Journal of Statistical Research *ISSN 0256 - 422 X*
2006, Vol. 40, No. 2, pp. 1-11
Bangladesh

COMPARING RATIO ESTIMATORS BASED ON SYSTEMATIC SAMPLES

Hasan N. Hamdan

Mathematics and Statistics Department
James Madison University, Harrisonburg, VA, 22807, U.S.A

Email: hamdanhx@jmu.edu

John L. Szarka III

Department of Statistics
406-A Hutcheson Hall, Blacksburg, VA 24061-0439, U.S.A.

Email: szarkajl@vt.edu

Amy Marie Jacks

Department of Mathematics and Statistics
University of Missouri-Rolla, 1870 Miner Circle, Rolla, Missouri 65409, U.S.A.

Email: amj139@umr.edu

SUMMARY

The purpose of this study is to evaluate competitive ratio estimators, and introduce a new approach to estimation when a systematic sample of size n with a random start is used. The competitive ratio estimators are the mean of ratios, ratio of means, and the conditional best linear unbiased estimator. These estimators are used to measure the population proportion of the total of a variable Y with respect to another variable, X. A new approach is suggested, a bootstrap estimate using a non-linear additive regression technique, in which a Monte-Carlo simulation is done using the predicted values from the fitted model to find estimates for the variances. This new approach has yielded small mean square errors.

Keywords and phrases: Ratio estimator, variance estimator, bootstrapping

AMS Classification: 62D05, 62G99

1 Introduction

We consider a three-dimensional object where the interest is to estimate $R = \frac{V(Y)}{V(X)}$, the ratio of the volume of object Y inside object X. For example, in stereology this might be the ratio by volume of mitochondria in a liver cell or the proportion of a mineral in a sample of a rock. The classic approach in stereology is one in which a systematic sample of size n of equally-spaced parallel planes $\{P_1, P_2, ..., P_n\}$ is taken and are perpendicular to an a *priori* fixed axis. For every sample plane, P_i, x_i and y_i are observed where $x_i = A(X \bigcap P_i)$ and $y_i = A(Y \bigcap P_i)$. A systematic sample consists of n pairings $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$. Three possible estimators are considered; R_1 (the mean of ratios), R_2 (the ratio of means), and R_3 (the conditional best linear unbiased estimator).

Since there is no explicit formula for the variance of a ratio estimator based on a systematic sample, several current variance estimators are presented and compared using simulated objects. Let $A_y(t)$ represent the cross-sectional area of Y at a point t along the fixed axis, and $V(Y) = \int_\alpha^\beta A_y(t)dt$, where $[\alpha,\beta]$ are the endpoints. Similarly, $V(X) = \int_\alpha^\beta A_x(t)dt$. R can be written as the ratio of two integrals, and are approximated by Riemann sums.

In the following sections, ratio estimators, their biases, and their variance estimators are discussed. To compare estimators, simulations of artificial objects used to evaluate the ratios and variance estimators. In section two, the different ratio estimators and their variances are examined. In section three, another estimator is based on a bootstrapping technique is introduced. In section four, the artifical objects used in the study, Perfect, Imperfect, and Generic are described. Section five explains the use of Monte Carlo simulations. Finally, in section six, recommendations are given along with some comments and suggestions.

2 Ratio Estimators

2.1 R_1, The Mean of Ratios

A way to estimate the true ratio of $R = \frac{V(Y)}{V(X)}$ is given by $R_1 = \frac{1}{n}\sum_i^n \frac{y_i}{x_i}$, which is the sample mean of the ratios. This estimator proves to be biased, (Cochran, 1977) but the bias can be approximated using an application of a Taylor Series expansion. This sum, $V(Y)$ can approximate the following integrand f:

$$\int_\alpha^\beta f(x)dx = \Delta_N [\lim_{N\to\infty} \sum_{i=1}^N f(x_0 + (i-1)\Delta_N)] \tag{2.1}$$

$$\approx \Delta_N \sum_{i=1}^N f(x_0 + (i-1)\Delta_N) \tag{2.2}$$

for large N, where $x_0 \sim U(\alpha, \alpha + \Delta_N)$ and $\Delta_N = \frac{\beta-\alpha}{N}$.

