# Sufficient Weighted Bootstrapping

Khairul Islam[1] and Sarjinder Singh[2]

[1]Department of Mathematics and Statistics
Eastern Michigan University, Ypsilanti, MI 48197

[2]Department of Mathematics
Texas A&M University-Kingsville, TX 78363

**Abstract**
Singh and Sedory [1] have introduced the idea of sufficient bootstrapping, which is based on retaining only distinct individual responses from a simple random and with replacement sampling (SRSWR). The idea of usage of the distinct units in SRSWR at the estimation stage is well known from the work of Raj and Khamis [2]. Basu [3] has investigated the theory of distinct units under probability proportional to size and with replacement sampling (PPSWR). Singh [4] has studied a weighted bootstrapping in the presence of auxiliary information and named as saddlestrapping because there exists a saddlepoint based on the correlation between the study and auxiliary variable. In this paper, we extend the idea of distinct units in PPSWR sampling to construct a new Sufficient Weighted Bootstrapping.

**Key words:** Bootstrapping, estimation of mean, auxiliary variable, sufficient weighted bootstrapping, simulation.

## 1. Introduction

Due to Efron [5, 6, 7], the bootstrapping is a computationally intensive statistical technique that replaces traditional algebraic derivations with databased computer simulation. The method is called bootstrapping because it involves resampling from the original data set, and hence is widely known as resampling procedure. For details, one could also refer to the books on bootstrapping by Efron and Tibshirani [8] and Chernick [9]. Casella [10] provides an introduction to the Silver Anniversary of the Bootstrap. Efron [11] discusses a second thought on bootstrapping. Davison et al. [12] have a critical review on recent developments in bootstrap methodology during the year 2003. Beran [13], Lele [14], Shao [15], Lahiri [16], and Politis [17] explain the impact of bootstrap on statistical algorithms and theory, estimating functions, sample surveys, small area estimation and time series, respectively. Ernst and Hutson [18] and Rueda et al. [19, 20] discuss application of bootstrapping for quantile estimation. Holmes [21] and Soltis and Soltis [22] discuss applications of bootstrapping in phylogenetic trees and phylogeny reconstruction respectively. Holmes et al. [23] provide an overview of a conversation on bootstrap between Bradley Efron and other good friends.

The use of auxiliary information in survey sampling has an eminent role to improve methods of sample selection and use them to estimate various parameters of interest.

[1] kislam@emich.edu

Singh [4] introduced a new method called Saddlestrapping, which makes use of the auxiliary information correlated with the study variable as weight, constructed by means of the probability proportional to size and with replacement (PPSWR), to select different bootstrapping samples from the given original sample. To learn more about the PPSWR sampling one can refer to Lahiri [24], Hansen and Hurvitz [25] and Singh [26]. It appears that the bootstrapping method due to Efron [5] is a special case of the saddlestrapping method. Barbe and Bertail [27] have published a monograph on weighted bootstrapping which presents an account of the asymptotic behavior of the weighted bootstrap – a new technique and powerful statistical technique. Researchers and advanced graduate students studying bootstrap method will find this a valuable technical survey, which is thorough and rigorous. The main aim of their monograph is to answer two questions: how well does the generalized bootstrap work? What are the differences between all the different weighted schemes? Lee and Young [28] have discussed pre-pivoting by weighted bootstrap iteration. Johnson [29] has provided a nice introduction to bootstrapping. The Saddlestrapping [4] is a kind of weighted bootstrapping, which aims to see the effect of correlation between the study and auxiliary variable on the distributions of sample means; this kind of study has recently been performed by Johnson [29]. Likewise, Hesterberg [30] taught a very valuable course entitled, “Bootstrap methods and permutation tests” during the conference of Statisticians at San Antonio, TX. Sing and Sedory [1] proposed sufficient bootstrapping by retaining only distinct individual responses from the bootstrap samples. They demonstrated that the sufficient bootstrapping could have better relative efficiency than the conventional bootstrapping in certain situations.

In this paper, we extend the idea of distinct units in PPSWR sampling to construct a new Sufficient Weighted Bootstrapping (SWB). The main objectives are to compare the means of various bootstrapping methods, empirically, and investigate the relative efficiency of the sufficient weighted bootstrapping over conventional bootstrapping due to Efron [5], recently proposed sufficient bootstrapping due to Singh and Sedory [1] and saddlestrapping due to Singh [4].

## 2. Proposed Methods

Let $Y$ and $X$ be the study and auxiliary variables in a finite population $\Omega$ of $N$ units having positive linear correlation with each other. Consider the problem of estimation of population mean $\bar{Y} = \frac{1}{N}\sum_{i\in\Omega} y_i$ of the study variable $y$ while using information on the auxiliary variable $x$. Note that the distribution of the study variable $Y$ will depend on the value of its correlation with the auxiliary variable $X$.

Let $s_n = \{(y_i, x_i): i = 1, 2, \ldots, n\}$ be a random sample of size $n$ taken from the population $\Omega$ by using any sampling scheme among the list of 50 sampling schemes available in Brewer and Hanif [31].

Let $\bar{y} = \frac{1}{n}\sum_{k\in s_n} y_k$ be the mean of the original sample and

$$s_y^2 = \frac{1}{n-1}\sum_{k\in s_n}(y_k - \bar{y})^2$$

be the sample variance of the original sample.

Let $s_b = \{(y_{b_i}): i = 1, 2, \ldots, n; b = 1, 2, \ldots, n^n\}$, be the $b$th bootstrap sample from the original sample $s_n$ and

$$p_i = \frac{x_i}{\sum_{i \in s_n} x_i}$$

be the probability of selecting the $i$th unit from the original sample $s_n$.

We define a new estimation method using only unique values of $y_i$ from the bootstrap sample with $i$th unit of $Y$ chosen with the weight $p_i$. This method is termed as a sufficient weighted bootstrap method, and the estimate of $\bar{y}$ is given by

$$\bar{y}_{swb} = \frac{1}{n\nu} \sum_{i \in \mathrm{s}_\nu} \frac{y_i}{p_i} \quad (1)$$

where $\mathrm{s}_\nu$ is the weighted sufficient bootstrapping sample of $\nu$ distinct units of $Y$ in $\mathrm{s}_b$ and $\bar{y}_{swb}$ is the mean of the sufficient weighted bootstrap sample of size $\nu$.

