Int. J. Contemp. Math. Sciences, Vol. 8, 2013, no. 5, 243 - 248
HIKARI Ltd, www.m-hikari.com

Geometric Distribution as a Randomization Device: Implemented to the Kuk's Model

Sarjinder Singh

Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363, USA
sarjinder@yahoo.com

Inderjit Singh Grewal

Department of Mathematics, Statistics and Physics
Punjab Agricultural University
Punjab, 141004, India
Isg1969@pau.edu

Abstract

In this paper, an interesting improvement in the Kuk (1990) randomized response model has been suggested. Recently, Guerriero and Sandri (2007) have shown that the family of estimators proposed by Kuk (1990) fairs better than the Simmons' family in terms of efficiency and protection. The proposed improved method is shown to have more protection and efficiency than the Kuk (1990) model while doing real surveys in practice.

Mathematics Subject Classification: 62D05

Keywords: Estimation of proportion; Randomized response sampling, Respondents' protection.

1. INTRODUCTION

The problem of estimation of proportion of a sensitive character using a randomization device in survey sampling is well known since Warner (1965). A detailed review and

applications of such techniques can be had from Fox and Tracy (1986). Kuk (1990) introduced an ingenious randomized response model in which if a respondent belongs to a sensitive group A, then he/she is instructed to use a deck of cards having θ_1 proportion of cards with the statement, "I belong to group A" and, and if the respondent belongs to non-sensitive group A^c then the respondent is requested to use a different deck of cards having θ_2 proportion of cards with the statement, "I do not belong to group A". Assume π be the true proportion of persons belonging to the sensitive group A. Obviously, the probability of 'yes' answer in the Kuk (1990) model is given by:

$$\theta_{kuk} = \theta_1\pi + (1-\pi)\theta_2 \tag{1.1}$$

Further assume a simple random with replacement sample of n respondents is selected from the population, and n_1 be the number of observed "yes" answers. The number of people n_1 that answer "yes" is binomially distributed with parameters $\theta_{kuk} = \theta_1\pi + (1-\pi)\theta_2$ and n. For the Kuk (1990) model, an unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_{kuk} = \frac{n_1/n - \theta_2}{\theta_1 - \theta_2},\ \theta_1 \neq \theta_2 \tag{1.2}$$

with variance, given by:

$$V(\hat{\pi}_{kuk}) = \frac{\theta_{kuk}(1-\theta_{kuk})}{n(\theta_1 - \theta_2)^2} \tag{1.3}$$

A variety of randomized response models such as the Warner (1965), Mangat (1994), Mangat and Singh (1990) are special cases of Kuk (1990) model. Mangat (1994) model is further improved by Gestvang and Singh (2006). Guerriero and Sandri (2007) have shown that the family of randomized response models proposed by Kuk (1990) fairs better than the Simmons' family in terms of efficiency and protection. In the next section, we suggest an interesting improvement in the Kuk (1990) model. Chaudhuri (2011) has given a decent review on randomized response sampling.

2. PROPOSED RANDOMIZED RESPONSE MODEL

In the proposed randomized response model, each respondent selected in the sample is provided with two decks of cards in the same way as in Kuk (1990) model. In the first deck of cards, let θ_1^* be the proportion of cards with the statement, "I belong to group *A*" and $(1-\theta_1^*)$ be the proportion of cards with the statement, "I do not belong to group *A*". In the second deck of cards, let θ_2^* be the proportion of cards with the statement, "I do not belong to group A" and $(1-\theta_2^*)$ be the proportion of cards with the statement, "I belong to group *A*". Up to here, it is same as the Kuk (1990) randomized response model. Now in the proposed model, if a respondent belongs to group A, he/she is instructed to draw cards, one-by-one using *with replacement*, from the first deck of cards until he/she gets the first card bearing the

statement of his/her own status, and requested to report the total number of cards, say X, drawn by him/her to obtain the first card of his/her own status. If a respondent belongs to group A^c, he/she is instructed to draw cards, one-by-one using with replacement, from the second deck of cards until he/she gets the first card bearing the statement of his/her own status, and requested to report the total number of cards, say Y, drawn by him/her to obtain the first card of his/her own status.

Obviously $X \sim G(\theta_1^*)$ and $Y \sim G(\theta_2^*)$, that is, X and Y follow Geometric distributions with parameters θ_1^* and θ_2^* respectively, because cards are drawn based on with replacement sampling. Let Z_i be the number of cards reported by the ith respondent, then we have the following theorem.

