Lecture 6: Minimum Variance Unbiased Estimators

(LaTeX prepared by Ben Vondersaar)
April 27, 2015

This lecture note is based on ECE 645(Spring 2015) by Prof. Stanley H. Chan in the School of Electrical and Computer Engineering at Purdue University.

1 Introduction

Suppose that we observe a random variable $\boldsymbol{Y}$ with a density $f_{\boldsymbol{Y}}(\boldsymbol{y};\theta)$ where θ is a deterministic but unknown parameter. Our goal is to estimate θ using one or more observations of $\boldsymbol{Y}$. To understand how the estimation is solved, we should first consider the mean-squared error:

$$\begin{aligned}
\text{MSE} &\overset{\text{def}}{=} \mathbb{E}_{\boldsymbol{Y}}[(\widehat{\theta}(\boldsymbol{Y})-\theta)^2] \\
&= \mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})^2 - 2\widehat{\theta}(\boldsymbol{Y})\theta + \theta^2] \\
&= \left[\mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})^2] - \mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})]^2\right] + \left[\mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})]^2 - 2\widehat{\theta}(\boldsymbol{Y})\theta + \theta^2\right] \\
&= \text{Var}_{\boldsymbol{Y}}(\widehat{\theta}(\boldsymbol{Y})) + \left(\mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})] - \theta\right)^2 \\
&= \text{Var}_{\boldsymbol{Y}}(\widehat{\theta}(\boldsymbol{Y})) + (\text{Bias}(\theta))^2.
\end{aligned}$$

The problem now simplifies to minimizing the variance of $\widehat{\theta}$ over all values of $\boldsymbol{Y}$, and minimizing the newly defined *bias*.

2 Unbiased Estimator

As shown in the breakdown of MSE, the bias of an estimator is defined as

$$b(\widehat{\theta}) = \mathbb{E}_{\boldsymbol{Y}}[\widehat{\theta}(\boldsymbol{Y})] - \theta. \tag{1}$$

An estimator is said to be *unbiased* if $b(\widehat{\theta}) = 0$. If multiple unbiased estimates of θ are available, and the estimators can be averaged to reduce the variance, leading to the true parameter θ as more observations are available.

Placing the unbiased restriction on the estimator simplifies the MSE minimization to depend only on its variance. The resulting estimator, called the **Minimum Variance Unbiased Estimator** (MVUE), have the smallest variance of all possible estimators over all possible values of θ, i.e.,

$$\text{Var}_{\boldsymbol{Y}}[\widehat{\theta}_{MVUE}(\boldsymbol{Y})] \leq \text{Var}_{\boldsymbol{Y}}[\widetilde{\theta}(\boldsymbol{Y})], \tag{2}$$

for all estimators $\widetilde{\theta}(\boldsymbol{Y}) \in \Lambda$ and all parameters $\theta \in \Lambda$.

It is important to note that a uniformly minimum variance unbiased estimator may not always exist, and even if it does, we may not be able to find it. There is not a single method that will always produce the MVUE. One useful approach to finding the MVUE begins by finding a sufficient statistic for the parameter.

3 Sufficient Statistic

Definition 1. Sufficient Statistic
A sufficient statistic for a parameter θ is a function $T : \Gamma \longrightarrow \mathbb{R}$ such that the conditional distribution

$$f_{\boldsymbol{Y}|T(\boldsymbol{Y})}(\boldsymbol{y} \mid t) = \frac{f_{\boldsymbol{Y},T(\boldsymbol{Y})}(\boldsymbol{y}, t\,;\,\theta)}{f_{T(\boldsymbol{Y})}(t,\ \theta)}, \tag{3}$$

is independent of θ, for all $\theta \in \Lambda$, where $t = T(\boldsymbol{y})$.

This definition says that if $T(\boldsymbol{Y})$ is a sufficient statistic, then $T(\boldsymbol{Y})$ should contain all the information needed to estimate θ. Or in other words, $T(\boldsymbol{Y})$ is sufficient if

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \mid T(\boldsymbol{Y}) = T(\boldsymbol{y}), \theta) = \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \mid T(\boldsymbol{Y}) = T(\boldsymbol{y})), \tag{4}$$

i.e., if we know $T(\boldsymbol{Y})$, then there is no need to know θ.

