A class of estimators for population median in two occasion rotation sampling

Kumari Priyanka* and Richa Mittal †

Abstract

The present work deals with the problem of estimation of finite population median at current occasion, in two occasion successive (rotation) sampling. A class of estimators has been proposed for the estimation of population median at current occasion, which includes many existing estimators as a particular case. Asymptotic properties including the asymptotic convergence of proposed class of estimators are elaborated. Optimum replacement strategies are also discussed. The proposed class of estimators at optimum condition is compared with the sample median estimator when there is no matching from the previous occasion as well as with some other members of the class. Theoretical results have been justified through empirical interpretation with the help of some natural populations.

1. Introduction

When both, the characteristic and the composition of the population change over time, then the cross-sectional surveys at a particular point of time become important. The survey estimates are therefore time specific, a feature that is particularly important in some context. For example, the unemployment rate is a key economic indicator that varies over time, the rate may change from one month to the next because of a change in the economy (with business laying off or recruiting new employees). To deal with such kind of circumstances, sampling is done on successive occasions with partial replacement of the units.

---

*Department of Mathematics, Shivaji College (University of Delhi), New Delhi 110 027, India
Email: `priyanka.ism@gmail.com`
†
Email:`sovereignricha@gmail.com`

The problem of sampling on two successive occasions was first considered by Jessen (1942), and latter this idea was extended by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1977), Singh and Singh (2001), Singh and Priyanka (2008), Singh et al. (2012) and many others. All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling.

Often, there are many practical situations where variables involved, consists of extreme values and resulting strong influence on the value of mean. In such cases the study variable is having a highly skewed distribution. For example, the study of environmental issues, the study of income as well as expenditure, the study of social evils such as abortions etc.. In these situations, the mean as a measure of central tendency may not be representative of the population because it moves with the direction of asymmetry leaving the median as a better measure since it is not affected by extreme values.

Most of the studies related to median have been developed by assuming simple random sampling or its ramification in stratified random sampling (Gross (1980), Sedransk and Meyer (1978), Smith and Sedransk (1983)).

As noted earlier, a large number of estimators for estimating the population mean at current occasion have been proposed by various authors, but only a few efforts (namely Martinez-Miranda et al. (2005), Singh et al. (2007) and Rueda and Munoz (2008)) have been made to estimate the population median on current occasion in two occasion successive sampling.

The present work develops a one-parameter class of estimators that estimate the population median on the current occasion in two-occasion successive sampling. The proposed class of estimators includes some of the estimators proposed by Singh et al. (2007) for second quantile as particular cases.

Asymptotic expressions for bias and mean square error including the asymptotic convergence of the proposed class of estimators are derived. The optimum replacement strategies are discussed. The proposed class of estimators at optimum conditions is compared with sample median estimator when there is no matching from the previous occasion as well as with some of the estimators due to Singh et al. (2007) and few other members of its class. Theoretical results are justified by empirical interpretation with the help of some natural populations.

2. Sample Structure and Notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of $u = (n - m) = n\mu$ units is drawn afresh from the remaining $(N - n)$ units of the population so that the sample size on the second occasion is also n. μ and λ, $(\mu + \lambda = 1)$ are the fractions of fresh and matched samples respectively at the second (current) occasion. The following notations are considered for the further use:

M_x, M_y : Population median of the variables x and y, respectively.

$\widehat{M}_x(n), \widehat{M}_x(m), \widehat{M}_y(m), \widehat{M}_y(u)$: Sample medians of the respective variables shown in suffices and based on the sample sizes given in braces.

$f_x(M_x), f_y(M_y)$ $\quad$: The marginal densities of variables x and y, respectively.

3. Proposed Class of Estimators

To estimate the population median M_y on the current (second) occasion, two independent estimators are suggested. One is based on sample of the size $u = n\mu$ drawn afresh on the current (second) occasion and which is given by

$$(3.1) \qquad T_u = \widehat{M}_y(u).$$

Second estimator is a one-parameter class of estimators based on the sample of size $m = n\lambda$ common to the both occasions and is defined as

$$(3.2) \qquad T_m(d) = \widehat{M}_y(m) \left[\frac{(A+C)\widehat{M}_x(n) + fB\widehat{M}_x(m)}{(A+fB)\widehat{M}_x(n) + C\widehat{M}_x(m)} \right],$$

$$A = (d-1)(d-2), \quad B = (d-1)(d-4),$$

$$C = (d-2)(d-3)(d-4) \quad \text{and} \quad f = \frac{n}{N},$$

where d is a non-negative constant, identified to minimize the mean square error of the estimator $T_m(d)$.

