Comparison between ratio estimation using sampling with and without replacement via the MSE criterion

Taejoo Chae

Oklahoma State University

December 9, 2013

Overview

- Sampling Methods
- Other Methods of Ratio Estimate
- Simulations
- Results

Sampling Methods

- Simple Random Sampling With Replacement(SRSWR)
- Simple Random Sampling Without Replacement(SRSWOR)
- Simple Random Sampling With Replacement Based Only on distinct Units(SRSWR(D))

Sample data: $y_1, ..., y_n$

Auxiliary sample data: $x_1, ..., x_n$

We estimate $\bar{Y}$ by $\bar{y}_n$ and $\bar{X}$ by $\bar{x}_n$ where $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Hence,

$$E(\bar{y}_n) = \bar{Y},\ V(\bar{y}_n) = (\tfrac{N+n-1}{Nn} - \tfrac{1}{N})S_Y^2 = (\tfrac{N-1}{Nn})S_Y^2$$

and

$$E(\bar{x}_n) = \bar{X},\ V(\bar{x}_n) = (\tfrac{N+n-1}{Nn} - \tfrac{1}{N})S_X^2 = (\tfrac{N-1}{Nn})S_X^2,$$

where

$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$

and

$S_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$

Thus $R = \frac{\bar{Y}}{\bar{X}}$ may be estimated by $\hat{r} = \frac{\bar{y}_n}{\bar{x}_n}$

Sampling Methods - SRSWOR I

Sample data: $y_1, ..., y_n$

Auxiliary sample data: $x_1, ..., x_n$

Hence,

$E(\bar{y}_n) = \bar{Y},\ V(\bar{y}_n) = (\frac{N-n}{Nn})S_Y^2$

and

$E(\bar{x}_n) = \bar{X},\ V(\bar{x}_n) = (\frac{N-n}{Nn})S_X^2,$

Sampling Methods - SRSWOR II

Thus, $R = \frac{\bar{Y}}{\bar{X}}$ can be estimated by $\hat{r} = \frac{\bar{y}_n}{\bar{x}_n}$

Hence, approximation of MSE for $\hat{r}$ would be:

$$MSE(\hat{r}) \simeq R^2(\frac{N-n}{N-1})\frac{(S_X^2 - 2S_{XY} + S_Y^2)}{n},$$

where

$S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$

Sampling Methods - SRSWR(D) I

Sample data: $y_1, ..., y_n$

Auxiliary sample data: $x_1, ..., x_n$

From these samples, duplicate sample units are deleted. Sample of distinct responses would be $y_1^*, ..., y_{v_1}^*$ and corresponding auxiliary sample would be $x_1^*, ..., x_{v_2}^*$. Then,

$\bar{y}_{v_1} = \frac{1}{v_1} \sum_{i=1}^{v_1} y_i^*$, where $y_1^* \neq y_2^* \neq ... \neq y_{v_1}^*$

and

$\bar{x}_{v_2} = \frac{1}{v_2} \sum_{i=1}^{v_2} x_i^*$, where $x_1^* \neq x_2^* \neq ... \neq x_{v_2}^*$.

Note that $v_1 \leq n$, and $v_2 \leq n$, and v_1 and v_2 are integer-valued random variables.

Sampling Methods - SRSWR(D) II

$E(\bar{y}_{v_1}) = \bar{Y}$, $V(\bar{y}_{v_1}) = (E(\frac{1}{v_1} - \frac{1}{N}))S_Y^2$

and

$E(\bar{x}_{v_2}) = \bar{X}$, $V(\bar{x}_{v_2}) = (E(\frac{1}{v_2} - \frac{1}{N}))S_X^2$,

where

$E(\frac{1}{v_1}) = E(\frac{1}{v_2}) = \frac{1}{N^n} \sum_{i=1}^{N} i^{(n-1)}$

Thus, $R = \frac{\bar{Y}}{\bar{X}}$ can be estimated by $\hat{r}(v_1, v_2) = \frac{\bar{y}_{v_1}}{\bar{x}_{v_2}}$.

