King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 270

September 2001

Estimation of Distribution Function and Median in Two Phase Sampling

Sarjinder Singh and Anwar H. Joarder

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

Estimation of Distribution Function and Median in Two Phase Sampling

Sarjinder Singh

Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon, SK, Canada S7N 5E6
Email: sarjinder@yahoo.com

and

Anwar H Joarder

Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia 31261
Email: anwarj@kfupm.edu.sa

Abstract. In this paper, we consider the problem of estimating the distribution function of a random variable in a finite population using two phase sampling, which is an extension of recent work by Singh, Joarder and Tracy (2001). Besides, three types of estimators for estimating the median are also considered. The variances of the proposed estimators are minimised for a fixed total cost and the expressions for the optimum values of the first and second phase sample sizes are obtained. Numerical illustrations are also given to have an idea about the relative efficiency of the proposed estimators of median over the usual one for the fixed total cost.

Key words: Distribution function, median, ratio and product type estimators, position and stratification estimators, two-phase sampling, asymptotic mean square error, auxiliary information.

1. Introduction

Assume that the population of interest U consists of N identifiable and distinct units labelled $U=(1,2,3,\dots,N)$. Associated with each unit i, there are two numbers, namely X_i and Y_i. First assume that $X_i's$ are known for all the units and $Y_i's$ are fixed but unknown. A sample $s=(s_1,s_2,\dots,s_n)$ of size n is selected and the values associated with the sample units are observed. The objective is to estimate the distribution function $F_y(t)$ of Y. Let $\Delta(x)$ be an indicator function such that

$$\Delta(x)=\begin{cases}1, \text{ if } x\geq 0\\ 0, \text{ otherwise}\end{cases}$$

Then the distribution function of Y is defined as

$$F_y(t)=\frac{1}{N}\sum_{i\in U}\Delta(t-y_i), \qquad t\in R \tag{1.1}$$

Kuk (1988) has considered a simple and unbiased estimator of $F_y(t)$ as

$$\hat{F}_y(t)=\frac{1}{N}\sum_{i\in s}\frac{\Delta(t-y_i)}{\pi_i} \tag{1.2}$$

where π_i's are the inclusion probabilities according to the sampling plan used. Similarly, for the auxiliary variable, one can define

$$\hat{F}_x(t)=\frac{1}{N}\sum_{i\in s}\frac{\Delta(t-x_i)}{\pi_i} \tag{1.3}$$

as an unbiased estimator of the distribution function

$$F_x(t)=\frac{1}{N}\sum_{i\in U}\Delta(t-x_i). \tag{1.4}$$

A ratio type estimator $\hat{F}_R(t)$ for estimating $F_y(t)$ can then be defined as

$$\hat{F}_R(t)=\hat{F}_y(t)\left[\frac{F_x(t)}{\hat{F}_x(t)}\right] \tag{1.5}$$

with variance

$$V\left[\hat{F}_R(t)\right]=\frac{1}{N^2}\sum_{i<}^{N}\sum_{j}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)\left[\left\{\frac{\Delta(t-y_i)}{\pi_i}-\frac{\Delta(t-y_j)}{\pi_j}\right\}^2+\left\{\frac{F_y(t)}{F_x(t)}\right\}^2\left\{\frac{\Delta(t-x_i)}{\pi_i}-\frac{\Delta(t-x_j)}{\pi_j}\right\}^2\right.$$

$$\left.-2\left\{\frac{F_y(t)}{F_x(t)}\right\}\left\{\frac{\Delta(t-x_i)}{\pi_i}-\frac{\Delta(t-x_j)}{\pi_j}\right\}\left\{\frac{\Delta(t-y_i)}{\pi_i}-\frac{\Delta(t-y_j)}{\pi_j}\right\}\right].$$

Rao *et al.* (1990) have also proposed ratio and difference type estimators of a population distribution function under a general sampling design using auxiliary information for the population, and compared them with a model-based estimator of Chambers and Dunstan (1986) through a simulation study. The model-based approach for estimating distribution function from survey data, studied by Chambers and Dunstan

(1986), allowed the population auxiliary information to be directly incorporated in the estimation process. This approach was found to offer significant gains when the auxiliary information is linearly related to the survey variable of interest. Later, Dunstan and Chambers (1989) have concluded through a Monte Carlo study that the method using "limited information" for the auxiliary variable is at most efficient as the "full information" method proposed by Chambers and Dunstan (1986). Kuo (1988) has introduced a classical as well as prediction approach to estimate distribution functions from survey data. Some recent developments on the estimation of distribution function are due to Chambers *et al.* (1992), Chambers *et al.* (1993), Kuk (1993) and Dorfman (1993).