Now considering the convex linear combination of the estimators T_u and $T_m(d)$, a class of estimators for M_y is proposed as

$$(3.3) \qquad \widehat{T}_d = \varphi T_u + (1-\varphi)T_m(d),$$

where φ is an unknown constant to be determined so as to minimize the mean square error of the class of the estimators $\widehat{T}_d$.

3.1. Remark. For estimating the median on each occasion, the estimator T_u is suitable, which implies that more belief on T_u could be shown by choosing φ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimator $T_m(d)$ could be more useful so φ might be chosen 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choice of φ is desired.

3.2. Remark. The following estimators can be identified as a particular case of the suggested class of estimators $\widehat{T}_d$ to estimate population median on the current occasion in two occasion successive (rotation) sampling for different values of the unknown parameter 'd':

(i) $\widehat{T}_1 = \varphi_1 T_u + (1-\varphi_1)T_m(1)$; (Ratio type estimator)
(ii) $\widehat{T}_2 = \varphi_2 T_u + (1-\varphi_2)T_m(2)$; (Product type estimator)
(iii) $\widehat{T}_3 = \varphi_3 T_u + (1-\varphi_3)T_m(3)$; (Dual to Ratio type estimator)

where

$$T_m(1) = \widehat{M}_y(m) \left[\frac{\widehat{M}_x(n)}{\widehat{M}_x(m)} \right],$$

$$T_m(2) = \widehat{M}_y(m) \left[\frac{\widehat{M}_x(m)}{\widehat{M}_x(n)} \right],$$

$$T_m(3) = \widehat{M}_y(m) \left[\frac{n\widehat{M}_x(n) - m\widehat{M}_x(m)}{(n-m)\widehat{M}_x(n)} \right]$$

and φ_i $(i = 1, 2, 3)$ are unknown constants to be determined so as to minimize the mean square error of the estimators $\widehat{T}_i$ $(i = 1, 2, 3)$.

3.3. Remark. The Ratio and Product type estimators, proposed by Singh et al. (2007) for second quantile become particular cases of the proposed family of the estimators $\widehat{T}_d$ for $d = 1$ and 2, respectively.

4. Properties of the Proposed Class of Estimators

The properties of the proposed class of estimators $\widehat{T}_d$ are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored.

(ii) As $N \to \infty$, the distribution of the bivariate variable (x, y) approaches a continuous distribution,which depend on population under consideration with marginal densities $f_x(\cdot)$ and $f_y(\cdot)$, respectively (see Kuk and Mak (1989)).

(iii) The marginal densities $f_x(\cdot)$ and $f_y(\cdot)$ are positive.

(iv) The sample medians $\widehat{M}_y(u)$, $\widehat{M}_y(m)$, $\widehat{M}_x(m)$ and $\widehat{M}_x(n)$ are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), P_{yx} is assumed to be the proportion of elements in the population such that $x \le \widehat{M}_x$ and $y \le \widehat{M}_y$.

(vi) The following large sample approximations are assumed:

$$\widehat{M}_y(u) = M_y(1 + e_0), \quad \widehat{M}_y(m) = M_y(1 + e_1), \quad \widehat{M}_x(m) = M_x(1 + e_2),$$

$$\widehat{M}_x(n) = M_x(1 + e_3) \text{ such that } |e_i| < 1 \ \forall \ i = 0, 1, 2 \text{ and } 3.$$

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003). Under the above transformations, the estimators T_u and $T_m(d)$ takes the following forms:

(4.1) $\qquad T_u = M_y(1 + e_0),$

$$T_m(d) = M_y[1 + e_1 + d_1 e_3 + d_2 e_2 - d_3 e_3 - d_4 e_2 - d_1 d_3 e_3^2$$

$$- d_1 d_4 e_2 e_3 - d_2 d_3 e_2 e_3 - d_2 d_4 e_2^2 + d_3^2 e_3^2 + d_4^2 e_2^2$$

(4.2) $$+ 2 d_3 d_4 e_2 e_3 + (d_1 - d_3) e_1 e_3 + (d_2 - d_4) e_1 e_2]$$

where $d_1 = \dfrac{A + C}{A + fB + C}$, $d_2 = \dfrac{fB}{A + fB + C}$, $d_3 = \dfrac{A + fB}{A + fB + C}$ and $d_4 = \dfrac{C}{A + fB + C}$.