Let $\bar{Y} = \Delta_N \sum_i^N A_Y(t_0 + (i-1)\Delta_N)$ and $\bar{X} = \Delta_N \sum_i^N A_X(t_0 + (i-1)\Delta_N)$. When N

is large, $R = \frac{\bar{Y}}{\bar{X}}$. Let $a_i = \frac{y_i}{x_i}$, then a bias-corrected estimate of R is found below:

$$\frac{1}{N}\sum_{i=1}^{N} a_i(x_i - \bar{X}') = \frac{1}{N}\sum_{i=1}^{N} y_i - (\frac{1}{N}\sum_{i=1}^{N} a_i)\bar{X}' \tag{2.3}$$

$$= \bar{Y}' - \bar{X}' E(a_i) \tag{2.4}$$

$$= \bar{X}'[R - E[R_1]], \tag{2.5}$$

where $\bar{Y}' = \frac{\bar{Y}}{d}$ and $\bar{X}' = \frac{\bar{X}}{d}$. R_1 corrected for the bias becomes

$$R_{1c} = \bar{a} + \frac{n(N-1)}{(n-1)N\bar{X}'}(\bar{y} - \bar{a}\bar{x}) \tag{2.6}$$

$$\approx \bar{a} + \frac{n}{(n-1)\bar{X}'}(\bar{y} - \bar{a}\bar{x}), \tag{2.7}$$

since $\frac{N-1}{N} \to 1$ as $N \to \infty$. This is called the Hartley and Ross estimator.

In general, the form of the variance for R_1 is as follows:

$$var(R_1) = \frac{1}{n^2}[\sum_{i=1}^{N} var(a_i) + 2\sum_{i=2}^{N}\sum_{j<i} cov(a_i, a_j)]. \tag{2.8}$$

To find the variance of R_1, note that it is the sample mean of a_i of the n ratios. Hence, its variance is $var(\bar{a}) = \frac{\sigma^2}{n}[1 + (n-1)\rho]$, where $\rho = \frac{cov(a_i,a_j)}{\sigma^2}$ is the correlation. Unfortunately, it is extremely difficult to obtain the true values of σ^2 and ρ since the sampling distribution of a_i is unknown, and the ratios are dependant.

However, several estimators of the variance exist (Wolter, 1985). The most basic example is to consider the systematic sample as a simple random sample. The variance estimator is given by the sample variance

$$v_1 R_1 = \frac{\sum_{i=1}^{n}(a_i - R_1)^2}{n(n-1)}. \tag{2.9}$$

This is a biased estimator. The bias is upward or downward depending on the interclass corrleation coefficient (Wolter, 1985).

Another estimator is based on non-overlapping differences, which is a modification of equation 7.2.4 in Wolter (1985, p. 251).

$$v_2 R_1 = \frac{1}{2n(n-1)}\sum_{j=2}^{n}(a_j - a_{j-1})^2. \tag{2.10}$$

An additional estimator is a modification of equation 7.2.9 in Wolter (1985, p. 252) using a sample estimate of a serial correlation to estimate ρ. The estimator is as follows:

$$v_3 R_1 = \begin{cases} v_1 R_1[1 + \frac{2}{ln(\widehat{\rho})} + \frac{2}{\widehat{\rho}^{-1} - 1}] & \widehat{\rho} > 0 \\ v_1 R_1 & \widehat{\rho} \leq 0, \end{cases} \tag{2.11}$$

where $\widehat{\rho} = \frac{1}{(n-1)s^2} \sum_{j=2}^{n} (a_j - R_1)(a_{j-1} - R_1)$.