In practice, situations are often encountered where the investigator is to study variables with a highly skewed distribution. In such situations median is regarded as a more appropriate measure of location than mean. In case all the population units are difficult to observe, one is to estimate the median from the sample data. Kuk and Mak (1989) have suggested estimators of median using auxiliary information. Francisco and Fuller (1991) have also considered the problem of estimating the median as a part of the estimation of finite population distribution function.

Let Y_i and X_i, $i=1,2,\ldots,N$, be the values of the population units for the study variable Y and auxiliary variable X, respectively. Further let y_i and x_i, $i=1,2,\ldots,n$, be the corresponding values for the units included in a WOR simple random sample of size n. Assuming the knowledge of median M_x of variable X, Kuk and Mak (1989) considered a ratio estimator

$$\hat{M}_{YR}=\hat{M}_y\left(\frac{M_x}{\hat{M}_x}\right) \tag{1.6}$$

for the median M_y. Further let p denote the proportion of Y values in the sample that are less than or equal to the median value M_y, which is an unknown parameter and is to be estimated from the sample observations. Thus p is also an unknown parameter. If $\hat{p}$ is an estimator of p, the sample median $\hat{M}_y$ can be written as the quantile estimator $\hat{Q}_y(\hat{p})$, where $\hat{p}=0.5$. Kuk and Mak (1989) have defined a matrix of proportions $[P_{ij}]$ as,

	$Y \le M_y$	$Y > M_y$	Total
$X \le M_x$	P_{11}	P_{21}	$P_{\cdot 1}$
$X > M_x$	P_{12}	P_{22}	$P_{.2}$
Total	$P_{1.}$	$P_{2.}$	1

They considered the position estimator of median as

$$\hat{M}_{YP}=\hat{Q}_Y(\hat{p}_1) \tag{1.7}$$

where

$$\hat{p}_1 = \frac{1}{n}\left[\frac{n_x p_{11}}{p_{.1}} + \frac{(n-n_x)p_{12}}{p_{.2}}\right] \approx \left[n_x p_{11} + (n-n_x)p_{12}\right]\left(2n^{-1}\right) \tag{1.8}$$

and $p_{ij}'s$ are sample analogue of population $P_{ij}'s$. An alternative estimator of the median proposed by them is termed as stratification estimator. If $\tilde{F}_{Y1}(y)$ and $\tilde{F}_{Y2}(y)$ denote the proportion of units in the sample with $X \le M_x$ and $X > M_x$, respectively, for the value y that have Y values less than or equal to y, then their estimator of M_y becomes

$$\hat{M}_{YS} = Inf\left\{y : \tilde{F}_Y(y) \ge 0.5\right\} \tag{1.9}$$

where

$$\hat{F}_Y(y) = 0.5\left\{\tilde{F}_{Y1}(y) + \tilde{F}_{Y2}(y)\right\}.$$

We see that the estimators at (1.5), (1.6), (1.7) and (1.9) are based on prior knowledge of population median M_x for the auxiliary character X. Many times the population median M_x for the auxiliary character may not be known. In such cases the investigator has to use two phase sampling, which necessitates the consideration of two phase sampling analogous to the above estimators.

In two phase sampling, we select a preliminary large sample of n' units using SRSWOR sampling in the first phase. Only the auxiliary character X is measured on these sample units. In the second phase, a sub-sample of n units is drawn from the preliminary sample of n' units through SRSWOR sampling and both the study variable Y and the auxiliary variable X are measured on the selected units.