Thus we have the following theorems:

4.1. Theorem. *The bias of the estimator $\widehat{T}_d$ to the first order of approximation is obtained as*

(4.3) $\quad B(\widehat{T}_d) = (1 - \varphi) B\{T_m(d)\}$

where

(4.4) $\quad B\{T_m(d)\} = \dfrac{1}{n} Q_1 + \dfrac{1}{m} Q_2,$

$$Q_1 = (-d_1 d_3 - d_1 d_4 - d_2 d_3 + d_3^2 + 2 d_3 d_4) \frac{\{f_x(M_x)\}^{-2}}{4 M_x^2}$$

$$+ (d_1 - d_3)(P_{yx} - 0.25) \frac{\{f_y(M_y)\}^{-1} \{f_x(M_x)\}^{-1}}{M_y M_x}$$

and

$$Q_2 = (-d_2 d_4 + d_4^2)\frac{\{f_x(M_x)\}^{-2}}{4M_x^2}$$

$$+ (d_2 - d_4)(P_{yx} - 0.25)\frac{\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}}{M_y M_x}.$$

Proof. The bias of the estimator $\widehat{T}_d$ is given by

$$B\{\widehat{T}_d\} = E\{\widehat{T}_d - M_y\}$$

(4.5) $$= \varphi B\{T_u\} + (1 - \varphi)B\{T_m(d)\}.$$

Since, the estimator T_u is unbiased for M_y and $T_m(d)$ is biased for M_y, so the bias of the estimator $T_m(d)$is given by

$$B\{T_m(d)\} = E\{T_m(d) - M_y\}.$$

Now, substituting the value of $T_m(d)$ from equation (4.2) in the above equation we get the expression for bias of $T_m(d)$ as in equation (4.4).

Finally substituting the value of $B\{T_m(d)\}$ in equation (4.5), we get the expression for the $B\{\widehat{T}_d\}$as in equation (4.3). $\qquad\square$

4.2. Theorem. *The mean square error of the estimator $\widehat{T}_d$ is given by*

(4.6) $$M(\widehat{T}_d) = \varphi^2 V(T_u) + (1 - \varphi)^2 M(T_m(d))_{opt.}$$

where

(4.7) $$V(T_u) = \frac{1}{u}\frac{\{f_y(M_y)\}^{-2}}{4}$$

and

(4.8) $$M(T_m(d))_{opt.} = \frac{1}{m}A_1 + \left(\frac{1}{m} - \frac{1}{n}\right)\{\alpha^{*2}A_2 + 2\alpha^* A_3\}$$

where

$$A_1 = \frac{\{f_y(M_y)\}^{-2}}{4}, \qquad A_2 = \frac{\{f_x(M_x)\}^{-2}}{4}\left[\frac{M_y^2}{M_x^2}\right],$$

$$A_3 = (P_{yx} - 0.25)\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}\left[\frac{M_y}{M_x}\right],$$

$$\alpha^* = [\alpha]_{d=d_0},$$

$$\alpha = (d_2 - d_4) = (d_3 - d_1) = \frac{fB - C}{A + fB + C} \text{ and } d_0 \text{ is the optimum value of } d.$$

Proof. The mean square error of the estimator $\widehat{T}_d$ is given by

$$\widehat{M}(T_d) = E[\widehat{T}_d - M_y]^2$$

$$= E[\varphi(T_u - M_y) + (1 - \varphi)\{T_m(d) - M_y\}]^2$$

(4.9) $$= \varphi^2 V(T_u) + (1 - \varphi)^2 M[T_m(d)] + 2\varphi(1 - \varphi)\operatorname{cov}(T_u, T_m(d))$$

where

(4.10) $$V(T_u) = E[T_u - M_y]^2$$

and

(4.11) $$M[T_m(d)] = E[T_m(d) - M_y]^2.$$

As T_u and $T_m(d)$ are based on two independent samples of sizes u and m respectively, hence $\operatorname{cov}(T_u, T_m(d)) = 0$. Now, substituting the values of T_u and $T_m(d)$ from equations

(4.1) and (4.2) in equation (4.10) and (4.11) respectively, taking expectations and ignoring finite population corrections we get the expression for $V(T_u)$ as in equation (4.7) and mean square error of $T_m(d)$ is obtained as

$$M[T_m(d)] = \left[\frac{1}{m} A_1 + \left(\frac{1}{m} - \frac{1}{n} \right) \{\alpha^2 A_2 + 2\alpha A_3\} \right]$$

where

$$A_1 = \frac{\{f_y(M_y)\}^{-2}}{4}, \quad A_2 = \frac{\{f_x(M_x)\}^{-2}}{4} \left[\frac{M_y^2}{M_x^2} \right],$$

$$A_3 = (P_{yx} - 0.25)\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1} \left[\frac{M_y}{M_x} \right]$$

and

$$\alpha = (d_2 - d_4) = (d_3 - d_1) = \frac{fB - C}{A + fB + C}.$$

The mean square error of the $T_m(d)$ is a function of α, which in turns is a function of d, hence it can be minimized for d, and therefore we have

$$\frac{\partial \{M[T_m(d)]\}}{\partial d} = 0.$$

This gives $\alpha = \frac{-A_3}{A_2}$, assuming $\frac{\partial \alpha}{\partial d} \neq 0$ which in turns yields a cubic equation in 'd' given by

