Statistics and Applications **{ISSN 2452-7395 (online)}**
Volume 15, Nos. 1&2, 2016 (New Series), pp 101-116

Empirical Bayes Estimation Method in Some Randomized Response Techniques

Purnima Shaw and Arijit Chaudhuri
Indian Statistical Institute, Kolkata, India

Received: August 24, 2016; Revised: January 20, 2017; Accepted: January 21, 2017

Abstract

The proportion of people bearing a stigmatizing characteristic is usually estimated utilizing survey data gathered through randomized response techniques. Using such data it is shown how alternatively the population mean of the unknowable prior probabilities assignable to the people anticipated to bear such a sensitive characteristic may be estimated promisingly with an empirical Bayesian approach.

Keywords: Empirical Bayesian approach, General sampling scheme, Randomized Response

1. Introduction

Empirical Bayes estimation of finite population proportion of a sensitive characteristic using Warner's (1965) Randomized Response(RR) data was initiated by Winkler and Franklin (1979) by postulating a Beta prior probability distribution. Pitz (1980) extended this method to Simmons'(1969) RR model. Following this, there have been several published works regarding this all of which are based on the assumption of a Beta prior or a truncated Beta prior distribution and the selection of units is according to Simple Random Sampling With Replacement. Chaudhuri and Christofides (2013) in their monograph illustrated a new procedure to estimate a finite population proportion of a sensitive characteristic using Warner's RR data by applying empirical Bayes approach when the units are sampled from a finite population using a general sampling scheme. In this paper, the above mentioned method is extended to Warner's (1965) follow-up Randomized Response Techniques (RRT's), which are Kuk's (1990), Christofides'(2003), Mangat and Singh's (1990), Forced Response, Unrelated Question and Hussain and Shabbir's(2009) Models. Comparison of the present method with the original method of estimation has been reported, when applicable.

Consider a finite population denoted by $U=(1,2,\dots,i,\dots N)$ of N units. A sample s is chosen from U with a pre-assigned probability $p(s)$ with first order inclusion probability $\pi_i = \sum_{s \ni i} p(s)>0 \ \forall\ i$ and second order inclusion probability $\pi_{ij}=\sum_{s\ni i,j} p(s)>0 \ \forall\ i \neq j$.

Let y_i = 1 if unit i bears a stigmatizing characteristic A
= 0 if i bears the complement of A.

E_P and E_R denote the design and RR based Expectation operators.

V_P and V_R denote the design and RR based Variance operators. Also, $E = E_R E_P = E_P E_R$ and

$$V = V_R E_P + E_R V_P = V_P E_R + E_P V_R$$

are overall Expectation and Variance operators respectively.

Corresponding Author: Purnima Shaw
E-mail: purnimashaw2011@gmail.com

Let us describe a method of empirical Bayes estimation using RR data obtained from Warner's(1965) RRT.

A box containing similar cards marked A and A^c in proportions $p(\neq 0.5)$ and (1-p) respectively is provided to a sampled person i. The participant's response is:
I_i = 1 if the card type drawn by i matches his/her characteristic
= 0 if there is no match

Then $r_i = \frac{I_i-(1-p)}{2p-1}$ (1.1)

is an unbiased estimator of y_i such that $E_R(r_i)=y_i$ and

$V_R(r_i) = \frac{p(1-p)}{(2p-1)^2} = v_R(r_i)$, (1.2)

vide Chaudhuri (2011).

Let L_i be the prior Prob [y_i = 1]

$L_i(1)$ = Posterior Prob $[y_i = 1|I_i = 1] = \frac{P[I_i=1|y_i=1]L_i}{P[I_i=1|y_i=1]L_i+P[I_i=1|y_i=0](1-L_i)}$

$= \frac{pL_i}{pL_i+(1-p)(1-L_i)} = \frac{pL_i}{(1-p)+(2p-1)L_i}$ which leads to $\frac{1}{L_i(1)} = \frac{1-p}{pL_i} + \frac{2p-1}{p}$

giving $L_i = \frac{1-p}{p}\left(\frac{1}{L_i(1)} - \frac{2p-1}{p}\right)^{-1}$. (1.3)

By using a somewhat simplified empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$ and hence,

$\hat{L}_i = \frac{1-p}{p}\left(\frac{1}{r_i} - \frac{2p-1}{p}\right)^{-1}$ (1.4)

$=\alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$ where $\alpha=\frac{1-p}{p}$ and $\beta=\frac{2p-1}{p}$

$= \alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i) \cong L_i$.

Expanding $\hat{L}_i$ (=$f(r_i)$) about L_i (=$f(L_i(1))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i=L_i(1)}(r_i - L_i(1))$

$=L_i + \frac{\alpha}{(1-L_i(1)\beta)^2}(r_i - L_i(1))$.

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1)\beta)^4} V_R(r_i)$. (1.5)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ v_R(r_i) = \frac{p^3(1-p)^3}{[p-r_i(2p-1)]^4(2p-1)^2} = v_R(\hat{L}_i)$. (1.6)

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible.

Proceeding similarly as above,

$L_i(0)$ = Posterior Prob$[y_i = 1|I_i = 0] = \frac{P[I_i=0|y_i=1]L_i}{P[I_i=0|y_i=1]L_i+P[I_i=0|y_i=0](1-L_i)}$

$= \frac{(1-p)L_i}{p+(1-2p)L_i}$

$$\Rightarrow\ L_i = \frac{p}{1-p}\left(\frac{1}{L_i(0)} - \frac{1-2p}{1-p}\right)^{-1}. \qquad (1.7)$$

$$\hat{L}_i = \frac{p}{1-p}\left(\frac{1}{r_i} - \frac{1-2p}{1-p}\right)^{-1}. \qquad (1.8)$$

$$E_R(\hat{L}_i) \cong L_i.$$

$$\hat{V}_R(\hat{L}_i) = \frac{p^3(1-p)^3}{[(1-p)+r_i(2p-1)]^4(2p-1)^2} = v_R(\hat{L}_i). \qquad (1.9)$$

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible.

So, the above empirical Bayes estimation method is not applicable with Warner's RR device.

But it may be interesting to examine how the method may work out with other RRT's like Kuk's (1990), Christofides'(2003), Mangat and Singh's (1990), Forced Response, Unrelated Question and Hussain and Shabbir's (2009) Models.

