On Simple Confidence Intervals for the Normal Mean with a Known Coefficient of Variation

Suparat Niwitpong, Sa-aat Niwitpong

Abstract—In this paper we proposed the new confidence interval for the normal population mean with a known coefficient of variation. In practice, this situation occurs normally in environment and agriculture sciences where we know the standard deviation is proportional to the mean. As a result, the coefficient of variation is of known. We propose the new confidence interval based on the best unbiased estimator for the population mean in the sense of minimum variance and this new confidence interval will compare with the existing confidence interval. We derive analytic expressions for the coverage probability and the expected length of each confidence interval. A numerical method will be used to assess the performance of these intervals based on their expected lengths.

Keywords—Confidence interval, coverage probability, expected length, known coefficient of variation.

I. INTRODUCTION

The confidence interval for the normal population mean has been studied recently. In the routine text such as Walpole et al. [8] shows that the confidence intervals for the normal means can be constructed in two cases; a) variance is known b) variance is unknown. However, in practice, there are situations in area of environmental and physical sciences that coefficients of variation are known. For example, in environmental studies, Bhat and Rao [1] argued that there are some situations that show the standard deviation of a pollutant is directly related to the mean, that means the τ is known. Furthermore in clinical chemistry, when the batches of some substance (chemicals) are to be analyzed, if sufficient batches of the substances are analyzed, their coefficients of variation will be known. Brazauskas and Ghorai [2] also gave some examples in medical, biological and chemical experiments shown that in practice there are problems concerning that coefficients of variation are known. Most of this statistical problem is due to the estimation of the mean of normal distribution with known coefficient of variation see e.g. Khan [3], Searls [6] and the confidence interval constructed based on the relation between the coefficient of variation and the population mean and it standard deviation that will show in the next section. In this paper, we also proposed the new confidence interval for the mean of normal population mean with a known coefficient of variation based on the best unbiased estimator for the mean with a known coefficient of variation based on Khan [3]. We assess these three confidence intervals using the coverage probabilities and their expected lengths. Typically, we prefer confidence interval with coverage probability at least the nominal value (1−α) and its expected length is short.

II. CONFIDENCE INTERVALS FOR NORMAL POPULATION MEAN

Let $X = [X_1, \ldots, X_n]$ be a random sample from the normal distribution with mean $\mu$ and standard deviation $\sigma$. The sample mean and variance for $X$ are, respectively, denoted as $\bar{X}$ and $S^2$ when $X = n^{-1} \sum_{i=1}^{n} X_i$, and $S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. We are interested in $100(1-\alpha)\%$ confidence interval for $\mu$. In practice, $\sigma^2$ is an unknown parameter and $S^2$ is an unbiased estimator for $\sigma^2$. In this case, a well-known $100(1-\alpha)\%$ confidence for $\mu$ is

$$CI_\mu = \left[ \bar{X} - c \frac{S}{\sqrt{n}}, \bar{X} + c \frac{S}{\sqrt{n}} \right]$$

where $c$ is $t_{1-\alpha/2}$, an upper $1 - \alpha/2$ percentiles of the $t$-distribution with n-1 degrees of freedom. In the next section we review a method to construct a confidence interval for $\mu$ with known coefficient of variation.

III. CONFIDENCE INTERVALS FOR NORMAL POPULATION MEAN WITH A KNOWN COEFFICIENT OF VARIATION

It is known that the unbiased estimator for $\mu$ is the mean $\bar{X}$, however, when we known a prior information, for example the coefficient of variation ($\tau = \sigma/\mu$), the estimator $\bar{X}$ is not appropriate in terms of the mean square error. Searls [6] proposed the estimator $\bar{X}^* = (n + \tau^2)^{-1} \sum_{i=1}^{n} X_i$ where he showed that this estimator has lower mean squares error than that of the unbiased estimator. Further, it is easy to show that the variance of the estimator $\bar{X}^* = \frac{nS^2}{(n + \tau^2)}. Hence, a 100(1-\alpha)\%$ confidence for $\mu$ with known coefficient of variation $\tau$ is

$$CI_\mu = \left[ \bar{X}^* - d \sqrt{\frac{nS^2}{(n + \tau^2)^2}}, \bar{X}^* + d \sqrt{\frac{nS^2}{(n + \tau^2)^2}} \right]$$

where $d$ is $z_{1-\alpha/2}$, an upper $1 - \alpha/2$ percentiles of the standard normal-distribution.