(4.12) $z_1 d^3 + z_2 d^2 + z_3 d + z_4 = 0$

where

$$z_1 = \left(\frac{A_3}{A_2} - 1 \right), \qquad z_2 = (f + 9) + \frac{A_3}{A_2}(f - 8),$$

$$z_3 = (-5f - 26) + \frac{A_3}{A_2}(23 - 5f)$$

and

$$z_4 = (4f + 24) + \frac{A_3}{A_2}(4f - 22).$$

Now for given values of M_x, M_y, $f_x(M_x)$ and $f_y(M_y)$ one will get the three optimum values of d for which $M[T_m(d)]$ attains the minimum value. The possibility of getting negative or imaginary roots cannot be ruled out. However, Singh and Shukla (1987) has pointed out that for any choice of f, M_x, M_y, $f_x(M_x)$ and $f_y(M_y)$, there exists at least one positive real root of the equation (4.12) ensuring that $M[T_m(d)]$ attaints its minimum within the parameter space $(0, \infty)$. Since, there may exist at most three optimum values of d, a criterion for suitable value of optimum d may be set as follows: "Out of all possible values of optimum d, choose $d = d_0$ as an adequate choice, which makes $|B[T_m(d)]|$ smallest".

Hence, the minimum mean square error of $T_m(d)$ is given by

(4.13) $M[T_m(d)]_{\text{opt.}} = \frac{1}{m} A_1 + \left(\frac{1}{m} - \frac{1}{n} \right) A_4$

where $A_1 = \frac{\{f_y(M_y)\}^{-2}}{4}$, $A_4 = \alpha^{*2} A_2 + 2\alpha^* A_3$, and $\alpha^* = [\alpha]_{d=d_0}$.

Further, substituting the expression for $V(T_u)$ and $M[T_m(d)]_{\text{opt.}}$ in equation (4.9) we get the expression for $M(\widehat{T}_d)$ as in equation (4.6). □

4.3. Remark. The cubic equation (4.12) depends on the population parameters P_{yx}, $f_y(M_y)$ and $f_x(M_x)$. If these parameters are known, the proposed estimator can be easily applied. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their sample estimates. The population proportion P_{yx} can be replaced by the sample estimate $\widehat{P}_{yx}$ and the marginal densities $f_y(M_y)$ and $f_x(M_x)$ can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator (Silverman (1986)). Here, the marginal densities $f_y(M_y)$ and $f_x(M_x)$ are replaced by $\widehat{f}_y(\widehat{M}_y(m))$ and $\widehat{f}_x(\widehat{M}_x(n))$ respectively, which are obtained by method of generalized nearest neighbour density estimation related to kernel estimator.

To estimate $f_y(M_y)$ and $f_x(M_x)$, by generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{\frac{1}{2}}$ and define the distance $\delta(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\widehat{M}_x(n)$, define $\delta_1(\widehat{M}_x(n)) \leq \delta_2(\widehat{M}_x(n)) \leq \cdots \leq \delta_n(\widehat{M}_x(n))$ to be the distances, arranged in ascending order, from $\widehat{M}_x(n)$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\widehat{f}(\widehat{M}_x(n)) = \frac{1}{n\delta_h(\widehat{M}_x(n))} \sum_{i=1}^{n} K\left[\frac{\widehat{M}_x(n) - x_i}{\delta_h(\widehat{M}_x(n))}\right]$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = \frac{1}{2\pi}e^{-(\frac{1}{2}x^2)}$.

The estimate of $f_y(M_y)$ can be obtained by the above explained procedure in similar manner.

4.4. Theorem. *The estimator $\widehat{T}_d$, its bias and mean square error are asymptotically convergent to the estimator $\widehat{T}_1$, its bias and mean square error respectively for large d.*

Proof. Taking limit as $d \to \infty$ in equation (3.3) we get

$$\lim_{d\to\infty} \widehat{T}_d = \varphi T_u + (1-\varphi) \lim_{d\to\infty} T_m(d).$$

Since, $d \neq 0$, dividing numerator and denominator of the second term in right hand side of above equation by d^3 and taking limit as $d \to \infty$, we have

$$\lim_{d\to\infty} \widehat{T}_d = \varphi T_u + (1-\varphi)T_m(1) = \widehat{T}_1.$$