2.(a) Kuk's Model: Two boxes containing red coloured identical cards in proportions p_1 and $p_2(\neq p_1)$, the remaining cards being non-red are provided to a sampled person *i*. If *i* bears A then he/she is to draw K cards by SRSWR method from the first box; otherwise from the second one, but the investigator should not be informed about the box used by *i*. The participant's response is $I_i=f_i$, number of red cards out of the K cards drawn.

Then $r_i = \frac{\frac{f_i}{K}-p_2}{p_1-p_2}$ (2.a.1)

is an unbiased estimator of y_i such that $E_R(r_i)=y_i$ and

$$V_R(r_i) = \frac{1-p_1-p_2}{K(p_1-p_2)}\, y_i + \frac{p_2(1-p_2)}{K(p_1-p_2)^2} \qquad (2.a.2)$$

and $\hat{V}_R(r_i) = v_R(r_i) = \frac{1-p_1-p_2}{K(p_1-p_2)}\, r_i + \frac{p_2(1-p_2)}{K(p_1-p_2)^2}$ (2.a.3)

vide Chaudhuri (2011).

Let L_i be the prior Prob$[y_i = 1]$

$$L_i(f_i) = \text{Posterior Prob}[y_i = 1 | I_i = f_i] = \frac{P[I_i=f_i|y_i=1]L_i}{P[I_i=f_i|y_i=1]L_i + P[I_i=f_i|y_i=0](1-L_i)}$$

$$= \frac{[p_1^{f_i}(1-p_1)^{k-f_i}]L_i}{[p_1^{f_i}(1-p_1)^{k-f_i}]L_i + [p_2^{f_i}(1-p_2)^{k-f_i}](1-L_i)}$$

$$= \frac{[p_1^{f_i}(1-p_1)^{k-f_i}]L_i}{[p_2^{f_i}(1-p_2)^{k-f_i}] + \left[p_1^{f_i}(1-p_1)^{k-f_i} - p_2^{f_i}(1-p_2)^{k-f_i}\right]L_i}$$ which leads to

$$\frac{1}{L_i(f_i)} = \frac{p_2^{f_i}(1-p_2)^{k-f_i}}{[p_1^{f_i}(1-p_1)^{k-f_i}]L_i} + \frac{p_1^{f_i}(1-p_1)^{k-f_i} - p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}$$

giving $$L_i = \frac{p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}\left(\frac{1}{L_i(f_i)} - \frac{p_1^{f_i}(1-p_1)^{k-f_i} - p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}\right)^{-1}. \qquad (2.a.4)$$

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(f_i)$ and hence,

$$\hat{L}_i = \frac{p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}\left(\frac{1}{r_i} - \frac{p_1^{f_i}(1-p_1)^{k-f_i} - p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}\right)^{-1} \quad (2.a.5)$$

$= \alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}$ and $\beta = \frac{p_1^{f_i}(1-p_1)^{k-f_i} - p_2^{f_i}(1-p_2)^{k-f_i}}{p_1^{f_i}(1-p_1)^{k-f_i}}$

$= \alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i) \cong L_i$.

Expanding $\hat{L}_i$ ($=f(r_i)$) about L_i ($=f(L_i(f_i))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i = L_i(f_i)}(r_i - L_i(f_i))$

$$= L_i + \frac{\alpha}{(1-L_i(f_i)\beta)^2}(r_i - L_i(f_i))$$

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(f_i)\beta)^4} V_R(r_i)$ (2.a.6)

$$\hat{V}_R(\hat{L}_i) = \upsilon_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4} \quad \upsilon_R(r_i) = \frac{p_1^{2f_i} p_2^{2f_i}(1-p_1)^{2(k-f_i)}(1-p_2)^{2(k-f_i)}}{\left[p_1^{f_i}(1-p_1)^{k-f_i}(1-r_i) + r_i p_2^{f_i}(1-p_2)^{k-f_i}\right]^4}\left[\frac{1-p_1-p_2}{K(p_1-p_2)} r_i + \frac{p_2(1-p_2)}{K(p_1-p_2)^2}\right]. \quad (2.a.7)$$

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

2.(b) Christofides' Model: A box containing $M(\geq 2)$ identical cards marked $1,2\ldots,j,\ldots,M$ in proportions $p_1, p_2,\ldots,p_j,\ldots,p_M$ such that $\sum_{j=1}^{M} p_j = 1$. A sampled person is asked to choose a card and report the value (say K) if he/she bears A^C; otherwise M+1-K is to be reported without disclosing the original value obtained in the card. The participant's response is :
$z_i = (M+1-K)y_i + K(1-y_i)$, $K=1,2,\ldots,M$.

Consider $\sum_{K=1}^{M} K p_K = \mu$ and $\sum_{K=1}^{M} K^2 p_K - \mu^2 = \sigma^2$.

Then

$$r_i = \frac{z_i - \mu}{M+1-2\mu} \quad (2.b.1)$$

is an unbiased estimator of y_i such that $E_R(r_i) = y_i$ and

$$V_R(r_i) = \frac{\sigma^2}{(M+1-2\mu)^2} = \upsilon_R(r_i), \quad (2.b.2)$$

vide Chaudhuri (2011).

let L_i be the prior Prob$[y_i = 1]$

$L_i(K)$ = Posterior Prob$[y_i = 1 | z_i = K] = \frac{P[z_i = K|y_i = 1]L_i}{P[z_i = K|y_i=1]L_i + P[z_i = K|y_i = 0](1-L_i)}$

$= \frac{p_{M+1-K}L_i}{p_{M+1-K}L_i + p_K(1-L_i)} = \frac{p_{M+1-K}L_i}{p_K + (p_{M+1-K} - p_K)L_i}$ which leads to $\frac{1}{L_i(K)} = \frac{p_K}{p_{M+1-K}L_i} + \frac{p_{M+1-K} - p_K}{p_{M+1-K}}$

giving $L_i = \frac{p_K}{p_{M+1-K}}\left(\frac{1}{L_i(K)} - \frac{p_{M+1-K} - p_K}{p_{M+1-K}}\right)^{-1}$. (2.b.3)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$ and hence,

$$\hat{L}_i = \frac{p_K}{p_{M+1-K}}\left(\frac{1}{r_i} - \frac{p_{M+1-K}-p_K}{p_{M+1-K}}\right)^{-1} \quad (2.b.4)$$

$= \alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{p_K}{p_{M+1-K}}$ and $\beta = \frac{p_{M+1-K}-p_K}{p_{M+1-K}}$

$= \alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i) \cong L_i$.

