ELEG 5633 Detection and Estimation
Minimum Variance Unbiased Estimators (MVUE)

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

Outline

- Minimum Variance Unbiased Estimators (MVUE)
- Cramer-Rao Lower Bound (CRLB)
- Best Linear Unbiased Estimators (BLUE)

Minimum Variance Unbiased Estimators (MVUE)

- Recall $\text{MSE}(\hat{\boldsymbol{\theta}}) = \|\text{bias}(\hat{\boldsymbol{\theta}})\|_2^2 + \text{var}(\hat{\boldsymbol{\theta}})$
- It is usually impossible to design $\hat{\boldsymbol{\theta}}$ to minimize the MSE because the bias depends on the true value $\boldsymbol{\theta}^*$, which is unknown.
- Restrict to unbiased estimators, $\mathbb{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}^*$. Then $\text{MSE}(\hat{\boldsymbol{\theta}}) = \text{var}(\hat{\boldsymbol{\theta}})$
- Note $\text{var}(\hat{\boldsymbol{\theta}})$ does not depend on $\boldsymbol{\theta}^*$.

- A realizable approach: optimize the MSE with respect to all unbiased estimators.
- Minimum Variance UnBiased (MVUB) estimator is defined as

$$\hat{\boldsymbol{\theta}} = \underset{\hat{\boldsymbol{\theta}}:\mathbb{E}(\hat{\boldsymbol{\theta}})=\boldsymbol{\theta}^*}{\text{argmin}} \; \mathbb{E}[\|\hat{\boldsymbol{\theta}} - \mathbb{E}(\hat{\boldsymbol{\theta}})\|_2^2]$$

Example

$X_1, X_2, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta^*, \sigma^2)$. Let $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i$. We have

$$\mathbb{E}\hat{\theta} = \theta^*$$

$$\mathsf{MSE}(\hat{\theta}) = \frac{1}{n^2}\sum_{i=1}^n \mathsf{var} X_i = \frac{\sigma^2}{n}$$

Is this the MVUB estimator?

This question can be answered by using Cramer-Rao Lower Bound (CRLB).

MVUE

- Does a MVUE always exist?
- If it does, can we always find it?
- Can we say anything about MVUE?

Cramer-Rao Lower Bound (CRLB)

- The CRLB gives a lower bound on the variance of ANY UNBIASED estimator
- Does NOT guarantee the bound can be achieved.
- Can be used to verify that a particular estimator is MVUB.
- Otherwise we can use other tools to construct a better estimator from any unbiased one – Possibly the MVUE if conditions are met.

CRLB for Scalar Parameters

Theorem (Cramer-Rao Lower Bound (CRLB))

Let $p(x|\theta)$ satisfy the regularity condition

$$\mathbb{E}\left[\frac{\partial \ln p(x|\theta)}{\partial \theta}\right] = 0$$

Then the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\mathit{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(x|\theta)}{\partial \theta^2} \ \big|_{\theta=\theta^*}\right]} = \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln p(x|\theta)}{\partial \theta}\right)^2 \ \big|_{\theta=\theta^*}\right]}$$

Furthermore an unbiased estimator may be found that attains the bound for all θ iff

$$\frac{\partial \ln p(x|\theta)}{\partial \theta} = I(\theta)[g(x) - \theta]$$

for some $g(\cdot)$ and I. That estimator, which is the MVUE, is $\hat{\theta} = g(x)$ and the minimum variance is $1/I(\theta)$.

Example

$X_1, X_2, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta^*, \sigma^2)$. Let $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i$. Is it MVUB?