This is the ratio type estimator to estimate population median in two occasion rotation sampling as given in Remark 3.2. Similarly, using the expressions of bias and mean square error of the estimator $\widehat{T}_d$, it is easy to see that

$$\lim_{d\to\infty} B\{\widehat{T}_d\} = B\{\widehat{T}_1\}$$

and

$$\lim_{d\to\infty} M\{\widehat{T}_d\} = M\{\widehat{T}_1\}.$$

Thus the proposed class of estimators converges to a well-defined estimator even if one chooses arbitrary, a larger value of the unknown parameter d. The bias and mean square error also tends asymptotically to that of ratio type estimator to estimate finite population median. There is no need to bother about the existence of the estimator while choosing a larger value of d. $\square$

5. Minimum Mean Square Error of the Proposed Class of Estimators $\widehat{T}_d$

Since, mean square error of $\widehat{T}_d$ in equation (4.6) is function of unknown constant φ, therefore, it is minimized with respect to φ and subsequently the optimum value of φ is obtained as

$$(5.1) \qquad \varphi_{\text{opt.}} = \frac{M\{T_m(d)\}_{\text{opt.}}}{V(T_u) + M\{T_m(d)\}_{\text{opt.}}}$$

and substituting the value of $\varphi_{\text{opt.}}$ from equation (5.1) in equation (4.6), we get the optimum mean square error of the estimator $\widehat{T}_d$ as

$$(5.2) \qquad M(\widehat{T}_d)_{\text{opt.}} = \frac{V(T_u) \cdot M\{T_m(d)\}_{\text{opt.}}}{V(T_u) + M\{T_m(d)\}_{\text{opt.}}} \, .$$

Further, by substituting the values from equation (4.7) and equation (4.8) in equation (5.2), we get the simplified value of $M(\widehat{T}_d)_{\text{opt.}}$ as

$$(5.3) \qquad M(\widehat{T}_d)_{\text{opt.}} = \frac{A_1[A_1 + \mu A_4]}{n[A_1 + \mu^2 A_4]},$$

where $\mu\,(=u/n)$ is the fraction of fresh sample drawn on the current (second) occasion. Again $M(\widehat{T}_d)_{\text{opt.}}$ derived in equation (5.3) is the function of μ. To estimate the population median on each occasion the better choice of μ is 1 (case of no matching); however, to estimate the change in median from one occasion to the other, μ should be 0 (case of complete matching). But intuition suggests that an optimum choice of μ is desired to devise the amicable strategy for both the problems simultaneously.

6. Optimum Replacement Policy

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fraction of sample to be drawn afresh on current occasion) so that M_y may be estimated with maximum precision, we minimize $M(\widehat{T}_d)_{\text{opt.}}$ in equation (5.3) with respect to μ.

The optimum value of μ so obtained is one of the two roots given by

$$(6.1) \qquad \widehat{\mu} = \frac{-A_1 \pm \sqrt{A_1(A_1 + A_4)}}{A_4} \, .$$

The real value of $\widehat{\mu}$ exists, iff $A_1(A_1 + A_4) \geq 0$. For any situation, which satisfies this condition, two real values of $\widehat{\mu}$ may be possible, hence in choosing a value of $\widehat{\mu}$, care should be taken to ensure that $0 \leq \widehat{\mu} \leq 1$, all other values of $\widehat{\mu}$ are inadmissible. If both the real values of $\widehat{\mu}$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of $\widehat{\mu}$ say μ_0 from equation (6.1) in equation (5.3), we get the optimum value of the mean square error of the estimator $\widehat{T}_d$ with respect to φ and μ both as

$$M(\widehat{T}_d)_{\text{opt.}^*} = \frac{A_1[A_1 + \mu_0 A_4]}{n[A_1 + \mu_0^2 A_4]} \, .$$

7. Efficiency Comparison

To evaluate the performance of the estimator $\widehat{T}_d$, the estimator $\widehat{T}_d$ at optimum conditions is compared with respect to the estimator $\widehat{M}_y(n)$(the sample median), when there is no matching from previous occasion. Since, $\widehat{M}_y(n)$ is unbiased for population median, its variance for large N is given by

$$(7.1) \qquad V[\widehat{M}_y(n)] = \frac{1}{n} \frac{\{f_y(M_y)\}^{-2}}{4} \, .$$

The percent relative efficiency of the estimator $\widehat{T}_d$ (under optimal condition) with respect to $\widehat{M}_y(n)$ is given by

$$(7.2) \qquad \text{P.R.E.}(\widehat{T}_d, \widehat{M}_y(n)) = \frac{V[\widehat{M}_y(n)]}{M(\widehat{T}_d)_{\text{opt.*}}} \times 100 \, .$$