Expanding $\hat{L}_i$ (=$f(r_i)$) about L_i (=$f(L_i(K))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i=L_i(K)}(r_i - L_i(K))$

$=L_i + \frac{\alpha}{(1-L_i(K)\beta)^2}(r_i - L_i(K))$.

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(K)\beta)^4} V_R(r_i)$ (2.b.5)

$$\hat{V}_R(\hat{L}_i) = \upsilon_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ \upsilon_R(r_i) = \frac{{p_K}^2 {p_{M+1-K}}^2}{[p_{M+1-K}(1-r_i)+r_i p_K]^4} \frac{\sigma^2}{(M+1-2\mu)^2}. \quad (2.b.6)$$

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

2.(c) Mangat and Singh' Model: A box containing identical cards marked T and R in proportions p_1 and $(1-p_1)$ respectively and another box containing A and A^C in proportions $p_2(\neq 0.5)$ and $(1-p_2)$ respectively are provided to a sampled person i. If a 'T' marked card is chosen from the first box then i reports the true value y_i and if an 'R' marked card is chosen then i chooses another card from the second box and reports 1 if there is a match otherwise 0 without disclosing the box number. The participant's response is:

$z_i = y_i$ if the card chosen from the first box is marked 'T'

$= I_i$ if the card chosen from the first box is marked 'R'

where $I_i = 1$ if the card type drawn from the second box matches the individual's characteristic

$= 0$ if there is no match

Then $r_i = \frac{z_i-(1-T)(1-p)}{T+(1-T)(2p-1)}$ (2.c.1)

is an unbiased estimator of y_i such that $E_R(r_i)=y_i$ and

$V_R(r_i) = \frac{(1-T)(1-p)[T+p(1-T)]}{[T+(1-T)(2p-1)]^2} = \upsilon_R(r_i)$, (2.c.2)

vide Chaudhuri(2011).

let L_i be the prior Prob$[y_i = 1]$

$L_i(1)$ = Posterior Prob$[y_i = 1|z_i = 1] = \frac{P[z_i=1|y_i=1]L_i}{P[z_i=1|y_i=1]L_i+P[z_i=1|y_i=0](1-L_i)}$

$= \frac{[T+(1-T)p]L_i}{[T+(1-T)p]L_i+[(1-T)(1-p)](1-L_i)} = \frac{[T+(1-T)p]L_i}{(1-T)(1-p)+[T+(1-T)(2p-1)]L_i}$ which leads to

$$\frac{1}{L_i(1)} = \frac{(1-T)(1-p)}{[T+(1-T)p]L_i} + \frac{[T+(1-T)(2p-1)]}{[T+(1-T)p]}$$

giving $L_i = \frac{(1-T)(1-p)}{[T+(1-T)p]}\left(\frac{1}{L_i(1)} - \frac{[T+(1-T)(2p-1)]}{[T+(1-T)p]}\right)^{-1}$. (2.c.3)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$

Hence, $\hat{L}_i=\frac{(1-T)(1-p)}{[T+(1-T)p]}\left(\frac{1}{r_i}-\frac{[T+(1-T)(2p-1)]}{[T+(1-T)p]}\right)^{-1}$ (2.c.4)

$=\alpha\left(\frac{1}{r_i}-\beta\right)^{-1}$, where $\alpha=\frac{(1-T)(1-p)}{[T+(1-T)p]}$ and $\beta=\frac{(1-T)(1-p)}{[T+(1-T)p]}$

$=\alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i)\cong L_i$.

Expanding $\hat{L}_i$ (=f(r_i)) about L_i (=f($L_i(1)$)) using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i=L_i(1)}(r_i - L_i(1))$

$=L_i + \frac{\alpha}{(1-L_i(1)\beta)^2}(r_i - L_i(1))$

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1)\beta)^4} V_R(r_i)$ (2.c.5)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ \upsilon_R(r_i) = \frac{(1-T)^2(1-p)^2[T+(1-T)p]^2}{[T+(1-T)p-r_i\{T+(1-T)(2p-1)\}]^4}\frac{(1-T)(1-p)[T+p(1-T)]}{[T+(1-T)(2p-1)]^2}$ (2.c.6)

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible.

Proceeding similarly as above,

$L_i(0)$ = Posterior Prob$[y_i = 1|z_i = 0] = \frac{P[z_i=0|y_i=1]L_i}{P[z_i=0|y_i=1]L_i+P[z_i=0|y_i=0](1-L_i)}$

$=\frac{(1-T)(1-p)L_i}{(1-T)(1-p)L_i+[T+(1-T)p](1-L_i)} = \frac{(1-T)(1-p)L_i}{T+(1-T)p+[(1-T)(1-2p)-T]L_i}$ which leads to

$\frac{1}{L_i(0)} = \frac{T+(1-T)p}{(1-T)(1-p)L_i} + \frac{[(1-T)(1-2p)-T]}{(1-T)(1-p)}$

giving $L_i = \frac{T+(1-T)p}{(1-T)(1-p)}\left(\frac{1}{L_i(0)} - \frac{[(1-T)(1-2p)-T]}{(1-T)(1-p)}\right)^{-1}$. (2.c.7)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$.

Hence, $\hat{L}_i=\frac{T+(1-T)p}{(1-T)(1-p)}\left(\frac{1}{r_i}-\frac{[(1-T)(1-2p)-T]}{(1-T)(1-p)}\right)^{-1}$ (2.c.8)

$=\alpha\left(\frac{1}{r_i}-\beta\right)^{-1}$, where $\alpha=\frac{T+(1-T)p}{(1-T)(1-p)}$ and $\beta=\frac{[(1-T)(1-2p)-T]}{(1-T)(1-p)}$

$=\alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i)\cong L_i$.

Expanding $\hat{L}_i$ (=$f(r_i)$) about L_i (=$f(L_i(0))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i=L_i(0)}(r_i - L_i(0))$

$=L_i + \frac{\alpha}{(1-L_i(0)\beta)^2}(r_i - L_i(0))$

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(0)\beta)^4} V_R(r_i)$ (2.c.9)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ \upsilon_R(r_i) = \frac{(1-T)^2(1-p)^2[T+(1-T)p]^2}{[(1-T)(1-p)-r_i\{(1-T)(1-2p)-T\}]^4}\frac{(1-T)(1-p)[T+p(1-T)]}{[T+(1-T)(2p-1)]^2}$ (2.c.10)

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible.

