Use of jack – Knife technique and unbiased class of product type estimators in circular systematic sampling (C.S.S.) scheme

Chetan Patel, Vinit Jaiswal, RA Sharma and KS Kushwaha

DOI: https://doi.org/10.22271/chemi.2020.v8.i2b.9588

Abstract

A general class of product type estimators in C.S.S. scheme is proposed to estimate the population mean of a response variable y. Jack – knife technique pioneered by Quenouille (1949, 1956) is applied to make the class unbiased. The explicit expression for sampling variance of the class Tpu is derived to first order (n – 1) of approximation. Minimum Variance Unbiased Estimator (M.V.U.E.) in the class Tpu is investigated. An empirical illustration is provided to examine the applicability of the results derived.

Keywords: Product estimator, Jack–knife technique, Circular System sampling, MVU estimator.

Introduction

The classical product estimator under linear systematic sampling scheme was proposed by Shukla (1971) [6] and its properties were studied. In general, the product estimator is biased. A weighted class of product type estimator was proposed and made exactly unbiased by Kushwaha and Singh (1989) [2] using Jack – knife technique.

A serious demerit of linear systematic sampling scheme is that it is not estimate the sampling variance of the estimator under study but it can be estimated unbiasedly with the use of interpenetrating systematic sampling technique with independent random start.

A sampling modification of linear systematic (L.S.S.) makes it possible to ensure a fixed sample size n and to make the sample mean unbiased for population mean, even if in class if N/Kn. This sampling scheme is known as ‘circular systematic sampling’ (CSS) scheme. Murthy (1977) [3] and Sukhatame (1970) [8] have suggested to use CSS in situation when N = Kn, where K is positive integer.

The main steps involved in selecting a sample using CSS are as follow:

1. Select the random number ‘r’ from 1 to N and name it as ‘random start’.
2. Choose some integer value of K = N/n or take integer nearest to N/n and name it as skip or sampling span.
3. Select all units in the sample with serial numbers given on.

\[ r + jk \text{ if } (r+jk) < N, \ j = 0,1,2,3, \ldots \ldots \ldots \ldots \ldots \ldots (n-1), \ 1 < r < N \]

\[ r + jk \text{ if } (r+jk) > N, \ j = 0,1,2,3, \ldots \ldots \ldots \ldots \ldots \ldots (n-1), \ 1 < r < N \]

Sudhakar (1978) pointed out that the use of skip or span of sampling as an integer nearest to N/n in C.S.S. does not draw a sample of desired size. Sudhakar (1978) has also mentioned that if we take the span of sampling as nearest integer < N/n, we do not encounter the above cited difficulty although this depends upon n.

Jack–knife technique has been successfully employed in several estimation and testing problems. Here we have utilized Jake–knife technique to get rid of bias in usual product type estimators.

The expression for the sampling variance of the proposed class of estimators is derived. An empirical illustration is provided to examine the performance of the derived estimators with respect to efficiency point or views over other estimators existing in literature.
The circular systematic sample means (y, x) are defined as:
\[
\bar{y} = \frac{1}{n} \sum_{j=0}^{n-1} y_{r+jk} \quad j = 0, 1, \ldots, (n-1)
\]
\[
\bar{x} = \frac{1}{n} \sum_{j=0}^{n-1} x_{r+jk} \quad k = \text{sampling span}
\]

Sample means (\(\bar{y}, \bar{x}\)) are respectively unbiased estimators for the population means (\(\bar{Y}, \bar{X}\)). The X of covariate X is assumed to be known in prior.

The usual product estimator \(\bar{Y}_p\) for \(\bar{Y}\) based on a circular systematic sample of size n is defined as:
\[
\bar{Y}_p = \frac{\bar{y}}{\bar{x}} \bar{X}
\]

To reduce the bias of \(\bar{Y}_p\), we take \(n = gm\) and split the sample into g sub samples each of size \(m = n/g\) in a systematic manner as this avoids the need for selecting the samples in the form of sub sample of smaller size m, thereby retaining the efficiency generally obtained by taking a large circular systematic sample of size n.

