 919.0%, 306.8% and 562.5%. Thus, changing the value of $\gamma_{300} = \gamma_{030} = \gamma_{003}$ from –2 to 0, the $RE(1)$ increases from 615.3% to 919.0%, but the value of $RE(2)$ decreases from 451.1% to 306.8%, and the value of $RE(3)$ decreases from 713.4% to 562.5%. As the values of $\gamma_{300} = \gamma_{030} = \gamma_{003}$ increase to 3, there is huge increase in the value of $RE(1)$ to 3543.1%, whereas a decrease in the values of $RE(2)$ and

Table 1: Relative efficiencies, $RE(k)$, $k = 1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.1	0.1	0.8	-2	-2	-2	10	10	10	615.3	451.1	713.4
			0	0	0	10	10	10	919.0	306.8	562.5
			3	3	3	10	10	10	3543.1	207.3	426.9
	0.2	0.7	-2	-2	-2	2	2	2	977.6	358.7	1368.6
			-2	-2	-2	3	3	3	907.5	338.1	843.5
			-2	-2	-2	5	5	5	793.7	303.4	477.3
			-2	-2	-2	10	10	10	604.3	241.4	228.9
			0	0	0	2	2	2	1633.8	246.8	548.7
			0	0	0	3	3	3	1447.0	236.9	439.1
			0	0	0	5	5	5	1177.7	219.3	313.8
			0	0	0	10	10	10	803.8	185.0	183.1
			3	3	3	3	3	3	13358.9	163.5	255.4
			3	3	3	5	5	5	4294.3	154.9	207.3
			3	3	3	10	10	10	1592.6	136.9	140.9
			5	5	5	10	10	10	4606.9	116.7	122.1
	0.3	0.6	-2	-2	-2	2	2	2	950.1	220.8	282.2
			-2	-2	-2	3	3	3	883.8	213.1	250.7
			-2	-2	-2	5	5	5	775.5	199.2	204.9
			-2	-2	-2	10	10	10	593.6	171.2	140.6
			0	0	0	2	2	2	1301.9	166.3	187.7
			0	0	0	3	3	3	1180.5	161.9	173.2
			0	0	0	5	5	5	994.9	153.7	150.0
			0	0	0	10	10	10	714.2	136.5	112.4
			3	3	3	2	2	2	2928.3	121.3	124.9
			3	3	3	3	3	3	2378.1	119.0	118.3
			3	3	3	5	5	5	1728.6	114.5	107.0
	0.4	0.5	-2	-2	-2	2	2	2	924.1	163.6	161.3
			-2	-2	-2	3	3	3	861.2	159.5	150.5
			-2	-2	-2	5	5	5	758.1	151.7	132.8
			0	0	0	2	2	2	1082.1	127.8	115.2
			0	0	0	3	3	3	996.9	125.2	109.6
	0.5	0.4	-2	-2	-2	2	2	2	899.5	131.5	113.5
			-2	-2	-2	3	3	3	839.8	128.8	108.0

Table 1 (continued): Relative efficiencies, $RE(k)$, $k=1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.2	0.1	0.7	-2	-2	-2	2	2	2	529.7	657.2	1375.4
			-2	-2	-2	3	3	3	509.4	588.5	846.1
			-2	-2	-2	5	5	5	473.1	486.6	478.1
			-2	-2	-2	10	10	10	401.6	339.7	229.0
			0	0	0	3	3	3	786.8	431.9	2202.0
			0	0	0	5	5	5	703.5	374.4	733.2
			0	0	0	10	10	10	556.3	280.9	274.9
			3	3	3	5	5	5	2610.9	278.1	3673.9
			3	3	3	10	10	10	1317.1	223.0	392.7
			5	5	5	10	10	10	14931.8	196.1	549.8
	0.2	0.6	-2	-2	-2	2	2	2	521.9	287.3	282.5
			-2	-2	-2	3	3	3	502.1	274.0	250.9
			-2	-2	-2	5	5	5	466.9	250.7	205.0
			-2	-2	-2	10	10	10	397.1	206.8	140.7
			0	0	0	2	2	2	743.9	230.5	282.5
			0	0	0	3	3	3	704.4	221.8	250.9
			0	0	0	5	5	5	636.9	206.3	205.0
			0	0	0	10	10	10	513.8	175.7	140.7
			3	3	3	2	2	2	2055.6	177.8	282.5
			3	3	3	3	3	3	1780.2	172.6	250.9
			3	3	3	5	5	5	1404.0	163.1	205.0
			3	3	3	10	10	10	918.7	143.3	140.7
			5	5	5	5	5	5	7125.6	143.0	205.0
			5	5	5	10	10	10	1935.8	127.6	140.7
	0.3	0.5	-2	-2	-2	2	2	2	514.3	192.3	161.3
			-2	-2	-2	3	3	3	495.1	186.5	150.6
			-2	-2	-2	5	5	5	460.8	175.7	132.8
			0	0	0	2	2	2	669.8	159.0	144.8
			0	0	0	3	3	3	637.7	154.9	136.0
			0	0	0	5	5	5	581.8	147.4	121.4
			3	3	3	2	2	2	1226.0	126.1	125.5
			3	3	3	3	3	3	1122.4	123.6	118.9
			3	3	3	5	5	5	960.2	118.8	107.5
			5	5	5	2	2	2	2745.7	110.9	115.2
			5	5	5	3	3	3	2275.6	108.9	109.6
	0.4	0.4	-2	-2	-2	2	2	2	506.9	147.8	113.5
			-2	-2	-2	3	3	3	488.3	144.3	108.1

