Estimation: chapter 4

Linear Models

Lifang Feng
lffeng@ustb.edu.cn

北京科技大学
University of Science and Technology Beijing

Spring 2022

MVUE and linear models

The minimum variance unbiased estimator (MVUE) is difficult to find (if it exists) in general. One exception to this rule is when the parameters to be estimated are related to the observations in a linear fashion, i.e. we have a *linear model.*

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

- $\mathbf{x} : N \times 1$ observation vector
- $\mathbf{H} : N \times p$ known observation matrix with $N > p$ and rank p
- $\theta : p \times 1$ vector of unknown parameters to be estimated
- $\mathbf{w} : N \times 1$ noise vector, $\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

One way of finding the MVUE is using the CRLB, where we know that $\hat{\theta} = g(\mathbf{x})$ is the MVUE if

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \mathbf{I}(\theta)\,(g(\mathbf{x}) - \theta)\,,$$

furthermore the covariance of the estimator will be $\mathbf{I}^{-1}(\theta)$.

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

for some function $\mathbf{g}$. Furthermore, the covariance matrix of $\hat{\boldsymbol{\theta}}$ will be $\mathbf{I}^{-1}(\boldsymbol{\theta})$. determine if this condition is satisfied for the linear model of (4.1), we have

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\ln(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right] \\
&= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right].
\end{aligned}$$

Using the identities

$$\begin{aligned}
\frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= \mathbf{b} \\
\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= 2\mathbf{A}\boldsymbol{\theta}
\end{aligned}$$

for $\mathbf{A}$ a symmetric matrix, we have

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2}[\mathbf{H}^T\mathbf{x} - \mathbf{H}^T\mathbf{H}\boldsymbol{\theta}].$$

Assuming that $\mathbf{H}^T\mathbf{H}$ is invertible

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2}[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \boldsymbol{\theta}]$$

which is exactly in the form of (4.2) with

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \\
\mathbf{I}(\boldsymbol{\theta}) &= \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2}.
\end{aligned}$$

Hence, the MVU estimator of $\boldsymbol{\theta}$ is given by (4.5), and its covariance matrix is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}.$$

MVUE for the linear model

If the data follows the linear model specified above (notice the condition on the rank of the matrix $\mathbf{H}$!) then the MVUE is given by

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

and the covariance matrix of $\hat{\theta}$ is

$$\mathbf{C}_{\hat{\theta}} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}.$$

In fact, we have a full statistical description of the estimator $\hat{\theta}$ as we know its pdf, which is

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right)$$

Examples -- Curve fitting

Consider the following polynomial curve fitting problem, where we wish to find $\theta_1, \theta_2, \cdots, \theta_p$ so as to best fit the experimental data points $(t_n, x(t_n))$ for $n = 0, 1, \cdots, N-1$ by the polynomial curve

$$x(t_n) = \theta_1 + \theta_2 t_n^1 + \theta_3 t_n^2 + \cdots + \theta_p t_n^{p-1} + w(t_n)$$

for $w(t_n)$ are WGN samples. Find the MVUE and its statistics when estimating $\theta_1, \theta_2, \cdots, \theta_p$.

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

$$\begin{aligned} \mathbf{x} &= [x(t_0)\, x(t_1) \ldots x(t_{N-1})]^T \\ \boldsymbol{\theta} &= [\theta_1\, \theta_2\, \theta_3]^T \end{aligned}$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}. \qquad \hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

General linear model

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

$\mathbf{x}$: $N \times 1$ observation vector
$\mathbf{H}$: $N \times p$ known observation matrix with $N > p$ and rank p
θ : $p \times 1$ vector of unknown parameters to be estimated
$\mathbf{w}$: $N \times 1$ noise vector, $\sim \mathcal{N}(\mathbf{0}, \mathbf{C_w})$

Re-do the derivation

Key idea: *whiten* the observations, then apply previous results

MVUE for general linear model

Under the general linear model, by "whitening" ($\mathbf{x}' = \mathbf{D}\mathbf{x}, \mathbf{H}' = \mathbf{D}\mathbf{H}, \mathbf{w}' = \mathbf{D}\mathbf{w}$ where $\mathbf{C}^{-1} = \mathbf{D}^T\mathbf{D}$ for $\mathbf{D}$ invertible) we see that the MVUE is given by

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

and the covariance matrix of $\hat{\theta}$ is

$$\mathbf{C}_{\hat{\theta}} = \sigma^2\left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}.$$

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\mathbf{H}'^T\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{x}' \\ &= (\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{x}\end{aligned}$$

$$\begin{aligned}\mathbf{x}' &= \mathbf{D}\mathbf{x} \\ &= \mathbf{D}\mathbf{H}\boldsymbol{\theta} + \mathbf{D}\mathbf{w} \\ &= \mathbf{H}'\boldsymbol{\theta} + \mathbf{w}'\end{aligned}$$

$$\begin{aligned}E\left[(\mathbf{D}\mathbf{w})(\mathbf{D}\mathbf{w})^T\right] &= \mathbf{D}\mathbf{C}\mathbf{D}^T \\ &= \mathbf{D}\mathbf{D}^{-1}\mathbf{D}^{T^{-1}}\mathbf{D}^T = \mathbf{I}\end{aligned}$$

Examples

Consider $x[n] = A + w[n]$, for $n = 0, 1, \cdots, N-1$ where $w[n]$ is colored Gaussian noise with covariance matrix $\mathbf{C}$ and A is to be estimated. Formulate the general linear model and determine the MVUE and its statistics when estimating A.

$$\begin{aligned} \hat{A} &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \\ &= \frac{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} \end{aligned} \qquad \begin{aligned} \operatorname{var}(\hat{A}) &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1} \\ &= \frac{1}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}. \end{aligned}$$

What if $\mathbf{x} = \mathbf{H}\theta + \mathbf{s} + \mathbf{w}$, where $\mathbf{s}$ is a known deterministic signal?

To determine the MVU estimator let $\mathbf{x}' = \mathbf{x} - \mathbf{s}$, so that

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

which is now in the form of the linear model. The MVU estimator follows as

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

with covariance

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}.$$

Theorem 4.2 (Minimum Variance Unbiased Estimator for General Linear Model) *If the data can be modeled as*

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w} \tag{4.30}$$

where $\mathbf{x}$ *is an* $N \times 1$ *vector of observations,* $\mathbf{H}$ *is a known* $N \times p$ *observation matrix* $(N > p)$ *of rank* p*,* $\boldsymbol{\theta}$ *is a* $p \times 1$ *vector of parameters to be estimated,* $\mathbf{s}$ *is an* $N \times 1$ *vector of known signal samples, and* $\mathbf{w}$ *is an* $N \times 1$ *noise vector with PDF* $\mathcal{N}(\mathbf{0}, \mathbf{C})$*, then the MVU estimator is*

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{s}) \tag{4.31}$$

and the covariance matrix is

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}. \tag{4.32}$$

For the general linear model the MVU estimator is efficient in that it attains the CRLB.

This theorem is quite powerful in practice since many signal processing problems can be modeled by (4.30).

Examples

Consider $x[n] = A + r^n + w[n]$ for $n = 0, 1, \cdots, N-1$ where $w[n]$ is WGN, r is known, and A is to be estimated. Find the MVUE and its statistics when estimating A.

$$\mathbf{x} = A \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \mathbf{s} + \mathbf{w} \qquad \text{where } \mathbf{s} = [1\, r \ldots r^{N-1}]^T.$$

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1}(x[n] - r^n) \qquad \operatorname{var}(\hat{A}) = \frac{\sigma^2}{N}.$$