2.(d) Forced Response Model: A box containing identical cards marked "Yes", "No" and "Genuine" in proportions p_1(>0), p_2(>0) and (1-p_1-p_2) respectively such that $p_1+p_2<1$, is provided to a sampled person i. The person is to reply "yes" if a "Yes" marked card is drawn, "no" if a "No" marked card is drawn and genuinely says "yes" or "no" accordingly as he/she bears A or A^C, if a "Genuine" marked card is drawn. The participant's response is:

I_i = 1 if i responds "yes"

= 0 if i responds "no"

Then $r_i = \frac{I_i - p_1}{(1-p_1-p_2)}$ (2.d.1)

is an unbiased estimator of y_i *such that* $E_R(r_i)=y_i$ and

$$V_R(r_i) = \frac{p_1(1-p_1)+y_i(1-p_1-p_2)(p_2-p_1)}{(1-p_1-p_2)^2} \quad (2.d.2)$$

and $\hat{V}_R(r_i) = \upsilon_R(r_i) = \frac{p_1(1-p_1)+r_i(1-p_1-p_2)(p_2-p_1)}{(1-p_1-p_2)^2}$ (2.d.3)

vide Chaudhuri (2011).

let L_i be the prior Prob $[y_i = 1]$

$L_i(1)$ = Posterior Prob $[y_i = 1|I_i = 1] = \frac{P[I_i=1|y_i=1]L_i}{P[I_i=1|y_i=1]L_i+P[I_i=1|y_i=0](1-L_i)}$

$= \frac{(1-p_2)L_i}{(1-p_2)L_i+p_1(1-L_i)} = \frac{(1-p_2)L_i}{p_1+(1-p_1-p_2)L_i}$ which leads to

$$\frac{1}{L_i(1)} = \frac{p_1}{(1-p_2)L_i} + \frac{(1-p_1-p_2)}{(1-p_2)}$$

giving $L_i = \frac{p_1}{(1-p_2)}\left(\frac{1}{L_i(1)} - \frac{(1-p_1-p_2)}{(1-p_2)}\right)^{-1}$. (2.d.4)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$ and hence,

$$\hat{L}_i = \frac{p_1}{(1-p_2)}\left(\frac{1}{r_i} - \frac{(1-p_1-p_2)}{(1-p_2)}\right)^{-1} \quad (2.d.5)$$

$=\alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha=\frac{p_1}{(1-p_2)}$ and $\beta=\frac{(1-p_1-p_2)}{(1-p_2)}$

$= \alpha\left(\frac{r_i}{1-r_i\beta}\right)$ *such that* $E_R(\hat{L}_i) \cong L_i$.

Expanding $\hat{L}_i$ (= $f(r_i)$) about L_i (=$f(L_i(1))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i} f(r_i)|_{r_i=L_i(1)}(r_i - L_i(1))$

$= L_i + \frac{\alpha}{(1-L_i(1)\beta)^2}(r_i - L_i(1))$.

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1)\beta)^4} V_R(r_i)$. (2.d.6)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ \upsilon_R(r_i) = \frac{{p_1}^2(1-p_2)^2}{[(1-p_2)-r_i(1-p_1-p_2)]^4}\ \frac{p_1(1-p_1)+r_i(1-p_1-p_2)(p_2-p_1)}{(1-p_1-p_2)^2}$. (2.d.7)

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

Proceeding similarly as above,

$L_i(0)$ = Posterior Prob$[y_i = 1|I_i = 0] = \frac{P[I_i=0|y_i=1]L_i}{P[I_i=0|y_i=1]L_i+P[I_i=0|y_i=0](1-L_i)}$

$= \frac{p_2L_i}{p_2L_i+(1-p_1)(1-L_i)} = \frac{p_2L_i}{(1-p_1)+(p_2+p_1-1)L_i}$ which leads to $\frac{1}{L_i(0)} = \frac{(1-p_1)}{p_2L_i} + \frac{(p_2+p_1-1)}{p_2}$

giving $L_i = \frac{(1-p_1)}{p_2L_i}\left(\frac{1}{L_i(0)} - \frac{(p_2+p_1-1)}{p_2}\right)^{-1}$. (2.d.8)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(0)$ and hence,

$\hat{L}_i = \frac{(1-p_1)}{p_2L_i}\left(\frac{1}{r_i} - \frac{(p_2+p_1-1)}{p_2}\right)^{-1}$ (2.d.9)

$=\alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{(1-p_1)}{p_2L_i}$ and $\beta = \frac{(p_2+p_1-1)}{p_2}$

$= \alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i) \cong L_i$.

Expanding $\hat{L}_i$ ($=f(r_i)$) about L_i ($=f(L_i(0))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i}f(r_i)|_{r_i=L_i(0)}(r_i - L_i(0))$

$=L_i + \frac{\alpha}{(1-L_i(0)\beta)^2}(r_i - L_i(0))$

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(0)\beta)^4} V_R(r_i)$. (2.d.10)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ v_R(r_i) = \frac{{p_2}^2(1-p_1)^2}{[p_2-r_i(p_2+p_1-1)]^4}\frac{p_1(1-p_1)+r_i(1-p_1-p_2)(p_2-p_1)}{(1-p_1-p_2)^2}$. (2.d.11)

$E_R(v_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

2.(e) Unrelated Question Model: This RRT is used when both A and AC are sensitive. In addition to yi consider xi which takes value 1 if an individual i bears a predecided innocuous characteristic and 0 if not. There are two boxes of which the first one contains identical cards marked A and B(any innocuous characteristic) in proportions $p1$ and (1-$p1$) respectively and the second box contains identical cards marked A and B in proportions $p2(\neq p_1)$ and (1-$p2$) respectively. An individual i chooses a card from the first box and another card from the second box. From Chaudhuri (2011), the participant's response is:

Ii = 1 if the card type drawn from the first box matches i's characteristic
= 0 if there is no match and,
Ji = 1 if the card type drawn from the second box matches i's characteristic
= 0 if there is no match

Then $ri = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ (2.e.1)

is an unbiased estimator of yi such that $E_R(r_i)=y_i$ and

$V_R(r_i) = \frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}(y_i - x_i)^2$ (2.e.2)

and $\hat{V}_R(r_i) = v_R(r_i)=r_i(r_i-1)$ (2.e.3)

vide Chaudhuri (2011).