Solution:

$$\log p(\mathbf{x}|\theta) = N \log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - A)^2$$

$$\frac{\partial}{\partial\theta}\log p(\mathbf{x}|\theta) = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - A)$$

$$\frac{\partial^2}{\partial^2\theta}\log p(\mathbf{x}|\theta) = -\frac{n}{\sigma^2}$$

$$I(\theta^*) = \mathbb{E}\left[\left(\frac{\partial \log p(\mathbf{x}|\theta)}{\partial\theta}\right)^2 |_{\theta=\theta^*}\right]$$

$$= \frac{1}{\sigma^4}\sum_{i=1}^n \mathbb{E}[(x_i - \theta^*)^2] = \frac{n}{\sigma^2} = -\mathbb{E}\left[\frac{\partial^2 \ln p(x|\theta)}{\partial\theta^2} \ |_{\theta=\theta^*}\right]$$

$$\mathrm{var}\hat{\theta} \geq \frac{1}{I(\theta^*)} = \frac{\sigma^2}{n}$$

More about $I(\theta)$

- Fisher Information

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln p(x|\theta)}{\partial \theta^2}|_{\theta=\theta^*}\right] = \mathbb{E}\left[\left(\frac{\partial \ln p(x|\theta)}{\partial \theta}\right)^2 |_{\theta=\theta^*}\right]$$

- It is always non-negative.
- It is additive for independent observations.
- The CRLB for N i.i.d. observations is $1/N$ times that for one observation.

Theorem (Vector Form of the Cramer-Rao Lower Bound (CRLB))

Assume $p(\mathbf{x}|\boldsymbol{\theta})$ *satisfy the regularity condition* $\mathbb{E}\left[\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}, \forall \boldsymbol{\theta}$*. Let* $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{x})$ *be an unbiased estimator of* $\boldsymbol{\theta}^*$*. Then the error covariance satisfies*

$$\mathbb{E}[(\hat{\boldsymbol{\theta}} - \mathbb{E}\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \mathbb{E}\hat{\boldsymbol{\theta}})^T] - \mathbf{I}^{-1}(\boldsymbol{\theta}^*) \succeq \mathbf{0}$$

where $\succeq \mathbf{0}$ *means the matrix is positive semi-definite.* $\mathbf{I}(\boldsymbol{\theta}^*)$ *is the Fisher-Information matrix with* (i, j)*th element*

$$\mathbf{I}_{ij}(\theta^*) = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]$$

Furthermore an unbiased estimator may be found that attains the bound iff

$$\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})[g(\mathbf{x}) - \boldsymbol{\theta}]$$

In that case, $\hat{\boldsymbol{\theta}} = g(\mathbf{x})$ *is the MVUE with covariance matrix* $\mathbf{I}^{-1}(\boldsymbol{\theta})$*.*

Example

Consider $x[n] = A + w[n], n = 0, 1, \ldots, N$, where $w[n]$ is WGN with variance σ^2. What is the CRLB for the vector parameter $\boldsymbol{\theta} = [A, \sigma^2]^T$?

Solutions:

Let $\theta_1 = A$ and $\theta_2 = \sigma^2$.

- $\log p(\mathbf{x}|\boldsymbol{\theta}) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i-1}^{N}(X_i - A)^2$

$$\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_1} = \frac{1}{\sigma^2}\sum_{i-1}^{N}(X_i - A)$$

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_1^2}\right] = \mathbb{E}\left[-\frac{N}{\sigma^2}\right] = -\frac{N}{\sigma^2}$$

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2}\right] = \mathbb{E}\left[-\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i - A)\right] = 0$$

Solution:(Cont'd)

$$\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_2} = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i-1}^{N}(X_i - A)^2$$

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_2^2}\right] = \mathbb{E}\left[\frac{N}{2}\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\sum_{i-1}^{N}(X_i - A)^2\right] = -\frac{N}{2\sigma^4}$$

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1}\right] = \mathbb{E}\left[-\frac{1}{\sigma^4}\sum_{i-1}^{N}(X_i - A)\right] = 0$$

- Fisher Information matrix $I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$
- $I^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$

Solution:(Cont'd)

- If an estimator can achieve the CRLB, then it must saitsfy $\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(g(x) - \boldsymbol{\theta})$

$$\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta}) \left(\begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} X_i \\ \frac{1}{N}\sum_{i=1}^{N}(X_i - A)^2 \end{bmatrix} - \begin{bmatrix} A \\ \sigma^2 \end{bmatrix} \right)$$

- Thus

$$\hat{A} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - A)^2$$

  Therefore CRLB cannot be achieved because $\hat{\sigma}^2$ depends on the unknown parameter A.