Singh [4] used $p_i$ as weights to select $i$th unit from the original sample and called such a method as saddlestrapping or weighted bootstrapping where all units of a given bootstrap sample $\mathrm{s}_b$ are kept.

Now, we provide some properties of the sufficient weighted bootstrap mean $\bar{y}_{swb}$ as theorems and particular cases of the theorems as lemmas.

**Theorem 1:** The sufficient weighted bootstrap sample mean $\bar{y}_{swb} = \frac{1}{n\nu} \sum_{i \in \mathrm{s}_\nu} y_i / p_i$ is unbiased for the mean of the original sample for the study variable $Y$.
***Proof:*** Let $\pi_i^* = \nu p_i,\ i = 1, 2, \ldots,\ \nu$ be the induced inclusion probability for a given sufficient bootstrap sample.
Let $E_2$ denote the expected value over all possible sufficient weighted bootstrap samples, each of size $\nu \geq 1$.
Let $E_1$ be the expected value of selecting $n$ distinct units, from the given original sample $s_n$ of $n$ distinct units.
The estimator $\bar{y}_{swb}$ can be written as

$$\bar{y}_{swb} = \frac{1}{n\nu} \sum_{i \in \mathrm{s}_\nu} \frac{y_i}{p_i} = \frac{1}{n} \sum_{i \in \mathrm{s}_\nu} \frac{y_i}{\nu p_i} = \frac{1}{n} \sum_{i \in \mathrm{s}_\nu} \frac{y_i}{\pi_i^*}$$

Taking expected value on both sides we have

$$E(\bar{y}_{swb}) = E_1 E_2 \left[ \frac{1}{n} \sum_{i \in \mathrm{s}_\nu} \frac{y_i}{\pi_i^*} \right] = E_1 \left[ \frac{1}{n} \sum_{i=1}^{n} y_i \right] = E_1[\bar{y}_n] = \bar{y}_n$$

which proves the unbiasedness of the weighted sufficient bootstrap sample mean.

**Theorem 2:** The variance of the sufficient weighted bootstrap sample mean $\bar{y}_{swb}$ for fixed $\nu \geq 2$ is given by

$$V(\bar{y}_{swb}) = \frac{1}{2n^2} \sum_{i \neq j} \sum_{\in \mathrm{s}_n} \left( \pi_i^* \pi_j^* - \pi_{ij}^* \right) \left( \frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*} \right)^2$$

***Proof:*** Let $\pi_{ij}^* = P[(i, j) \in \mathrm{s}_\nu]$ be the induced inclusion probabilities for the two distinct units to be included in the sufficient bootstrap sample $\mathrm{s}_\nu$.
Let $V_2$ be the variance over all possible sufficient weighted bootstrap samples $\mathrm{s}_\nu$, each of size $\nu \geq 2$. Note that if $\nu = 1$, then $V(\bar{y}_{swb}) = 0$.
The variance of the estimator $\bar{y}_{swb}$ for fixed $\nu \geq 2$ is given by

$$\begin{aligned}
V(\bar{y}_{swb}) &= E_1 V_2\left[\frac{1}{n\nu}\sum_{i\in s_\nu}\frac{y_i}{p_i}\right] + V_1 E_2\left[\frac{1}{n\nu}\sum_{i\in s_\nu}\frac{y_i}{p_i}\right] \\
&= E_1 V_2\left[\frac{1}{n}\sum_{i\in s_\nu}\frac{y_i}{\pi_i^*}\right] + V_1 E_2\left[\frac{1}{n}\sum_{i\in s_\nu}\frac{y_i}{\pi_i^*}\right] \\
&= E_1\left[\frac{1}{2n^2}\sum_{i\neq j}\sum_{\in s_n}\left(\pi_i^*\pi_j^* - \pi_{ij}^*\right)\left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*}\right)^2\right] + V_1\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] \\
&= \frac{1}{2n^2}\sum_{i\neq j}\sum_{\in s_n}\left(\pi_i^*\pi_j^* - \pi_{ij}^*\right)\left(\frac{y_i}{\pi_i^*} - \frac{y_j}{\pi_j^*}\right)^2 + 0
\end{aligned}$$

where $V_1\left[\frac{1}{n}\sum_{i\in s_n} y_i\right] = 0$ because $V_1$ is the variance over all possible sample of $n$ distinct units out of $n$ distinct units in sample $s_n$. Hence the theorem.

**Lemma 1:** If we let $\pi_i^* = \frac{\nu}{n}$ and $\pi_{ij}^* = \frac{\nu(\nu-1)}{n(n-1)}$, then

$$V_{SRS}(\bar{y}_{swb}) = \left(\frac{1}{\nu} - \frac{1}{n}\right) s_y^2$$

***Proof:***. Note that

$$\pi_i^*\pi_j^* - \pi_{ij}^* = \left(\frac{\nu}{n}\right)^2 - \frac{\nu(\nu-1)}{n(n-1)} = \left(\frac{\nu}{n}\right)\left[\frac{\nu}{n} - \frac{\nu-1}{n-1}\right] = \left(\frac{\nu}{n}\right)\left(\frac{n-\nu}{n(n-1)}\right)$$

So,

$$\begin{aligned}
V(\bar{y}_{swb}) &= \frac{1}{2n^2}\frac{n-\nu}{\nu(n-1)}\sum_{i\neq j}\sum_{\in s_n}\left(y_i - y_j\right)^2 \\
&= \frac{1-\frac{\nu}{n}}{2n\nu(n-1)}\sum_{i\neq j}\sum_{\in s_n}\left(y_i - y_j\right)^2 \\
&= \left(\frac{1}{\nu} - \frac{1}{n}\right)\frac{1}{2n(n-1)}\sum_{i\neq j}\sum_{\in s_n}\left(y_i - y_j\right)^2 \\
&= \left(\frac{1}{\nu} - \frac{1}{n}\right) s_y^2
\end{aligned}$$

where

$$s_y^2 = \frac{1}{2n(n-1)}\sum_{i\neq j}\sum_{\in s_n}\left(y_i - y_j\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

which proves the lemma.

It appears that $V_{SRS}(\bar{y}_{swb})$ is same as the variance of the sufficient bootstrap sample mean, $V(\bar{y}_{sb})$, derived by Singh and Sedory [1].