Table 1 (continued): Relative efficiencies, $RE(k)$, $k = 1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.3	0.1	0.6	-2	-2	-2	2	2	2	383.8	444.7	282.8
			-2	-2	-2	3	3	3	373.4	412.2	251.1
			-2	-2	-2	5	5	5	354.1	359.5	205.2
			-2	-2	-2	10	10	10	313.7	272.4	140.8
			0	0	0	2	2	2	567.7	410.2	570.8
			0	0	0	3	3	3	545.2	382.3	454.9
			0	0	0	5	5	5	505.1	336.5	323.6
			0	0	0	10	10	10	426.7	259.1	188.0
			3	3	3	5	5	5	1401.7	307.1	2416.8
			3	3	3	10	10	10	928.3	241.3	378.2
			5	5	5	10	10	10	4290.6	230.7	1163.0
	0.2	0.5	-2	-2	-2	2	2	2	379.8	239.7	161.4
			-2	-2	-2	3	3	3	369.6	230.3	150.7
			-2	-2	-2	5	5	5	350.7	213.6	132.9
			0	0	0	2	2	2	524.9	216.3	194.9
			0	0	0	3	3	3	505.6	208.6	179.4
			0	0	0	5	5	5	470.9	194.9	154.8
			0	0	0	10	10	10	402.0	167.3	115.3
			3	3	3	2	2	2	1229.3	188.6	282.9
			3	3	3	3	3	3	1128.4	182.8	251.4
			3	3	3	5	5	5	969.2	172.1	205.6
			3	3	3	10	10	10	716.5	150.2	141.2
			5	5	5	2	2	2	11669.3	173.8	404.6
			5	5	5	3	3	3	6310.2	168.9	343.1
			5	5	5	5	5	5	3289.2	159.7	263.1
			5	5	5	10	10	10	1497.2	140.7	166.2
	0.3	0.4	-2	-2	-2	2	2	2	375.9	170.4	113.6
			-2	-2	-2	3	3	3	365.9	165.8	108.1
			0	0	0	2	2	2	488.1	152.3	118.4
			0	0	0	3	3	3	471.3	148.6	112.5
			3	3	3	2	2	2	883.5	131.3	126.4
			3	3	3	3	3	3	830.1	128.5	119.7
			3	3	3	5	5	5	740.6	123.4	108.1
			5	5	5	2	2	2	1921.1	120.3	132.4
			5	5	5	3	3	3	1685.4	118.0	125.0
			5	5	5	5	5	5	1353.4	113.6	112.5

Table 1 (continued): Relative efficiencies, $RE(k)$, $k=1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.4	0.1	0.5	-2	-2	-2	2	2	2	313.4	336.1	161.5
			-2	-2	-2	3	3	3	306.5	317.2	150.7
			-2	-2	-2	5	5	5	293.7	285.0	133.0
			0	0	0	2	2	2	456.2	365.2	298.1
			0	0	0	3	3	3	441.9	342.9	263.3
			0	0	0	5	5	5	415.6	305.7	213.5
			0	0	0	10	10	10	362.0	240.4	144.9
			3	3	3	5	5	5	1102.4	342.9	2330.8
			3	3	3	10	10	10	791.2	262.8	378.1
	0.2	0.4	-2	-2	-2	2	2	2	310.8	205.5	113.6
			-2	-2	-2	3	3	3	304.0	198.6	108.2
			0	0	0	2	2	2	428.7	203.7	150.2
			0	0	0	3	3	3	416.0	196.9	140.8
			0	0	0	5	5	5	392.6	184.6	125.0
			3	3	3	2	2	2	994.8	200.9	290.1
			3	3	3	3	3	3	928.8	194.3	256.9
			3	3	3	5	5	5	820.1	182.3	209.0
			3	3	3	10	10	10	634.4	157.9	142.5
			5	5	5	2	2	2	8315.3	199.1	765.8
			5	5	5	3	3	3	5218.4	192.6	570.8
			5	5	5	5	5	5	2990.7	180.8	378.2
			5	5	5	10	10	10	1446.7	156.8	205.2
	0.3	0.3	3	3	3	2	2	2	759.0	136.9	127.8
			3	3	3	3	3	3	720.0	133.9	120.8
			3	3	3	5	5	5	652.9	128.3	108.8
			5	5	5	2	2	2	1829.2	131.4	157.1
			5	5	5	3	3	3	1617.9	128.7	146.6
			5	5	5	5	5	5	1314.4	123.5	129.3