- Recall MLE: $\hat{\sigma}^2_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{A})^2$, $\mathbb{E}(\hat{\sigma}^2_{\text{ML}}) = \frac{N-1}{N}\sigma^2$. It is a biased estiamtor.

Efficiency

- An unbiased estimator that achieved the CRLB is said to be efficient.
- Efficient estimators are MVUB, but not all MVUB estimators are necessarily efficient.
  - An MVUB could minimimize the MSE, but the minimium achievable MSE is larger than the CRLB.
- An estimator $\hat{\theta}_n$ is said to be asymptotically efficient if it achieves the CRLB, as $n \to \infty$.
- Recall that under mild regularity conditions, the MLE has an asymptotic distribution

$$\hat{\boldsymbol{\theta}}_n \sim \mathcal{N}\left(\boldsymbol{\theta}^*, \frac{1}{n} I^{-1}(\boldsymbol{\theta}^*)\right) \quad asymptotically$$

  so $\hat{\boldsymbol{\theta}}_n$ is asymptotically unbiased.

$$\text{var}(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n} I^{-1}(\boldsymbol{\theta}^*)$$

  so it is asymptotically efficient.

Best Linear Unbiased Estimators (BLUE)

So far, we have learned

- CRLB, may give you the MVUE
- MVUE still may be tough to find

Best Linear Unbiased Estimators

- Find the MVUE by constraining the estimators to be linear, i.e.,

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{x}$$

- Only need the first and second moments of $p(\mathbf{x}|\boldsymbol{\theta})$, which is fairly practical.
- Trading optimality for practicality. There is no reason to believe that a linear estimator is efficient, an MVUE, or optimal in any sense.

BLUE Assumptions

- In order to employ BLUE, the relationship between $\mathbf{x}$ and θ must be linear, i.e.,

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

  This ensures that we can find a linear unbiased estimator.
- $\mathbf{H}$ is known
- $\mathbf{C} = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]$ is known.

We wish to find the linear unbiased estimator with the minimum variance for each $\theta_i \in \boldsymbol{\theta}$.

- $\hat{\boldsymbol{\theta}} = \mathbf{A}^T\mathbf{x}$
- $\mathbf{A}^T\mathbf{H} = \mathbf{I}$
- Find $\mathbf{A}^T$ to minimize $\sum_{i=1}^{N} \mathrm{var}(\hat{\theta}_i) = \mathrm{tr}(\mathbf{A}^T\mathbf{C}\mathbf{A})$

BLUE Formulation

$$\begin{aligned} \min_{\mathbf{A}} \quad & \mathrm{tr}(\mathbf{A}^T\mathbf{C}\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{A}^T\mathbf{H} = \mathbf{I} \end{aligned}$$

Solutions: This is a convex optimization problem.

- $J(\mathbf{A}) = \mathrm{tr}\left(\mathbf{A}^T\mathbf{C}\mathbf{A}) - (\mathbf{A}^T\mathbf{H} - \mathbf{I})\boldsymbol{\lambda}\right)$
- $\frac{\partial J(\mathbf{A})}{\partial \mathbf{A}} = 2\mathbf{C}\mathbf{A} - \mathbf{H}\boldsymbol{\lambda} = 0 \quad \Rightarrow \quad \mathbf{A} = \frac{1}{2}\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\lambda}$
- Determine $\boldsymbol{\lambda}$ by using the constraint $\mathbf{A}^T\mathbf{H} = \mathbf{I} = \frac{1}{2}\boldsymbol{\lambda}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H} = \mathbf{I}$

$$\frac{1}{2}\boldsymbol{\lambda}^T = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$$