The following Theorem provides a necessary and sufficient condition for having a sufficient statistic.

Theorem 1. Factorization Theorem
A statistic $T(\boldsymbol{Y})$ is sufficient if and only if there exists two functions $g_\theta(T(\boldsymbol{y}))$ and $h(\boldsymbol{y})$, such that

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) = g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y}), \tag{5}$$

for all$\boldsymbol{y} \in \Gamma$ and $\theta \in \Lambda$.

Proof.

Suppose that $T(\boldsymbol{Y})$ is a sufficient statistic for θ. Then, by definition, $f_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}(\boldsymbol{y}\,|\,t;\theta)$ is independent of θ, where $t = T(\boldsymbol{y})$. Applying the conditional probability:

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta) &\overset{(a)}{=} f_{\boldsymbol{Y},T(\boldsymbol{Y})}(\boldsymbol{y},\,t\,;\,\theta) \\ &= \underbrace{f_{\boldsymbol{Y}|T(\boldsymbol{Y})}(\boldsymbol{y}\,|\,t;\theta)}_{h(\boldsymbol{y})}\underbrace{f_{T(\boldsymbol{Y})}(t\,;\,\theta)}_{g_\theta(t)}, \end{aligned}$$

where (a) holds because $T(\boldsymbol{Y})$ is a function of $\boldsymbol{Y}$. Thus we have determined the two function $g_\theta(t)$ and $h(\boldsymbol{y})$.

Conversely, suppose that $f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta) = g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y})$. Then, since

$$f_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}(\boldsymbol{y}\,|\,t;\theta) = \frac{f_{\boldsymbol{Y},T(\boldsymbol{Y})}(\boldsymbol{y},\,t;\theta)}{f_{T(\boldsymbol{Y})}(t\,;\,\theta)}$$

and since $T(\boldsymbol{Y})$ is a function of $\boldsymbol{Y}$, we have

$$f_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}(\boldsymbol{y}\,|\,t\,;\,\theta) = \begin{cases} \frac{f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta)}{f_{T(\boldsymbol{Y})}(t;\theta)}, & \text{if} \quad T(\boldsymbol{y}) = t, \\ 0, & \text{otherwise.} \end{cases}$$

Note that

$$f_{T(\boldsymbol{Y})}(t\,;\,\theta) = \int_{T(\boldsymbol{y})=t} f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta)d\boldsymbol{y}.$$

Therefore,

$$\begin{aligned} f_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}(\boldsymbol{y}\,|\,t;\theta) &= \begin{cases} \frac{g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y})}{\int_{T(\boldsymbol{y})=t} f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta)d\boldsymbol{y}}, & \text{if} \quad T(\boldsymbol{y}) = t, \\ 0, & \text{otherwise} \end{cases} \\ &= \frac{g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y})}{\int_{T(\boldsymbol{y})=t} g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y})d\boldsymbol{y}} \cdot \mathbf{1}\left(T(\boldsymbol{y}) = t\right) \\ &= \frac{h(\boldsymbol{y})}{\int_{T(\boldsymbol{y})=t} h(\boldsymbol{y})d\boldsymbol{y}} \cdot \mathbf{1}\left(T(\boldsymbol{y}) = t\right), \end{aligned}$$

which is independent of θ.

Example: DC Level in White Gaussian Noise

We observe a random signal $\boldsymbol{Y}$ such that

$$\boldsymbol{Y} = \theta + \boldsymbol{V},$$

where $\boldsymbol{V} \sim N(0, \sigma^2 \boldsymbol{I})$, and θ is an unknown but deterministic number.

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) &= (2\pi\sigma)^{-n/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right] \\ &= \underbrace{(2\pi\sigma^2)^{-n/2} \exp\left[\frac{-1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum_{i=1}^{n} y_i\right)\right]}_{g_\theta(T(\boldsymbol{y}))} \underbrace{\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right]}_{h(\boldsymbol{y})}. \end{aligned}$$

Thus, by the factorization theorem, $T(\boldsymbol{y}) = \sum_{i=1}^{n} y_i$, is a sufficient statistic.