**Lemma 2:** If we let $\pi_i^* = \nu p_i$ and $\pi_{ij}^* = \nu(\nu-1)p_i p_j$, then

$$V(\bar{y}_{swb}) = \left(\frac{1}{2n^2}\right)\left(\frac{1}{\nu}\right)\sum_{i\neq j}\sum_{\in s_n} p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

***Proof:*** By writing $\pi_i^* = \nu p_i$ and $\pi_{ij}^* = \nu(\nu-1)p_i p_j$, we have

$$V(\bar{y}_{swb}) = \frac{1}{2n^2}\sum_{i\neq j}\sum_{\in s_n}\left(\nu p_i \nu p_j - \nu(\nu-1)p_i p_j\right)\left(\frac{y_i}{\nu p_i} - \frac{y_j}{\nu p_j}\right)^2$$
$$= \frac{1}{2n^2}\sum_{i\neq j}\sum_{\in s_n}\frac{1}{\nu}\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 p_i p_j$$
$$= \left(\frac{1}{2n^2}\right)\left(\frac{1}{\nu}\right)\sum_{i\neq j}\sum_{\in s_n} p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

Note that this result is not true in case of distinct units for traditional PPSWR sampling. For example, one can see Desraj [32] and Adhikary [33]. We show through simulation that $V(\bar{y}_{swb})$ is minimum.

## 3. Evaluation of the Proposed Methods

In order to evaluate the performance of the sufficient weighted bootstrapping, in this section, we first undertake a graphical comparison among means of bootstrapping samples (**mboot**), weighted bootstrapping samples (**mwboot**), sufficient bootstrapping samples (**msboot**) and sufficient weighted bootstrapping samples (**mswboot**) for varying values of correlation between the study variable $y$ and the auxiliary variable $x$ in the given original sample. Secondly, to evaluate the performance numerically, we investigate the relative efficiency of sufficient weighted bootstrapping mean compared to bootstrapping, sufficient bootstrapping and weighted bootstrapping means.

To control the correlation between the study variable $y$ and the auxiliary variable $x$ in the given original sample, we follow the technique similar to Singh et al. [34]. We generate $n$ independent pairs of random numbers $y_i^*$ and $x_i^*$, (say), $i = 1, 2, \ldots, n$ from the standard normal distribution by using the rnorm() function available in R. For fixed values of $\sigma_y$= 15, $\sigma_x$= 10, $\mu_y$= 415, and $\mu_x$= 335, chosen arbitrarily, we generated transformed variables $y_i$ and $x_i$ as follows:

$$y_i = \mu_y + \sqrt{\sigma_y^2(1-\rho_{xy}^2)}y_i^* + \rho_{xy}\,\sigma_y x_i^* \quad (2)$$

$$x_i = \mu_x + \sigma_x x_i^* \quad (3)$$

for different values of the population correlation coefficient $\rho_{xy}$ between the study variable $Y$ and the auxiliary variable $X$ to form original sample $s_n = \{(y_i, x_i): i = 1, 2, \ldots, n\}$. For graphical comparisons, and numeric comparisons using the relative efficiency, we consider the population correlation coefficient $\rho_{xy}$ between $x$ and $y$, to be 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 and 0.95, chosen arbitrarily, for samples of sizes $n$ =10, 15, 20, 30, 50 and 100, chosen arbitrarily as well. Given an original sample $s_n$ of size $n$, with the specified values of means, standard deviations and correlations, we generate $M$ = 50, 000 resamples of size $n$ from $s_n$.

The relative efficiency of the sufficient weighted bootstrap sample mean over $M = 50{,}000$ resamples is computed using the following formula:

$$re(swb, b) = \frac{\hat{\sigma}_b}{\hat{\sigma}_{swb}} \times 100$$

$$re(swb, sb) = \frac{\hat{\sigma}_{sb}}{\hat{\sigma}_{swb}} \times 100$$

$$re(swb, wb) = \frac{\hat{\sigma}_{wb}}{\hat{\sigma}_{swb}} \times 100$$

where

$$\hat{\sigma}_{swb} = \left\{ \frac{1}{M-1} \sum_{i=1}^{M} \left( \bar{y}_{(swb)_i} - \bar{\bar{y}}_{swb} \right)^2 \right\}^{1/2}$$

$$\hat{\sigma}_{b} = \left\{ \frac{1}{M-1} \sum_{i=1}^{M} \left( \bar{y}_{(b)_i} - \bar{\bar{y}}_{b} \right)^2 \right\}^{1/2}$$

$$\hat{\sigma}_{sb} = \left\{ \frac{1}{M-1} \sum_{i=1}^{M} \left( \bar{y}_{(sb)_i} - \bar{\bar{y}}_{sb} \right)^2 \right\}^{1/2}$$

$$\hat{\sigma}_{wb} = \left\{ \frac{1}{M-1} \sum_{i=1}^{M} \left( \bar{y}_{(wb)_i} - \bar{\bar{y}}_{wb} \right)^2 \right\}^{1/2}$$

$re(swb,.)$ is the relative efficiency of mean of sufficient weighted bootstrapping sample as compared to bootstrapping/ sufficient bootstrapping/ weighted bootstrapping sample means.

$\bar{y}_{(.)_i}$ is the mean of the $i$th bootstrap/ sufficient bootstrap/ weighted bootstrap/ sufficient weighted bootstrap sample of the original sample.

$\bar{\bar{y}}_{(.)}$ is the mean of $M$ (=50,000) bootstrap/ sufficient bootstrap/ weighted bootstrap/ sufficient weighted bootstrap sample means.

$\hat{\sigma}_{(.)}$ is the estimate of standard error of bootstrap/ sufficient bootstrap/ weighted bootstrap/ sufficient weighted bootstrap sample means over $M$ (=50,000) samples from the original sample.

The sample R code used for the estimation and simulation of results is provided in the Appendix.

### 3.1 Graphical Comparisons

It appears that the simulated results are less sensitive to the sample sizes than to the values of correlation between the study and auxiliary variables. Therefore, in Figures 1-4, we provide normal density curves drawn using the means and standard deviations of the simulated bootstrap, sufficient bootstrap, weighted bootstrap and sufficient weighted bootstrap sample means, for sample sizes varying at n=10, 15, 20, 30, 50 and 100 and correlation between the study and auxiliary variables varying at $\rho = 0.65$, 0.75, 0.85 and 0.95. Other values of correlation in the range $0.65 \leq \rho \leq 0.95$ utilized in the study provide the similar pattern as reported in Figures 1-4.