Table 1 (continued): Relative efficiencies, $RE(k)$, $k=1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.5	0.1	0.4	-2	-2	-2	2	2	2	272.2	270.1	113.7
			-2	-2	-2	3	3	3	267.0	257.7	108.2
			0	0	0	2	2	2	398.2	329.1	205.3
			0	0	0	3	3	3	387.3	310.9	188.1
			0	0	0	5	5	5	367.1	280.0	161.1
			0	0	0	10	10	10	324.8	224.2	118.5
			3	3	3	5	5	5	1020.9	388.1	3115.4
			3	3	3	10	10	10	749.6	288.6	392.0
	0.2	0.3	0	0	0	2	2	2	377.2	192.5	122.2
			0	0	0	3	3	3	367.3	186.4	115.8
			0	0	0	5	5	5	349.2	175.3	104.7
			3	3	3	2	2	2	928.5	214.9	305.8
			3	3	3	3	3	3	871.1	207.3	268.4
			3	3	3	5	5	5	775.3	193.7	215.7
			3	3	3	10	10	10	608.2	166.4	144.7
			5	5	5	3	3	3	10163.3	224.1	2219.8
			5	5	5	5	5	5	4162.8	208.3	735.2
			5	5	5	10	10	10	1681.3	177.1	275.1
	0.3	0.2	3	3	3	2	2	2	720.8	143.1	129.9
			3	3	3	3	3	3	685.8	139.8	122.4
			3	3	3	5	5	5	625.0	133.7	109.8
			5	5	5	2	2	2	2216.4	144.9	197.4
			5	5	5	3	3	3	1915.3	141.5	180.6
			5	5	5	5	5	5	1506.2	135.3	154.4
			5	5	5	10	10	10	981.8	121.8	113.3

Table 1 (continued): Relative efficiencies, $RE(k)$, $k=1,2,3$, of Proposed Estimators for Difference Choices of Known Higher Order Moments of the Three Scrambling Variables

π_1	π_2	π_3	γ_{300}	γ_{030}	γ_{003}	γ_{400}	γ_{040}	γ_{004}	$RE(1)$	$RE(2)$	$RE(3)$
0.6	0.1	0.3	0	0	0	2	2	2	365.7	299.5	157.2
			0	0	0	3	3	3	356.4	284.4	146.7
			0	0	0	5	5	5	339.2	258.3	129.5
			3	3	3	5	5	5	1068.8	447.1	12053.1
			3	3	3	10	10	10	773.7	320.0	424.2
	0.2	0.2	3	3	3	2	2	2	967.3	230.9	334.1
			3	3	3	3	3	3	904.8	222.2	288.7
			3	3	3	5	5	5	801.3	206.7	227.1
			3	3	3	10	10	10	623.1	175.9	148.0
			5	5	5	10	10	10	2651.1	203.3	446.3
	0.3	0.1	3	3	3	2	2	2	742.9	149.8	132.7
			3	3	3	3	3	3	705.5	146.2	124.6
			3	3	3	5	5	5	641.0	139.5	110.9
			5	5	5	2	2	2	4297.9	161.4	279.0
			5	5	5	3	3	3	3289.0	157.2	245.3
			5	5	5	5	5	5	2238.2	149.5	197.4
			5	5	5	10	10	10	1244.3	133.2	132.7
0.7	0.1	0.2	0	0	0	2	2	2	348.7	274.8	126.9
			0	0	0	3	3	3	340.1	262.0	119.7
			0	0	0	5	5	5	324.1	239.7	107.6
			3	3	3	10	10	10	880.7	359.0	488.1
	0.2	0.1	3	3	3	2	2	2	1147.2	249.6	385.8
			3	3	3	3	3	3	1058.8	239.4	324.1
			3	3	3	5	5	5	917.4	221.5	245.5
			3	3	3	10	10	10	687.8	186.5	152.9
			5	5	5	10	10	10	36398.8	238.8	1620.7
0.8	0.1	0.1	0	0	0	2	2	2	343.5	253.9	105.3
			3	3	3	10	10	10	1177.1	408.8	624.8

$RE(3)$ is 207.3% and 426.9%. Thus, for this case the minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 615.3%, 207.3% and 426.9%, while the maximum values are 3543.1%, 451.1% and 713.4% respectively.

Consider another situation: If one of the attributes is rare, $\pi_1 = 0.1$, and the second attribute is moderate, $\pi_2 = 0.3$, then $\pi_3 = 0.6$ and based on 11 observations, the average relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 1311.77%, 161.59%, 168.35% with standard deviations 742.70%, 37.46% and 58.19% respectively as the values different parameters of the scrambling variables changes. The medians of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain 994.9%, 161.9% and 150.0%. The minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 593.6%, 114.5% and 107.0% while the maximum values are 2928.3%, 220.8% and 282.2% respectively.