- Optimum solution

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

- Error covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbb{E}[(\mathbf{A}^T\mathbf{x} - \boldsymbol{\theta})(\mathbf{A}^T\mathbf{x} - \boldsymbol{\theta})^T] = \mathbb{E}[\mathbf{A}^T\mathbf{w}\mathbf{w}^T\mathbf{A}] = \mathbf{A}^T\mathbf{C}\mathbf{A} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$$

Theorem (Gauss-Markov)

If the data are of the general linear model form

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $\mathbf{H}$ *is a known* $N \times p$ *matrix,* $\boldsymbol{\theta}$ *is a* $p \times 1$ *vector of parameters to be estimated, and* $\mathbf{w}$ *is a* $N \times 1$ *noise vector with zero mean and covariance* $\mathbf{C}$*, then the BLUE of* $\boldsymbol{\theta}$ *is*

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

In addition, the covariance matrix of $\hat{\boldsymbol{\theta}}$ *is*

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$$

Example

$x_n = A + w_n, n = 1, \ldots, N$, w_n not Gaussian, but independent, identically distributed with zero mean and variance σ^2.

Solutions:

- $\mathbf{H} = \mathbf{1}_N$
  $\hat{A}_{\mathsf{BLUE}} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \mathsf{var}(\hat{A}_{\mathsf{BLUE}}) = \frac{\sigma^2}{N}$
- The sample mean is the BLUE independent of the PDF of the data. It is the MVUE for Gaussian noise.
- Recall MMSE: $\hat{A}_{\mathsf{MMSE}} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{1}{N}\sigma^2}\bar{X} \qquad \mathsf{var}(\hat{A}_{\mathsf{MMSE}}) = \frac{\sigma_A^2\sigma^2}{N\sigma_A^2 + \sigma^2} = \frac{\sigma^2}{N + \frac{\sigma^2}{\sigma_A^2}}$
-
$$\mathsf{var}(\hat{A}_{\mathsf{MMSE}}) \leq \mathsf{var}(\hat{A}_{\mathsf{BLUE}})$$
$$\lim_{\sigma_A^2 \to \infty} \mathsf{var}(\hat{A}_{\mathsf{BLUE}}) = \mathsf{var}(\hat{A}_{\mathsf{MMSE}})$$
- In MMSE, we have prior information about A ($\mu_A = 0$ and σ_A^2). In BLUE and MVUE, no prior information of A is avaiable $\sigma_A^2 = \infty$).

Example

$x[n] = A + w[n], n = 0, 1, \ldots, N-1$, $w[n]$ not Gaussian, but independent with zero mean and variance σ_n^2.

Example

For the general linear model

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w}$$

where $\mathbf{s}$ is a known $N \times 1$ vector and $\mathbb{E}[\mathbf{w}] = 0$, $\mathbb{E}[\mathbf{w}\mathbf{w}^T] = \mathbf{C}$. Find the BLUE.

Solutions: Let $\mathbf{y} = \mathbf{x} - \mathbf{s}$.

Example

Consider the following curve fitting problem, where we wish to find $\theta_0, \cdots, \theta_{p-1}$ so as to best fit the experimental data points $(t_n, x(t_n))$ for $n = 0, \cdots, N-1$ by the polynomial curve

$$x(t_n) = \theta_0 + \theta_1 t_n + \theta_2 t_n^2 + \cdots + \theta_{p-1} t_n^{p-1} + w(t_n) \tag{1}$$

where $w(t_n)$ are i.i.d. with zero mean and variance σ^2. Find the BLUE of $\boldsymbol{\theta} = [\theta_0, \theta_1, \cdots, \theta_{p-1}]^T$.