In all simulations, the bootstrapping sample means seem to have the widest spread and sufficient weighted bootstrap sample means seem to have the least spread. Overall, spread of the sufficient weighted bootstrap sample means (mswboot) over 50,000 resamples are less than the means due to bootstrap (mboot), sufficient bootstrap (msboot) and weighted bootstrap sample means (mwboot).

We exclude reporting histograms of the simulated sample means because four histograms added together provide overlapping and non-distinctive images that do not provide adequate information about the spread of the distribution of respective means.

**Figure 1:** Estimated normal density of simulated mboot, msboot, mwboot and mswboot for varying sample size with the population correlation $\rho = 0.65$.

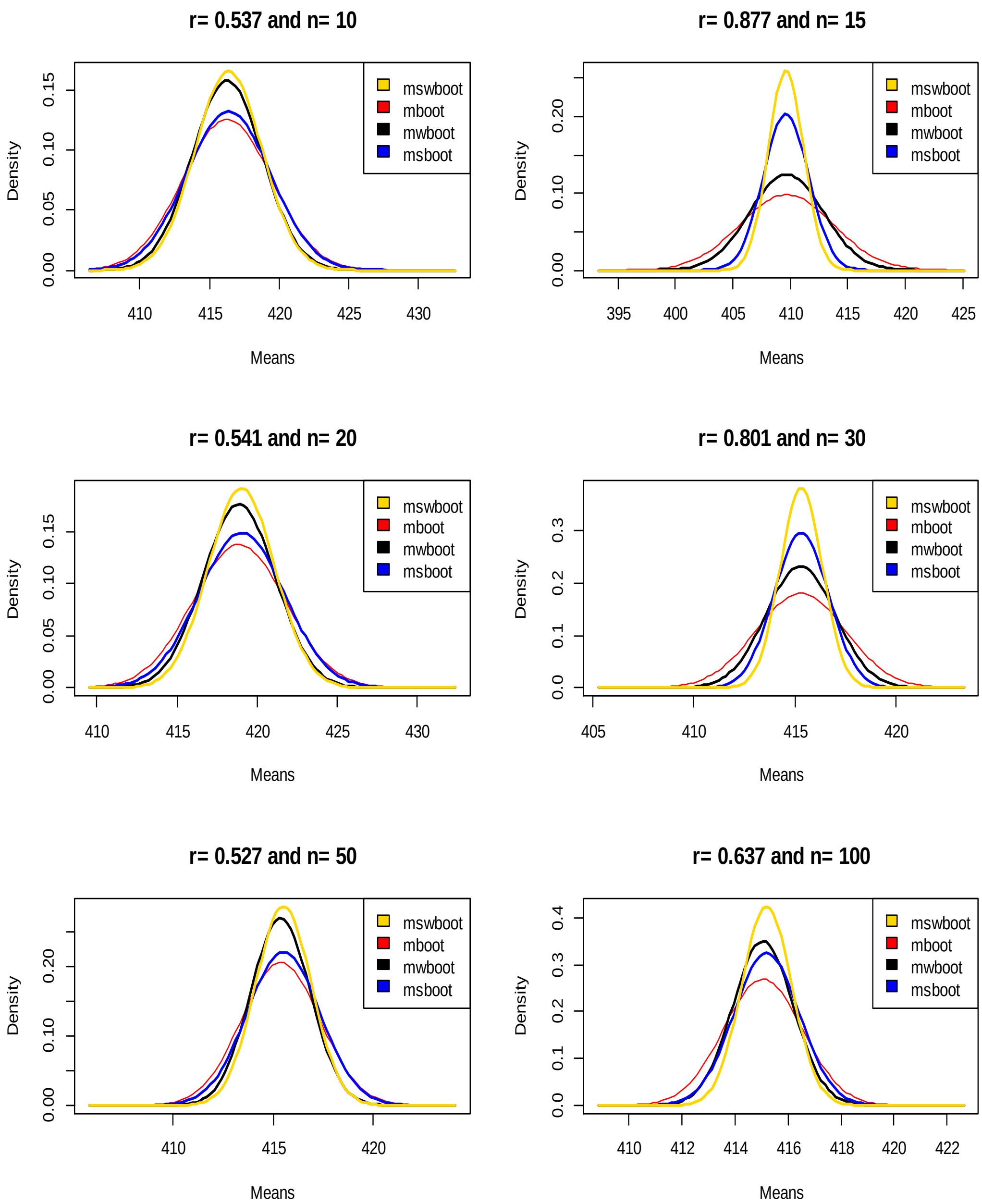


**Figure 2:** Estimated normal density of simulated mboot, msboot, mwboot and mswboot for varying sample size with the population correlation $\rho = 0.75$.