The fourth and last estimator of the variance of R_1 is modified from an estimator that is based on a quadratic approximation of the covariogram derived by Gundersen and Jensen (1987). It is based on the assumption that the covariogram of the (x_i, y_i) values can be written as a quadratic function using Riemann sums to approximate the volumes. They extended this variance estimation of R_1 for a systematic sample. It is similar to Wolter's $v_3 R_1$ since it uses auto-regressive terms to account for potential autocorrelation of the a_i terms along the principal axis. Let us define the terms A, B, and C as

$$A = \sum_{i=1}^{n} a_i a_i, B = \sum_{i=1}^{n-1} a_i a_{i+1}, C = \sum_{i=1}^{n-2} a_i a_{i+2}.$$

The coefficient of variation of R_1 is defined as:

$$\widehat{CE}(R_1) = \frac{\sqrt{\frac{3A-4B+C}{12}}}{nR_1}. \tag{2.12}$$

The estimator for R_1 is given as follows:

$$v_4 R_1 = \widehat{CE}(R_1)^2 * {R_1}^2 \tag{2.13}$$

$$= \frac{3A - 4B + C}{12n^2}. \tag{2.14}$$

2.2 R_2, Ratio of Means

Another widely used estimator for the true ratio R is the ratio of means $R_2 = \frac{\sum y_i}{\sum x_i} = \frac{\bar{y}}{\bar{x}}$. Like R_1, R_2 is also a biased estimator for R_1, though it tends to be smaller than the bias of R_1. Cochran's (1977) estimate of the bias of R_2 is based on the leading term in the Taylor series expansion. $E(R_2 - R) \approx \frac{1}{n\bar{X'}^2}(R{S_x}^2 - \rho S_x S_y)$ where $\bar{X}' = \frac{\bar{X}}{d_2}, R_2 = \frac{\bar{y}}{\bar{x}} = \frac{\bar{y'}}{\bar{x'}}$, ρ represents the correlation coefficient $\frac{E[(X-\bar{X'})(Y-\bar{Y'})]}{\sqrt{{S_x}^2 {S_y}^2}}$, and ${S_x}^2$ and ${S_y}^2$ are the variance terms.

Our approximation of the bias of R_2 is:

$$biasR_2 = E[R_2 - R] \approx \frac{\bar{y}V(X) - \bar{x}V(Y)}{d_2 \bar{x}^2}, \tag{2.15}$$

where d_2 is the length of the principal axis along which the systematic samples are taken. Recall, $V(X)$ and $V(Y)$ refer to the volumes, not variances of X and Y. This equation holds true since $E[\bar{x}] = \frac{V(x)}{d_2}$ and $E[\bar{y}] = \frac{V(y)}{d_2}$.

The first variance estimator of R_2 is to use a variance based on a simple random sample. It is suggested by Cochran (1977) and Thompson (1992) as

$$v_1 R_2 = \frac{\sum (y_i - R_2 x_i)^2}{n(n-1)\bar{x}^2}. \tag{2.16}$$

Gundersen and Jensen (1982, 1987) obtained an estimate of the variance of R_2 based on a systematic sample that is similar to the one taken for R_1. Let

$$D = \sum_{i=1}^{n} x_i y_i, E = \frac{1}{2}\sum_{i=1}^{n-1}[x_i y_{i+1} + x_{i+1} y_i], F = \frac{1}{2}\sum_{i=1}^{n-2}[x_i y_{i+2} + x_{i+2} y_i].$$

They showed that $\widehat{cov}(\sum x, \sum y) = \frac{3D+F-4E}{12}$ and concluded

$$\widehat{CE}(R_2)^2 = CE^2(\sum x) + CE^2(\sum y) - \frac{2\widehat{cov}(\sum x, \sum y)}{\sum x \sum y}, \tag{2.17}$$

where $CE(\sum x)$ and $CE(\sum y)$ are found using D, E, and F for x_i and y_i. The second variance estimator of R_2 is

$$v_2 R_2 = (\widehat{CE}(R_2))^2 {R_2}^2. \tag{2.18}$$

2.3 R_3, The Conditional Best Linear Unbiased Estimator

R_3 is derived from the identity $Y = RX$. This identity, along with X being known, implies R and Y can be indirectly estimated by $\widehat{Y} = \widehat{R}X$. This equation states if x_i are y_i are directly proportional, $\widehat{R}$ is an accurate estimator of R. Baddeley and Jensen (2005) suggest a method for a conditional best linear unbiased estimator.