The estimator $\widehat{T}_d$(at optimal conditions) is also compared with respect to the estimators $\widehat{T}_1$, $\widehat{T}_2$ and $\widehat{T}_3$, respectively. Hence for large N, the expressions for optimum mean square errors of $\widehat{T}_1$, $\widehat{T}_2$ and $\widehat{T}_3$ are given by

$$M(\widehat{T}_1)_{\text{opt.*}} = \frac{A_1[A_1 + \mu_1 A_5]}{n[A_1 + \mu_1^2 A_5]},$$

$$M(\widehat{T}_2)_{\text{opt.*}} = \frac{A_1[A_1 + \mu_2 A_6]}{n[A_1 + \mu_2^2 A_6]}$$

and

$$M(\widehat{T}_3)_{\text{opt.*}} = \frac{A_1[A_1 + \mu_3 A_7]}{n[A_1 + \mu_3^2 A_7]}$$

where

$$\mu_1 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_5}}{A_5}, \qquad \mu_2 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_6}}{A_6},$$

$$\mu_3 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_7}}{A_7}, \qquad A_1 = \frac{\{f_y(M_y)\}^{-2}}{4},$$

$$A_5 = A_2 - 2A_3, \quad A_6 = A_2 + 2A_3 \ \text{ and } \ A_7 = \left(\frac{f}{1+f}\right)^2 A_2 + 2\left(\frac{f}{1+f}\right) A_3 \, ,$$

where $A_2 = \frac{\{f_x(M_x)\}^{-2}}{4}\left[\frac{M_y^2}{M_x^2}\right]$ and $A_3 = (P_{yx} - 0.25)\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}\left[\frac{M_y}{M_x}\right].$

The percent relative efficiencies of $\widehat{T}_d$ at optimum conditions with respect to the estimators $\widehat{T}_i$ for $i = 1, 2$ and 3 at optimum conditions are given by

$$\text{P.R.E.}(\widehat{T}_d, \widehat{T}_i) = \frac{M(\widehat{T}_i)_{\text{opt.*}}}{M(\widehat{T}_d)_{\text{opt.*}}} \times 100 \text{ for } i = 1, 2 \text{ and } 3 \, .$$

## 8. Numerical Illustrations

The various results obtained in previous sections are now illustrated using two natural populations.

Population Source. (Free access to the data by Statistical Abstracts of the United States) In the first case, a real life situation consisting $N = 51$ states of United States has been considered. Let y_i represent the number of abortions during 2007 in the ith state of U.S. and x_i be the number of abortions during 2005 in the ith state of U.S. The data are presented pictorially in Figure 8.1 as under:
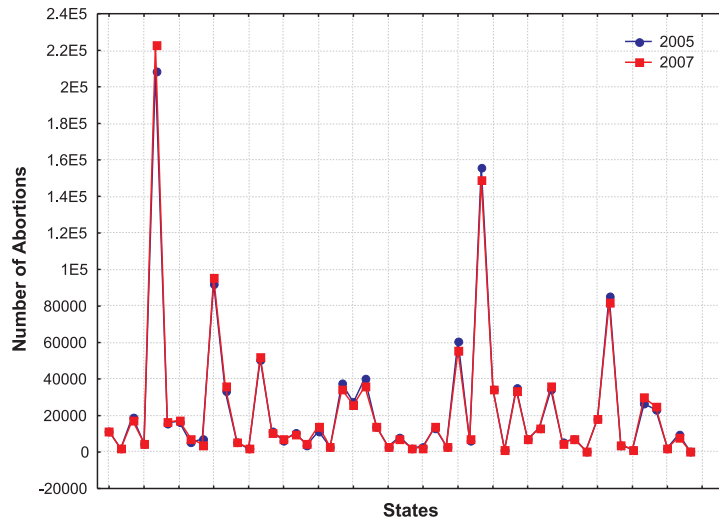
Figure 8.1. Number of Abortions during 2005 and 2007 versus different states of U.S.

Similarly in the second case, the study population consist of $N = 51$ states of United States for year 2004. Let y_i (study variable) be the percent of bachelor degree holders or more in the year 2004 in the ith state of U.S. and x_i be the percent of bachelor degree holders or more in the year 2000 in the ith state of U.S. The data are represented pictorially in Figure 8.2 as under:
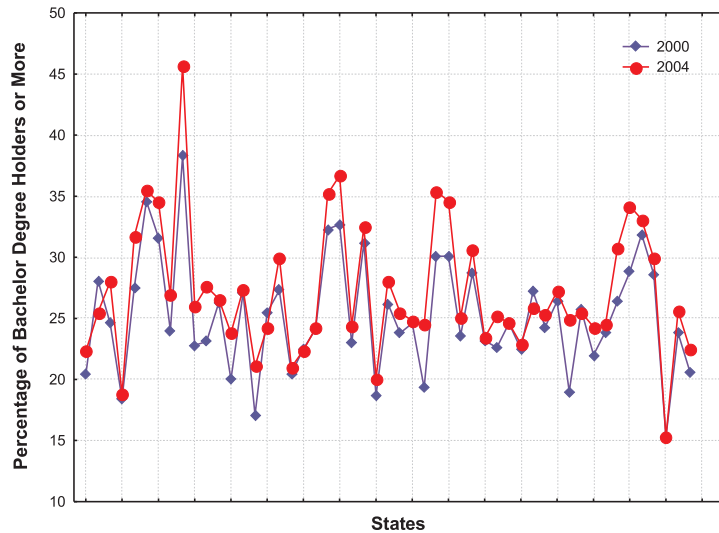
Figure 8.2. Percent of Bachelor Degree Holders or More during 2000 and 2004 versus Different States of U.S.