Hence, the MSE of $\hat{r}(v_1, v_2)$ is approximated by:

$$E(\hat{r}(v_1, v_2) - R)^2 \simeq V(\bar{y}_{v_1}) + \frac{V(\bar{x}_{v_2})}{R} + (\frac{\bar{X}}{R})^2 V(\bar{y}_{v_1}) V(\bar{x}_{v_2})$$

For ease of presentation, set

$$\delta\bar{y}_{v_1} = \frac{\bar{y}_{v_1} - \bar{Y}}{\bar{Y}},$$

and

$$\delta\bar{x}_{v_2} = \frac{\bar{x}_{v_2} - \bar{X}}{\bar{X}}$$

and

$$\bar{y}_{\hat{r}(v_1,v_2)} = \bar{X}\frac{\bar{y}_{v_1}}{\bar{x}_{v_2}}$$

Sampling Methods - SRSWR(D) IV

Then,

$$
\begin{aligned}
(\bar{y}_{\hat{r}(v_1,v_2)} - \bar{Y})^2 \quad & \simeq \bar{Y}^2[(\delta\bar{y}_{v_1} - \delta\bar{x}_{v_2}) + (\delta\bar{x}_{v_2})^2(\delta\bar{y}_{v_1} - \delta\bar{x}_{v_2})^2] \\
& = \bar{Y}^2[(\delta\bar{y}_{v_1})^2 + (\delta\bar{x}_{v_2})^2 + (\delta\bar{y}_{v_1})^2(\delta\bar{x}_{v_2})^2]
\end{aligned}
$$

Hence,

$$
(\hat{r}(v_1, v_2) - R)^2 \simeq (\delta\bar{y}_{v_1})^2 + (\delta\bar{x}_{v_2})^2 + (\delta\bar{y}_{v_1})^2(\delta\bar{x}_{v_2})^2
$$

Taking expectations we get that

$$
E((\hat{r}(v_1, v_2) - R)^2) \simeq (E(\frac{1}{v}) - \frac{1}{N})(\frac{S_Y^2}{\bar{Y}^2} + \frac{S_X^2}{\bar{X}^2} + \frac{S_Y^2 S_X^2}{\bar{Y}^2\bar{X}^2})
$$

Quenouille's ratio estimator (Jack-knife):

$$t_Q = 2\hat{r} - \frac{1}{2}(\hat{r}_1 + \hat{r}_2)$$

where $\hat{r} = \frac{\bar{y}}{\bar{x}}$ or $\hat{r}(v_1.v_2) = \frac{\bar{y}_{v1}}{\bar{x}_{v2}}$, and $\hat{r}_1$ and $\hat{r}_2$ would be ratios from two random halves of sample data.

Note that we can instead define

$\hat{r}_{-i} = \hat{r}$ based on all y's except y_i and all x's except x_i.

and define

$$t_Q = n\hat{r} - \frac{1}{n}\sum_{i=1}^{n} \hat{r}_{-i}$$

Other Ratio Estimate Methods - Chakrabarty Estimate

Chakrabarty's ratio estimators:

$$\bar{y}_{c1} = (1 - W)\bar{y} + W\bar{y}_{\hat{r}} = (1 - W)\bar{y} + W\frac{\bar{y}}{\bar{x}}\bar{X} = \bar{y}((1 - W) + W\frac{\bar{X}}{\bar{x}})$$

$$\bar{y}_{c2} = (1 - W)\bar{y} + Wt_Q\bar{X} = (1 - W)\bar{y} + W[2r - \frac{1}{2}(r_1 + r_2)]\bar{X}$$

where $0 \leq W \leq 1$ and $\hat{r}_{c1} = \frac{\bar{y}_{c1}}{\bar{x}_{c1}}$ and $\hat{r}_{c2} = \frac{\bar{y}_{c2}}{\bar{x}_{c2}}$.

Simulation I

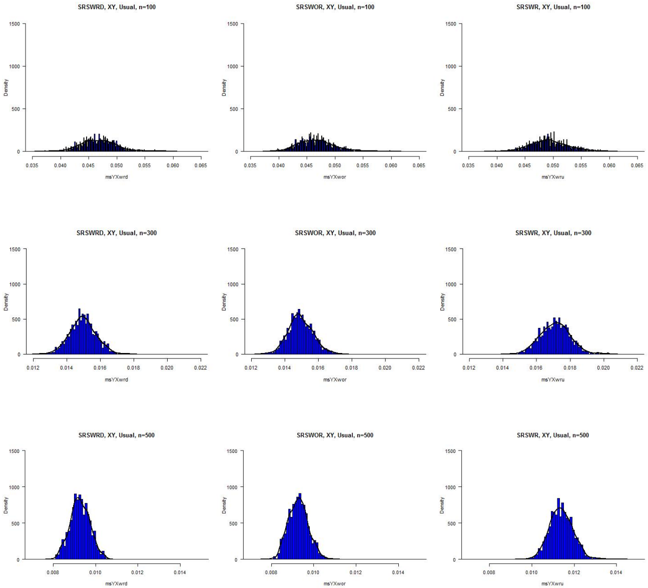
1. Generate data of population size N from each of the following distributions:
   1. Normal(X) population
   2. LogNormal(Y) population
   3. Gamma(Z) population
2. Select a pair of the above populations
3. Draw SRSWR of size n from a selected populations.
4. Delete duplicates from this SRSWR to get SRSWR(D), and calculate ratio estimate of any two of the above 3 populations using these SRSWR(D).
5. Sample size changes: $n \to v_1$ and v_2.

