Estimation: chapter 5

General Minimum Variance Unbiased Estimation

Lifang Feng
lffeng@ustb.edu.cn

北京科技大學
University of Science and Technology Beijing

Spring 2022

Finding MVUE so far

When the observations and the data are related in a linear way $\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$ and the noise was Gaussian, then the MVUE was easy to find:

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

with covariance matrix

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

Using the CRLB, if you got lucky then you could write

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

where $\hat{\theta} = g(\mathbf{x})$ would be your MVUE, which would also be an efficient estimator (meeting the CRLB).

Even if no efficient estimator exists, a MVUE may exist. In this section we try to find a systematic way of determining the MVUE if it exists.

Finding the MVUE

We wish to estimate the parameter(s) θ from the observations $\mathbf{x}$.

1. Determine (if possible) a *sufficient statistic* $T(\mathbf{x})$ for the parameter to be estimated θ. This may be done using the Neyman-Fisher factorization theorem.

2. Determine whether the sufficient statistic is also *complete.* This is generally hard to do. If it is not complete, we can saying nothing more about the MVUE. If it is, continue to step 3.

3. Find the MVUE $\hat{\theta}$ from $T(\mathbf{x})$ is one of two ways using the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem:

   - Find a function $g(\cdot)$ of the sufficient statistic that yields an unbiased estimator $\hat{\theta} = g(T(\mathbf{x}))$, the MVUE! By definition of completeness of the statistic, this will yield the MVUE.
   - Find any unbiased estimator $\breve{\theta}$ for θ, and then determine $\hat{\theta} = E[\breve{\theta}|T(\mathbf{x})]$. This is usually very tedious/difficult to do. The expectation is taken over the distribution $p(\breve{\theta}|T(\mathbf{x}))$.

Sufficient statistics

A *statistic* is the result of applying a function to a set of data - our observations. A single statistic is a single function of the observations, say $T(\mathbf{x})$ and it is called a *sufficient statistic* if the PDF $p(\mathbf{x}|T(\mathbf{x}) = T_0; \theta)$ is independent of θ.

Smaller, easier to keep, deal with!

While you can go and determine whether a statistic is sufficient using that definition, an easier way may be to use the following theorem.

Theorem: Neyman-Fisher Factorization Theorem: If we can factor the PDF $p(\mathbf{x}; \theta)$ as

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}); \theta)h(\mathbf{x})$$

where $g(\cdot)$ is a function depending only on the $\mathbf{x}$ through $T(\mathbf{x})$ and $h(\cdot)$ is a function depending only on $\mathbf{x}$, then $T(\mathbf{x})$ is a sufficient statistic for θ. Conversely, if $T(\mathbf{x})$ is a sufficient statistic for θ then the PDF can be factored as above.

Example of sufficient statistics

Consider our classical problem of estimation the DC level of a signal in WGN:, i.e.

$$x[n] = A + w[n], \ \ n = 0, 1, \cdots, N-1, \ \ w[n] \text{ i.i.d. } WGN \text{ of variance } \sigma^2.$$

What is the MVUE? What are some examples of sufficient statistics?

$$p(\mathbf{x}; A) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(NA^2 - 2A\sum_{n=0}^{N-1}x[n]\right)\right]}_{g(T(\mathbf{x}),A)} \underbrace{\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]}_{h(\mathbf{x})}$$

Example of sufficient statistics

Consider estimating the noise variance σ^2 from N samples

$$x[n] = w[n], \ \ n = 0, 1, \cdots, N-1$$

Find a sufficient statistic for σ^2.

$$p(\mathbf{x}, \sigma^2) = \underbrace{\frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]}_{g\left(T(\mathbf{x}),\sigma^2\right)} \times \underbrace{1}_{h(\mathbf{x})}$$