Theorem 2.1. An unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_p = \frac{\frac{\theta_1^*\theta_2^*}{n}\sum_{i=1}^{n} Z_i - \theta_1^*}{\theta_2^* - \theta_1^*}\ ,\ \theta_1^* \neq \theta_2^* \tag{2.1}$$

Proof. Given $X \sim G(\theta_1^*)$ and $Y \sim G(\theta_2^*)$, therefore, we have:

$$E(Z_i) = \frac{\pi}{\theta_1^*} + \frac{(1-\pi)}{\theta_2^*} \tag{2.2}$$

By taking the expected value on both sides of (2.1) and using (2.2) we have:

$$E(\hat{\pi}_p) = E\left[\frac{\frac{\theta_1^*\theta_2^*}{n}\sum_{i=1}^{n} Z_i - \theta_1^*}{\theta_2^* - \theta_1^*}\right] = \frac{\frac{\theta_1^*\theta_2^*}{n}\sum_{i=1}^{n} E(Z_i) - \theta_1^*}{\theta_2^* - \theta_1^*}$$

$$= \frac{\frac{\theta_1^*\theta_2^*}{n}\sum_{i=1}^{n}\left\{\frac{\pi}{\theta_1^*} + \frac{(1-\pi)}{\theta_2^*}\right\} - \theta_1^*}{\theta_2^* - \theta_1^*} = \frac{\theta_2^*\pi + (1-\pi)\theta_1^* - \theta_1^*}{\theta_2^* - \theta_1^*} = \pi$$

which proves the theorem.

Theorem 2.2. The variance of the proposed estimator $\hat{\pi}_p$ is given by:

$$V(\hat{\pi}_p) = \frac{\pi(1-\pi)}{n} + \frac{\theta_2^{*2}(1-\theta_1^*)\pi + \theta_1^{*2}(1-\theta_2^*)(1-\pi)}{n(\theta_2^* - \theta_1^*)^2} \tag{2.3}$$

Proof. Since the responses are independent, thus we have:

$$V(\hat{\pi}_p) = V\left[\frac{\frac{\theta_1^*\theta_2^*}{n}\sum_{i=1}^{n} Z_i - \theta_1^*}{\theta_2^* - \theta_1^*}\right] = V\left[\frac{(\theta_1^*\theta_2^*)^2 \sum_{i=1}^{n} V(Z_i)}{n^2(\theta_2^* - \theta_1^*)^2}\right] \tag{2.4}$$

Now we have:

$$V(Z_i) = E(Z_i^2) - (E(Z_i))^2 \tag{2.5}$$

Given $X \sim G(\theta_1^*)$ and $Y \sim G(\theta_2^*)$, therefore, we have

$$E(Z_i^2) = \pi\left(\frac{2-\theta_1^*}{\theta_1^{*2}}\right) + (1-\pi)\left(\frac{2-\theta_2^*}{\theta_2^{*2}}\right) \tag{2.6}$$

Using (2.2) and (2.6) in (2.5), and then using it in (2.4) we have (2.3). Hence the theorem.

In the next section, we make rules about the choice of θ_1^* and θ_2^* in the proposed randomized response device by keeping the efficiency of the proposed estimator and the protection of respondents as the major issues to be considered in real surveys.

3. EFFICIENCY COMPARISONS

The percent relative efficiency (RE) of the proposed model over the Kuk (1990) model is defined as:

$$\text{RE} = \frac{V(\hat{\pi}_{kuk})}{V(\hat{\pi}_p)} \times 100\% \tag{3.1}$$

Keep in mind that if $\theta_1 + \theta_2 = 1$, then the Kuk (1990) model is same as the Warner (1965) model, and if $\theta_1 = 1$ and $0 < \theta_2 < 1$, then the Kuk (1990) model is same at the Mangat (1994) model. Now in the Kuk (1990) model, the choice of θ_1 and θ_2 is such that the difference $|\theta_2 - \theta_1|$ should be maximum to reduce the variance of the estimator $\hat{\pi}_{kuk}$ in (1.2) by keeping the respondents' cooperation in mind. The RE value is free from the sample size. In an investigation, we decided to keep $\theta_1 = 0.7$ and $\theta_2 = 0.2$ which could be considered as a good choice of randomization device parameters in the Kuk's model by keeping the respondents' cooperation in mind and maximum difference between θ_1 and θ_2. We searched for choice of θ_1^* and θ_2^* for different values of π such that the percent relative efficiency value remains higher than 100%. The results so obtained are presented in Table 3.1.

Table 3.1. The values of θ_1^*, θ_2^* and RE values for $\theta_1 = 0.7$, $\theta_2 = 0.2$ and $0.1 \le \pi \le 0.9$.