4 Rao-Blackwell Theorem

Theorem 2. Rao-Blackwell
Suppose that $\widehat{g}(\boldsymbol{y})$ is an unbiased estimate of $g(\theta)$ and that $T(\boldsymbol{y})$ is a sufficient statistic for θ. Then:

1. $\widetilde{g}\,(T(\boldsymbol{y})) \stackrel{\text{def}}{=} \mathbb{E}_{\boldsymbol{Y}|T(\boldsymbol{Y})}[\widehat{g}(\boldsymbol{Y})|T(\boldsymbol{Y}) = T(\boldsymbol{y})]$ is also an unbiased estimator for $g(\theta)$;
2. $\mathrm{Var}_{T(\boldsymbol{Y})}[\widetilde{g}\,(T(\boldsymbol{Y}))] \le \mathrm{Var}_{\boldsymbol{Y}}[\widetilde{g}(\boldsymbol{Y})]$ with equality if and only if $\mathbb{P}(\widetilde{g}(T(\boldsymbol{Y})) = \widetilde{g}(\boldsymbol{Y})) = 1$.

The Rao-Blackwell Theorem can be seen as a procedure to "improve" any unbiased estimator.
Proof.
(1) To show $\widetilde{g}\,(T(\boldsymbol{y}))$ is an unbiased estimator, we note that

$$\begin{aligned} \mathbb{E}_{T(\boldsymbol{Y})}\left[\widetilde{g}\,(T(\boldsymbol{Y}))\right] &\stackrel{(a)}{=} \mathbb{E}_{T(\boldsymbol{Y})}\left[\mathbb{E}_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}\left[\widehat{g}(\boldsymbol{Y})|T(\boldsymbol{Y}) = T(\boldsymbol{y})\right]\right] \\ &\stackrel{(b)}{=} \mathbb{E}_{\boldsymbol{Y}}\left[\widehat{g}(\boldsymbol{Y})\right] \\ &= g(\theta), \end{aligned}$$

where (a) holds because $\widetilde{g}\,(T(\boldsymbol{y})) \stackrel{\text{def}}{=} \mathbb{E}_{\boldsymbol{Y}|T(\boldsymbol{Y})}[\widehat{g}(\boldsymbol{Y})|T(\boldsymbol{Y}) = T(\boldsymbol{y})]$, and (b) holds because of the iterated expectation.

(2): First, by using the definition of variance we note that

$$\mathrm{Var}_{T(\boldsymbol{Y})}[\widetilde{g}\,(T(\boldsymbol{Y}))] = \mathbb{E}_{T(\boldsymbol{Y})}\left[\widetilde{g}\,(T(\boldsymbol{Y}))^2\right] - g(\theta)^2,$$

and

$$\mathrm{Var}_{\boldsymbol{Y}}[\widehat{g}\,(T(\boldsymbol{Y}))] = \mathbb{E}_{\boldsymbol{Y}}\left[\widehat{g}(\boldsymbol{Y})^2\right] - g(\theta)^2.$$

Consequently, we can show that

$$\begin{aligned} \mathbb{E}_{T(\boldsymbol{Y})}\left[\widetilde{g}\,(T(\boldsymbol{Y}))^2\right] &= \mathbb{E}_{T(\boldsymbol{Y})}\left[\mathbb{E}_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}[\widehat{g}(\boldsymbol{Y})]^2\right] \\ &\stackrel{(c)}{\le} \mathbb{E}_{T(\boldsymbol{Y})}\left[\mathbb{E}_{\boldsymbol{Y}\,|\,T(\boldsymbol{Y})}[\widehat{g}(\boldsymbol{Y})^2]\right] \\ &\stackrel{(d)}{=} \mathbb{E}_{\boldsymbol{Y}}\left[\widehat{g}(\boldsymbol{Y})^2\right], \end{aligned}$$

where (c) holds because of Jensen's inequality and (d) is the result of the iterated expectation. Equality holds when Jensen's equality holds, i.e., $\mathbb{P}(\widetilde{g}(T(\boldsymbol{Y})) = \widetilde{g}(\boldsymbol{Y})) = 1$.