Naturally extended Neyman-Fisher factorization

There exists a natural extension to the Neyman-Fisher factorization theorem for multiple statistics. The r statistics $T_1(\mathbf{x}), T_2(\mathbf{x}), \cdots, T_r(\mathbf{x})$ are joint sufficient statistics if the conditional pdf $p(\mathbf{x}|T_1(\mathbf{x}), \cdots, T_r(\mathbf{x}); \theta)$ does NOT depend on θ. They are joint sufficient statistics if and only if the pdf may be factored as

$$p(\mathbf{x}; \theta) = g(T_1(\mathbf{x}), T_2(\mathbf{x}), \cdots, T_r(\mathbf{x}); \theta) h(\mathbf{x}).$$

Example of sufficient statistics

Consider estimating the phase ϕ of a sinusoid from the samples

$x[n] = A\cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, 1, 2 \cdots N-1,$ where σ^2, A, f_0 are known.

Find sufficient statistic(s) for estimating ϕ.

$$p(\mathbf{x}; \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x[n] - A\cos(2\pi f_0 n + \phi)]^2 \right\}$$

$$\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos(2\pi f_0 n + \phi) + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi)$$

$$\begin{aligned} = \; & \sum_{n=0}^{N-1} x^2[n] - 2A \left(\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right) \cos\phi \\ & + 2A \left(\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \right) \sin\phi + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi). \end{aligned}$$

Rao-Blackwell-Lehmann-Scheffe (RBLS) Theorem

Finding a sufficient statistic may be difficult determining whether it is complete even more difficult. In practice many sufficient statistics we will deal with are complete. We are now ready for the RBLS theorem.

The Rao-Blackwell-Lehmann-Scheffe (RBLS) Theorem: If $\breve{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ then $\hat{\theta} = E[\breve{\theta}|T(\mathbf{x})]$ is

1. a valid estimator for θ (i.e. it does not depend on θ)
2. unbiased
3. of lesser of equal variance than that of $\breve{\theta}$ for all θ.

if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

Sufficient statistics and MVUEs

If we have determined a sufficient statistic $T(\mathbf{x})$ for θ then we can use this to improve *any* unbiased estimator of θ, as is proven in the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem. If we're lucky and the statistic is also *complete*, we can use it to find the MVUE!

There are many definitions for the completeness of a statistic. One that is easy for us to use in the context of estimation is the following: a statistic is called *complete* if only 1 function of it yields an unbiased estimator of θ. This is generally difficult to check but is easy conceptually.

$$v(T) = g(T) - h(T)$$

Another way of checking if a statistic is complete is the following: A statistic T is complete if $\int_{-\infty}^{\infty} v(T)p(T;\theta)dT = 0$ for all θ is only satisfied by $v(T) = 0$ for all T.

Example 5.5 - DC Level in WGN

We will continue Example 5.2. Although we already know that $\hat{A} = \bar{x}$ is the MVU estimator (since it is efficient), we will use the RBLS theorem, which can be used even when an efficient estimator does not exist and hence the CRLB method is no longer viable. The procedure for finding the MVU estimator $\hat{A}$ may be implemented in two different ways. They are both based on the sufficient statistic $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$.

1. Find *any* unbiased estimator of A, say $\check{A} = x[0]$, and determine $\hat{A} = E(\check{A}|T)$. The expectation is taken with respect to $p(\check{A}|T)$.

2. Find some function g so that $\hat{A} = g(T)$ is an unbiased estimator of A.

Example 5.6 - Completeness of a Sufficient Statistic

For the estimation of A, the sufficient statistic $\sum_{n=0}^{N-1} x[n]$ is complete or there is but one function g for which $E[g(\sum_{n=0}^{N-1} x[n])] = A$. Suppose, however, that there exists a second function h for which $E[h(\sum_{n=0}^{N-1} x[n])] = A$. Then, it would follow that with $T = \sum_{n=0}^{N-1} x[n]$,

$$E\left[g(T) - h(T)\right] = A - A = 0 \qquad \text{for all } A$$

or since $T \sim \mathcal{N}(NA, N\sigma^2)$

$$\int_{-\infty}^{\infty} v(T) \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{1}{2N\sigma^2}(T - NA)^2\right] dT = 0 \qquad \text{for all } A$$

