A New Formulation of Minimum Risk Fixed-Width Confidence Interval (MRFWCI) for a Normal Mean

Swathi Venkatesan

Department of Statistics
University of Connecticut

Construction of Loss and Risk Functions

Suppose $X_1, X_2, \ldots, X_n \ldots$ are *independent and identically distributed (i.i.d)* observations from a common $N(\mu, \sigma^2)$ distribution where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $\theta = (\mu, \sigma)$ and both μ, σ are unknown.

Having pre-assigned numbers, $d(> 0)$ and $0 < \alpha < 1$, the FWCI for μ is:

$$\textbf{FWCI: } J_n \equiv \left[\bar{X}_n \pm d\right],$$

$$\begin{aligned} &\text{with confidence coefficient } P_\theta\{\mu \in J_n\} = 2\Phi(n^{1/2}d/\sigma) - 1 \geq 1 - \alpha \\ &\text{when } n \text{ is the smallest integer } \geq z_{\alpha/2}^2 \sigma^2 d^{-2} \equiv n^*. \end{aligned} \tag{1}$$

Generically, we may initiate the following idea of a loss function in practice:

$$L_n(\theta) = EstErr_n(\theta) + c_n(\theta), \tag{2}$$

where $EstErr_n(\theta)$ is the estimation error involving $I\left[\tau(\theta) \notin FWCI; n\right]$ plus the $c_n(\theta) \equiv$ sampling cost$_n$

We combine (1) and (2) to express:

$$p_n(\theta; J_n) \equiv E_\theta\{I[\mu \notin J_n; n]\} = 2\{1 - \Phi(n^{1/2}d/\sigma)\}. \tag{3}$$

We incorporate the following expression of the cost function:

$$c_n(\theta) = 2(d^\rho \sigma)^{-1}\phi(z_{\alpha/2})n^{1/2}. \tag{4}$$

Then, in the spirits of (1) and (3)-(4), we construct the loss function (for fixed $0 < \rho \leq 1$) as:

$$\textbf{Loss:}\ L_n(\mu, J_n) \equiv d^{-\rho-1} I[\mu \notin J_n; n] + c_n(\theta). \tag{5}$$

The corresponding Risk is given by:

$$\textbf{Risk:}\ R_n(\mu, J_n) \equiv E_\theta[L_n(\mu, J_n)]$$

$$= 2d^{-\rho-1}[1 - \Phi(n^{1/2}d/\sigma)] + 2(d^\rho\sigma)^{-1}\phi(z_{\alpha/}$$

General setup of MRFWCI-Normal mean Estimation

We propose the MRFWCI for μ based on the recorded observation $\{N, X_1, X_2, \ldots, X_N\}$ as:

$$\textbf{MRFWCI: } J_N = [\bar{X}_N \pm d] \text{ where } \bar{X}_N = N^{-1}\sum_{i=1}^{N} X_i.$$

The risk function $R_N(\mu, J_N)$ with the terminal strategy (N, J_N) for μ under (5) is:

$$R_N(\mu, J_N) = 2\{d^{-\rho-1}E_\theta[1 - \Phi(N^{1/2}d/\sigma)] + (d^\rho\sigma)^{-1}\phi(z_{\alpha/2})E_\theta(N^{1/2})\}.$$

Without specifying a formal sampling strategy , we assume a number of key associated properties as $d \to 0$:

A1 $E_\theta(N) = n^* + a_1 + o(1)$;

A2 $H \equiv \frac{N-n^*}{n^{*1/2}} \xrightarrow{\text{Ł}} N(0, a_2), a_2 > 0$;

A3 $P_\theta\{N \le \epsilon^*\} = O(n^{*-a_3}), \epsilon \in (0,1), a_3 > 0$;

A4 $|H|^{a_4}$ is uniformly integrable , $a_4 > 0$.

where the expressions of the real numbers $a_i' s$ would not involve d.

Asymptotic first-order properties

For the general MRFWCI estimation methodology carried out under the loss function (5), for every fixed α, ρ and θ, with a_i defined via assumption A_i, we have the following asymptotic first-order conclusions as $d \to 0$:

(i) $n^{*-1}N \xrightarrow{P_\theta} 1$;

(ii) $E_\theta[(n^{*-1}N)^\kappa] \to 1$ [*asymptotic first-order efficiency property*];

(iii) $P_\theta\{\mu \in J_N\} \to (1-\alpha)$ [*asymptotic consistency property*];

(iv) $\text{RiskEff}_d = \frac{R_N(\mu, J_N)}{R_{n^*}(\mu, J_{n^*})} \to 1$ [*asymptotic risk efficiency property*].