□

Implication of Rao-Blackwell:

1. With a sufficient statistic, we can improve *any* unbiased estimator that is not already a function of T by conditioning on $T(\boldsymbol{Y})$

2. If T is sufficient for θ, and if there is only one function of T that is an unbiased estimator of $g(\theta)$ (i.e., $\widehat{g}(\boldsymbol{Y})$) then the function must be MVUE. To see this, we suppose that $g^*(T(\boldsymbol{Y}))$ is the only function of $T(\boldsymbol{Y})$ such that
$$\mathbb{E}_{T(\boldsymbol{Y})}[g^*T((\boldsymbol{Y}))] = g(\theta).$$
Let $\widehat{g}(\boldsymbol{Y})$ be any unbiased estimator. Then, Rao-Blackwell says
$$\widetilde{g}(T(\boldsymbol{y})) \overset{\text{def}}{=} \mathbb{E}_{\boldsymbol{Y}}[\widehat{g}(\boldsymbol{Y}) \,|\, T(\boldsymbol{Y}) = T(\boldsymbol{y})]$$
is also unbiased and has a variance
$$\text{Var}_{T(\boldsymbol{Y})}[\widetilde{g}\,(T(\boldsymbol{Y}))] \leq \text{Var}_{\boldsymbol{Y}}[\widetilde{g}(\boldsymbol{Y})].$$
But since $\widehat{g}$ is arbitrary, we have
$$\text{Var}_{T(\boldsymbol{Y})}[g^*(T(\boldsymbol{Y}))] \leq \text{Var}_{T(\boldsymbol{Y})}[\widetilde{g}(T(\boldsymbol{Y}))].$$
So $g^*(T(\boldsymbol{Y}))$ is an MVUE.

5 Complete Sufficient Statistic

Definition 2. A
family of distributions $\{\mathbb{P}_\theta : \theta \in \Lambda\}$ is *complete* if the condition
$$\mathbb{E}_{\boldsymbol{Y}}[\varphi(\boldsymbol{Y})] = 0 \tag{6}$$
implies that
$$\mathbb{P}_\theta[\varphi(\boldsymbol{Y}) = 0] = 1, \tag{7}$$
for any function φ and any $\theta \in \Lambda$.

Interpretation: We can consider the above definition of completeness in the discrete domain. For example, let $\boldsymbol{\varphi}$ and $\boldsymbol{f} \geq 0$ be two vectors. Then, $\boldsymbol{\varphi}^T\boldsymbol{f} = 0$ implies that $\boldsymbol{\varphi} = 0$ because $\boldsymbol{f} = 0$. Now, if we imagine that the inner product is defined through expectation, then $\mathbb{E}_{\boldsymbol{Y}}[\varphi(\boldsymbol{Y})] = \int \phi(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y},\, \theta)\boldsymbol{y} \approx \boldsymbol{\varphi}^T\boldsymbol{f}_{\boldsymbol{Y}}$.

Example 1. COMPLETENESS OF BINOMIAL
Let $f_Y(y\,;\,\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$. Then,
$$\begin{aligned}
\mathbb{E}_Y[\varphi(Y)] &= \sum_{y=0}^{n} \binom{n}{y}\varphi(y)\theta^y(1-\theta)^{n-y} \\
&= (1-\theta)^n \sum_{y=0}^{n} \underbrace{\binom{n}{y}\varphi(y)}_{\overset{\text{def}}{=} a_y} \underbrace{\left(\frac{\theta}{1-\theta}\right)^y}_{\overset{\text{def}}{=} x^y}.
\end{aligned}$$
So, $\boldsymbol{a}^T\boldsymbol{x} = 0$ implies that $\boldsymbol{a} = 0$ because $\boldsymbol{x} > 0$. So the family of binomial distributions is complete.

Definition 3. COMPLETE SUFFICIENT STATISTIC
Given $Y \sim \mathbb{P}_\theta$. Define a sufficient statistic $T(Y)$ for θ. If the distribution of $T(Y)$, denoted by $\mathbb{Q}_\theta$, is complete, then T is said to be a *complete sufficient statistic.*

We now prove two properties of complete sufficient statistic.