By contrast, when all three variables have moderate prevalence over the population as $\pi_1 = 0.3$, $\pi_2 = 0.3$ and $\pi_3 = 0.4$ then, based on 10 observations, the average relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 911.53%, 137.22%, 117.67% with standard deviations 558.53%, 20.52% and 8.19% respectively as the values different parameters of the scrambling variables changes. The medians of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain 785.35%, 129.90% and 166.00%. The minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 365.90%, 113.60% and 108.1% while the maximum values are 1921.10%, 170.40% and 132.40% respectively (see Table 1).

Note that in Table 1 the $RE(1)$, $RE(2)$ and $RE(3)$ for $\pi_1 = 0.1$, $\pi_2 = 0.1$ and $\pi_3 = 0.8$ are not the same as for $\pi_1 = 0.8$, $\pi_2 = 0.1$ and $\pi_3 = 0.1$ because of different choices of mean and variances of the scrambling variables for the three categories. Further note that the choice of parameters considered herein, shows in majority when π_1 remains close to zero, for example $\pi_1 = 0.1$ and $\pi_2 = 0.1$ there are three situations where the proposed method remains efficient and as soon as π_1 becomes 0.8 the proposed method shows efficiency only in two situations.

Thus, the proposed randomization device should be considered for a situation when the first attribute is rare, the second attribute is moderate and the third attribute is widespread. It may be concluded that the proposed randomized response technique based on higher order moments of the scrambling variables can be used to estimate multinomial proportions. The choice of the scrambling variables for a particular study may require an expert to decide based on simulation studies or past experience. The FORTRAN codes used in the simulation study are provided in Appendix A.

Generalization to the Case of a Multinomial Distribution

Consider a population Ω consisting of m mutually exclusive groups such that $\Omega = \bigcup_{k=1}^{m} A_k$. Let π_k be the proportion of a sensitive attribute is the k^{th} group. Then extending the proposed randomized response model from Section 2, that a respondent belonging to the k^{th} group is requested to report a random number from the k^{th} scrambling variable S_k for $k = 1,2,....,m$, then $(m-1)$ unbiased estimates of the population proportion π_k for $k = 1,2,..,(m-1)$ are given by

$$\hat{\Pi}_{(m-1)\times 1} = (A^{-1})_{(m-1)\times(m-1)} Z_{(m-1)\times 1} \qquad (4.1)$$

where

$$\hat{\Pi} = \begin{bmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \\ \hat{\pi}_{m-1} \end{bmatrix},$$

$$Z = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} Z_i - E(S_m) \\ \frac{1}{n}\sum_{i=1}^{n} Z_i^2 - E(S_m^2) \\ \\ \frac{1}{n}\sum_{i=1}^{n} Z_i^{m-1} - E(S_m^{m-1}) \end{bmatrix}$$

and

$$A =$$
$$\begin{bmatrix} E(S_1) - E(S_m), E(S_2) - E(S_m), ...E(S_{m-1}) - E(S_m) \\ E(S_1^2) - E(S_m^2), E(S_2^2) - E(S_m^2), ...E(S_{m-1}^2) - E(S_m^2) \\ \\ E(S_1^{m-1}) - E(S_m^{m-1}), E(S_2^{m-1}) - E(S_m^{m-1}), ...E(S_{m-1}^{m-1}) - E(S_m^{m-1}) \end{bmatrix}$$

and, the unbiased estimate of the proportion π_m is given by

$$\hat{\pi}_m = 1 - \sum_{k=1}^{m-1} \hat{\pi}_k \, . \tag{4.2}$$

The variance of $\hat{\Pi}$ is given by

$$V(\hat{\Pi}) = (A^{-1}) V(Z) (A^{-1})^t \tag{4.3}$$

and

$$V(\hat{\pi}_m) = \sum_{k=1}^{m-1} V(\hat{\pi}_k) + 2 \sum_{k<j=1}^{m-1} Cov(\hat{\pi}_k, \hat{\pi}_j) \tag{4.4}$$

where $V(Z)$ denotes the variance-covariance matrix of the scrambled responses which utilizes the higher order moments of the scrambling variables Z_i^l, $t = 1,2,3,....,m$.

Acknowledgements
The authors are thankful to the Editor Professor Sawilowsky and to Dr. Julie M. Smith to bring the original manuscript in the present form. The authors are also thankful to Ms. Rebecca West at TAMUK for editing the original version of the manuscript.

Appendix A: Theorems and Proofs

Theorem 2.1

An unbiased estimator of the population proportion π_1 is given by

$$\hat{\pi}_1 = \frac{\left\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right\}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_3\right\} - (\theta_2 - \theta_3)\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - (\gamma_{002} + \theta_3^2)\right\}}{\Delta} \tag{2.6}$$

Theorem 2.1 Proof

Solving (2.2) and (2.4) for π_1 and using the method of moments proves the theorem.

Theorem 2.2

An unbiased estimator of the population proportion π_2 is given by

$$\hat{\pi}_2 = \frac{(\theta_1 - \theta_3)\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - (\gamma_{002} + \theta_3^2)\right\} - \left\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\right\}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \theta_3\right)}{\Delta} \tag{2.7}$$

Theorem 2.2 Proof

Solving (2.2) and (2.4) for π_2 and using the method of moments proves the theorem.