The following two conditions must be satisfied. First, for any given fixed X_i, the Y_i's are conditionally uncorrelated. Second, $E[Y_i|X_i = x_i] = Rx_i$ and $var[Y_i|X_i = x_i] = ax_i^b, i = 1, ..., n$. Constants a and b are estimated from the data. Under these assumptions, $\frac{E[Y_i]}{E[X_i]} = E[\frac{Y_i}{X_i}] = R$, making R_3 an unbiased estimator. In this case, the conditional best linear unbiased estimator is

$$R_3 = \frac{\sum X_i^{1-b} Y_i}{\sum X_i^{2-b}} = \frac{\sum X_i^{2-b} R_i}{\sum X_i^{2-b}}. \tag{2.19}$$

Note that R_1 and R_2 are special cases of R_3. When $b = 2$, $R_3 = R_1$, and when $b = 1$, $R_3 = R_2$.

A conditional variance of R_3 can be found by Baddeley and Jensen (2005) and is given by

$$var[R_3|X_i, i \in n] = \frac{a}{\sum X_i^{2-b}}. \tag{2.20}$$

To estimate a and b, we used the new simple approach: The data are divided into disjoint subsets or probes using the closest k values. For each subest, the sample mean $\bar{x}_i$ and the corresponding sample variance of the y_i's are found. In this case, the data was divided into five probes and new pairs $(\bar{x}_i, {s_{y_i}}^2), i = 1, ..., 5$ were created. Then, $log({s_{y_i}}^2)$ is regressed against $\log(\bar{x}_i)$ and since $s^2 \approxeq ax^b$, the slope of this regression will be an estimate of b and the intercept will be an estimate of $log(a)$.

3 Bootstrap Estimator

The variance estimators used for R_1 and R_2 depend on an advanced covariance structure assumption. We try to find a way to provide another estimator that does not have any covariance assumptions. One possibility is a non-parametric approach which provides variance estimators for R_1, R_2, and R_3.

A model can be formed based on functions of distance of locations for x_i and y_i along the principal axis using non-linear additive regression techniques. In this case, the LOESS function in the statistical software package R, is used to obtain $\widehat{Y} = \widehat{f}(d)$ and $\widehat{X} = \widehat{g}(d)$, where d represents the distance along the object. The functions $\widehat{f}$ and $\widehat{g}$ produce simulations of all three objects.

The following steps are used to find bootstrapped estimators of the true ratio and the corresponding variance estimators based on R_1, R_2, and R_3. First, identify a principal axis along which the systematic sample is taken, and label one end of this axis as the origin. Second, take a single random starting systematic sample of size n, and observe $\{(x_i, y_i), i = 1, ..., n\}$ where x_i and y_i represent the areas of X and Y at the i^th location along the axis. Third, convert i to the distance from the origin. A non-parametric smoothed fit of x_i against d_i and y_i against d_i are obtained. Using the LOESS function with a small span, a locally-quadratic fitting of the curve is applied. Lastly, our two curves $\widehat{Y} = \widehat{f}(d)$ and $\widehat{X} = \widehat{g}(d)$ represent the estimates of cross-sectional areas of Y and X respectively. Then, take a repeated single random start systematic samples of size n and observe $\{(\widehat{x_{ib}}, \widehat{y_{ib}}), i = 1, ..., n, b = 1, ..., B\}$.

$$\bar{R}_i = mb_i = \frac{\sum_{b=1}^{B} R_{ib}}{B}, \tag{3.1}$$

and

$$vb_i = \frac{\sum_{b=1}^{B} (R_{ib} - \bar{R}_i)^2}{B-1}, i = 1, 2, 3. \tag{3.2}$$

The bootstrap ratio and variance estimates are based on a sample mean and sample variance.