Simulation II

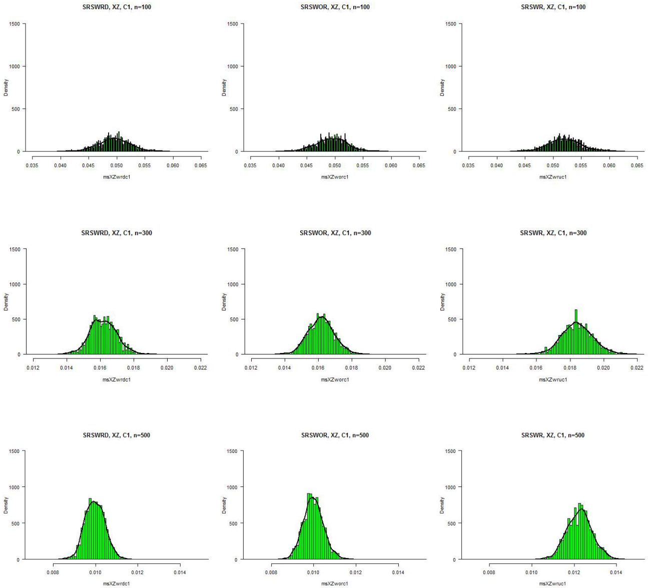
6. Select SRSWR of size v_1 and v_2 and a SRSWOR of size v_1 and v_2.
7. Repeat the above steps, a number of replications(reps).
8. Calculate the MSE, $s_j^2 = \frac{1}{reps-1}\sum_{i=1}^{reps}(\hat{r}_i - \bar{r})^2$, where $\bar{r} = \frac{1}{reps}\sum_{i=1}^{reps}\hat{r}_i$, where j is repitition number j.
9. Repeat the above process for the remaining two other pair choices.

N = 2000, n = 100, 300, 500				
R \ Pop	XY	XZ	YZ	
Usual	$\odot$			(1)
Quenouille			$\odot$	(4)
C(1)		$\odot$		(2)
C(2)		$\odot$		(3)

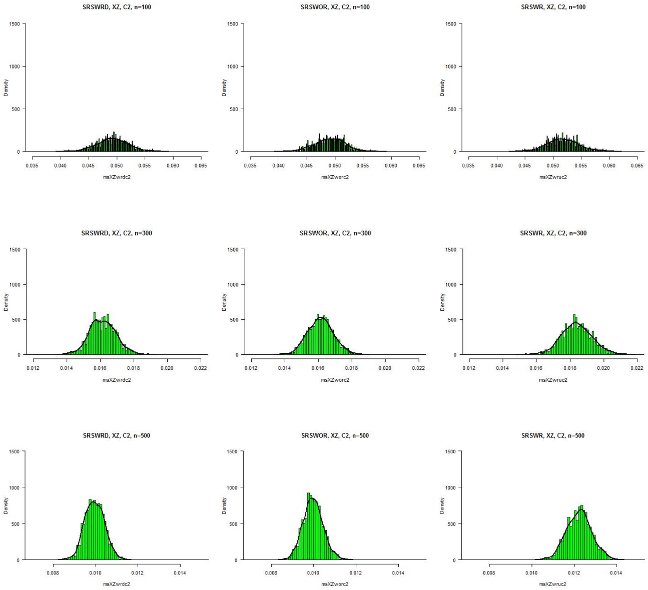
Result - (1)


SRSWRD, XY, Usual, n=100
SRSWOR, XY, Usual, n=100
SRSWR, XY, Usual, n=100
SRSWRD, XY, Usual, n=300
SRSWOR, XY, Usual, n=300
SRSWR, XY, Usual, n=300
SRSWRD, XY, Usual, n=500
SRSWOR, XY, Usual, n=500
SRSWR, XY, Usual, n=500
Density
msYXwrd
msYXwor
msYXwru