Proposition 1. UNIQUENESS
Let $T(\boldsymbol{Y})$ be a a complete sufficient statistic, and $\widetilde{g}(T(\boldsymbol{Y}))$ and $\widehat{g}(T(\boldsymbol{Y}))$ are two unbiased estimators. Then,

$$\mathbb{P}\left[\widetilde{g}(T(\boldsymbol{Y})) = \widehat{g}(T(\boldsymbol{Y}))\right] = 1. \tag{8}$$

Proof.

$$\begin{aligned}\mathbb{E}_{T(\boldsymbol{Y})}[\widetilde{g}(T(\boldsymbol{Y})) - \widehat{g}(T(\boldsymbol{Y}))] &= \mathbb{E}_{T(\boldsymbol{Y})}[\widetilde{g}(T(\boldsymbol{Y}))] - \mathbb{E}_{T(\boldsymbol{Y})}[\widehat{g}(T(\boldsymbol{Y}))] \\ &= g(\theta) - g(\theta) = 0,\end{aligned}$$

which implies by the definition of completeness that $\mathbb{P}\left[\widetilde{g}(T(Y)) = \widehat{g}(T(Y))\right] = 1$.

□

Proposition 2. MVUE
Let $T(\boldsymbol{Y})$ be a complete sufficient statistic. If $\widetilde{g}(T(\boldsymbol{Y}))$ is an unbiased estimator, then $\widetilde{g}(T(\boldsymbol{Y}))$ is an MVUE.

Proof.
By Rao-Blackwell, if $\widehat{g}(\boldsymbol{Y})$ is an unbiased estimator, we can always find another estimator

$$\widetilde{g}(T(\boldsymbol{Y})) = \mathbb{E}_{\boldsymbol{Y} \mid T(\boldsymbol{Y})}\left[\widehat{g}(\boldsymbol{Y})\right]. \tag{9}$$

Since $T(\boldsymbol{Y})$ is complete, $\widetilde{g}(T(\boldsymbol{Y}))$ is unique. So it must be MVUE.

□

A General Procedure to obtain MVUE

Approach 1:

1. Find a complete sufficient statistic $T(\boldsymbol{Y})$.
2. Find an unbiased estimator, $\widehat{g}(\boldsymbol{Y})$.
3. Use Rao-Blackwell Theorem to define $\widetilde{g}(\boldsymbol{Y}) = \mathbb{E}_{\boldsymbol{Y} \mid T(\boldsymbol{Y})}[\widehat{g}(\boldsymbol{Y})]$.

Then, $\widetilde{g}(T(\boldsymbol{Y}))$ is MVUE.

Approach 2:

1. Find a complete sufficient statistic $T(\boldsymbol{Y})$.
2. Find an estimator that only depends on $T(\boldsymbol{Y})$ and not $\boldsymbol{Y}$, $\widetilde{g}(T(\boldsymbol{Y}))$.
3. Show that $\widetilde{g}(T(\boldsymbol{Y}))$ is unbiased.

Then, $\widetilde{g}(T(\boldsymbol{Y}))$ is MVUE.

6 Exponential Families

Finding the complete sufficient statistic could be a challenging task for arbitrary distributions. However, for many distributions including Gaussian, Poisson, Laplacian, binomial, geometric and certain multivariate forms, finding the complete sufficient statistic could be manageable. These distributions belongs to a family called the **exponential family** which we now discuss.

Definition 4. Exponential Family

A family of distributions $\{\mathbb{P}_\theta; \theta \in \Lambda\}$ is said to be an exponential family if its density can be placed in the form

$$f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta) = C(\theta)\exp\left\{\sum_{l=1}^{m} Q_l(\theta)T_l(\boldsymbol{y})\right\}h(\theta), \tag{10}$$

where $C, Q_1, ..., Q_m, T_1, ..., T_m$, and h are real-valued functions.