Theorem 2.3

An unbiased estimator of the population proportion π_3 is given by

$$\hat{\pi}_3 = 1 - \hat{\pi}_1 - \hat{\pi}_2 \, . \tag{2.8}$$

Theorem 2.3 Proof

Theorem 2.3 is proven by taking expected values on both sides of (2.8).

Based on these theorems and proofs, the following lemmas are put forth.

Lemma 2.1

The variance of Z_i is given by

$$\begin{aligned} V(Z_i) &= A_1\pi_1 + A_2\pi_2 + A_{11}\pi_1(1-\pi_1) \\ &\quad + A_{22}\pi_2(1-\pi_2) + A_{12}\pi_1\pi_2 + A_{00} \end{aligned} \tag{2.9}$$

where

$$A_1 = (\gamma_{200} - \gamma_{002}),$$

$$A_2 = (\gamma_{020} - \gamma_{002}),$$

$$A_{00} = \gamma_{002},$$

$$A_{11} = (\theta_1 - \theta_3)^2,$$

$$A_{22} = (\theta_2 - \theta_3)^2$$

and

$$A_{12} = -2(\theta_1 - \theta_3)(\theta_2 - \theta_3).$$

Lemma 2.2

The variance of Z_i^2 is given by

$$\begin{aligned} V(Z_i^2) &= B_1\pi_1 + B_2\pi_2 + B_{11}\pi_1(1-\pi_1) \\ &\quad + B_{22}\pi_2(1-\pi_2) + B_{12}\pi_1\pi_2 + B_{00} \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} B_1 &= (\gamma_{400} - \gamma_{004}) + 4(\gamma_{300}\theta_1 - \gamma_{003}\theta_3) \\ &\quad + 6(\gamma_{200}\theta_1^2 - \gamma_{002}\theta_3^2) + (\theta_1^4 - \theta_3^4) \\ &\quad + (\gamma_{002} + \theta_3^2)^2 - (\gamma_{200} + \theta_1^2)^2 \end{aligned}$$

$$\begin{aligned} B_2 &= (\gamma_{040} - \gamma_{004}) + 4(\gamma_{030}\theta_2 - \gamma_{003}\theta_3) \\ &\quad + 6(\gamma_{020}\theta_2^2 - \gamma_{002}\theta_3^2) + (\theta_2^4 - \theta_3^4) \\ &\quad + (\gamma_{002} + \theta_3^2) - (\gamma_{020} + \theta_2^2)^2 \end{aligned}$$

$$B_{11} = \left[(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\right]^2,$$

$$B_{22} = \left[(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right]^2$$

$$B_{12} = \\ -2\left\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\right\}\left\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right\}$$

and

$$B_{00} = (\gamma_{004} + 4\gamma_{003}\theta_3 + 4\gamma_{002}\theta_3^2 - \gamma_{002}^2).$$

Lemma 2.2 Proof

Based on the definition of variance,

$$\begin{aligned} V(Z_i^2) &= E(Z_i^4) - \{E(Z_i^2)\}^2 \\ &= (\gamma_{400} + 4\gamma_{300}\theta_1 + 6\gamma_{200}\theta_1^2 + \theta_1^4)\pi_1 \\ &\quad + (\gamma_{040} + 4\gamma_{030}\theta_2 + 6\gamma_{020}\theta_2^2 + \theta_2^4)\pi_2 \\ &\quad + (1-\pi_1-\pi_2)(\gamma_{004} + 4\gamma_{003}\theta_3 + 6\gamma_{002}\theta_3^2 + \theta_3^4) \\ &\quad - \begin{bmatrix} \pi_1(\gamma_{200} + \theta_1^2) + \pi_2(\gamma_{020} + \gamma_2^2) \\ +(1-\pi_1-\pi_2)(\gamma_{002} + \theta_3^2) \end{bmatrix}^2 \end{aligned}$$

which, on rearranging reduces to (2.10), and proves the lemma.

Lemma 2.3

The covariance between Z_i and Z_i^2 is given by

$$\begin{aligned} Cov(Z_i, Z_i^2) &= C_1\pi_1 + C_2\pi_2 + C_{11}\pi_1(1-\pi_1) \\ &\quad + C_{22}\pi_2(1-\pi_2) + C_{12}\pi_1\pi_2 + C_{00} \end{aligned} \tag{2.11}$$

where

$$\begin{aligned} C_1 &= (\gamma_{300} - \gamma_{003}) + 3(\theta_1\gamma_{200} - \theta_3\gamma_{002}) \\ &\quad + (\theta_1^3 - \theta_3^2) - \theta_3\{(\gamma_{200} + \theta_1^3) \\ &\quad - (\gamma_{002} + \theta_3^2)\} - (\theta_1 - \theta_3)(\gamma_{200} + \theta_1^2) \end{aligned}$$