2. Proposed Estimator

We consider an analogue of (1.5) in two phase sampling as

$$\hat{F}_{RD}(t) = \hat{F}_y(t)\left(\frac{\hat{F}_x'(t)}{\hat{F}_x(t)}\right) \tag{2.1}$$

where

$$\hat{F}_x'(t) = \frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i),\ \ \hat{F}_x(t) = \frac{1}{n}\sum_{i=1}^{n}\Delta(t-x_i) \text{ and } \hat{F}_y(t) = \frac{1}{n}\sum_{i=1}^{n}\Delta(t-y_i).$$

The following lemmas are needed to find the variance of the estimator in 2.1.

Lemma 2.1: The variance of $\hat{F}_y(t)$ is given by

$$V\left[\hat{F}_y(t)\right] = \frac{(N-n)}{nN^2}\left[\sum_{i=1}^{N}\{\Delta(t-y_i)\}^2 - \frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-y_i)\Delta(t-y_j)\right]. \tag{2.2}$$

Proof: Let E_1 and E_2 denote the expected values over all possible first phase samples and for a given first phase sample, respectively. Also let the variances V_1 and V_2 be similarly defined. Then we have

$$V\left[\hat{F}_y(t)\right] = E_1V_2\left[\hat{F}_y(t)\right] + V_1E_2\left[\hat{F}_y(t)\right]$$

$$= E_1V_2\left[\frac{1}{n}\sum_{i=1}^{n}\Delta(t-y_i)\right] + V_1E_2\left[\frac{1}{n}\sum_{i=1}^{n}\Delta(t-y_i)\right]$$

$$= \frac{N-n}{nN^2}\left[\sum_{i=1}^{N}\{\Delta(t-y_i)\}^2 - \frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-y_i)\Delta(t-y_j)\right].$$

Hence the lemma.

Lemma 2.2: The variance of $\hat{F}_x(t)$ is given by

$$V\left[\hat{F}_x(t)\right] = \frac{(N-n)}{nN^2}\left[\sum_{i=1}^{N}\{\Delta(t-x_i)\}^2 - \frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-x_i)\Delta(t-x_j)\right]. \tag{2.3}$$

Lemma 2.3: The variance of $\hat{F}_x'(t)$ is given by

$$V\left[\hat{F}_x'(t)\right] = \frac{(N-n')}{n'N^2}\left[\sum_{i=1}^{N}\{\Delta(t-x_i)\}^2 - \frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-x_i)\Delta(t-x_j)\right]. \tag{2.4}$$

The proof is obvious.

Lemma 2.4: The covariance between $\hat{F}_y(t)$ and $\hat{F}_x(t)$ is given by

$$Cov\left[\hat{F}_y(t), \hat{F}_x(t)\right] = \frac{(N-n)}{nN^2}\left[\sum_{i=1}^{N}\Delta(t-x_i)\Delta(t-y_i) + \frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-x_i)\Delta(t-y_j)\right]. \tag{2.5}$$

Proof: We have

$$Cov\left[\hat{F}_y(t), \hat{F}_x(t)\right] = E_1\left[C_2\left\{\hat{F}_y(t), \hat{F}_x(t)\right\}\right] + C_1\left[E_2\left\{\hat{F}_y(t)\right\}, E_2\left\{\hat{F}_x(t)\right\}\right]$$

$$= E_1C_2\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta(t-y_i)\right\}, \left\{\frac{1}{n}\sum_{i=1}^{n}\Delta(t-x_i)\right\}\right]$$

$$+ C_1\left[E_2\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta(t-y_i)\right\}, E_2\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta(t-x_i)\right\}\right]$$

$$= E_1\left[\frac{(n'-n)}{nn'(n'-1)}\left\{\sum_{i=1}^{n'}\Delta(t-y_i)\Delta(t-x_i)+\frac{1}{n'}\left(\sum_{i=1}^{n'}\Delta(t-y_i)\right)\left(\sum_{i=1}^{n'}\Delta(t-x_i)\right)\right\}\right]$$

$$+C_1\left[\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i),\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-y_i)\right]$$

$$=\frac{(N-n)}{nN^2}\left[\sum_{i=1}^{N}\Delta(t-x_i)\Delta(t-y_i)+\frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-x_i)\Delta(t-y_j)\right]$$

which is the same as (2.5).