4 Simulation Study

4.1 Three Study Objects

Simulations were performed on three different artificial objects, referred to as Perfect, Imperfect, and Generic. The Perfect object is an ellipsoid embedded inside a hyperboloid. The Imperfect object was created as a variation to the Perfect object by taking a large systematic sample of n_1=300 from the Perfect object and adding a random noise to each x_i and y_i value. The Generic object was created by taking random x_i values and multiplying them by random uniform values between 0 and 1 to create the y_i values.

4.2 A Perfect Object

This object gives a simple example with known exact ratios. If R_1, R_2, and R_3 do not perform well for a simple object, they can not be expected to work in more complex cases. The formulas for the ellipsoid and hyperboloid are:

$$\frac{x^2}{{a_1}^2} + \frac{y^2}{{b_1}^2} + \frac{z^2}{{d_1}^2} = 1 \tag{4.1}$$

$$\frac{x^2}{{a_2}^2} + \frac{y^2}{{b_2}^2} - \frac{z^2}{{d_2}^2} = 1. \tag{4.2}$$

The proportion of the ellipsoid inside the hyperboloid can be set by manipulating the values of a, b, and d. The limits are set so that $a_1 \le a_2$, $b_1 \le b_2$, $d_1 \le d_2$, and $\min\{d_1, d_2\} \ge \max\{a_2, b_2\}$. This way, the ellipsoid is guaranteed to be completely embedded inside the hyperboloid. The last condition ensures that the major axis of this object is the z-axis. Finally, $a \ge b$ so that the major axis of elliptical cross-section at any height is a line parallel to the x-axis.

If the sample size is given by n, then the height for each section is defined as $h = \frac{2d_2}{n}$. Let P_i represent a plane at height z_i along the principal axis. Let $t \sim U(-d_2, -d_2 + h)$, and $z_i = t + (i-1)h$, $i = 1, ..., n$. Also define A_i and B_i to be the cross-sectional areas at height z_i. In order to determine A_i and B_i, fix the height at z_i and integrate over the x and y coordinate axes. For A_i, the equation is solved for y as a function of x to obtain the following:

$$-\sqrt{{b_2}^2[1 + \frac{{z_i}^2}{{d_2}^2} - \frac{x^2}{{a_2}^2}]} \le y \le \sqrt{{b_2}^2[1 + \frac{{z_i}^2}{{d_2}^2} - \frac{x^2}{{a_2}^2}]}$$

The range of x is $-\sqrt{{a_2}^2(1 + \frac{{z_i}^2}{{d_2}^2})} \le x \le \sqrt{{a_2}^2(1 + \frac{{z_i}^2}{{d_2}^2})}$. Let $q = 1 + \frac{{z_i}^2}{{d_2}^2}$. The area of the hyperboloid cross-section at height z_i is given by

$$A_i = 4\int_0^{\sqrt{q{a_2}^2}} \sqrt{{b_2}^2(q - \frac{x^2}{{a_2}^2})}dx \tag{4.3}$$

$$= \frac{4b_2}{a_2}\int_0^{\sqrt{q{a_2}^2}} \sqrt{q{a_2}^2 - x^2}dx. \tag{4.4}$$

By letting $x = \sqrt{q{a_2}^2}sin\theta$, $A_i = a_2 b_2(1 + \frac{{z_i}^2}{{d_2}^2})\pi$. Using a similar formula, $B_i = a_1 b_1(1 - \frac{{z_i}^2}{{d_1}^2})\pi$. The volume of the ellipsoid, $V(Y)$, is $\frac{4a_1 b_1 d_1 \pi}{3}$ and the volume of the hyperboloid, $V(X)$ is $\frac{8a_2 b_2 d_2 \pi}{3}$. Therefore, the true ratio is given by $R = \frac{V(Y)}{V(X)} = \frac{a_1 b_1 d_1}{2a_2 b_2 d_2}$. The case explored in this study is when $a_1 = a_2, b_1 = b_2, d_1 = d_2$, where the true ratio is R=.5.