Table 1. Descriptive Statistics for Population-I and Population-II

	Population-I		Population-II	
	Number of Abortions in 2005	Number of Abortions in 2007	% of Bachelor Degree Holders or More in 2000	% of Bachelor Degree Holder or More in 2004
Mean	23651.76	23697.65	27.19	27.17
Standard Error	5389.35	5510.75	0.65	0.75
Median	10410.00	9600.00	24.60	25.50
Standard Deviation	38487.71	39354.65	4.66	5.40
Kurtosis	12.39	14.42	0.29	1.67
Skewness	3.31	3.52	0.40	0.89
Minimum	70.00	90.00	15.30	15.30
Maximum	208430.00	223180.00	30.30	45.70

The graph in Figure 8.1 shows that the distribution of number of abortions in different states is skewed towards right. Similar graph is obtained for Population-II as indicated in Figure 8.2. One reason of skewness may be the distribution of population in different states, that is, the states having larger populations are expected to have larger number of abortion cases and the larger percent of bachelor degree holders or more for the second case as well. Thus skewness of the data indicates that the use of median may be a good measure of central location than mean in such a situation.

Based on the above description, the descriptive statistics for both populations have been computed and are presented in Table 1.

For the two populations under consideration, the cubic equation (4.12) is solved for "d" for some choices of "f". The optimum mean square errors of the proposed class of estimators are found to be same for all the three values of "d" obtained. So, using the criteria set in the proof of Theorem 4.1, Table 2 shows the best choice of the optimum value of "d" for different choices of "f" for both, Population-I and Population-II.

Table 2. Best choice of d for Population-I and Population-II, for different choices of f

f	Population-I			Population-II		
	d	|Bias|	d_0	d	|Bias|	d_0
	10.0002	3.6526	**2.4170**	22.8356	0.1419	**2.3553**
0.9800	2.4170	0.3097		2.3533	0.1089	
	1.4705	4.1206		1.2030	0.1467	
	10.7520	1.8948	**2.6449**	2.5878	1.3940	**25.5834**
0.1960	2.6449	1.2919		25.5834	0.0702	
	1.3740	2.1515		1.1537	0.0748	
	11.5280	1.3005	**11.5280**	28.3715	0.0486	**28.3715**
0.2941	2.8115	1.5131		2.7621	0.1526	
	1.3146	1.4675		1.1244	0.0504	
	12.3230	0.9984	**12.3230**	31.1885	0.0367	**31.1885**
0.3922	2.9414	1.5271		2.8979	0.1562	
	1.2729	1.1168		1.1047	0.0381	
	13.1327	0.8141	**13.1327**	34.0268	0.0296	**34.0268**
0.4902	3.0462	1.4584		3.0070	0.1532	
	1.2417	0.9026		1.0905	0.0306	

Table 3. Optimum value of μ and percent relative efficiencies of $\widehat{T}_d$ at optimum conditions with respect to $\widehat{M}_y(n)$ and $\widehat{T}_i$ for $i = 1, 2$ and 3 at optimum conditions

	Population-I	**Population-II**
f	0.9800	0.9800
d_0	2.4170	2.3553
μ_0	0.6800	0.6271
P.R.E.$(\widehat{T}_d, \widehat{M}_y(n))$	136.00	125.41
P.R.E.$(\widehat{T}_d, \widehat{T}_1)$	103.33	100.16
P.R.E.$(\widehat{T}_d, \widehat{T}_2)$	206.73	173.48
P.R.E.$(\widehat{T}_d, \widehat{T}_3)$	128.93	120.81

9. Interpretation of Results and Conclusion

(1) From Table 2, it can clearly be seen that the real optimum value of 'd' always exists for both the considered populations. This justifies the feasibility of the proposed class of estimators $\widehat{T}_d$.