$$\begin{aligned} C_2 &= (\gamma_{030} - \gamma_{003}) + 3(\theta_2\gamma_{020} - \theta_3\gamma_{002}) \\ &\quad + (\theta_2^3 - \theta_3^2) - \theta_3\{(\gamma_{020} + \theta_2^2) \\ &\quad - (\gamma_{002} + \theta_3^2)\} - (\theta_2 - \theta_3)(\gamma_{020} + \theta_2^2) \end{aligned}$$

$$C_{11} = (\theta_1 - \theta_3)\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\},$$

$$C_{22} = (\theta_2 - \theta_3)\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\}$$

$$\begin{aligned} C_{12} &= (\theta_2 - \theta_3)\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\} \\ &\quad + (\theta_1 - \theta_3)\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\} \end{aligned}$$

and

$$C_{00} = (\gamma_{003} + 2\theta_3\gamma_{002}).$$

Lemma 2.3 Proof

Based on the definition of covariance,

$$\begin{aligned} Cov(Z_i, Z_i^2) &= E(Z_i^3) - E(Z_i)E(Z_i^2) \\ &= (\gamma_{300} + 3\theta_1\gamma_{200} + \theta_1^3)\pi_1 \\ &\quad + (\gamma_{030} + 3\theta_2\gamma_{020} + \theta_2^3)\pi_2 \\ &\quad + (\gamma_{003} + 3\theta_3\gamma_{002} + \theta_3^3)(1 - \pi_1 - \pi_2) \\ &\quad - \left[\begin{array}{l} \{\pi_1\theta_1 + \pi_2\theta_2 + (1 - \pi_1 - \pi_2)\theta_3\} \\ \times \left\{\begin{array}{l} \pi_1(\gamma_{200} + \theta_1^2) + \pi_2(\gamma_{020} + \theta_2^2) \\ +(1 - \pi_1 - \pi_2)(\gamma_{002} + \theta_3^2) \end{array}\right\} \end{array}\right] \end{aligned}$$

which, on rearranging, reduces to (2.11) and proves the lemma.

Consider the following theorems:

Theorem 2.4

The variance of the unbiased estimator $\hat{\pi}_1$ of the population proportion π_1 is given by

$$V(\hat{\pi}_1) =$$

$$\frac{1}{n\Delta^2}\left[\begin{array}{l} \pi_1\left\{\begin{array}{l} \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)^2 A_1 + B_1(\theta_2 - \theta_3)^2 \\ -2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_1 \end{array}\right\} \\ +\pi_2\left\{\begin{array}{l} \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)^2 A_2 + B_2(\theta_2 - \theta_3)^2 \\ -2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_2 \end{array}\right\} \\ +\pi_1(1 - \pi_1)\left\{\begin{array}{l} \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{11} + \\ (\theta_2 - \theta_3)^2 B_{11} - 2(\theta_2 - \theta_3) \\ \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{11} \end{array}\right\} \\ +\pi_2(1 - \pi_2)\left\{\begin{array}{l} \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{22} \\ +(\theta_2 - \theta_3)^2 B_{22} - 2(\theta_2 - \theta_3) \\ \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{22} \end{array}\right\} \\ +\pi_1\pi_2\left\{\begin{array}{l} \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{12} + \\ (\theta_2 - \theta_3)^2 B_{12} - 2(\theta_2 - \theta_3) \\ \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{12} \end{array}\right\} \\ +\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{00} + \\ (\theta_2 - \theta_3)^2 B_{00} - 2(\theta_2 - \theta_3) \\ \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{00} \end{array}\right] \tag{2.12}$$

Theorem 2.4 Proof

Based on the definition of variance,

$$V(\hat{\pi}_1) =$$

$$\frac{1}{n^2\Delta^2}\left[\begin{array}{l} \{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\}^2 \sum_{i=1}^{n} V(Z_i) \\ +(\theta_2 - \theta_3)^2 \sum_{i=1}^{n} V(Z_i^2) - 2(\theta_2 - \theta_3)\{(\gamma_{020} + \theta_2^2) \\ -(\gamma_{002} + \theta_3^2)\} \sum_{i=1}^{n} Cov(Z_i, Z_i^2) \end{array}\right]$$

Using the lemmas, the following theorems are put forth.

Theorem 2.5

The variance of the unbiased estimator $\hat{\pi}_2$ of the population proportion π_2 is given by

$$V(\hat{\pi}_2)=$$

$$\frac{1}{n\Delta^2}\left[\begin{array}{l}
\pi_1\left\{\begin{array}{l} A_1\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\ +B_1(\theta_1-\theta_3)^2-2(\theta_1-\theta_3) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)C_1 \end{array}\right\} \\
+\pi_2\left\{\begin{array}{l} \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\ A_2+(\theta_1-\theta_3)^2B_2-2(\theta_1-\theta_3) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)C_2 \end{array}\right\} \\
+\pi_1(1-\pi_1)\left\{\begin{array}{l} \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\ A_{11}+(\theta_1-\theta_3)^2B_{11}-2(\theta_1-\theta_3) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2\Big)C_{11} \end{array}\right\} \\
+\pi_2(1-\pi_2)\left\{\begin{array}{l} \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\ A_{22}+(\theta_1-\theta_3)^2B_{22}-2(\theta_1-\theta_3) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2\Big)C_{22} \end{array}\right\} \\
+\pi_1\pi_2\left\{\begin{array}{l} (\theta_1-\theta_3)^2B_{12}+\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\ A_{12}-(\theta_1-\theta_3)\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)C_{12} \end{array}\right\} \\
+A_{00}\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2 \\
+B_{00}(\theta_1-\theta_3)^2-2(\theta_1-\theta_3) \\
\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)C_{00}
\end{array}\right] \tag{2.13}$$