where $v(T) = g(T) - h(T)$. Letting $\tau = T/N$ and $v'(\tau) = v(N\tau)$, we have

$$\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{N}{2\sigma^2}(A - \tau)^2\right] d\tau = 0 \qquad \text{for all } A \tag{5.7}$$

which may be recognized as the convolution of a function $v'(\tau)$ with a Gaussian pulse $w(\tau)$ (see Figure 5.3). For the result to be zero for all A, $v'(\tau)$ must be identically zero.

Example of an incomplete statistic

Consider estimating A from

$$x[0] = A + w[0], \quad w[0] \sim \mathcal{U}[-1/2, 1/2].$$

Is $x[0]$ a sufficient statistic? Is it complete?

example, we suppose that there exists another function h with the unbiased property $h(x[0]) = A$ and attempt to prove that $h = g$. Again letting $v(T) = g(T) - h(T)$, we examine the possible solutions for v of the equation

$$\int_{-\infty}^{\infty} v(T)p(\mathbf{x}; A)\, d\mathbf{x} = 0 \qquad \text{for all } A.$$

For this problem, however, $\mathbf{x} = x[0] = T$, so that

$$\int_{-\infty}^{\infty} v(T)p(T; A)\, dT = 0 \qquad \text{for all A.}$$

But

$$p(T; A) = \begin{cases} 1 & A - \frac{1}{2} \leq T \leq A + \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Finding the MVUE

We wish to estimate the parameter(s) θ from the observations $\mathbf{x}$.

1. Determine (if possible) a *sufficient statistic* $T(\mathbf{x})$ for the parameter to be estimated θ. This may be done using the Neyman-Fisher factorization theorem.

2. Determine whether the sufficient statistic is also *complete.* This is generally hard to do. If it is not complete, we can saying nothing more about the MVUE. If it is, continue to step 3.

3. Find the MVUE $\hat{\theta}$ from $T(\mathbf{x})$ is one of two ways using the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem:

   - Find a function $g(\cdot)$ of the sufficient statistic that yields an unbiased estimator $\hat{\theta} = g(T(\mathbf{x}))$, the MVUE! By definition of completeness of the statistic, this will yield the MVUE.
   - Find any unbiased estimator $\breve{\theta}$ for θ, and then determine $\hat{\theta} = E[\breve{\theta}|T(\mathbf{x})]$. This is usually very tedious/difficult to do. The expectation is taken over the distribution $p(\breve{\theta}|T(\mathbf{x}))$.

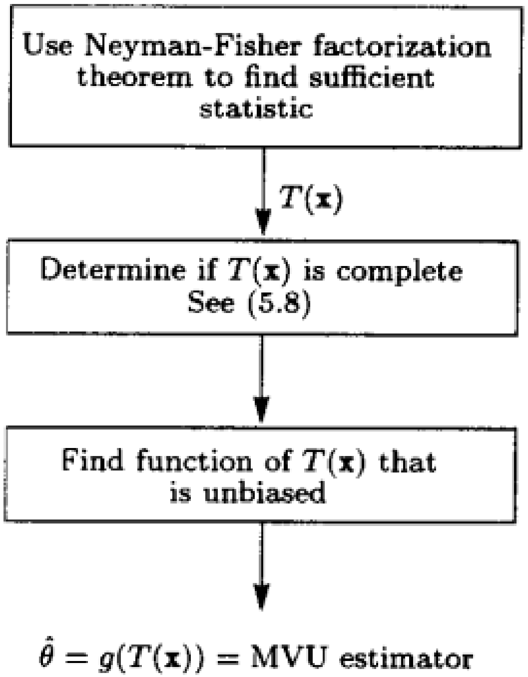

Figure 5.5 Procedure for finding MVU estimator (scalar parameter)