Lemma 2.5: The covariance between $\hat{F}_x(t)$ and $\hat{F}_x'(t)$ is given by

$$Cov\left[\hat{F}_x(t),\hat{F}_x'(t)\right]=V\left[\hat{F}_x'(t)\right]. \tag{2.6}$$

Proof: We have

$$Cov\left[\hat{F}_x(t),\hat{F}_x'(t)\right]=E_1\left[C_2\left\{\hat{F}_x(t),\hat{F}_x'(t)\right\}\right]+C_1\left[E_2\left\{\hat{F}_x(t)\right\},E_2\left\{\hat{F}_x'(t)\right\}\right]$$

$$=0+C_1\left[\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i),\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i)\right]=V\left[\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i)\right]=V\left[\hat{F}_x'(t)\right].$$

Hence the lemma.

Lemma 2.6: The covariance between $\hat{F}_y(t)$ and $\hat{F}_x'(t)$ is given by

$$Cov\left[\hat{F}_y(t),\hat{F}_x'(t)\right]=\frac{(N-n')}{n'N^2}\left[\sum_{i=1}^{N}\Delta(t-y_i)\Delta(t-x_i)+\frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-y_i)\Delta(t-x_j)\right]. \tag{2.7}$$

Proof: We have

$$Cov\left[\hat{F}_y(t),\hat{F}_x'(t)\right]=E_1\left[C_2\left\{\hat{F}_y(t),\hat{F}_x'(t)\right\}\right]+C_1\left[E_2\left\{\hat{F}_y(t)\right\},E_2\left\{\hat{F}_x'(t)\right\}\right]$$

$$=0+C_1\left[\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-y_i),\frac{1}{n'}\sum_{i=1}^{n'}\Delta(t-x_i)\right]$$

$$=\frac{(N-n')}{n'N^2}\left[\sum_{i=1}^{N}\Delta(t-y_i)\Delta(t-x_i)+\frac{1}{(N-1)}\sum_{i\neq j=1}^{N}\Delta(t-y_i)\Delta(t-x_j)\right]$$

which proves the lemma. Thus we have the following theorem.

Theorem 2.1: The variance of the proposed estimator $\hat{F}_{RD}(t)$ at (2.1) is approximately

$$V\left[\hat{F}_{RD}(t)\right]=V\left[\hat{F}_y(t)\right]+\left\{\frac{F_y(t)}{F_x(t)}\right\}^2\left[V\left\{\hat{F}_x'(t)\right\}+V\left\{\hat{F}_x(t)\right\}-2Cov\left\{\hat{F}_y(t),\hat{F}_x(t)\right\}\right]$$

$$+2\left\{\frac{F_y(t)}{F_x(t)}\right\}\left[Cov\left\{\hat{F}_y(t),\hat{F}_x'(t)\right\}-Cov\left\{\hat{F}_y(t),\hat{F}_x(t)\right\}\right]. \tag{2.8}$$

Proof: The proof of this theorem follows directly from elementary concepts. The algebraic expression for $V\left[\hat{F}_{RD}(t)\right]$ may be obtained by the use lemmas 2.1 to 2.6.

The finite population median M_y is closely related to the distribution function $F_y(t)$ since $M_y = Inf.\{t : F_y(t) \geq 0.5\}$. This relationship can be used to obtain an estimator of the median of y as $F_y(t)$ is a monotone non-decreasing function of t. If $G_y(.)$ denotes the inverse function of $F_y(.)$, then we have the median $M_y = G_y(0.5)$. We, however, do not pursue this option and instead consider certain other estimators of median M_y using two phase simple random sampling. In order to condense the article the proofs of further theorems are omitted. Let $\hat{M}_x'$ denote the estimator of M_x based on the preliminary large sample. Also, let $\hat{M}_x$ and $\hat{M}_y$ denote the estimators of M_x and M_y, respectively, based on the sample drawn at the second phase.

3. Proposed Estimator for Median

Product Estimator: First we consider a product estimator for the estimation of median in two phase sampling as

$$\hat{M}_{y1} = \hat{M}_y\left(\frac{\hat{M}_x}{\hat{M}_x'}\right). \tag{3.1}$$

The variance of this estimator is given in the following theorem without proof.