Let \((\bar{y}_t, \bar{x}_t, t = 1, 2, \ldots, g)\) be unbiased estimators for \((\bar{y}, \bar{x})\) based on circular systematic sub – sample each of size m units. Let us define complimentary sub – sample means \((\bar{y}'_t, \bar{x}'_t, t = 1, 2, \ldots, g)\) as:
\[
\bar{y}'_t = \frac{(n\bar{y} - m\bar{y}_t)}{m}
\]
\[
\bar{x}'_t = \frac{(n\bar{x} - m\bar{x}_t)}{m}
\]

Which are the sample means based on subsample, each of size \(m = (n-m)\) units obtaining by omitting \(i^{th}\) \((t=1, 2, \ldots, g)\) subsample from the circular systematic sample of size n. The estimators \((\bar{y}'_t, \bar{x}'_t)\) are also unbiased for \((\bar{Y}, \bar{X})\) respectively. With this background, we define another product estimator reported as
\[
\bar{y}_{pt} = \frac{\bar{y}'_t}{\bar{x}_t}
\]

With its Jack-Knife version written as
\[
\bar{y}_{pt}^* = \frac{1}{g} \sum_{t=1}^{n} \bar{y}_{pt}
\]

The expression for bias of \((y_p, \bar{y}_p)\) to the terms of order 0(n-1), are respectively made available as
\[
B_1(\bar{y}_p) = \frac{\bar{y}}{n} (1 + (n - 1)\rho_{yw}) k_c c^2_k
\]
\[
B_1(\bar{y}_p^*) = \frac{\bar{y}}{n} (g + (n - g)\rho_{wy}) k_c c^2_k
\]

Where \(v(\bar{y})_{css} = \frac{1}{n} \sum_{r=1}^{N} (\bar{y}_r - \bar{Y})^2\)
\[
= \frac{1}{N} \sum_{r=1}^{N} \left( \frac{1}{n} \sum_{j=0}^{n-1} y_{r+jk} - \bar{Y} \right)^2
\]
\[
= \frac{\sigma^2}{n} \left[ 1 + (n - 1)\rho_{yw} \right]
\]

And
\[
v(\bar{x})_{css} = \frac{\sigma^2}{n} \left[ 1 + (n - 1)\rho_{xy} \right]
\]

By definition, the intra class correlation within the sample for the same character is define as
\[
\rho_{yy} = \rho_{yw} = \frac{E(y_{r+jk} - \bar{Y})(y_{r+jk} - \bar{Y})}{E(y_{r+jk} - \bar{Y})^2}
\]
\[
\rho_{yw} = \frac{E(y_{r+jk} - \bar{Y})}{E(y_{r+jk} - \bar{Y})(y_{r+jk} - \bar{x})}
\]
\[
\rho_{yy} = \rho_{wy} = \frac{E(y_{r+jk} - \bar{Y})(y_{r+jk} - \bar{x})}{\sqrt{E(y_{r+jk} - \bar{Y})^2 E(y_{r+jk} - \bar{x})^2}}
\]
cov \((\bar{y}, \bar{x}) = \frac{\rho_{yx} \sigma_y \sigma_x}{n} \sqrt{(1 + (n - 1)\rho_{ywx})(1 + (n - 1)\rho_{xw})}

c_z = \frac{2x}{z}, \quad (Z = y, x)

k' = \rho \frac{c_z}{c_x}, \quad \rho = \rho_{yx}

Here \(\rho_{yx}\) is the population correlation coefficient and \(\rho_{wx}\) is the intra class correlation coefficient for both the variables (y,x) and has been assumed to be same for the both the variates (See Murthy, 1977, pp-374-375), \((c_z, c_x)\) are the C. V's for \((y, x)\) respectively.

Now, taking the linear combination of \(\bar{y}, \bar{y}_p\) and \(\bar{y}_p'\) we propose a weighted class of estimator \(T_p\) written as:

\[ T_p = \omega_1 \bar{y} + \omega_2 \bar{y}_p + \omega_3 \bar{y}_p' + \sum_{i=1}^{n} \omega_{i=1} \quad (2.6) \]

Where \((i=1, 2, 3, ...)\) are suitably chosen weights attached to different estimators. Now we have the following theorem - Theorem (2.1): - the weighted class \(T_p\) of the estimator proposed in (2.6) would be unbiased if:

\[ hw_1 + w_0 = 0 \]

For \(h = \frac{\sigma^2 + \mu - \mu}{(1+n-1)\rho_{xy}} \quad (2.7) \]

Which can be proved easily.