4.3 An Imperfect Object

In practice, there are few perfect shapes, so we took the perfect object and added noise to make it imperfect and called it the Imperfect object. A large number n_1=300 of pairs

(x_i, y_i) was sampled systematically from the Perfect object. Then two random numbers were generated from a normal distribution, $N(0, 1.5)$. The smaller of the two numbers was added to the Y value, and the larger added to the X value so that $0 \leq Y \leq X$ always. Then two estimated regression functions were created, $\widehat{f}(d)$ and $\widehat{g}(d)$, based on fitting a LOESS curve in R to the 300 values. Then, a systematic sample of size n is taken from this population using the *predict* function in R, based on these curves. The exact value of R is the ratio of the area under the curve $\widehat{g}(d)$ to the area under the curve $\widehat{f}(d)$.

4.4 A Generic Object

This is an arbitrary object with two randomly generated parts, the inside (Y = phase of interest) and the outside (X = the reference phase). A large sample of n_1=300 where each x_i was a realization from $N(1 + (\frac{z_i}{d_2})^2, {\sigma_i}^2)$, where $d_2 = 2$, $t \sim U(-2, -2 + \frac{d_2}{n_1})$ and $z_i = t + (i-1)\frac{d_2}{n_1}$ and ${\sigma_i}^2 = |\frac{z_i}{2}|$. The y_i values are obtained as a scalar multiple of x_i as $y_i = \{y_i = u_i x_i, i = 1, 2, ..., n_1\}$ where every $u_i \sim U(0, 1)$. The procedure for the Imperfect object is replicated for the Generic object.

5 Monte Carlo Study

The work actually involoved running three simulations for each object. The first was done to obtain an approximation of the true variances in all cases and for each of the three objects. In the same simulation, an approximations of the true ratio was also found for each of the Imperefct and Generic objects (Note that the exact ratio is only known in the Perfect Object case, and it has a value of .5). These approximations were found based on 300 Monte Carlo simulations of sampling from the object (i.e. the approximations are based on estimating the sampling distribution). The second was done to bootstrap the estimates of the ratios and variances based on a single start from systematic samples of various sizes; and finally, the process was repeated with a fixed sample size, to see how the bootstrap behaves as a function of the number of replications B. In order to obtain the bootstrap estimates, sample sizes of $\{n = 10, 15, 20, 30, 40, 50, 75, 100\}$ with $B = 100$ replications were chosen, this particular B was chosen because the estimators appear to be stable after such number of replications. To check this, bootsrap replications $\{B = 20, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000\}$ were run for $n = 20$. Seventy five Monte Carlo simualtions were done for each bootstrap combiantion. The simulations were all done using the statistical package R.

6 Results and Discussion

The purpose of this study is to compare and evaluate ratio estimators based on systematic samples. This is done by looking at the mean square error of R_1, R_2, and R_3. Approximations to the biases of R_1 and R_2 are estimated, and it is known that R_3 is unbiased. The true distributions of these estimators is unknown for all n because the samples are systematic (This is the case because not only do the x_i's and y_i's covary, but depend on one another).

In Table 1, R_2 and R_3 are presented along with the corresponding bootstrap estimates and mean square errors of each. Only samples of size 10, 20, 50, and 100 are presented for space considerations. R_1 is not included in Table 1 because the bias was too large (around .07). Even when corrected for the bias, R_{1c} was still larger than R_2 and R_3 in the Perfect object case.

In total, there are four variance estimators for R_1, two for R_2, and one for R_3. In each case, the variance estimator with smallest values was chosen. Then, the mean square error was estimated based on this variance estimate. For the Perfect object, Gundersen and Jensen variance estimators given by (2.14) or v_4R_1, and (2.18), which is v_2R_2 produced the smallest estimates.

These results are found to be the same when the Imperfect object is used. For this reason, the table for the Imperfect object is not included.