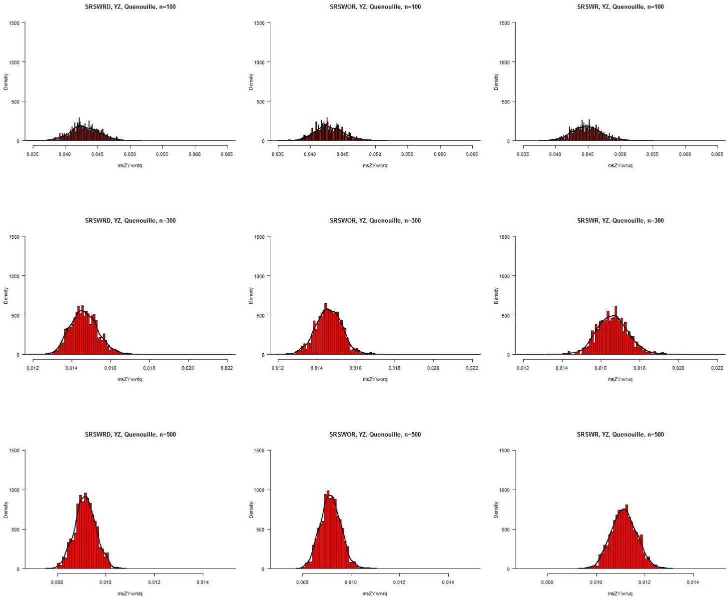
Result - (2)


SRSWRD, XZ, C1, n=100
SRSWOR, XZ, C1, n=100
SRSWR, XZ, C1, n=100
msXZwrdc1
msXZworc1
msXZwruc1
SRSWRD, XZ, C1, n=300
SRSWOR, XZ, C1, n=300
SRSWR, XZ, C1, n=300
SRSWRD, XZ, C1, n=500
SRSWOR, XZ, C1, n=500
SRSWR, XZ, C1, n=500
Density

Result - (3)


SRSWRD, XZ, C2, n=100
SRSWOR, XZ, C2, n=100
SRSWR, XZ, C2, n=100
SRSWRD, XZ, C2, n=300
SRSWOR, XZ, C2, n=300
SRSWR, XZ, C2, n=300
SRSWRD, XZ, C2, n=500
SRSWOR, XZ, C2, n=500
SRSWR, XZ, C2, n=500
Density
msXZwrdc2
msXZworc2
msXZwruc2

Result - (4)


SRSWRD, YZ, Quenouille, n=100
SRSWOR, YZ, Quenouille, n=100
SRSWR, YZ, Quenouille, n=100
SRSWRD, YZ, Quenouille, n=300
SRSWOR, YZ, Quenouille, n=300
SRSWR, YZ, Quenouille, n=300
SRSWRD, YZ, Quenouille, n=500
SRSWOR, YZ, Quenouille, n=500
SRSWR, YZ, Quenouille, n=500
Density
msZYwrdq
msZYworq
msZYwruq

Conclusion

•SRSWR(D) is outstanding method that deserves further investigation.

•SRSWOR sometimes is better and most of the time is second. Thus a good competitor to SRSWR(D).

•SRSWR falls third uniformly.

References I

- Asok, C., A Note on the comparison between simple mean and mean based on distinct units in sampling with replacement, *The American Statistician* **34** 3 (1980) 158.
- Basu, D., On sampling with and without replacement, *Sankhya: The Indian Journal of Statistics*, **20** 3 (1958), 287-294.
- Cochran, W. G., Sampling techniques 3^{rd} ed., New York, NY: Wiley and Sons, 1977

References II

- Chaudhuri, A., Stenger H., Survey sampling: Theory and methods., 2^{nd} ed., New York, NY: Chapman and Hoff, 2005
- Fuller, W.A., Sampling statistics., New York, NY: Wiley and Sons, 2009
- Korwar, R. M. and Serfling, R. J., On averaging over distinct units in sampling with replacement, *The Annals of Mathematical Statistics*, **41** 6 (1970), 2132-2134.
- Mussa, A. S., A new bias-reducing modification of the finite population ratio estimator and a comparison among proposed alternatives, *Journal of Official Statistics*, **15** 1 (1999), 25-38.

References III

Pathak, P. K., On the evaluation of moments of distinct units in a sample, *Sankhya: The Indian Journal of Statistics*, **23** 4 (1961) 415-420.

Raj, D., Khamis, S. H., Some remarks on sampling with replacement, *Ann. Math. Statist.*, **29** 2 (1958) 550-557.

Sukhatme, P. V., Sukhatme, B. V., Sampling theory of surveys applications., Ames, IA: Iowa State University Press, 198

THANK YOU