Theorem 3.1: The variance of the estimator $\hat{M}_{y1}$

$$V\left(\hat{M}_{y1}\right)=\frac{1}{4}\left(\frac{1}{n'}-\frac{1}{N}\right)\left\{f_y\left(M_y\right)\right\}^{-2}+\left(\frac{1}{n}-\frac{1}{n'}\right)\left[\frac{1}{4}\left\{f_y\left(M_y\right)\right\}^{-2}+\frac{1}{4}\left(\frac{M_y}{M_x}\right)^2\left\{f_x\left(M_x\right)\right\}^{-2}\right.$$

$$\left.+2\left(\frac{M_y}{M_x}\right)\left(P_{11}-0.25\right)\left\{f_x\left(M_x\right)f_y\left(M_y\right)\right\}^{-1}\right]. \tag{3.2}$$

Also, we have

$$V\left(\hat{M}_y\right)=V\left(\hat{M}_y-M_y\right)=\left\{f_y\left(M_y\right)\right\}^{-2}V\left(p_y\right)=(1-f)(4n)^{-1}\left\{f_y\left(M_y\right)\right\}^{-2}$$

where $f=n/N$ denotes the finite population correction factor. Then the estimator $\hat{M}_{y1}$ is more efficient than the usual estimator $\hat{M}_y$ if

$$\rho_c<-\left\{f_x(M_x)\right\}^{-1}M_x^{-1}\Big/\left[2\left\{f_y\left(M_y\right)\right\}^{-1}M_y^{-1}\right] \tag{3.3}$$

where $\rho_c=4(P_{11}-0.25)$ goes from -1 to +1 as P_{11} increases from 0 to 0.5. This condition is analogous to the condition under which the product estimator (for mean) is superior to the sample mean in two phase sampling. In the same way, the ratio estimator $\hat{M}_{YR}$ in two phase sampling is straightforward.

Position Estimator: Suppose n_x^* be the number of units in the second phase sample s_n with $X\le\hat{M}_x'$. Thus, if P_{ij} are known, we can estimate p by

$$\hat{p}_0'=n^{-1}\left\{n_x^*P_{11}/P_{.1}+\left(n-n_x^*\right)P_{12}/P_{.2}\right\}\approx\left(2n^{-1}\right)\left\{n_x^*P_{11}+\left(n-n_x^*\right)\left(0.5-P_{11}\right)\right\} \tag{3.4}$$

because $P_{.j}=P_{1j}+P_{2j}\approx 0.5$ for $j=1,2$. If we replace P_{ij} by p_{ij} from the sample in (3.4), we get an estimator of p as

$$\hat{p}_1'=n^{-1}\left\{n_x^*p_{11}/p_{.1}+\left(n-n_x^*\right)p_{12}/p_{.2}\right\}\approx\left(2n^{-1}\right)\left\{n_x^*p_{11}+\left(n-n_x^*\right)p_{12}\right\}. \tag{3.5}$$

Then we define a position estimator in two phase sampling as

$$\hat{p}_{y2}=\hat{Q}(\hat{p}_1'). \tag{3.6}$$

This leads to the result stated in the theorem below:

Theorem 3.2: The variance of the estimator $\hat{M}_{y2}$ is given by

$$V\left(\hat{M}_{y2}\right)=\left\{f_y\left(M_y\right)\right\}^{-2}\left[\frac{1}{4}\left(\frac{1}{n}-\frac{1}{N}\right)-4\left(\frac{1}{n}-\frac{1}{n'}\right)\left(P_{11}-0.25\right)^2\right]. \tag{3.7}$$

Relation (3.7) shows that the estimator $\hat{M}_{y2}$ is always more efficient than the usual one $\hat{M}_y$ provided the sample size n remains the same for both the cases.