If we take \(\alpha_1 = \omega_1, \alpha_2 = \omega = \lambda\) and \((1-\omega - \lambda)\) the unbiasedness condition (2.7) reduced to:

\[ \lambda = - \left(\frac{(1-\omega)}{(h-1)}\right) \]

Where \((\omega, \lambda)\) are constants to be chosen suitably. Thus, we obtain a general class \(T_p\) reduced to exactly unbiased product type estimators in C.S.S. scheme written as:

\[ T_{pu} = w \bar{y} - \left(\frac{1-w}{h-1}\right) \bar{y}_p + \left(\frac{1-w}{h-1}\right) h \bar{y}_p \quad (2.8) \]

Properties of class \(T_p\)

The expression for sampling variance of the proposed class \(T_p\) in (2.6) can be written as

\[ V(T_p) = \omega_1^2 v(\bar{y}) + \omega_2^2 v(\bar{y}_p) + \omega_3^2 v(\bar{y}_p') + 2\omega_1\omega_2 cov(\bar{y}, \bar{y}_p) + 2\omega_1\omega_3 (\bar{y}_p, \bar{y}_p') 2\omega_2\omega_3 (\bar{y}_p, \bar{y}_p') \quad (3.1) \]

Following the standard technique reported by Sukhatame et al. (1970, p. 162-64) [8] under large sample approximation to the terms or under o (n-1), the variance and covariance expressions for the various estimators in (3.1) can be derived easily and cited in the following lemma (3.1):

\[ v(\bar{y}) = \frac{\bar{y}^2}{n} \left(1 + (n - 1)\rho_{xy} \right) c_y^2 \]

\[ v(\bar{y}_p) = \left(\bar{y}_p^2 \right) = cov(\bar{y}_p, \bar{y}_p) \]

\[ \bar{y}^2 \left(1 + (n - 1)\rho_{xy} \right) \left[c_y^2 + (1 + 2k^*)c_x^2 \right] \quad (3.2) \]

\[ Cov(\bar{y}, \bar{y}_p') = Cov(\bar{y}, \bar{y}_p') \]

\[ = \frac{\bar{y}^2}{n} \left(1 + (n - 1)\rho_{xy} \right) (C_x^2 + k^* + C_x^2) \]

Putting the results (3.2) and \(\omega_1 = \omega_1, \omega_2 = \omega, \omega_3 = (1 - \omega - \lambda)\) in simplifying it, the v(T_p) can be obtained as:

\[ V(T_{pu}) = \omega_1^2 V(\bar{y}) + \omega_2^2 V(\bar{y}_p) + 2\omega_1(1-\omega) cov(\bar{y}, \bar{y}_p) + \omega_3^2 \left[1 + (n - 1)\rho_{xy}\right] \left[c_y^2 + (1 - \omega)\left(1 + 2k^*\right)c_x^2 \right] \quad (3.3) \]

The V(T_{pu}) in (3.3) will be minimum for :-

\[ \omega = (1 + k^*) = \omega_0 \quad (stay) \]

Thus, minimum V(T_{pu}) is given as :-

\[ MinV(T_{pu}) = \frac{\bar{y}^2}{n} \left(1 + (n - 1)\rho_{xy}\right) (C_x^2 - k^*2C_x^2) \quad (3.5) \]

Which is equivalent to the approximate variance of usual unbiased linear regression estimator in C.S.S. scheme written as:

\[ \bar{y}_pu = \bar{y} + b_{y\bar{x}}(\bar{x} - \bar{x}) \quad (3.6) \]

Substituting \(\omega_0 = (1 + k^*), \omega_2 = \lambda^* = k^*/(h-1)\) and \(\omega_3 = (1 - \omega^* - \lambda^*) = -k^*/(h-1)\) in (2.6) we obtain as optimum estimator \(T_{pu}\) written as:

\[ T_{pu} = (1 + k^*) \bar{y} + \left(\frac{k^*}{h-1}\right) \bar{y}_p + \left(\frac{k^*}{h-1}\right) \bar{y}_p' \quad (3.7) \]

With variance given in (3.5)

It is to be pointed out that the class T_{pu} in (2.8) would be more efficient than the unbiased sample estimator y and the product estimator y defined under C.S.S. scheme according if either

\[ 1 < \omega < (1+2k^*) \]

\[ Or (1+2k^*) < \omega < 1 \]

And either \(0 < \omega < 2(1+k^*) \]

\[ 2(1+k^*) < \omega < 0 \]

\[ ~74~ \]
Remark 3.1: The variance of any estimator obtained from the class $T_{pu}$ (2.8) can be derived by substituting the appropriate value of constant in $V(T_{pu})$ in (3.3).

Numerical illustration
To see the performance of the class $T_{pu}$, we consider the data on $y$: the pounds of steam used monthly and $X$: the average atmospheric temperature in Fahrenheit. A machine point out (25 values) have been taken from Draper and smith (1966, p.p. 615-16). The summarized data are compiled as follows.