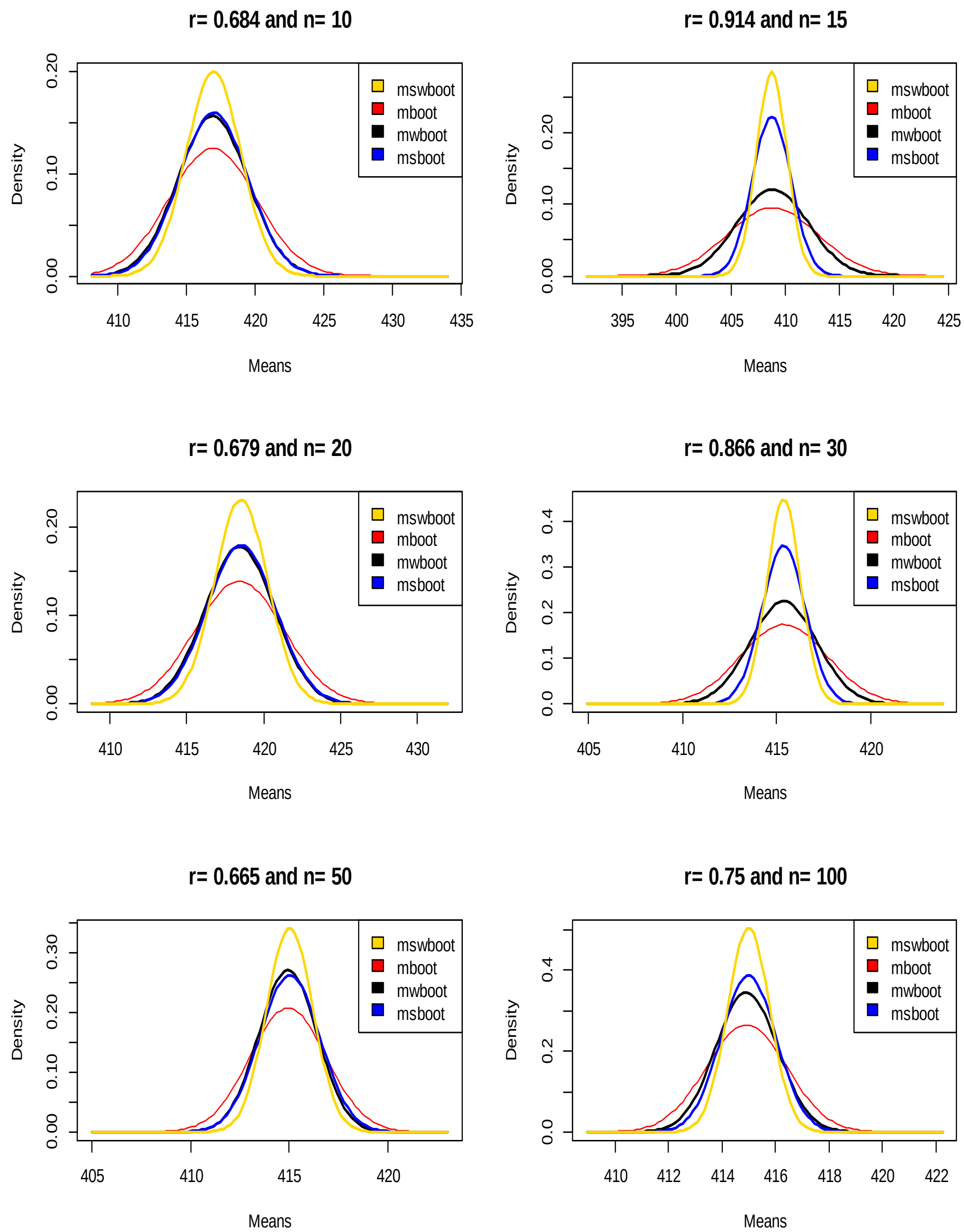


**Figure 3:** Estimated normal density of simulated mboot, msboot, mwboot and mswboot for varying sample size with the population correlation $\rho = 0.85$.

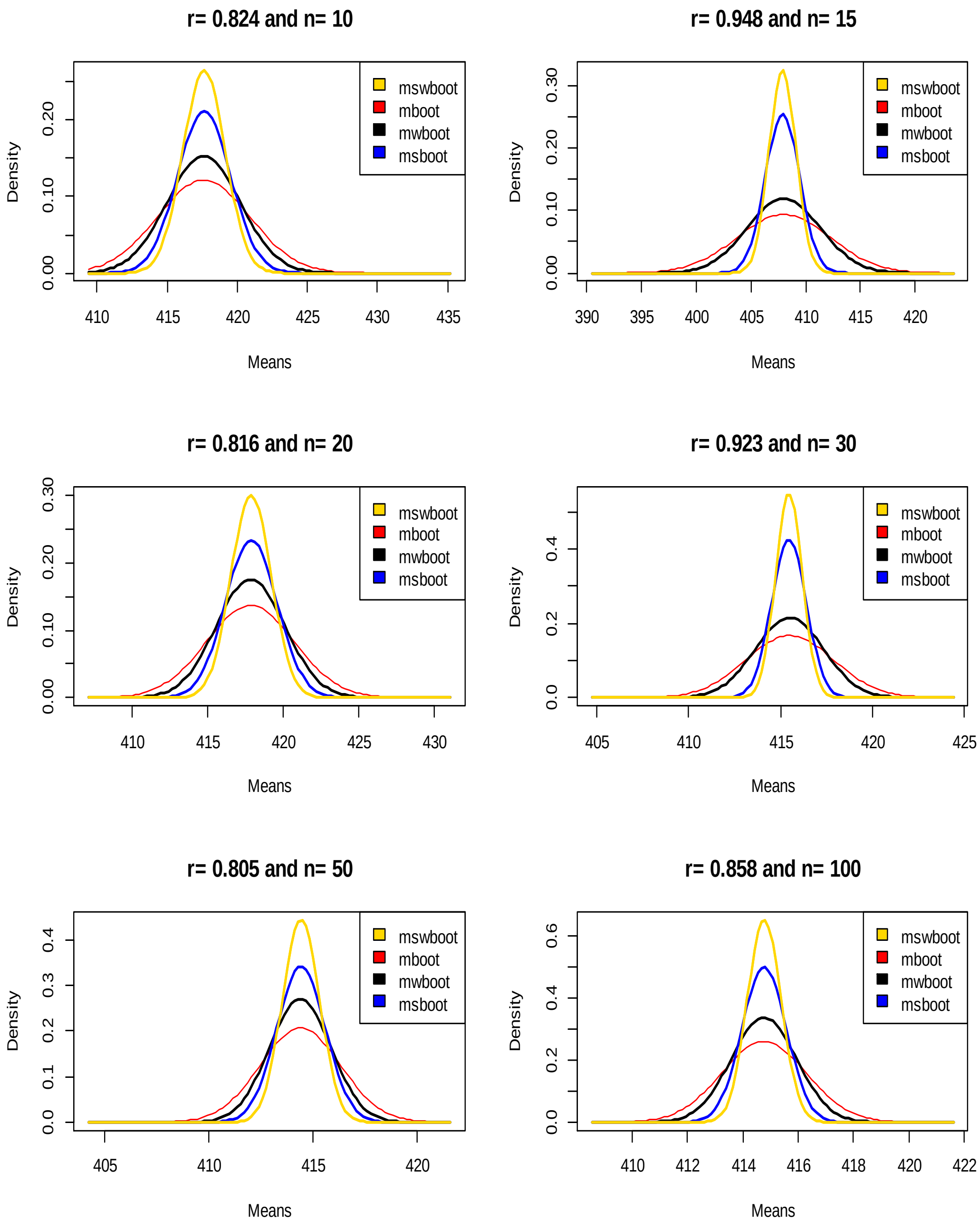
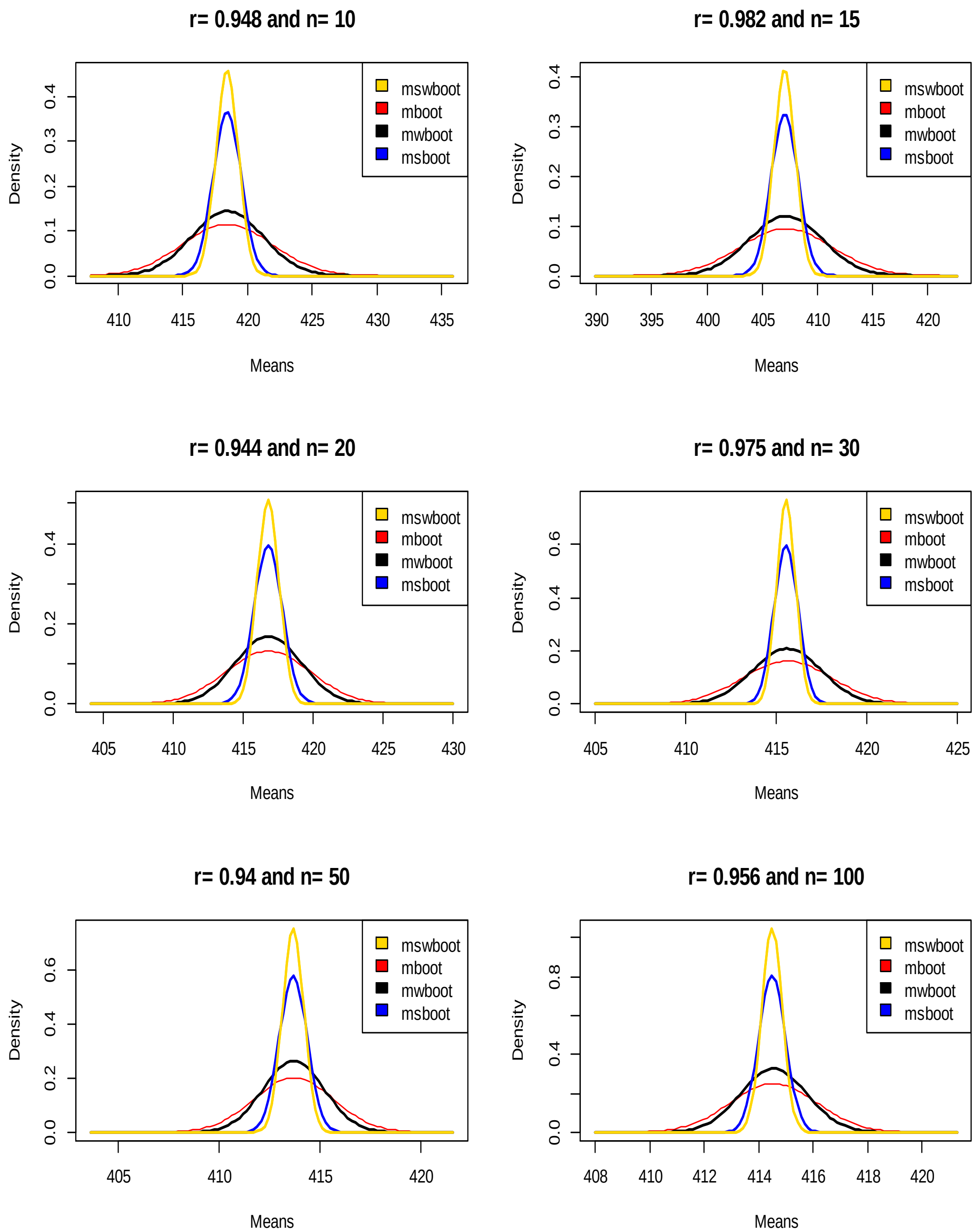