We now propose the new confidence interval for $\mu$ based on
a prior information $\tau = \sigma/\mu$. It is easy to see that $\mu = \sigma/\tau$, hence our proposed confidence interval for $\mu$ is equal to the confidence interval of $\sigma/\tau$ which is

$$CI_{1\mu} = \left[ \delta - d\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}, \delta + d\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}} \right]$$

where $\delta = \alpha\delta_2 + (1-\alpha)\delta_1$, $0 < \alpha < 1$ and $\alpha = \tau^2/(\tau^2 + n(k_1/k_2-1))$, $k_1 = (n-1)\Gamma^2((n-1)/2)$, $k_2 = 2\Gamma^2(n/2)$, $\delta_1 = X$, $\delta_2 = c_n\sqrt{nS_n}$ and $S_n = n^{-1}\sum_{i=1}^n(X_i - \bar{X})^2$.

Note that confidence intervals (2), (3) and (4) have a priori information $\tau$ in their confidence intervals whereas confidence interval (1) has no a priori information in its confidence interval. Niwitpong [5] reported that confidence interval $CI_{1\mu}$ is not appropriate to used when $\tau$ is known. She also reported that confidence intervals $CI_{3\mu}$ and $CI_{4\mu}$ are shorter than that of confidence interval $CI_{1\mu}$ when $\tau$ is known. In this paper, we therefore, consider confidence intervals $CI_{3\mu}$ and $CI_{4\mu}$ compared to our proposed confidence interval $CI_{2\mu}$. Like Niwitpong [5], we compare all proposed confidence intervals based on their coverage probabilities and their expected lengths.

**IV. COVERAGE PROBABILITIES AND EXPECTED LENGTH OF EACH CONFIDENCE INTERVAL**

Following Niwitpong and Niwitpong [4], we now derive an analytic expression for the coverage probability for $CI_{1\mu}$. Let $P(\mu \in CI_{1\mu})$ be the coverage probability of confidence interval $CI_{1\mu}$ then an analytic expression for this confidence coverage is given in Theorem 1, below.

**Theorem 1** [Niwitpong [5]] The coverage probability and expected length for $CI_{1\mu}$ are respectively

$$E[\Phi(A_2) - \Phi(A_1)], A_1 = \frac{X}{\bar{X}}, A_2 = \frac{\hat{S}}{\bar{S}}, Z is the standard normal distribution and \Phi(\cdot) is a standard normal function and$$

$$E(CI_{1\mu}) = 2\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}$$

**Proof** See Niwitpong [5].

Similarly to $CI_{1\mu}$, we now derive an analytic expression for the coverage probability for $CI_{2\mu}$. Let $P(\mu \in CI_{2\mu})$ be the coverage probability of confidence interval $CI_{2\mu}$ then an analytic expression for this confidence coverage is given in Theorem 2, below.

**Theorem 2** [Niwitpong [5]] The coverage probability and expected length for $CI_{2\mu}$ are respectively

$$E[\Phi(B_2) - \Phi(B_1)], B_1 = \frac{d}{\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}}, B_2 = \frac{d}{\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}}$$

**Proof** See Niwitpong [5].

Similarly to $CI_{1\mu}$, we now derive an analytic expression for the coverage probability for $CI_{3\mu}$. Let $P(\mu \in CI_{3\mu})$ be the coverage probability of confidence interval $CI_{3\mu}$ then an analytic expression for this confidence coverage is given in Theorem 3, below.

