Efficient Minimum Mean Square Error (MMSE) Estimation Using Sample Standard Deviation Improving Searls’ Usual MMSE Mean Estimator for Normal Population with Known Coefficient of Variation

Raghunadh M. Acharya¹ and Ashok Sahai²,*

¹Director; Decision Tree Pvt. Ltd; Hyderabad. India.
²Department of Mathematics & Statistics; Faculty of Science & Technology. St. Augustine Campus; The University of The West Indies. Trinidad & Tobago; West Indies.

ABSTRACT

This paper is aimed at finding an optimal estimator of the normal population mean when the coefficient of variation is known. The paper proposes an “Efficient Minimum Mean Square Error (MMSE) Estimation Approach [Using ‘Sample Standard Deviation’, for improving the well-known Searls’ Usual MMSE Mean Estimator (1964)] for Normal Population with Known Coefficient of Variation”. The ‘Relative Efficiencies [as compared to the usual unbiased sample-mean estimator]’, per the proposed approach, has no simple algebraic form, and hence is not amenable to an analytical study determining its relative gainfulness, as compared to the usual unbiased sample mean estimator. Nevertheless, we examine the relative efficiency of our estimator with respect to the usual unbiased estimator \( \bar{y} \), using an illustrative simulation study with high replication. MATLAB R2013a is used in programming this illustrative “Simulated Empirical Numerical Study”.

© 2015 Elixir All rights reserved.

Introduction

This paper is aimed at finding an optimal estimator of the normal population mean when the coefficient of variation is known. This might well happen to be the case particularly if the population mean \( \theta \) is assumed to be positive. Let the population standard deviation be denoted as \( \sigma \) (\( = a \theta \); \( a \) being the known population coefficient of variation). Searls (1964) [7], Khan (1968) [3], Gleser & Healy (1976) [2], Amholt & Hebert (1995) [1], Miodrag M Lovric & Sahai, Ashok (2011), [4], Sahai, Ashok (2011) [5], Winston A. Richards, Robin Antoine, Ashok Sahai, and M. Raghunadh Acharya (2010) [9], Skrepnek, Grant H; Sahai, Ashok; Antoine, Robin (2015) [10], and Wright, Kimberly & Raghunadh M. Acharya (2009) [11] considered the problem of estimating the normal population mean and variance, when its coefficient of variation (c. v.) is known/unknown. In the context of this paper, the reader would benefit perusing [6] Samuel-Cahn, E. (1994), “Combining Unbiased Estimators”, The American Statistician, 48, 34-36 & Searls, Donald T. & Intarapanich, P. (1990) [8]. It is very well known that the Searles (1964) estimator for the normal population mean gets to be:

\[ \text{Say, } SE = \frac{\bar{y}}{(1+a^2/n)} = \frac{D}{D}; \text{ wherein, } (1+a^2/n) = D; \text{ Say.} \]

In (1.1), ‘a’ is the Known Coefficient of Variation (\( \sigma/\theta \) > \{1/2\} & ‘n’ is the size of the random sample [\( > 30 \)] from the Normal Population \( N(\theta, \theta^2) \).


The major fact is that the statistic \( (\bar{y}, S^2) \); Wherein,

\[ S^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)} = \text{SampleVariance} \]

{Well-known Estimator of \( \sigma^2 \)} based on the random sample of size ‘n’; from the Normal population \( N(\theta, \theta^2) \) is sufficient but NOT complete for \( \{\theta, \sigma^2\} \); rendering the unavailability of the “\( \text{Rao-Blackwellization}\)”. Hence, essentially, we are groping in dark for the ‘UMVU’ & ‘MMSE’ estimators for \( \theta \), or any function of \( \theta \), as such!

The Proposition of an “Efficient Minimum Mean Square Error (MMSE) Estimation Approach Using Sample Standard Deviation improving the well-known Searls’ (1964) MMSE Mean Estimator for Normal Population with Known Coefficient of Variation”

We start by considering the “Minimum Mean Square (MMSE)/Ordinary Searle’s Estimator (OSE)”, using the known Coefficient of Variation (‘C.V.’) \( a' = \sigma/\mu \), and the usual Unbiased Estimator Sample Mean \( \bar{y} \).

The Usual (well-known) Searles Estimator (1964) of the ‘Normal mean’ with known coefficient of variation ‘a’ is as follows:

\[ \text{SE} = \frac{\bar{y}}{(1+a^2/n)} = \text{OSE} = \text{USEARLESESTROFMEAN}; \text{ Say } = \frac{\bar{y}}{(1+a^2/n)}. \]
Wherein,

\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{Sample Variance} = \{\text{Well-known Estimator of 'population variance'}\}. \quad (2.2).
\]

The ‘Relative Efficiency of any Estimator ‘•’ (Relative to the Usual Unbiased Estimator Sample Mean ‘\(\bar{y}\)’)’ is defined as:

\[
\text{Reff (•)} = \left( \frac{\text{Var} (\bar{y})}{\text{MSE (•)} \times 100\%} \right) \times \text{[Defined in %]}.
\]  