π	θ_1^*	θ_2^*	RE	π	θ_1^*	θ_2^*	RE	π	θ_1^*	θ_2^*	RE	π	θ_1^*	θ_2^*	RE
0.1	0.1	0.3	166.7	0.2	0.5	0.9	106.2	0.5	0.8	0.1	117.9	0.8	0.5	0.2	122.7
		0.4	241.9	0.3	0.1	0.4	123.5			0.2	101.8		0.6	0.1	222.2
		0.5	289.9			0.5	139.2		0.9	0.1	120.7		0.6	0.2	160.0
		0.6	320.5			0.6	149.2			0.2	108.2		0.7	0.1	233.2
		0.7	340.9			0.7	156.0	0.6	0.1	0.7	102.2			0.2	187.5
		0.8	355.1			0.8	160.9			0.8	105.6			0.3	132.6
		0.9	365.3			0.9	164.6			0.9	108.2		0.8	0.1	241.0
	0.2	0.5	146.4		0.2	0.6	111.0		0.4	0.1	108.7			0.2	207.7
		0.6	208.3			0.7	127.5		0.5		121.8			0.3	166.7
		0.7	258.6			0.8	139.5		0.6		130.2			0.4	120.0
		0.8	297.4			0.9	148.6		0.7		136.1		0.9	0.1	246.8
		0.9	326.7		0.3	0.8	114.0			0.2	111.6			0.2	222.7
	0.3	0.6	102.7			0.9	130.0		0.8	0.1	140.3			0.3	193.9
		0.7	164.4		0.4	0.9	108.6			0.2	121.6			0.4	160.0
		0.8	224.6	0.4	0.1	0.4	104.4			0.3	100.0			0.5	121.3
		0.9	277.8			0.5	116.9		0.9	0.1	143.6	0.9	0.3	0.1	202.2
		0.8	147.1			0.6	125.0			0.2	129.2		0.4	0.1	293.6
		0.9	219.3			0.7	130.6			0.3	113.0		0.5	0.1	351.7
	0.5	0.9	155.0			0.8	134.7	0.7	0.3	0.1	105.3			0.2	177.7
0.2	0.1	0.3	119.2			0.9	137.8		0.4		134.4		0.6	0.1	388.9
		0.4	157.5		0.2	0.7	107.1		0.5		151.4			0.2	252.8
		0.5	180.2			0.8	116.8		0.6		162.3			0.3	124.7
		0.6	194.4			0.9	124.1			0.2	120.7		0.7	0.1	413.6
		0.7	204.1		0.3	0.9	108.5		0.7	0.1	169.7			0.2	313.8
		0.8	210.9		0.8	0.1	101.4			0.2	138.7			0.3	199.5
		0.9	215.9		0.9	0.1	103.9			0.3	101.9			0.4	101.5
	0.2	0.5	107.4	0.5	0.1	0.5	102.2		0.8	0.1	175.1		0.8	0.1	430.8
		0.6	140.0			0.6	109.3			0.2	151.8			0.2	360.8
		0.7	164.1			0.7	114.2			0.3	124.1			0.3	272.5
		0.8	181.7			0.8	117.9		0.9	0.1	179.1			0.4	178.4
		0.9	194.9			0.9	120.7			0.2	161.7		0.9	0.1	443.2
	0.3	0.7	116.1		0.2	0.8	101.8			0.3	141.4			0.2	396.4
		0.8	145.8			0.9	108.2			0.4	118.1			0.3	337.0
		0.9	169.7		0.5	0.1	102.2	0.8	0.3	0.1	136.2			0.4	266.1
	0.4	0.8	105.0		0.6	0.1	109.3		0.4		180.0			0.5	188.1
		0.9	140.0		0.7	0.1	114.2		0.5		205.9			0.6	112.4

Discussion of Results: Assume the investigator expects the value of π close to 0.1, then instead of choosing $\theta_1 = 0.7$ and $\theta_2 = 0.2$, if the investigator decides to do survey with $\theta_1^* = 0.1$ and $\theta_2^* = 0.3$, the relative efficiency remains 166.7%; and the relative efficiency may be increased to 365.3% by choosing $\theta_2^* = 0.9$. Note that the interviewer does not know which deck of cards has been used by the interviewee thus it will be hard to guess from which deck the number of reported cards is drawn. The choice $\theta_1^* = 0.3$ and $\theta_2^* = 0.7$ may give better cooperation with the respondents, but the percent relative efficiency remains 164.4%. It is very interesting to note that the choice $\theta_1^* = 0.7$ and $\theta_2^* = 0.2$ gives less efficient results. Again assume the investigator expects the value of π close to 0.2, then instead of choosing $\theta_1 = 0.7$ and $\theta_2 = 0.2$, if the investigator decides to do survey with $\theta_1^* = 0.1$ and $\theta_2^* = 0.3$, the relative efficiency remains 119.2%; and the relative efficiency may be increased to 215.9% by choosing $\theta_2^* = 0.9$. In the same way the rest of the results in Table 3.1 can be interpreted. We conclude that it is always possible to make the suggested geometric distribution randomization device estimator more efficient than the Kuk's model.

REFERENCES

[1] A. Chaudhuri. Randomized response and indirect questioning techniques in surveys. CRC Press, London, (2011).

[2] J. A. Fox. and P.E. Tracy. *Randomized response: A method for sensitive surveys.* SAGE Publications. (1986)

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

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

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

[6] N.S. Mangat An improved randomized response strategy. *J. R. Statist. Soc*., B, 56 (1994), 93-95.

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

[8] S.L. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Stat. Assoc.*, 60, (1965), 63-69.

Received: November, 2012