**Theorem 3** [Niwitpong [5]] The coverage probability and expected length for $CI_{3\mu}$ are respectively

$$E[\Phi(C_2) - \Phi(C_1)], C_1 = \frac{X}{\bar{X}}, C_2 = \frac{\hat{S}}{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}, Q = \sigma/\sqrt{n}$$

**Proof** See Niwitpong [5].

Similarly to $CI_{1\mu}$, we now derive an analytic expression for the coverage probability for $CI_{4\mu}$. Let $P(\mu \in CI_{4\mu})$ be the coverage probability of confidence interval $CI_{4\mu}$ then an analytic expression for this confidence coverage is given in Theorem 4, below.

**Theorem 4** The coverage probability and expected length for $CI_{4\mu}$ are respectively

$$E[\Phi(C_4) - \Phi(-C_4)], C_4 = \frac{d}{\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}}, C_3 = \frac{d}{\sqrt{\frac{\tau^2S_n^2}{n(1+2\tau^2)}}}$$

**Proof** Similarly to Niwitpong [5].
We wrote function in R program to generate the data which is shorter than the confidence interval $CI$. This is notified in the previous section: $CI$ notified in the previous section: $CI$ is preferable.

Table I. The expected lengths of $CI_b, CI_s$ and $CI_p$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$E(CI_b)$</th>
<th>$E(CI_s)$</th>
<th>$E(CI_p)$</th>
<th>$E(CI_b)$</th>
<th>$E(CI_s)$</th>
<th>$E(CI_p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$</td>
<td>$n = 30$</td>
<td>$n = 50$</td>
<td>$n = 100$</td>
<td>$n = 10$</td>
<td>$n = 30$</td>
<td>$n = 50$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

**V. SIMULATION FRAMEWORK**

In this section, since we have proved that the coverage probabilities of all confidence intervals are equal to $1 - \alpha$, we therefore only use numerical method to assess these confidence intervals notified in the previous section: $CI_b, CI_s$ and $CI_p$ based on their average length widths. We design our simulation framework without losing generality, by setting $\sigma = 1, \tau = 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.5, 0.8, 0.9, 1, 1.1, 1.3, 1.5, 1.9, 2.5$ and the samples sizes $n = 10, 30, 50, 100$. We wrote function in R program to generate the data which is normally distributed with zero mean and unit variance and then compute the average length width of each confidence interval: $CI_b, CI_s$ and $CI_p$. All results are illustrated in Table I.

**VI. SIMULATION RESULTS**

From Table I, we found that the expected length of the confidence interval $CI_b$ with a known coefficient of variation, $\tau \leq 0.10$ when $n = 10, \tau \leq 0.03$ when $n = 30$, $\tau \leq 0.02$ when $n = 50$ and $\tau \leq 0.01$ when $n = 100$, is shorter than other confidence intervals. The confidence interval $CI_s$ is shorter than the confidence interval $CI_p$ when $\tau \leq 0.1, n = 10, \tau \leq 0.5, n = 30, 50, 100$ otherwise the confidence interval $CI_p$ is preferable.

**VII. CONCLUSION**

In this paper we proposed new confidence intervals, $CI_b$, for the normal population mean with a known coefficient of variation based on the best unbiased estimator of Khan [3]. This new confidence interval will compare to our previous confidence intervals, see e.g. Niwitpong [5]. We derived, mathematically, coverage probabilities and expected lengths of these intervals. It is shown in sections IV that the coverage probabilities of $CI_b$ and $CI_a$ and $CI_p$ are equal to the nominal value $1 - \alpha$. From numerical results in Table I, it is recommended that for a known small coefficient of variation and a small sample size, our new confidence interval $CI_b$ is preferable. The confidence interval $CI_a$ outperforms other confidence intervals when $\tau \leq 0.5$ otherwise we choose the confidence interval $CI_p$.

**ACKNOWLEDGMENT**

The author is grateful to King Mongkut’s University of Technology North Bangkok for its financial support.

**REFERENCES**