**Figure 4:** Estimated normal density of simulated mboot, msboot, mwboot and mswboot for varying sample size with the population correlation $\rho = 0.95$.



### 3.2 Relative Efficiency Comparison

In Table 1, we provide the simulated relative efficiency of mean of the sufficient weighted bootstrap sample as compared to bootstrap, sufficient bootstrap and weighted bootstrap sample means for varying values of the positive correlation between study and auxiliary variables and sample sizes.

**Table 1:** Estimated relative efficiency of mean of the sufficient weighted bootstrap sample compared to the means of bootstrap, sufficient bootstrap and weighed bootstrap samples for varying values of the population correlation between study and auxiliary variables and sample sizes.

| $\rho$ | $n$ | $r$ | $re(swb, b)$ | $re(swb, sb)$ | $re(swb, wb)$ |
|---|---|---|---|---|---|
| $\rho = 0.65$ | | | | | |
| | 10 | 0.537 | 131.88 | 105.11 | 125.27 |
| | 15 | 0.877 | 264.72 | 208.08 | 128.03 |
| | 20 | 0.541 | 139.34 | 108.63 | 128.62 |
| | 30 | 0.801 | 211.66 | 164.18 | 129.25 |
| | 50 | 0.527 | 139.57 | 106.49 | 130.04 |
| | 100 | 0.637 | 157.93 | 121.12 | 130.89 |
| $\rho = 0.70$ | | | | | |
| | 10 | 0.611 | 144.03 | 114.74 | 125.25 |
| | 15 | 0.896 | 281.12 | 221.06 | 128.01 |
| | 20 | 0.610 | 150.93 | 117.69 | 128.61 |
| | 30 | 0.835 | 232.52 | 180.39 | 129.24 |
| | 50 | 0.596 | 150.50 | 114.86 | 130.11 |
| | 100 | 0.694 | 172.34 | 132.21 | 130.89 |
| $\rho = 0.75$ | | | | | |
| | 10 | 0.684 | 160.36 | 127.69 | 125.23 |
| | 15 | 0.914 | 299.31 | 235.45 | 127.97 |
| | 20 | 0.679 | 166.31 | 129.71 | 128.60 |
| | 30 | 0.866 | 257.55 | 199.83 | 129.23 |
| | 50 | 0.665 | 165.02 | 126.01 | 130.20 |
| | 100 | 0.750 | 190.91 | 146.50 | 130.89 |
| $\rho = 0.80$ | | | | | |
| | 10 | 0.755 | 183.30 | 145.89 | 125.22 |
| | 15 | 0.931 | 320.31 | 252.07 | 127.91 |
| | 20 | 0.748 | 187.71 | 146.41 | 128.58 |
| | 30 | 0.895 | 288.53 | 223.89 | 129.22 |
| | 50 | 0.735 | 185.25 | 141.56 | 130.31 |
| | 100 | 0.805 | 215.91 | 165.74 | 130.88 |