Let L_i be the prior Prob $[y_i = 1]$

$L_i(1,1) = \frac{p_1 p_2 L_i}{(1-p_1)(1-p_2)+(p_1+p_2-1)L_i}$

giving $L_i = \frac{(1-p_1)(1-p_2)}{p_1 p_2}\left(\frac{1}{L_i(1,1)} - \frac{(p_1+p_2-1)}{p_1 p_2}\right)^{-1}$ (2.e.4)

$\hat{L}_i = \frac{(1-p_1)(1-p_2)}{p_1 p_2}\left(\frac{1}{r_i} - \frac{(p_1+p_2-1)}{p_1 p_2}\right)^{-1}$ (2.e.5)

$= \alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{(1-p_1)(1-p_2)}{p_1 p_2}$ and $\beta = \frac{(p_1+p_2-1)}{p_1 p_2}$

$E_R(\hat{L}_i) \cong L_i$

$V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1,1)\beta)^4} V_R(r_i)$ (2.e.6)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{p_1{}^2 p_2{}^2 (1-p_1)^2 (1-p_2)^2}{[p_1 p_2 - r_i(p_1+p_2-1)]^4} r_i(r_i - 1) = \upsilon_R(\hat{L}_i)$ (2.e.7)

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$

Proceeding similarly as above, $L_i(1,0) = \frac{p_1(1-p_2)L_i}{(1-p_1)p_2+(p_1-p_2)L_i}$

giving $L_i = \frac{(1-p_1)p_2}{p_1(1-p_2)}\left(\frac{1}{L_i(1,0)} - \frac{p_1-p_2}{p_1(1-p_2)}\right)^{-1}$ (2.e.8)

$\hat{L}_i = \frac{(1-p_1)p_2}{p_1(1-p_2)}\left(\frac{1}{r_i} - \frac{p_1-p_2}{p_1(1-p_2)}\right)^{-1}$ (2.e.9)

$= \alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{(1-p_1)p_2}{p_1(1-p_2)}$ and $\beta = \frac{p_1-p_2}{p_1(1-p_2)}$ such that $E_R(\hat{L}_i) \cong L_i$

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1,0)\beta)^4} V_R(r_i)$ (2.e.10)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{p_1{}^2 p_2{}^2 (1-p_1)^2 (1-p_2)^2}{[p_1(1-p_2) - r_i(p_1-p_2)]^4} r_i(r_i - 1) = \upsilon_R(\hat{L}_i)$. (2.e.11)

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

Similarly, $L_i(0,1) = \frac{(1-p_1)p_2 L_i}{p_1(1-p_2)+(p_2-p_1)L_i}$

giving $L_i = \frac{p_1(1-p_2)}{(1-p_1)p_2}\left(\frac{1}{L_i(0,1)} - \frac{p_2-p_1}{(1-p_1)p_2}\right)^{-1}$ (2.e.12)

$\hat{L}_i = \frac{p_1(1-p_2)}{(1-p_1)p_2}\left(\frac{1}{r_i} - \frac{p_2-p_1}{(1-p_1)p_2}\right)^{-1}$ (2.e.13)

$= \alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha = \frac{p_1(1-p_2)}{(1-p_1)p_2}$ and $\beta = \frac{p_2-p_1}{(1-p_1)p_2}$

$E_R(\hat{L}_i) \cong L_i$.

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(0,1)\beta)^4} V_R(r_i)$. (2.e.14)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{p_1{}^2 p_2{}^2 (1-p_1)^2 (1-p_2)^2}{[p_2(1-p_1) - r_i(p_2-p_1)]^4} r_i(r_i - 1) = \upsilon_R(\hat{L}_i)$. (2.e.15)

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

Lastly, $L_i(0,0) = \frac{(1-p_1)(1-p_2)L_i}{p_1p_2+(1-p_1-p_2)L_i}$

giving $L_i = \frac{p_1p_2}{(1-p_1)(1-p_2)}\left(\frac{1}{L_i(0,0)} - \frac{1-p_1-p_2}{(1-p_1)(1-p_2)}\right)^{-1}$ (2.e.16)

$\hat{L}_i = \frac{p_1p_2}{(1-p_1)(1-p_2)}\left(\frac{1}{r_i} - \frac{1-p_1-p_2}{(1-p_1)(1-p_2)}\right)^{-1}$. (2.e.17)

$=\alpha\left(\frac{1}{r_i} - \beta\right)^{-1}$, where $\alpha=\frac{p_1p_2}{(1-p_1)(1-p_2)}$ and $\beta=\frac{1-p_1-p_2}{(1-p_1)(1-p_2)}$

such that $E_R(\hat{L}_i) \cong L_i$

$V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(0,0)\beta)^4} V_R(r_i)$ (2.e.18)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{{p_1}^2{p_2}^2(1-p_1)^2(1-p_2)^2}{[(1-p_1)(1-p_2)-r_i(1-p_2-p_1)]^4} r_i(r_i-1) = \upsilon_R(\hat{L}_i)$. (2.e.19)

$E_R(\upsilon_R(\hat{L}_i)) \cong V_R(\hat{L}_i)$.

2.(f) Hussain and Shabbir's Model: There are two types of boxes, the first type containing identical cards marked A and A^C in proportions $p_1(\neq 0.5)$and $(1-p_1)$ respectively and the second type containing identical cards marked A and A^C in proportions $p_2(\neq 0.5)$ and $(1-p_2)$ respectively, such that the probability of choosing the first box is a and that of the second box is $b(=1-a)$. Originally Hussain and Shabbir (2009) have considered only the Simple Random Sampling with replacement for the sampling of units but we can easily change the same into the general sampling scheme. So, we can write the participant's response as:

I_i = 1 if the card type drawn by i matches his/her characteristic
= 0 if there is no match

$E_R(I_i)=a[p_1y_i + (1-p_1)(1-y_i)] + b[p_2y_i + (1-p_2)(1-y_i)]$

$=1-ap_1-bp_2+(2ap_1 + 2bp_2 - 1)y_i$

Then $r_i = \frac{I_i-(1-ap_1-bp_2)}{2ap_1+2bp_2-1}$ (2.f.1)

is an unbiased estimator of y_i such that $E_R(r_i)=y_i$ and

$V_R(I_i) = E_R(I_i)(1-E_R(I_i))= (1-ap_1-bp_2)(ap_1+bp_2)$.