N=25 $C_2^y = 0.10299$ $\rho_{yx} = -0.8486$
N=10 $C_2^y = 0.1077$ $k^* = -0.4450$
$\bar{y} = 9.424$ $S_{xy} = -23.79$ $\rho_\omega = -0.218$
$\bar{x} = 52.6$ $k=2.5 \approx 3$(sampling span)

The interclass correlation coefficient $\rho_\omega$ has been worked out as:

$$\rho_\omega = 1 - \frac{n}{(n-1)\sigma^2}$$

Where $(\sigma^2, \sigma^2_\omega)$ are the population and within sample variance respectively defined under CSS scheme as:

$$\sigma^2 = \frac{1}{N} \sum_{r=1}^{N} \sum_{j=0}^{n-1} (y_{r+jk} - \bar{y})^2$$

$$\sigma^2_\omega = \frac{1}{N} \sum_{r=1}^{N} \sigma^2_{\omega r} = \frac{1}{N} \sum_{r=1}^{N} \sum_{j=0}^{n-1} (y_{r+jk} - \bar{y}_n)^2$$

We have worked out the value of

$$V_1(\cdot) = \frac{V(T_{pu})}{\frac{1}{n}(1 + (n-1)\rho_\omega)} = \left[C_2^y + (1 + \omega)((1 - \omega) + 2k^*)C_2^2\right]$$

And P.R.E. $(\cdot, \bar{y}) = \frac{V(y)}{V(T_{pu})} \times 100$, worked out and are compiled in the table (4.1).

| Values of $\omega$ | Estimators | $V_1(\cdot)$ | P.R.E. $(\cdot, \bar{y})$ |
|-------------------|------------|-------------|--------------------------|
| $\omega = 0.555$ | $\bar{y}$ | 29.9x10^{-3} | 100.00 |
| 0.950 | $T_{PU}$ | 8.6x10^{-3} | 347.67 |
| $-0.1100<\omega<1.0$ | $T_{PU}$ | 25.4x10^{-3} | 117.72 |
| $-0.075$ | $\bar{y}_p$ or $\bar{y}_p'$ | <$V_2(\bar{y})$ | >100.00 |
| $0<\omega<2(1+k^*)$ | $T_{PU}$ | 41.74x10^{-3} | >100.00 |
| Or | | 78.3x10^{-3} | 38.18 |
| $0<\omega<1.11$ | $T_{PU}$ | <$V(\bar{y})$ | >71.31 |
| $0.75$ | $T_{PU}$ | 12.6x10^{-3} | 237.31 |
| 1.05 | $T_{PU}$ | 24.8x10^{-3} | 120.05 |

Table (4.1) reveals that P.R.E. $(T_{puo})$ is maximum among the estimators considered in the study which indicates that $T_{puo}$ is the most efficient (optimum) estimator in the class $T_{pu}$. In practice. One may substitute the estimated values of variances and covariance’s in order to obtain a “near optimum values of $\omega$”. For the choice of $\omega$ in the interval ($-0.1100<\omega<1.0$), the class $T_{pu}$ is always more efficient than sample mean $\bar{y}$. It is also evident that the class $T_{pu}$ in interval ($0<\omega<1.11$), is more efficient than usual product estimator $\bar{y}_p$ as well as it jack-knife version $\bar{y}_p'$. 

References
1. Draper NR, Smith H. Applied regression analysis. Wiley series in probabilities and mathematical statistics, New York, 1966.
2. Kushwaha KS, Singh HP. class of almost unbiased ratio and product type estimator in systematic sampling. J and Soc. Agric statistics. 1989; 41(2):193-205.
3. Murthy MN. Sampling theory and methods. Statistical pub soc Calcutta India, 1977.
4. Quenouille MH. Notes in bias in estimation. Biometric. 1956; 43:353-360
5. Quenouille MH. Approximation test of correlation in time series. J Roy Stat Soc. 1949; 11:68.84
6. Shukla ND. Systematic Sampling and product method of estimation. Proceeding of all India Seminar on Demography and Stats. B.H.U. Varanasi, India, 1971.
7. Sudhakar K. A note on circular systematic sampling design. Somhra and J. Stats: 1978; 40(C1):723-73, Calcutta.
8. Sukhatame PV, Sukhatame BV. Sampling Theory of Surveys with Applications, Iowa State Univ Press Ames, Iowa USA, 1970.