**Table 1:** Continued

| $\rho$ | $n$ | $r$ | $re(swb, b)$ | $re(swb, sb)$ | $re(swb, wb)$ |
|---|---|---|---|---|---|
| $\rho = 0.85$ | | | | | |
| | 10 | 0.824 | 217.87 | 173.31 | 125.20 |
| | 15 | 0.948 | 346.13 | 272.50 | 127.83 |
| | 20 | 0.816 | 219.62 | 171.31 | 128.55 |
| | 30 | 0.923 | 328.69 | 255.07 | 129.18 |
| | 50 | 0.805 | 215.50 | 164.85 | 130.45 |
| | 100 | 0.858 | 251.76 | 193.30 | 130.85 |
| $\rho = 0.90$ | | | | | |
| | 10 | 0.889 | 276.31 | 219.70 | 125.19 |
| | 15 | 0.965 | 381.35 | 300.35 | 127.69 |
| | 20 | 0.882 | 272.99 | 212.93 | 128.51 |
| | 30 | 0.950 | 384.82 | 298.66 | 129.12 |
| | 50 | 0.874 | 266.35 | 204.05 | 130.64 |
| | 100 | 0.908 | 308.76 | 237.12 | 130.80 |
| $\rho = 0.95$ | | | | | |
| | 10 | 0.948 | 400.07 | 318.00 | 125.24 |
| | 15 | 0.982 | 440.07 | 346.77 | 127.48 |
| | 20 | 0.944 | 384.82 | 300.11 | 128.41 |
| | 30 | 0.975 | 475.63 | 369.15 | 129.02 |
| | 50 | 0.940 | 374.15 | 287.22 | 130.92 |
| | 100 | 0.956 | 419.00 | 321.79 | 130.66 |

On the basis of the results presented in Table 1, it is evident that the estimate of mean of the sufficient weighted bootstrap sample is more efficient compared to the means of bootstrap, sufficient bootstrap and weighted bootstrap samples. The estimates of relative efficiency are sensitive to the degree of positive correlation between the study and auxiliary variables. The relative efficiency shows an increasing trend with the increase of positive correlation between the study and auxiliary variables. For example, when $\rho$ =0.65 and $n = 10$, the relative efficiency of mean of the sufficient weighted bootstrap sample compared to the means of bootstrap, sufficient bootstrap and weighted bootstrap samples are 131.88%, 105.11% and 125.27%, respectively. On the other hand, when $\rho$ =0.95 and $n = 10$, the relative efficiency of mean of the sufficient weighted bootstrap sample compared to the means of bootstrap, sufficient bootstrap and weighted bootstrap samples are 400.07%, 318% and 125.24%, respectively. The relative efficiency of sufficient weighted bootstrap sample mean compared to the weighted bootstrap sample mean is relatively more stable compared to the other versions of underlying bootstrap sample means for varying correlation.

## 4. Concluding Remarks

Theoretically, the mean of sufficient weighted bootstrap sample is unbiased estimate of the original sample mean of the study variable. For estimating mean of the study variable in the presence of the auxiliary variable with higher positive correlation, the sufficient weighted bootstrap mean provides higher relative efficiency compared to bootstrap, sufficient bootstrap and weighted bootstrap sample means. The simulation results suggest that the relative efficiency of mean of the sufficient weighted bootstrap sample is sensitive to the increase of positive correlation between the study and auxiliary variables. It also appears that the relative efficiency of sufficient weighted bootstrap sample mean compared to the weighted bootstrap mean is more stable with the increase of the population correlation between the study and auxiliary variables.

## References

[1] Singh, S. and Sedory, S.A. Sufficient bootstrapping. Computational Statistics and Data Analysis 55 (2011) 1629–1637
[2] Raj, D. and Khamis, S.H. Some Remarks on Sampling with Replacement. Ann. Math. Statist. Volume 29, Number 2 (1958), 550-557.
[3] Basu, D. On sampling with and without replacement. Sankhya, 20, 287-298, 1958.
[4] Singh, S. Saddlestrapping. Nonlinear Analysis: Modelling and Control, 2009, Vol. 14, No. 3, 357–388, 2009.
[5] B. Efron, Bootstrap methods: Another look at the Jackknife, The Annals of Statistics, 7(1), pp. 1–26, 1979.
[6] B. Efron, Better bootstrap confidence intervals (with discussion), J. Amer. Statist. Assoc. 82 (1987), pp. 171–200.
[7] B. Efron, Six questions raised by the bootstrap. In Exploring the Limits of Bootstrap, edited by R. Lepage and L. Billard. Wiley Series in Probability and Mathematical Statistics, NY, 1992.
[8] B. Efron, R. J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, 1993.
[9] M. R. Chernick, Bootstrap Methods. A Practitioner's Guide. Wiley, NY, 1999.
[10] G. Casella, Introduction to the silver anniversary of the Bootstrap, Stat. Sci., 18(2), pp. 133–134, 2003.
[11] B. Efron, Second thoughts on the bootstrap. Stat. Sci., 18(2), pp. 135–140, 2003.
[12] A. C. Davison, D. V. Hinkley, G. A. Young, Recent developments in bootstrap methodology, Stat. Sci., 18(2), pp. 141–157, 2003.
[13] R. Beran, The impact of the bootstrap on statistical algorithms and theory, Stat. Sci., 18(2), pp. 175–184, 2003.
[14] S. R. Lele, Impact of bootstrap on the estimating functions, Stat. Sci., 18(2), pp. 185–190, 2003.
[15] J. Shao, Impact of the bootstrap on sample, surveys, Stat. Sci., 18(2), pp. 191–198, 2003.
[16] P. Lahiri, On the impact of bootstrap in survey sampling and small-area estimation, Stat. Sci., 18(2), pp. 199–210, 2003.
[17] D. N. Politis, The impact of bootstrap methods on time series analysis, Stat. Sci., 18(2), pp. 219–230, 2003.
[18] M. D. Ernst, A. D. Hutson, Utilizing a quantile function approach to obtain exact bootstrap solutions, Stat. Sci., 18(2), pp. 231–240, 2003.
[19] M. Rueda, A. Arcos, E. Artes, Quantile interval estimation in a finite population using a multivariate ratio estimator, Metrika, 47, pp. 203–213, 1998.
[20] M. Rueda, Martinez-Miranda, A. Arcos, Bootstrap confidence intervals for finite population quantiles in the presence of auxiliary information, Model Assisted