Stratification Estimator: Suppose that $\tilde{F}_{y1}(y)$ and $\tilde{F}_{y2}(y)$ denote the proportion of units in the second phase sample with $X \le \hat{M}_x'$ and $X > \hat{M}_x'$, respectively, for the value y that have Y values less than or equal to y. Then $F_y(y)$ can be estimated by

$$\tilde{F}_y(y) = n_x' \tilde{F}_{y1}(y)/n' + (n' - n_x')\tilde{F}_{y2}(y)/n' \approx 0.5\{\tilde{F}_{y1}(y) + \tilde{F}_{y2}(y)\}$$

where n_x' is the number of units in the preliminary large sample with $X \le \hat{M}_x'$. Then we consider an analogue of stratification estimator in two phase sampling as

$$\hat{M}_{y3} = Inf\{y : \tilde{M}_y(y) \ge 0.5\}. \tag{3.8}$$

For the variance of estimator $\hat{M}_{y3}$ we have the following theorem without proof.

Theorem 4.1: The variance of the estimator $\hat{M}_{y3}$

$$V(\hat{M}_{y3}) = \{f_y(M_y)\}^{-2}\left[\left(\frac{1}{n} - \frac{1}{N}\right)/4 - 4\left(\frac{1}{n} - \frac{1}{n'}\right)(P_{11} - 0.25)^2\right]. \tag{3.9}$$

Relation (3.7) and (3.9) show that the position estimator and stratification estimators remain equal efficient in two phase sampling.

4. Cost Aspect

Let C_1 and C_2 denote the costs per unit in the second phase and first phase, respectively. Further if C_0 denotes the over head cost then the fixed cost C is given by

$$C = C_0 + nC_1 + n'C_2. \tag{4.1}$$

Following Sukhatme *et al.* (1984), the variance of $\hat{M}_{y1}$ and $\hat{M}_{y2}$ (or hence of $\hat{M}_{y3}$) will be minimum for the fixed cost C given at (4.1) if the optimum values of n and n' are given as indicated in Table 1 and Table 2, respectively.

Table 1: Optimum values of n and n' for the estimator $\hat{M}_{y1}$.

$$n = \frac{(C - C_0)\sqrt{\{f_y(M_y)\}^{-2}/4 + (M_y/M_x)^2\{f_x(M_x)\}^{-2}/4 - T}}{W\sqrt{C_1}}$$

$$n' = \frac{(C - C_0)\sqrt{T\{f_x(M_x)f_y(M_y)\}^{-1} - (M_y/M_x)^2\{f_x(M_x)\}^{-2}/4}}{W\sqrt{C_2}}$$

where $T = 2(M_y/M_x)(0.25 - P_{11})\{f_x(M_x)f_y(M_y)\}^{-1}$ and

$$W = \sqrt{C_1\left[\{f_y(M_y)\}^{-2}/4 + (M_y/M_x)^2\{f_x(M_x)\}^{-2}/4 - T\right]} + \sqrt{C_2\left[T + (M_y/M_x)^2\{f_x(M_x)\}^{-2}/4\right]}$$

It should be noted that the optimum values of n and n' are real only if the condition (3.3) is satisfied. Otherwise one may use the usual ratio estimator of median.

Table 2: Optimum values of n and n' for the estimator $\hat{M}_{y2}$ (or hence $\hat{M}_{y3}$).

$$n = \frac{(C-C_0)\sqrt{2P_{11}(1-2P_{11})}}{C_1\sqrt{2P_{11}(1-2P_{11})}+2|P_{11}-0.25|\sqrt{C_1C_2}}$$

$$n' = \frac{2(C-C_0)|P_{11}-0.25|}{\sqrt{2C_1C_2P_{11}(1-2P_{11})}+2C_2|P_{11}-0.25|}$$

where $|u|$ denotes the absolute value of u. Moreover, in the case of simple random sampling without using the auxiliary variable, the minimum variance of $\hat{M}_y$ is given by

$$Min.V\left[\hat{M}_y\right] = \frac{1}{4}\left[\frac{C}{(C-C_0)} - \frac{1}{N}\right]\{f_y(M_y)\}^{-2}. \tag{4.2}$$

On putting the optimum values of n and n' from Tables in (3.2) and (3.7) [or hence in (3.9)], one may easily get the minimum variance of $\hat{M}_{y1}$ and $\hat{M}_{y2}$ (or hence of $\hat{M}_{y3}$). Theoretical comparison of Min. $V(\hat{M}_{y1})$ and Min. $V(\hat{M}_{y2})$ with Min. $V(\hat{M}_y)$ is difficult to handle algebraically. Some numerical examples have been, therefore, carried out to have an idea of the relative efficiency (R.E.) of the proposed estimators over the usual one.