(2) From Table 3, it can be seen that the optimum value of μ also exist for both the considered populations. Hence, it indicates that the proposed class of estimators $\widehat{T}_d$ is quite feasible under optimal conditions.

(3) Table 3 indicates that the proposed class of estimators $\widehat{T}_d$ at optimum conditions is highly preferable over sample median estimator $\widehat{M}_y(n)$. It also performs better than the estimators $\widehat{T}_1$ and $\widehat{T}_2$ which are the estimators proposed by Singh et al. (2007) for second quantile. It also proves to be highly efficient than the estimator $\widehat{T}_3$ which is a Dual to Ratio type estimator, a member of its own class.

Hence, it can be concluded that the estimation of median at current occasion is certainly feasible in two occasion successive sampling. The enchanting convergence property of proposed class of estimators $\widehat{T}_d$ justifies the incorporation of unknown parameter in the structure of proposed class of estimators, since the optimum value of the parameter always exists. Hence the proposed class of estimators $\widehat{T}_d$ can be recommended for its further use by survey practitioners.

References

[1] Allen, J., Saxena, S., Singh, H.P., Singh, S. and Smarandache, F. *Randomness and optimal estimation in data sampling*, American Research Press, 26–43, 2002.

[2] Arnab, R and Okafor, F.C. *A note on double sampling over two occasions*, Pakistan Journal of Statistics, 9–18, 1992.

[3] Eckler, A.R. *Rotation Sampling*, Ann. Math. Statist., 664–685, 1955.

[4] Feng, S. and Zou, G. *Sample rotation method with auxiliary variable*, Commun. Statist. Theo-Meth. **26**(6), 1497–1509, 1997.

[5] Gordon, L. *Successive sampling in finite populations*, The Annals of Statistics **11**(2), 702–706, 1983.

[6] Gross, S.T. *Median estimation in sample surveys*, Proc. Surv. Res. Meth. Sect. Amer. Statist. Assoc., 181–184, 1980.

[7] Jessen, R.J. *Statistical investigation of a sample survey for obtaining farm facts*, Iowa Agricultural Experiment Station Road Bulletin No. **304**, Ames, 1–104, 1942.

[8] Kuk, A.Y.C. and Mak, T.K. *Median estimation in presence of auxiliary information*, J.R. Statit. Soc. **B 51**, 261–269, 1989.

[9] Martinez-Miranda, M.D., Rueda-Garcia, M., Arcos-cebrian, A., Roman-Montoya, Y. and Gonzaez-Aguilera, S. *Quintile estimation under successive sampling*, Computational Statistics **20**, 385–399, 2005.

[10] Narain, R.D. *On the recurrence formula in sampling on successive occasions*, Journal of the Indian Society of Agricultural Statistics **5**, 96–99, 1953.

[11] Patterson, H.D. *Sampling on successive occasions with partial replacement of units*, Jour. Royal Statist. Assoc., Ser. **B 12**, 241–255, 1950.

[12] Rueda, M.D.M. and Munoz, J.F. *Successive sampling to estimate quantiles with P-Auxiliary variables*, Quality and Quantity **42**, 427–443, 2008.

[13] Sedransk, J. and Meyer, J. *Confidence intervals for n quantiles of a finite population, simple random and stratified simple random sampling*, J. R. Statist. Soc. **B 40**, 239–252, 1978.

[14] Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, (Chapman and Hall,London, 1986).

[15] Singh, G.N. and Singh, V.K. *On the use of auxiliary information in successive sampling*, Jour. Ind. Soc. Agri. Statist. **54**(1), 1–12, 2001.

[16] Singh, G.N. and Priyanka, K. *Search of good rotation patterns to improve the precision of estimates at current occasion*, Communications in Statistics (Theory and Methods) Vol. **37**(3), 337–348, 2008.

[17] Singh, G.N., Prasad, S. and Majhi, D. *Best Linear Unbiased Estimators of Population Variance in successive Sampling*, Model Assisted Statistics and Applications **7**, 169-178, 2012.

[18] Singh, H.P., Tailor, R., Singh, S. and Jong-Min Kim. *Quintile Estimation in Successive Sampling*, Journal of the Korean Statistical Society **36**(4), 543–556, 2007.

[19] Singh, S. *Advanced Sampling Theory with Applications*, How Michael 'selected' Amy. Vol. **1** and **2** (Kluwer Academic Publishers, The Netherlands, 2003) 1–1247.

[20] Singh, V.K. and Shukla, D. *One parameter family of factor-type ratio estimators*, Metron **45**, 1-2, 30, 273–283, 1987.

[21] Smith, P. and Sedransk, J. *Lower bounds for confidence coefficients for confidence intervals for finite population quantiles*, Commun. Statist. - Theory Meth. **12**, 1329–1344, 1983.