Table 1: Ratio and Mean Square Error Estimates for the Perfect Object (R =.5)

n	R_2	R_3	mb_2	mb_3	MSE_{R2}	MSE_{R3}	MSE_{mb2}	MSE_{mb3}
10	0.50031	0.50031	0.50012	0.50012	1.0135e-03	3.9797e-03	1.1753e-05	1.1008e-05
20	0.50026	0.50026	0.50001	0.50119	2.3883e-04	3.3060e-04	7.3939e-07	9.9323e-06
50	0.50002	0.50330	0.50000	0.50366	3.7633e-05	3.0710e-05	1.8826e-08	4.1455e-06
100	0.50000	0.50432	0.50000	0.50431	9.3827e-06	6.4160e-06	1.2128e-09	1.4088e-06

Table 2: Ratio and Mean Square Error Estimates for the Generic Object

n	R_1	R_2	R_3	mb_1	mb_2	mb_3	MSE_{R1}	MSE_{R2}	MSE_{R3}	MSE_{mb1}	MSE_{mb2}	MSE_{mb3}
10	0.49961	0.49421	0.49468	0.49956	0.49481	0.49483	4.431e-05	8.965e-06	2.779e-04	3.177e-05	9.017e-06	1.088e-05
20	0.49959	0.49410	0.49466	0.50026	0.49525	0.49543	3.458e-05	2.761e-06	5.966e-05	2.876e-05	4.649e-06	2.669e-06
50	0.49979	0.49434	0.49493	0.50033	0.49515	0.49549	3.108e-05	1.740e-06	8.380e-06	2.814e-05	2.963e-06	4.640e-07
100	0.49978	0.49432	0.49493	0.50048	0.49524	0.49565	3.044e-05	1.545e-06	2.003e-06	2.807e-05	2.895e-06	1.223e-07

In the Generic object case, the results are different as shown in Table 2. R_2 has the smallest mean square error. The variance estimator used for R_2 is based on v_2R_2 (which is the case in the Perfect and Imperfect object as well). However, v_3R_1 (11) outperforms v_4R_1, in this case.

Finally, bootstrapping seems to always improve the estimators, as shown in both tables $(mb_1, mb_2, mb_3, MSE_{mb1}, MSE_{mb2}, MSE_{mb3})$. For the Perfect and Imperfect objects, the bootstrap estimate mb_2 seems to be the best overall estimator. However, for the Generic case, even though R_2 is the best estimator, mb_3, the bootstrap estimator for R_3 is the best overall estimator. The bootstrap estimators of R_1 and R_2 seems to stabalize sooner than that of R_3 (i.e 75 replications needed to stablize R_1 and R_2, while 100 replications were needed to stablize the bootstrap of R_3).

Finally and for further research, these estimators should be compared and evaluated using more complicated and more realistic objects with different ratio sizes. For example,it would be interesting to know if these results will hold when small or large ratios are used or when different kinds of objects are considered (objects with wholes, objects that are not connected or have many parts and so on).

Aknowledgments

The authors would like to thank the National Science Foundation for their support through the REU grant 526043 and the authors also would like to thank the Department of Mathematics and Statistics at JMU for their support as well.

References

[1] Baddeley, Adrian. (1991). Stereology. In *Spatial Statistics and Digital Image Analyses*, National Academy Press, Washington D.C.

[2] Baddeley, Adrian and Jensen, E.B. (2005). *Stereology for Statisticians*, Chapman & Hall, Boca Raton.

[3] Cochran, William G. (1977). *Sampling Techniques*, John Wiley & Sons, Inc., New York.

[4] Cruz-Orive, Luis-M. and Weible, Ewald R. (1980). Sampling Designs for Stereology. *Journal of Microscopy,* **122***, pp. 240-241.*

[5] Gundersen, H.J.G. and Jensen, E.B. (1987). The Efficiency of Systematic Sampling and its Prediction. *Journal of Microscopy,* **147** *pp. 248-251.*

[6] Thompson, Steven K. (1992). *Sampling*, John Wiley & Sons, Inc., New York.

[7] Wolter, Kirk M. (1985). *Introduction to Variance Estimation*, Springer-Verlag, New York.