Theorem 2.5 Proof

Based on the definition of variance,

$$V(\hat{\pi}_2)=$$

$$\frac{1}{n^2\Delta^2}\left[\begin{array}{l}
(\theta_1-\theta_3)^2\sum_{i=1}^{n}V(Z_i^2)+ \\
\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)^2\sum_{i=1}^{n}V(Z_i) \\
-2(\theta_1-\theta_3)\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big) \\
\sum_{i=1}^{n}Cov(Z_i,Z_i^2)
\end{array}\right]$$

Theorem 2.6

The covariance between the unbiased estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ is given by

$$Cov(\hat{\pi}_1,\hat{\pi}_2)=$$

$$\frac{1}{n\Delta^2}\left[\begin{array}{l}
\pi_1\left\{\begin{array}{l} C_1\Psi-B_1(\theta_1-\theta_3)(\theta_2-\theta_3) \\ -A_1\Big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big) \end{array}\right\} \\
+\pi_2\left\{\begin{array}{l} C_2\Psi-B_2(\theta_1-\theta_3)(\theta_2-\theta_3) \\ -A_2\Big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big) \end{array}\right\} \\
+\pi_1(1-\pi_1)\left\{\begin{array}{l} C_{11}\Psi-B_{11}(\theta_1-\theta_3)(\theta_2-\theta_3) \\ -A_{11}\Big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big) \end{array}\right\} \\
+\pi_2(1-\pi_2)\left\{\begin{array}{l} C_{22}\Psi-B_{22}(\theta_1-\theta_3)(\theta_2-\theta_3) \\ -A_{22}\Big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big) \\ \Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big) \end{array}\right\} \\
+C_{00}\Psi-B_{00}(\theta_1-\theta_3)(\theta_2-\theta_3) \\
-A_{00}\left\{\begin{array}{l} (\gamma_{020}+\theta_2^2) \\ -(\gamma_{002}+\theta_3^2) \end{array}\right\}\left\{\begin{array}{l} (\gamma_{200}+\theta_1^2) \\ -(\gamma_{002}+\theta_3^2) \end{array}\right\}
\end{array}\right] \tag{2.14}$$

where

$$\begin{array}{l}\Psi=(\theta_1-\theta_3)\Big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big) \\ \qquad+(\theta_2-\theta_3)\Big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big)\end{array}.$$

Theorem 2.6 Proof

Based on the definition of covariance,

$$Cov\left(\hat{\pi}_1, \hat{\pi}_2\right) = \frac{Cov\left[\begin{array}{l} \left((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\right) \\ \sum_{i=1}^{n} Z_i - (\theta_2-\theta_3)\sum_{i=1}^{n} Z_i^2,\ (\theta_1-\theta_3) \\ \sum_{i=1}^{n} Z_i^2 - \begin{pmatrix} (\gamma_{200}+\theta_1^2) \\ -(\gamma_{002}+\theta_3^2) \end{pmatrix} \sum_{i=1}^{n} Z_i \end{array}\right]}{n^2\Delta^2}$$

$$= \frac{1}{n^2\Delta^2}\left[\begin{array}{l} (\theta_1-\theta_3)\left((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\right)Cov\left(Z_i, Z_i^2\right) \\ -(\theta_1-\theta_3)(\theta_2-\theta_3)\sum_{i=1}^{n} Cov\left(Z_i^2, Z_i^2\right) \\ -\begin{pmatrix} (\gamma_{020}+\theta_2^2) \\ -(\gamma_{002}+\theta_3^2) \end{pmatrix}\begin{pmatrix} (\gamma_{200}+\theta_1^2) \\ -(\gamma_{002}+\theta_3^2) \end{pmatrix} \\ \sum_{i=1}^{n} Cov\left(Z_i, Z_i\right) + (\theta_2-\theta_3) \\ \begin{pmatrix} (\gamma_{200}+\theta_1^2) \\ -(\gamma_{002}+\theta_3^2) \end{pmatrix}\sum_{i=1}^{n} Cov(Z_i, Z_i^2) \end{array}\right]$$

These lemmas result in theorem 2.7.

Theorem 2.7

The variance of the unbiased estimator $\hat{\pi}_3$ of the parameter π_3 is given by

$$V(\hat{\pi}_3) = V(\hat{\pi}_1) + V(\hat{\pi}_2) + 2Cov(\hat{\pi}_1, \hat{\pi}_2) \tag{2.15}$$

Theorem 2.7 Proof

Theorem 2.7 is proved based on the definition of variance.