$V_R(r_i) =\frac{(1-ap_1-bp_2)(ap_1+bp_2)}{(2ap_1+2bp_2-1)^2} = \upsilon_R(r_i)$. (2.f.2)

Let L_i be the prior Prob[$y_i = 1$]

$L_i(1)$ = Posterior Prob$[y_i = 1|I_i = 1]= \frac{P[I_i=1|y_i=1]L_i}{P[I_i=1|y_i=1]L_i+P[I_i=1|y_i=0](1-L_i)}$

$= \frac{(ap_1+bp_2)L_i}{(ap_1+bp_2)L_i+(1-ap_1-bp_2)(1-L_i)}$

$= \frac{(ap_1+bp_2)L_i}{(1-ap_1-bp_2)+(2ap_1+2bp_2-1)L_i}$

which leads to $\frac{1}{L_i(1)} = \frac{(1-ap_1-bp_2)}{(ap_1+bp_2)L_i} + \frac{2ap_1+2bp_2-1}{ap_1+bp_2}$

giving $L_i = \frac{(1-ap_1-bp_2)}{(ap_1+bp_2)}\left(\frac{1}{L_i(1)} - \frac{2ap_1+2bp_2-1}{ap_1+bp_2}\right)^{-1}$. (2.f.3)

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(1)$ and hence,

$$\hat{L}_i = \frac{(1-\mathrm{a}p_1-bp_2)}{(\mathrm{a}p_1+\mathrm{b}p_2)}\left(\frac{1}{r_i}-\frac{2\mathrm{a}p_1+2bp_2-1}{\mathrm{a}p_1+\mathrm{b}p_2}\right)^{-1} \tag{2.f.4}$$

$=\alpha\left(\frac{1}{r_i}-\beta\right)^{-1}$, where $\alpha=\frac{(1-\mathrm{a}p_1-bp_2)}{(\mathrm{a}p_1+\mathrm{b}p_2)}$ and $\beta=\frac{2\mathrm{a}p_1+2bp_2-1}{\mathrm{a}p_1+\mathrm{b}p_2}$

$=\alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i)\cong L_i$.

Expanding $\hat{L}_i$ (=$f(r_i)$) about L_i (=$f(L_i(1))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i}f(r_i)|_{r_i=L_i(1)}(r_i-L_i(1))$

$=L_i+\frac{\alpha}{(1-L_i(1)\beta)^2}(r_i-L_i(1))$.

So, $$V_R(\hat{L}_i)=\frac{\alpha^2}{(1-L_i(1)\beta)^4}V_R(r_i) \tag{2.f.5}$$

Hence, $$\hat{V}_R(\hat{L}_i)=\frac{\alpha^2}{(1-r_i\beta)^4}\ v_R(r_i)=\frac{(1-\mathrm{a}p_1-bp_2)^3(\mathrm{a}p_1+\mathrm{b}p_2)^3}{(2\mathrm{a}p_1+2bp_2-1)^2[\mathrm{a}p_1+\mathrm{b}p_2-r_i(2\mathrm{a}p_1+2bp_2-1)]^4}=v_R(\hat{L}_i) \tag{2.f.6}$$

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible. Proceeding similarly as above,

$L_i(0)$ = Posterior Prob$[y_i=1|I_i=0]=\frac{P[I_i=0|y_i=1]L_i}{P[I_i=0|y_i=1]L_i+P[I_i=0|y_i=0](1-L_i)}$

$=\frac{[\mathrm{a}(1-p_1)+\mathrm{b}(1-p_2)]L_i}{[\mathrm{a}(1-p_1)+\mathrm{b}(1-p_2)]L_i+(\mathrm{a}p_1+bp_2)(1-L_i)}$

$=\frac{(1-\mathrm{a}p_1-bp_2)L_i}{(\mathrm{a}p_1+bp_2)+(1-2\mathrm{a}p_1+2bp_2)L_i}$

which leads to $\frac{1}{L_i(0)}=\frac{(\mathrm{a}p_1+bp_2)}{(1-\mathrm{a}p_1-bp_2)L_i}+\frac{1-2\mathrm{a}p_1+2bp_2}{1-\mathrm{a}p_1-bp_2}$

giving $$L_i=\frac{(\mathrm{a}p_1+bp_2)}{(1-\mathrm{a}p_1-bp_2)}\left(\frac{1}{L_i(0)}-\frac{1-2\mathrm{a}p_1+2bp_2}{1-\mathrm{a}p_1-bp_2}\right)^{-1}. \tag{2.f.7}$$

By using an empirical Bayes approach, r_i can be taken as an estimator of $L_i(0)$ and hence,

$$\hat{L}_i=\frac{(\mathrm{a}p_1+bp_2)}{(1-\mathrm{a}p_1-bp_2)}\left(\frac{1}{r_i}-\frac{1-2\mathrm{a}p_1+2bp_2}{1-\mathrm{a}p_1-bp_2}\right)^{-1} \tag{2.f.8}$$

$=\alpha\left(\frac{1}{r_i}-\beta\right)^{-1}$, where $\alpha=\frac{(\mathrm{a}p_1+bp_2)}{(1-\mathrm{a}p_1-bp_2)}$ and $\beta=\frac{1-2\mathrm{a}p_1+2bp_2}{1-\mathrm{a}p_1-bp_2}$

$=\alpha\left(\frac{r_i}{1-r_i\beta}\right)$ such that $E_R(\hat{L}_i)\cong L_i$.

Expanding $\hat{L}_i$ (=$f(r_i)$) about L_i (=$f(L_i(1))$ using Taylor's Expansion and neglecting higher order terms we get, $\hat{L}_i = L_i + \frac{\partial}{\partial r_i}f(r_i)|_{r_i=L_i(1)}(r_i-L_i(0))$

$=L_i+\frac{\alpha}{(1-L_i(0)\beta)^2}(r_i-L_i(0))$.