Asymptotic second-order properties

For the general MRFWCI estimation methodology carried out under (5), for every fixed α, ρ and θ, with a_i defined via assumption A_i, we have the following asymptotic second-order conclusions when $a_3 > \frac{5}{2}$ and $a_4 = 2$ as $d \to 0$:

(i) $E_\theta[(n^{*-1}N)^{1/2}] = 1 + a_5 n^{*-1} + o(n^{*-1})$ where $a_5 = \frac{1}{2}(a_1 - \frac{1}{2}a_2)$;

(ii) $P_\theta\{\mu \in J_N\} = (1-\alpha) + 2a_6 n^{*-1} + o(n^{*-1})$
where $a_6 = \frac{1}{2}\{a_1 - (a_2/4)(1 - z^2_{\alpha/2})\}z_{\alpha/2}\phi(z_{\alpha/2})$;

(iii) $\text{RiskEff}_d = \omega = \frac{R_N(\mu, J_N)}{R_{n^*}(\mu, J_{n^*})} = 1 + a_7 n^{*-1} + o(n^{*-1})$
where $a_7 = 2\{\alpha + 2z_{\alpha/2}\phi(z_{\alpha/2})\}^{-1}\{z_{\alpha/2}\phi(z_{\alpha/2})a_5 - a_6\}$.

Illustrations of Multistage Sampling Strategies

- **Purely Sequential Sampling strategy**
  We begin with pilot observations $X_1, \ldots, X_m; m \geq 2$. Then, we record one additional observation at-a-time according to the following stopping time:
  $$N \equiv N_d = inf\{n \geq m; n \geq z^2_{\alpha/2} S^2_n / d^2\} \text{ and } J_N = [\bar{X}_N \pm d].$$
- **Two-stage Sampling strategy**
  We assume that there exists known $\sigma_L (> 0)$ such that $\sigma > \sigma_L$.The pilot sample size is defined as:
  $$m\ equiv m_d = max\{m_0(\geq 2), < z^2_{\alpha/2}\sigma^2_L/d^2 > +1\}.$$
  We begin with pilot observations $X_1, \ldots, X_m, m \geq 2$ and find the final sample size as:
  $$N \equiv N_d = max\{m, < t^2_{m-1,\alpha/2} S^2_m / d^2 > +1\} \text{ and } J_N = [\bar{X}_N \pm d].$$

The table below summarizes some properties of the sampling strategies.

Sampling Strategy	a_1	a_2
Purely Sequential	-1.1828	2
Two-stage	$\frac{1}{2}[(z_{\alpha/2}^2 + 1)\frac{\sigma^2}{\sigma_L^2} + 1]$	$2\frac{\sigma^2}{\sigma_L^2}$

Airquality Data Analysis

Now we will illustrate applications of the MRFWCI problem for the average in the context of *airquality dataset* describing daily air quality measurements in New York. The variable of interest (X) was *wind speed*,which was normally distributed. The mean and standard deviation from the full dataset were: $\mu = 9.957$ and $\sigma = 3.523$. The table below summarises the performance of the sampling strategies (fixing $\alpha = 0.05$ and $\rho = 1.0$). The notations are defined below:

- n: terminal sample size;
- $\hat{\mu}_n = n^{-1} \sum_{j=1}^{n} x_j$; terminal sample mean;
- $J_n = [\hat{\mu}_n \pm d]$; terminal 95% MRFWCI for μ.

Sampling Strategy	d	n	$\hat{\mu}_n$	J_n
Purely Sequential	0.7	102	10.034	$[9.334, 10.734]$
$m = 15$	0.8	83	10.427	$[9.626, 11.226]$
	0.9	65	10.795	$[9.895, 11.695]$
Two-stage	0.7	105	10.072	$[9.375, 10.775]$
$\sigma_L = 2, m_0 = 5$	0.8	79	10.534	$[9.734, 11.334]$
	0.9	69	10.559	$[9.965, 11.459]$
Two-stage	0.7	115	10.146	$[9.446, 10.846]$
$\sigma_L = 3, m_0 = 5$	0.8	90	10.321	$[9.521, 11.121]$
	0.9	56	10.993	$[10.093, 11.893]$

Simulation Study

Simulated performances of MRFWCI strategies under $10,000$ replications for a $N(25, 16)$ population with $\alpha = 0.05$ and $\rho = 1.0$ in (5).

n^*	d	$\bar{n}$	$\bar{p}$	$\bar{\omega}$
Purely Sequential $m = 15$				
200	0.5543	198.744	0.9484	1.004
1000	0.2479	998.048	0.9462	1.001
5000	0.1108	4997.536	0.9477	1.001
Two-Stage $\sigma_L = 3, m_0 = 10$				
200	0.5543	200.851	0.9459	1.007
1000	0.2479	1001.101	0.9459	1.003
5000	0.1108	4997.536	0.9477	1.001

Conclusion

When handling FWCI problems,the lack of a loss function creates a false impression that perhaps (i) observations may not cost at all and/or (ii) one's available budget may be unlimited. In the MRFWCI formulation,an FWCI problem has been casted in the light of an MRPE by balancing the estimation error with the cost of sampling.

Reference I

Nitis Mukhopadhyay and Swathi Venkatesan.

A New Formulation of Minimum Risk Fixed-Width Confidence Interval (MRFWCI) Estimation Problems for a Normal Mean with Illustrations and Simulations: Applications to Airquality Data.
Sequential Analysis, 41, 2022.