References

Abul-Ela, A. L. A., Greenberg, B. G., & Horvitz, D. G. (1967). A multi-proportion randomized response model. *Journal of the American Statistical Association*, *62*, 990-1008.

Blank, S. G. (2008). *Using the randomized response technique to investigate illegal fishing and contribute to abalone management in Northern California.* A 90 point thesis submitted to Victoria University of Wellington, as partial fulfillment for the degree of Master of Environmental Studies. School of Geography, Environment and Earth Sciences, Victoria University of Wellington.

Bourke, P. D. (1974). Multi-proportions randomized response using the unrelated question technique. *Report No. 74 of the Errors in Survey Research Project.* Institute of Statistics, University of Stockholm (Mimeo).

Bourke, P. D. (1981). On the analysis of some multivariate randomized response designs for categorical data. *Journal of Statistical Planning and Inference*, *5*, 165-170.

Bourke, P. D. (1982). RR multivariate designs for categorical data. *Communications in Statistics – Theory and Methods*, *A*, *11*, 2889-2901.

Bourke, P. D. (1990). Estimating a distribution function for each category of a sensitive variable. *Communications in Statistics – Theory and Methods*, *19*(*9*), 3233-3241.

Bourke, P. D., & Dalenius, T. (1973). Multi-proportions randomized response using a single sample. *Report No. 68 of the Errors in Survey Research Project.* Institute of Statistics, University of Stockholm. (Mimeo).

Bourke, P. D., & Dalenious, T. (1974). RR models with lying. *Technical Report –71.* Institute of Statistics, University of Stockholm. (Mimeo).

Christofides, T. C. (2003). A generalized randomized response technique. *Metrika*, *57*(*2*), 195-200.

Diana, G., & Perri, P. F. (2009a). Estimating sensitive proportion through randomized response procedures based on auxiliary information. *Statistical Papers*, *50*, 661-672.

Diana, G., & Perri, P. F. (2009b). A class of estimators for quantitative sensitive data. *Statistical Papers*, doi: 10.1007/s00362-009-0273-1. (Published online).

Drane, W. (1976). On the theory of randomized responses to two sentitive questions. *Communications in Statistics – Theory and Methods*, *A*, *5*, 565-574.

Esponda, F., & Guerrero, V. M. (2009). Surveys with negative questions for sensitive items. *Statistics and Probability Letters*, *79*, 2456-2461.

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*, 489-505.

Gjestvang, C. R., & Singh, S. (2006). A new randomized response model. *Journal of the Royal Statistical Society*, *Series B*, *68*, 523-530.

Greenberg, B.G., Abul-Ela, Abdel-Latif, A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question RR model: theoretical framework. *J. Amer. Statist. Assoc.*, 64, 520-539.

Guerriero, M., & Sandri, M. F. (2007). A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference*, *137*, 2184-2190.

Hochberg, Y. (1975). Two-stage randomized response scheme for estimating a multinomial. *Communications in Statistics – Theory and Methods*, *4*, 1021-1032.

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

Mangat, N. S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society*, *Series B*, *56*, 93-95.

Mangat, N. S., & Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, *77*(*2*), 439-442.

Mukherjee, R. (1981). Inference on confidential characters from survey data. *Calcutta Statistical Association Bulletin*, *30*, 77-88.

Mukhopadhyay, P. (1980). On the estimation of some confidential characters from survey. *Calcutta Statistical Association Bulletin*, *29*, 77-88.

Raghavarao, D., & Federer, W. T. (1979). Block total response as an alternative to the randomized response method. *Journal of the Royal Statistical Society*, *Series B*, *41*, 40-45.

Silva, L. C. (1983). On the generalized randomized response model with polychotomous variables. *Review of Investment Operac.*, *4*, *III*, 75-100.

Singh, S., & Chen, C. C. (2009). Utilization of higher order moments of scrambling variables in randomized response sampling. *Journal of Statistical Planning and Inference*, *139*, 3377-3380.

Singh, S., & Kim, J-M. (2011). A pseudo-empirical log-likelihood estimator using scrambled responses. *Statistics and Probability Letters*, *81*, 345-351.

Singh, S., Kim, J-M., & Grewal, I. S. (2008). Imputing and Jackknifing scrambled responses. *Metron*, *LXVI*(*2*), 183-204.

Tamhane, A. C. (1981). Randomized response techniques for multiple attributes. *Journal of the American Statistical Association*, *76*, 916-923.

Tan, M. T., Tian, G-L., & Tang, M-L. (2009). Sample surveys with sensitive questions: A nonrandomized response approach. *The American Statistician*, *63*(*1*), 9-16.

Tracy, D. S., & 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 der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, *28*, 505-537.

van den Hout, A., & van der Heijden, P. G. M. (2002). Randomized response, statistical disclosure control and misclassification: a review. *International Statistical Review*, *70*, 269-288.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2004). *Measurement in Nursing and Health Research*, *3rd Ed*. New York, NY: Springer.

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