Statistics and Applications, 1, pp. 279–290, 2005, 2006.
[21] S. Holmes, Bootstrapping phylogenetic trees: Theory and methods, Stat. Sci., 18(2), pp. 241–255, 2003.
[22] P. S. Soltis, D. E. Soltis, Applying the bootstrap in phylogeny reconstruction, Stat. Sci., 18(2), pp. 256–267. 2003.
[23] S. Holmes, C. Morris, R. Tibshirani, Bradley Efron: A Conversation with good friends, Stat. Sci., 18(2), pp. 268–281, 2003.
[24] D. B. Lahiri, A method for sample selection providing unbiased ratio estimates, Bull. Ins. Statist. Inst., 33(2), pp. 133–140, 1951.
[25] M. H. Hansen, V. N. Hurwitz, On the theory of sampling from finite populations, Ann. Math. Stat., 14, pp. 333–362, 1943.
[26] S. Singh, Advanced Sampling Theory with Applications: How Michael Selected Amy, Kluwer, The Netherlands, 2003.
[27] P. Barbe, P. Bertail, The Weighted Bootstrap, Lecture notes in Statistics, Vol. 98, SpringerVerlag, 1995.
[28] S. M. S. Lee, G. A. Young, Prepivoting by weighted bootstrap iteration, Biometrika, 90(2), pp. 393–410, 2003.
[29] R. W. Johnson, An introduction to the bootstrap. Teaching Statistics, Journal of the Royal Statistical Society D, 23(2), pp. 49–54, 2001.
[30] T. Hesterberg, Bootstrap methods and permutation tests, A course at the conference of Statisticians at San Antonio, Texas during March 2008, 2008.
[31] K. R. W. Brewer, M. Hanif, Sampling with Unequal Probabilities, Springer, 1983.
[32] D. Raj. Sampling Theory. MacGraw-Hill, New York, 1968.
[33] A. K. Adhikary. Improving the Hansen-Hurwitz estimator in PPSWR Sampling, Calcutta Statistical Association Bulletin, Vol. 61 Nos. 241-244, pp. 267-288, 2009.
[34] S. Singh, S. Horn, F. Yu, Estimation of variance of general regression estimator: Higher level calibration approach, Survey Methodology, 24(1), pp. 41–50, 1998.

## Appendix: Sample R Program used in Computation and Simulations

```
set.seed(12345);
R=50000;  #replications of samples from the original sample
sigmay=15 #sd of y
sigmax=10; #sd of x
muy=415; #mean of y
mux=335; #mean of x
rho=0.90; #correlation between y and x
n=c(10,15,20,30,50,100);
k=length(n);
tiff(file = "C:/Users/kislam/Desktop/A/gp90.tiff")
par(mfrow = c(3, 2))
for (j in 1:k){
ystar=rnorm(n[j]); #y* sample fron N(0,1)distribution
xstar=rnorm(n[j]); #x* sample fron N(0,1)distribution
y=muy+sqrt(sigmay^2*(1-rho^2))*ystar+rho*sigmay*xstar; # study variable y
x=mux+sigmax*xstar; # auxiliary variable x
corr=round(cor(x,y),digits=3); #correlation in generated sample sn
p=x/sum(x); #pi, prob proportioanl to sample size of auxiliary variable x
os=data.frame(y,p) #origianl sample with yi and pi values as a data.frame
mboot=c(); #empty storage for the means of bootstrap samples
mwboot=c(); #empty storage for the means of weighted bootstrap samples
```

```
msboot=c(); #empty storage for the means of sufficient bootstrap samples
mswboot=c();    #empty storage for the means of sufficient weighted bootstrap samples
#mboot, mwboot, msboot, mswboot will get updated in the following loop for each
values of R;
for (i in 1:R){
indx=sample(1:n[j], n[j], rep=T);  #indx for bootstrap sample
suf_indx=unique(indx); #indx for sufficient bootstrap sample
boot=os[indx,]   ; #bootstrap sample
mboot[i]=mean(boot$y); # mean of bootstrap sample
mwboot[i]=(1/n[j]^2)*sum(boot$y/boot$p);        #mean of weighted bootstrap sample
suf_boot=os[suf_indx,] #sufficient bootstrap sample
nu=nrow(suf_boot); #sufficient bootstrap sample size
msboot[i]=mean(suf_boot$y);   # mean of sufficient bootstrap sample
mswboot[i]=(1/(n[j]*nu))*sum(suf_boot$y/suf_boot$p); # mean of weighted sufficient
bootstrap sample
}
meanb=mean(mboot) # overall mean of means of bootstrap samples
sdb=sd(mboot) # SD of means of bootstrap samples
meansb=mean(msboot) # overall mean of means of sufficient bootstrap samples
sdsb=sd(msboot) # SD of means of sufficient bootstrap samples
meanwb=mean(mwboot) # overall mean of means of weighted bootstrap samples
sdwb=sd(mwboot) # SD of means of weighted bootstrap samples
meanswb=mean(mswboot) # overall mean of means of sufficient weighted bootstrap
samples
sdswb=sd(mswboot) # SD of means of sufficient weighted bootstrap samples
RE=c(rho, n[j],corr,sdb/sdswb*100, sdsb/sdswb*100, sdwb/sdswb*100)
# vector of relative efficiencies;
print(RE) # printing relative efficiency of underlying methods;
colors <- c("gold","red", "black", "blue")
labels <- colors
curve(dnorm(x, meanswb, sdswb), xlim=range(mboot), col="gold", lwd=2, xlab=
"Means", ylab= "Density",main=paste("r=", corr,"and", "n=",n[j]))
curve(dnorm(x, meanb, sdb), add=TRUE, col="red", lwd=1)
curve(dnorm(x, meansb, sdsb), add=TRUE, col="black", lwd=2)
curve(dnorm(x, meanwb, sdwb), add=TRUE, col="blue", lwd=2)
curve(dnorm(x, meanswb, sdswb), add=TRUE, col="gold", lwd=2)
legend("topright", legend=c("mswboot","mboot","mwboot","msboot"), lables,
fill=colors)
}
dev.off()
```