(2.3)

To prepare for the proposition of our “Efficient Minimum Mean Square Error (MMSE) Estimation Approach Using ‘Sample Standard Deviation’ improving the well-known Searls’ (1964) MMSE Mean Estimator (Say PPDSSDMMESESTROFMEAN) for Normal Population with Known Coefficient of Variation”, we note some elementary results:

**Lemma 2.1**

For a random sample of size ‘n’ from the Normal Population N (0, \(a^2\theta^2\)); say, \(y_1, y_2, y_3, \ldots, y_n\) we have: (i) \(E \left[ \frac{1}{S} \right] = \text{UCC/}\sigma = \text{UCC/ (a. \theta)}\); wherein, UCC = \(\sqrt{\frac{\Gamma(n/2 - 1)}{\Gamma((n-1)/2)}}\)

(2.4)

Hence, an ‘Unbiased Efficient Estimator’ of ‘\(\theta\)’ is: \(\left[\text{UCC}^*(S/a)\right]\).

AND (ii) \(E \left[ \frac{1}{S^2} \right] = \text{CC2}/(\theta^2\sigma^2)\); wherein, CC2 = (n-1)/ (n-3).

**Proof**: It follows from the well-known fact that \(\left\{\frac{(n-1)}{\sigma^2} S^2 \approx \chi^2_{(n-1)}\right\}\). \(\quad \text{Q.E.D.}\)

**Lemma 2.2**

For a random sample of size ‘n’ from the Normal Population N (0, \(a^2\theta^2\)); say, \(y_1, y_2, y_3, \ldots, y_n\) we have: (i) \(E \left[ \frac{1}{\bar{y}} \right] = (1 + a^2/n)\theta^2\)

(2.6)

AND (ii) \(E \left[ \frac{1}{\bar{y}^2} \right] = (1 + 6. (a^2/n) + 3. (a^4/n^2))\); \(\theta^2 = \text{CCC}. \theta^2\); Say.

(2.7)

**Proof**: It follows from the well-known fact that \(\frac{1}{\bar{y}} \approx N (0, a^2. \theta^2/n)\). \(\quad \text{Q.E.D.}\)

In view of the above simple results our “Efficient Minimum Mean Square Error (MMSE) Estimation Approach Significantly Improving over the Searls’ Usual Mean Estimator for Normal Population with Known Coefficient of Variation of Normal Mean (PPDSSDMMESESTROFMEAN)” gets to be derived as follows.

In the class of estimators “\(\hat{M}. \{\frac{1}{\bar{y}^2}\}/(S)\)”, \(\hat{S} = \sqrt{(S^2)} = \{\text{Sample Standard Deviation}\}\), the ‘MMSE’ estimator gets to be the one with the ‘Optimal Value; Say \(M^*\) of ‘\(\hat{M}\)’, as derived below, using (2.4) to (2.7):

Let, \(\hat{\alpha}(M) = E \left[ \frac{1}{\bar{y}^2} \right] / \hat{\text{MSE}} (M. \{\frac{1}{\bar{y}^2}\}/S)\)

The “Normal Equation” for the Stationary point of \(\hat{\alpha}(M): \sigma \hat{\alpha}(M) = 0\) yields: \(\rightarrow \quad \hat{M}^* = \{\text{CCC}^*\text{UCC}^*\text{a}/\text{CC2}\}; \text{wherein} \text{a = a/\theta; CCC, UCC}^* \quad \text{& 'CC2'} \text{ as are in (2.4), (2.5) & (2.7).} \)

(2.8)

The aforesaid is seminal to our proposed MMSESTROFMEAN of the Normal mean ‘\(\theta\)’, using design-parameter ‘\(\hat{M}\)’:

Say, \(\text{PPDSSDMMESESTROFMEAN} = \hat{M}^* \cdot \{\frac{1}{\bar{y}^2}\}/S\); \(\quad \text{Wherein; a = a/\theta,}\)

\[
S^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)[S = \text{Sqrt}(S^2)]
\]

AND, ‘CCC’, ‘UCC’ & ‘CC2’ are as in (2.4), (2.5) & (2.7)

**The Simulation Empirical Numerical Study**

In the preceding section, it is apparent, that the extent of ‘Relative Gain in Efficiency of Estimation’ will be algebraically rather too intricate, as the issue will depend on the values of the parameters like ‘\(n\)’, ‘\(a\)’, ‘\(\theta\)’, and ‘\(\sigma\)’.

Consequently, the answer to the question as to what is the relative gain/achievement in pursuing the “Efficient Minimum Mean Square Error (MMSE) Estimation Approach using ‘Sample Standard Deviation’ improving the well-known Searls’ Mean Estimator for Normal Population with Known Coefficient of Variation] of Normal Mean (PPDSSDMMESESTROFMEAN)” is pursued in trying to know it through an illustrative “Simulation Empirical Numerical Study”, as is attempted in this section.