5. Numerical Illustrations

For purposes of numerical illustration, we have considered various hypothetical joint probability density functions (p.d.fs.) of x and y. Suppose that $f(x, y)$ denotes the joint p.d.f. of x and y for $a \le x \le b$ and $c \le y \le d$, that is

$$\int_a^b\int_c^d f(x,y)dxdy = 1 \tag{5.1}$$

The values of $f_x(M_x)$, $f_y(M_y)$ and P_{11} for the different populations following different distributions were obtained. We have assumed that $C = \$1000$, $C_0 = \$200$, $C_1 = \$20$, $C_2 = \$0.75$ and $N = 1000$. The relative efficiency (R.E.) of proposed estimators over the usual one for the fixed cost of survey has been presented in Table 3 and Table 4.

Table 3: Relative efficiency (%) of product estimator $\hat{M}_{y1}$ over the $\hat{M}_y$ for different distributions.

Distribution	Range	1st phase sample	2nd phase sample	R.E. (%)
$f(x,y)=4(x+y)$	$0\le x\le 1$ $0\le y\le 0.5$	145	35	125.14
$f(x,y)=0.176x^2e^{-y}$	$0\le x\le 3$ $0\le y\le 1$	135	35	120.11
$f(x,y)=4.33\,x^2/y$	$0\le x\le 1$ $1\le y\le 2$	134	34	119.75

Table 4: Relative efficiency (%) of position estimator $\hat{M}_{y_2}$ (or stratification estimator $\hat{M}_{y_3}$) over $\hat{M}_y$ for different distributions.

Distribution	Range	1st phase sample	2nd phase sample	R.E. (%)
$f(x,y)=4(x+y)$	$0\le x\le 1$ $0\le y\le 0.5$	180	33	148.19
$f(x,y)=9x^2y^2/8$	$0\le x\le 1$ $0\le y\le 2$	84	37	103.66
$f(x,y)=0.176x^2e^{-y}$	$0\le x\le 3$ $0\le y\le 1$	215	32	177.87
$f(x,y)=4.33\,x^2/y$	$0\le x\le 1$ $1\le y\le 2$	158	34	132.47

Acknowledgement

The second author acknowledges the excellent research facilities available at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

References

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, **73**, 597-604.

Chambers, R.L., Dorfman, A.M. and Hall, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika*, **79**, 577-582.

Chambers, R.L., Dorfman, A.H. and Wehrly, T.E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *Jour. Amer. Statist. Assoc.*, **88**, 268-277.

Dorfman, A.H. (1993). A comparison of design-based and model based estimators of the finite population distribution function. *Austr. Jour. Statist.*, **35**, 29-41.

Dunstan, R. and Chambers, R.L. (1989). Estimating distribution function from survey data with limited auxiliary information. *Austr. Jour. Statist.*, **31**, 1-11.

Francisco, C.A. and Fuller, W.A. (1991). Quantile estimation with a complex survey design. *Ann. Statist.* **19**, 454-469.

Kuk, A.Y.C. (1988). Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika*, **75**, 97-103.

Kuk, A.Y.C. (1993). A kernel method for estimating finite population distribution functions using auxiliary information. *Biometrika,* **80**, 385-392.

Kuk, A.Y.C. and Mak, T.K. (1989). Median estimation in the presence of auxiliary information. *J. Roy. Statist. Soc.*,B, **51**, 261-269.

Kuo, L. (1988). Classical and prediction approaches to estimating distribution functions from survey data. *Proc. Sec. Survey Research Methods, Amer. Statist. Assoc.*, 280-285.

Rao, J.N.K., Kovar, J.G. and Mantel, H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika,* **77**, 365-375.

Singh, S., Joarder, A. and Tracy, D.S. (2001). Median estimation using double sampling. *Austral.& New Zealand J. Statistics,* **43**, 1, 33-46.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications.* 3rd ed., Iowa State University Press, Amess, Iowa.