So, $V_R(\hat{L}_i) = \frac{\alpha^2}{(1-L_i(1)\beta)^4} V_R(r_i)$. (2.f.9)

Hence, $\hat{V}_R(\hat{L}_i) = \frac{\alpha^2}{(1-r_i\beta)^4}\ \upsilon_R(r_i) = \frac{(1-ap_1-bp_2)^3(ap_1+bp_2)^3}{(2ap_1+2bp_2-1)^2[1-ap_1-bp_2-r_i(1-2ap_1+2bp_2)]^4} = \upsilon_R(\hat{L}_i)$ (2.f.10)

Now putting the expression for r_i in the expressions of $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$, it is seen that $\hat{L}_i$ and $\hat{V}_R(\hat{L}_i)$ do not exist, making numerical calculations impossible.

3. Estimation: So, an Empirical Bayes approach is used to estimate

$$\bar{L} = \frac{1}{N}\sum_{i=1}^{N} L_i \quad (3.1)$$

instead of $\theta = \frac{1}{N}\sum_{i=1}^{N} y_i$

by taking r_i as an initial estimator for $L_i(\mathbf{R})$ where **R** is the vector of responses of a sampled unit using an RRT.

$$\hat{\bar{L}} = \frac{1}{N}\sum_{i\in s}\frac{\hat{L}_i}{\pi_i} \quad (3.2)$$

such that $E(\hat{\bar{L}}) = E_P E_R(\hat{\bar{L}}) \cong \hat{\bar{L}}$, vide Chaudhuri (2011).

$$V(\hat{\bar{L}}) = V_P E_R(\hat{\bar{L}}) + E_P V_R(\hat{\bar{L}}) = \frac{1}{N^2}\left[\sum_{i<}^{N}\sum_{j}^{N}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{L_i}{\pi_i} - \frac{L_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{L_i^2}{\pi_i}\beta_i + \sum_{i=1}^{N}\frac{V_R(\hat{L}_i)}{\pi_i}\right] \quad (3.3)$$

(where $\beta_i = 1 + \frac{1}{\pi_I}\sum_{j\neq i}^{N}\pi_{ij} - \sum_{i=1}^{N}\pi_i$)

If every sample s contains a common number of distinct units in it, then,

$$V(\hat{\bar{L}}) = \frac{1}{N^2}\left[\sum_{i<}^{N}\sum_{j}^{N}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{L_i}{\pi_i} - \frac{L_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{V_R(\hat{L}_i)}{\pi_i}\right]. \quad (3.4)$$

$$\upsilon(\hat{\bar{L}}) = \frac{1}{N^2}\left[\sum_{i<}\sum_{j\in s}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{\hat{L}_i}{\pi_i} - \frac{\hat{L}_j}{\pi_j}\right)^2 + \sum_{i\in s}\frac{\hat{L}_i^{\,2} - \upsilon_R(\hat{L}_i)}{\pi_i}\beta_i + \sum_{i\in s}\frac{\upsilon_R(\hat{L}_i)}{\pi_i}\right]. \quad (3.5)$$

If every sample s contains a common number of distinct units in it, then,

$$\upsilon(\hat{\bar{L}}) = \frac{1}{N^2}\left[\sum_{i<}\sum_{j\in s}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{\hat{L}_i}{\pi_i} - \frac{\hat{L}_j}{\pi_j}\right)^2 + \sum_{i\in s}\frac{\upsilon_R(\hat{L}_i)}{\pi_i}\right]. \quad (3.6)$$

$E(\upsilon(\hat{\bar{L}})) = E_P E_R(\upsilon(\hat{\bar{L}})) \cong V(\hat{\bar{L}})$.

4. Numerical Calculations And Comparisons With The Original Estimation Method:

Data for 395 households in an Indian province called state Meghalaya from The Third National Family Health Survey in India was extracted for the variable:

y_i = 1 if the woman i of a household in Meghalaya had visited a health facility or camp in the past 3 months and she felt that the facility was clean

= 0 if the woman i of a household in Meghalaya had visited a health facility or camp in the past 3 months and she felt that the facility was not clean

Samples of size 73 were drawn each using the Hartley and Rao (HR)(1962) sampling scheme in which a systematic sample by Probability Proportional to Size (PPS) method is used after the random arrangement of the population units. The number of members in a household to which i belongs were utilized as size measures. It is unnecessary to check if the size-measure variable is well-correlated with the y-variable or not as Chaudhuri (2011) has discussed. He argues that a large-scale survey is implemented on taking a single sample is used to estimate several parameters of which a few may relate to sensitive features. As mentioned by him since an RR procedure is not "a sampling scheme-specific", an estimation method may be developed based on a general sampling scheme and using the RR's realized on hand. So, we illustrate employing the Hartley-Rao scheme as it is so well-known and handy with several properties as are classically known.

We take only one sample selected by the Hartley and Rao (1962) method. We collect the responses taking several combinations of RR device parameters of each of the RR devices
(a) Kuk's RRT
(b) Christofides' RRT
(c) Forced Response Model
for which $\hat{L}_i$ and $\hat{V}_R\left(\hat{L}_i\right)$ exist.

We calculate: (1)the coefficient of variation,

$$CV_{\text{Bayes}}=\frac{\sqrt{\upsilon(\hat{\hat{L}})}}{\hat{\hat{L}}}\,100 \tag{4.1}$$

and compare this with

$$CV_{\text{original}}=\frac{\sqrt{\upsilon(e)}}{e}\,100 \tag{4.2}$$

where $$e=\frac{1}{N}\sum_{i\in s}\frac{r_i}{\pi_i} \tag{4.3}$$

such that E(e) = $E_P E_R$(e) $=\theta$ and

$$\upsilon(e)=\frac{1}{N^2}\left[\sum_{i<}\sum_{j\in s}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{r_i}{\pi_i}-\frac{r_j}{\pi_j}\right)^2+\sum_{i\in s}\frac{{r_i}^2-\upsilon_R(r_i)}{\pi_i}\beta_i+\sum_{i\in s}\frac{\upsilon_R(r_i)}{\pi_i}\right]. \tag{4.4}$$

If every sample s contains a common number of distinct units in it, then, for the original estimator e of θ, an unbiased variance estimator is

$$\upsilon(e)=\frac{1}{N^2}\left[\sum_{i<}\sum_{j\in s}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{r_i}{\pi_i}-\frac{r_j}{\pi_j}\right)^2+\sum_{i\in s}\frac{\upsilon_R(r_i)}{\pi_i}\right] \tag{4.5}$$

such that E($\upsilon(e)$) = $E_P E_R$($\upsilon(e)$) $=$ V(e).