Example 5.8 - Mean of Uniform Noise

We observe the data

$$x[n] = w[n] \qquad n = 0, 1, \ldots, N-1$$

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$. Our initial approach of using the CRLB for finding an efficient and hence MVU estimator cannot even be tried for this problem. This is because the PDF does not satisfy the required regularity conditions (see Problem 3.1). A natural estimator of θ is the sample mean or

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

The sample mean is easily shown to be unbiased and to have variance

$$\begin{aligned} \operatorname{var}(\hat{\theta}) &= \frac{1}{N} \operatorname{var}(x[n]) \\ &= \frac{\beta^2}{12N}. \end{aligned} \tag{5.9}$$

Extension to a Vector Parameter

Theorem 5.3 (Neyman-Fisher Factorization Theorem (Vector Parameter)) *If we can factor the PDF $p(\mathbf{x};\boldsymbol{\theta})$ as*

$$p(\mathbf{x};\boldsymbol{\theta}) = g(\mathbf{T}(\mathbf{x}),\boldsymbol{\theta})h(\mathbf{x}) \tag{5.11}$$

where g is a function depending only on $\mathbf{x}$ through $\mathbf{T}(\mathbf{x})$, an $r \times 1$ statistic, and also on $\boldsymbol{\theta}$, and h is a function depending only on $\mathbf{x}$, then $\mathbf{T}(\mathbf{x})$ is a sufficient statistic for $\boldsymbol{\theta}$. Conversely, if $\mathbf{T}(\mathbf{x})$ is a sufficient statistic for $\boldsymbol{\theta}$, then the PDF can be factored as in (5.11).

Example 5.9 - Sinusoidal Parameter Estimation

Assume that a sinusoidal signal is embedded in WGN

$$x[n] = A\cos 2\pi f_0 n + w[n] \qquad n = 0,1,\ldots,N-1$$

where the amplitude A, frequency f_0, and noise variance σ^2 are unknown. The unknown parameter vector is therefore $\boldsymbol{\theta} = [A\, f_0\, \sigma^2]^T$. The PDF is

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2\right].$$

Expanding the exponent, we have

$$\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2 = \sum_{n=0}^{N-1} x^2[n] - 2A\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1}\cos^2 2\pi f_0 n.$$

Now because of the term $\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n$, where f_0 is *unknown*, we cannot reduce the PDF to have the form expressed in (5.11). If, on the other hand, the frequency were known, the unknown parameter vector would be $\boldsymbol{\theta} = [A\, \sigma^2]^T$ and the PDF could be expressed as

$$p(\mathbf{x};\boldsymbol{\theta}) =$$

$$\underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1} x^2[n] - 2A\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n + A^2\sum_{n=0}^{N-1}\cos^2 2\pi f_0 n\right)\right]}_{g(\mathbf{T}(\mathbf{x}),\boldsymbol{\theta})} \cdot \underbrace{1}_{h(\mathbf{x})}$$

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$

Theorem 5.4 (Rao-Blackwell-Lehmann-Scheffe (Vector Parameter)) *If* $\check{\boldsymbol{\theta}}$ *is an unbiased estimator of* $\boldsymbol{\theta}$ *and* $\mathbf{T}(\boldsymbol{\theta})$ *is an* $r \times 1$ *sufficient statistic for* $\boldsymbol{\theta}$*, then* $\hat{\boldsymbol{\theta}} = E(\check{\boldsymbol{\theta}}|\mathbf{T}(\mathbf{x}))$ *is*

1. *a valid estimator for* $\boldsymbol{\theta}$ *(not dependent on* $\boldsymbol{\theta}$*)*
2. *unbiased*
3. *of lesser or equal variance than that of* $\check{\boldsymbol{\theta}}$ *(each element of* $\hat{\boldsymbol{\theta}}$ *has lesser or equal variance)*

Additionally, if the sufficient statistic is complete, then $\hat{\boldsymbol{\theta}}$ *is the MVU estimator.*