The ‘Relative Efficiencies (Relative to the Usual Unbiased Estimator Sample Mean ‘\(\bar{y}\)’) for the TWO estimators, namely ‘USEARLESESTROFMEAN’ in (2.1) & for our proposed estimator ‘PPDSSDMMESESTROFMEAN’ in (2.9).

These, say \(\text{Reff (•)*s},\) have been calculated for SIX illustrative values of the ‘sample size \(n\)’: 35, 50, 75, 100, 150 & 200 & FIVE illustrative values of the ‘Normal Standard Deviation \(\sigma\)’: 4, 5, 6, 7 & 8.

We could assume that, without any loss of generality [Using the translation of the Parent Normal Population Data by “6-\(\bar{y}\)” without changing the Population Standard Deviation \(\sigma = a.\theta\)], & for the simplicity of the illustration, the NORMAL population mean \(\theta = 6\) [i.e. positive]. The values of the actual MSE’s are calculated by considering the random samples of size ‘\(n\)’ using ‘55,555’, a large number of replications (pseudo-random normal samples of the size ‘\(n\)’) for the TWO estimators as also for the usual unbiased estimator, namely the sample mean ‘\(\bar{y}\)’. 

\[36413\text{ Raghunadh M. Acharya and Ashok Sahai/ Elixir Dis. Math. 88 (2015) 36412-36414} \]
Hence values of Reff (•)’s are calculated, as per (2.3). These Reff (•)’s are reported to the closest SIX decimal places of their respective actual values in three tables in the APPENDIX. MATLAB R2013a is used in programming the calculations in this illustrative “Simulated Empirical Numerical Study”.

Conclusions

As expected, the “Relative Efficiency” of the proposed “Efficient Minimum Mean Square Error (MMSE) Estimation [Approach Significantly Improving over the Searls’ Usual Mean Estimator for Normal Population with Known Coefficient of Variation]” of Normal Mean (PPDSSDMSESTROFMEAN) estimator of the “Normal Population Mean” is way above that of the “Usual Minimum Mean Square (MMSE)/Ordinary Searle’s Estimator (OSE)/USEARLESESTROFMEAN” for all sample sizes, as also for all the illustrative value-combinations: ‘[0, σ]’.

As noted earlier, it is very significant to note again that, the fact that the “Rao-Blackwellization” is unavailable in the absence of a complete-sufficient statistic for ‘θ’, and hence, essentially, we are groping in dark for the ‘UMVU’ & ‘MMSE’ estimators for ‘θ’, or any function of ‘θ’, as such.

This fact has been seminal to the motivational zeal for us to look for a more efficient estimator of the “Normal Population Mean when Y ~ N (θ, σ²)” and we are working on it!!

References


### Table A.1

<table>
<thead>
<tr>
<th>RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ &amp; OUR IMPROVED MMSE MEAN ESTIMATORS</th>
<th>[ n = 35; \mu = 6 ].</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARLESESTROFMEAN</td>
<td>101.141573</td>
</tr>
<tr>
<td>PPDSSDMSESTROFMEAN</td>
<td>162.430521</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ &amp; OUR IMPROVED MMSE MEAN ESTIMATORS</th>
<th>[ n = 50; \mu = 6 ].</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARLESESTROFMEAN</td>
<td>100.910467</td>
</tr>
<tr>
<td>PPDSSDMSESTROFMEAN</td>
<td>168.445853</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ &amp; OUR IMPROVED MMSE MEAN ESTIMATORS</th>
<th>[ n = 75; \mu = 6 ].</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARLESESTROFMEAN</td>
<td>100.648765</td>
</tr>
<tr>
<td>PPDSSDMSESTROFMEAN</td>
<td>172.823340</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ &amp; OUR IMPROVED MMSE MEAN ESTIMATORS</th>
<th>[ n = 100; \mu = 6 ].</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARLESESTROFMEAN</td>
<td>100.445578</td>
</tr>
<tr>
<td>PPDSSDMSESTROFMEAN</td>
<td>177.059633</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ &amp; OUR IMPROVED MMSE MEAN ESTIMATORS</th>
<th>[ n = 150; \mu = 6 ].</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARLESESTROFMEAN</td>
<td>100.287921</td>
</tr>
<tr>
<td>PPDSSDMSESTROFMEAN</td>
<td>182.795983</td>
</tr>
</tbody>
</table>

| \( \sigma \)'s: \( \sigma = 4 \) | \( \sigma = 5 \) | \( \sigma = 6 \) | \( \sigma = 7 \) | \( \sigma = 8 \) |
|---------------------------------|---------------------------------|
| RELATIVE EFFICIENCIES [W.R.T 'θ'] OF SERALES’ & OUR IMPROVED MMSE MEAN ESTIMATORS | \[ n = 200; \mu = 6 \]. |
| USEARLESESTROFMEAN | 100.263177 | 100.334535 | 100.447803 | 100.776003 | 100.986859 |
| PPDSSDMSESTROFMEAN | 183.363022 | 214.230063 | 237.063900 | 242.873966 | 249.459199 |