Bayesian approach of estimation is better than the original method if $CV_{Bayes} < CV_{Original}$
(2) the estimated efficiency ($\hat{E}$) of the Bayesian approach of estimation with respect to the original method.

$$\hat{E}_{BO}=\frac{\upsilon(e)}{\upsilon(\hat{\hat{L}})}\,100\,. \tag{4.6}$$

Larger the $\hat{E}_{BO}$, the better it is to use the Bayesian approach of estimation.

Performances of the procedures of 'original' and 'empirical Bayes' methods of estimation are illustrated below for few RR device parameters of each of the RR devices (a), (b) and (c) mentioned above. Other cases are not shown here. But these cases showing the preferences in respect of CV and estimated Efficiency are noted within parentheses below in the consecutive tables, which follow.

Table 1.1: Findings:(1) Comparison for Kuk's RRT (K=5)

Parameters (p_1 , p_2)	$CV_{Original}$	CV_{Bayes}
(0.45,0.62)	65.15	36.032
(0.59,0.43)	70.037	29.615
(0.39,0.65)	61.122	44.929
(0.64,0.40)	65.786	32.313
(0.60,0.35)	40.110	24.656
60% cases are favourable		

Table 1.2

Parameters(p_1, p_2)	$v(e)$	$v(\hat{\hat{L}})$	$\hat{E}_{BO}$
(0.55,0.45)	0.100	0.077	130.99
(0.46,0.53)	0.203	0.158	128.85
(0.70,0.64)	0.332	0.223	148.54
(0.70,0.75)	0.358	0.238	150.52
(0.64,0.69)	0.319	0.299	106.91
52% cases are favourable			

Table 2.1: Comparison for Christofides' RRT (M=3)

Parameters (p_1, p_2, p_3)	$CV_{Original}$	CV_{Bayes}
(0.16,0.46,0.38)	80.59	15.04
(0.40,0.41,0.19)	66.59	29.42
(0.39,0.16,0.45)	59.55	22.00
(0.48,0.37,0.15)	42.78	27.67
(0.11,0.35,0.54)	35.25	21.49
59% cases are favourable		

Table 2.2

Parameters(p_1, p_2, p_3)	$v(e)$	$v(\hat{\hat{L}})$	$\hat{E}_{BO}$
(0.26,0.39,0.35)	0.263	0.193	136.34
(0.36,0.47,0.19)	0.052	0.011	486.32
(0.54,0.19,0.37)	0.062	0.027	230.25
(0.16,0.41,0.43)	0.026	0.023	114.16
(0.24,0.49,0.36)	0.144	0.092	155.80
55% cases are favourable			

Table 3.1: Comparison for Forced Response Model

Parameters (p_1, p_2, p_3)	$CV_{Original}$	CV_{Bayes}
(0.44,0.13,0.43)	43.77	29.78
(0.52,0.11,0.37)	61.24	18.03
(0.38,0.10,0.52)	164.44	47.27
(0.39,0.11,0.50)	95.08	43.60
(0.11,0.42,0.47)	151.35	37.76
86% cases are favourable		

Table 3.2

Parameters(p_1, p_2, p_3)	$\upsilon(e)$	$\upsilon(\hat{\hat{L}})$	$\hat{E}_{BO}$
(0.35,0.52,0.13)	0.202	0.022	906.92
(0.53,0.11,0.36)	0.030	0.014	214.64
(0.50,0.07,0.43)	0.022	0.010	211.09
(0.47,0.45,0.08)	0.021	0.013	159.06
(0.22,0.42,0.36)	0.031	0.023	134.78
54% cases are favourable			

Comparison for Unrelated Question Model: In the Third National Family Health Survey, it was found that in 110 households in the state of Jammu and Kashmir, the respondent's (woman interviewed) husband consumes alcohol either very often (here assumed as 'consumption of alcohol each day in a week') or less often (here assumed as 'not consuming alcohol at least in one day of a week').

Let y_i = 1 if individual i drinks alcohol very often(A say)
= 0 if i drinks alcohol less often(A^C)
x_i = 1 if individual i prefers music to painting(B say)
= 0 if i bears B^C

The number of members in a household to which i belongs were utilized as size measures. Samples of size 23 were drawn each using the Hartley and Rao (1962) sampling scheme (HR). Performances of the procedures of 'original' and 'empirical Bayes' methods of estimation are illustrated below for few RR device parameters. Other cases are not shown here. But these cases showing the preferences in respect of CV and estimated Efficiency are noted within parentheses below in the consecutive tables, which follow.

Table 4.1:Findings

Parameters(p_1 , p_2)	$CV_{Original}$	CV_{Bayes}
(0.68,0.35)	40.24	26.96
(0.70,0.45)	44.46	30.54
(0.35,0.59)	46.05	29.66
(0.36,0.67)	49.27	25.21
(0.71,0.37)	22.15	15.04
52% cases are favourable		

Table 4.2

Parameters(p_1 , p_2)	$\upsilon(e)$	$\upsilon(\hat{\hat{L}})$	$\hat{E}_{BO}$
(0.45,0.37)	0.169	0.049	340.86
(0.48,0.36)	0.064	0.023	280.04
(0.63,0.35)	0.031	0.025	125.18
(0.35,0.54)	0.059	0.025	230.79
(0.39,0.35)	0.942	0.138	684.77
69% cases are favourable			

5. Conclusion

Although the Bayesian approach of estimation of finite population proportion of a sensitive characteristic does not work out with some RRT's still the method is worth consideration for the models with which it is applicable.

References

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* Boca Raton: Chapman & Hall, CRC Press, Taylor & Francis Group.

Chaudhuri, A., Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys*. Springer-Verlag Berlin Heidelberg.

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

Greenberg, B.G., Abul-Ela, A.A., Siimons, W.R., and Horvitz, D.G. (1969). The unrelated question RR model: Theoretical framework. *JASA*, **64**, 520-539.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics,* **33(2)**, 350-374.

Hussain, Z. and Shabbir, J. (2009). Bayesian estimation of population proportion of a sensitive characteristic using simple beta prior. *Pakistan Journal of Statistics,* **25**, 27-35.

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

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

Pitz, G.F. (1980). Bayesian analysis of random response models. *Psychological Bulletin,* **87**, 209-212. *The Third National Family Health Survey.*

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

Winkler, R.L. and Franklin, L.A. (1979). Warner's randomized response model: a Bayesian approach. *Journal of the American Statistical Association*, **74**, 207-214.