Theorem 3. Completeness of Exponential Family

If $\Gamma = \mathbb{R}^n, \Lambda = \mathbb{R}^m$, $C, T_1, ..., T_m$, and h are real-valued functions, and $\mathbb{P}_\theta$ has a density of the form

$$f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta) = C(\theta)\exp\left\{\sum_{l=1}^{m} \theta_l T_l(\boldsymbol{y})\right\}h(\theta). \tag{11}$$

Then $T(\boldsymbol{Y}) = [T_1(\boldsymbol{Y}), ..., T_m(\boldsymbol{Y})]$ is a complete sufficient statistic for $\{\mathbb{P}_\theta; \theta \in \Lambda\}$ if Λ contains an m-dimensional rectangle.

Proof.

See Poor IV.C.3.

□

Example 2. Discrete Signal Gain in Gaussian Noise

Let $\boldsymbol{Y} = [Y_1, ..., Y_n]^T$ be an incoming discrete signal consisting of a known sequence with an unknown, deterministic gain plus white Gaussian noise such that $\boldsymbol{Y} = \theta\boldsymbol{s} + \boldsymbol{V}$ where $\boldsymbol{V} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$, $\boldsymbol{s} = [s_1, ..., s_n]^T$, and $\theta \in \mathbb{R}$. Find the MVUE of θ.

Step 1: Show that the conditional distribution $f_\theta(\boldsymbol{Y})$ belongs to the exponential family and identify the complete sufficient statistic.

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}\,;\,\theta) &= (2\pi\sigma^2)^{-n/2}\exp\left\{\frac{-1}{2\sigma^2}\|\boldsymbol{Y} - \theta\boldsymbol{s}\|^2\right\} \\ &= \underbrace{(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{\theta^2\|\boldsymbol{s}\|^2}{2\sigma^2}\right\}}_{C(\theta)}\exp\left\{\underbrace{\frac{\theta}{\sigma^2}}_{Q(\theta)}\underbrace{\boldsymbol{Y}^T\boldsymbol{s}}_{T(\boldsymbol{Y})}\right\}\underbrace{\exp\left\{\frac{-1}{2\sigma^2}\|\boldsymbol{Y}\|^2\right\}}_{h(\boldsymbol{Y})}. \end{aligned}$$

Thus, $T(\boldsymbol{Y}) = \boldsymbol{Y}^T\boldsymbol{s}$. Since $\theta \in \mathbb{R}$ and $\mathbb{R}$ contains a rectangle, $T(\boldsymbol{Y})$ is a complete sufficient statistic.

Step 2: Find an unbiased estimator. Consider $\mathbb{E}[T(\boldsymbol{Y})]$:

$$\begin{aligned} \mathbb{E}[T(\boldsymbol{Y})] &= \mathbb{E}[\boldsymbol{Y}^T\boldsymbol{s}] \\ &= \mathbb{E}[\theta\boldsymbol{s}^T\boldsymbol{s}] + \mathbb{E}[\boldsymbol{V}^T\boldsymbol{s}] \\ &= \theta\|\boldsymbol{s}\|^2. \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\frac{T(\boldsymbol{Y})}{\|\boldsymbol{s}\|^2}\right] = \theta,$$

and if we define $g(T(\boldsymbol{Y})) = \frac{T(\boldsymbol{Y})}{\|\boldsymbol{s}\|^2}$, then $g(T(\boldsymbol{Y}))$ is an unbiased estimator.

Therefore, we conclude that the estimator $g(T(\boldsymbol{Y})) = \frac{T(\boldsymbol{Y})}{\|\boldsymbol{s}\|^2}$ is MVUE.

Below is a numerical example to illustrate MVUE in MATLAB. We consider $\boldsymbol{s}$ as a two-period sine wave of frequency 100Hz. We let $\sigma^2 = 2$. We evaluate the MVUE for $\theta = 1, 5, 10, 50$.

```
f_s = 4000;
f = 100;
t = [0:1/f_s:.02];
s = sin(2*pi*f*t)’;
V = sqrt(2)*randn(size(s));
theta = [1,5,10,50];
for i = 1:4
    Y = theta(i) * s + V;
    theta_MVUE = Y’*s / (s’ * s);
    subplot(2,2,i)
    plot(t,Y)
    title([’\theta = ’ num2str(theta(i)) ’ , \theta_{MVUE} = ’ num2str(theta_MVUE)])
end
```