E5: General Minimum Variance Unbiased Estimation

Using the concept of sufficient statistics and the Rao-Blackwell-Lehmann-Scheffe Theorem, we will infer the MVU estimator from the pdf.

Sufficient Statistics

Recall the problem of estimating the DC level in WGN. $x[n] = A + w[n]$ and the MVU estimator is $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$, where $var(\hat{A}) = \sigma^2/N$.

Another estimator is $\bar{A} = x[0]$ with $var(\bar{A}) = \sigma^2$. Clearly $\bar{A}$ is unbiased, but omitting a number of samples (intuitively) resulted in a much higher variance than the variance of $\hat{A}$.

Question: which data points are pertinent to estimation?

In the example above, consider 3 sets:

$$S_1 = \{x[0], x[1], \ldots, x[N-1]\}$$
$$S_2 = \{x[0]+x[1], x[2], \ldots, x[N-1]\}$$
$$S_3 = \left\{\sum_{n=0}^{N-1} x[n]\right\}$$

We can use all of these to compute $\hat{A}$, the MVU estimator.

S_1 is the original data set so its elements are sufficient to compute $\hat{A}$. Similarly, the elements of S_2 and S_3 are also sufficient to compute $\hat{A}$. Among all such sufficient sets, S_3 is the one with the smallest number of elements.

We can consider a function of some subset of data points as a statistic. Then a set containing statistics may or may not be sufficient to compute the MVU estimator. There may exist multiple sufficient statistic sets. We are particularly interested in minimal sufficient statistics.

In the example above, we have:

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

Let $T(x) \triangleq \sum_{n=0}^{N-1} x[n] \triangleq T_0$. Consider the pdf of the data given T_0:

$$p\left(x \middle| \sum_{n=0}^{N-1} x[n] = T_0 ; A\right)$$

Since the knowledge of T_0 is sufficient for MVU estimation of A, this conditional pdf should not depend on A.

If it did depend on A, then that would mean there is still information on A that is not captured in the statistic.

$\therefore$ The conditional pdf of data given sufficient statistics must be ~~indep~~ constant w.r.t. parameters.

Ex Verification of a Sufficient Statistic

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

Let $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$.

$$p(\mathbf{x} \mid T(\mathbf{x}) = T_0; A) = \frac{p(\mathbf{x}, T(\mathbf{x}) = T_0; A)}{p(T(\mathbf{x}) = T_0; A)} \quad \text{(Bayes Rule)}$$

Since $T(\mathbf{x})$ is a function of $\mathbf{x}$, the joint pdf in the numerator is $p(\mathbf{x};A)\,\delta(T(\mathbf{x})-T_0)$; so

$$p(\mathbf{x} \mid T(\mathbf{x}) = T_0; A) = \frac{p(\mathbf{x};A)\,\delta(T(\mathbf{x})-T_0)}{p(T(\mathbf{x}) = T_0; A)}$$

Here $T(\mathbf{x}) \sim \mathcal{N}(NA, N\sigma^2)$, so

$$p(\mathbf{x};A)\delta(T(\mathbf{x})-T_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2} \delta(T(\mathbf{x})-T_0)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1} x^2[n] - 2AT(\mathbf{x}) + NA^2\right)} \delta(T(\mathbf{x})-T_0)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1} x^2[n] - 2AT_0 + NA^2\right)} \delta(T(\mathbf{x})-T_0)$$

Then

$$p(\mathbf{x} \mid T(\mathbf{x}) = T_0; A) = \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]} e^{-\frac{1}{2\sigma^2}(-2AT_0+NA^2)} \delta(T(\mathbf{x})-T_0)}{\frac{1}{(2\pi N\sigma^2)^{1/2}} e^{-\frac{1}{2N\sigma^2}(T_0-NA)^2}}$$

$$= \frac{N^{1/2}}{(2\pi\sigma^2)^{(N-1)/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]} e^{\frac{T_0^2}{2N\sigma^2}} \delta(T(\mathbf{x})-T_0)$$

which, as expected, does not depend on A.

Finding Sufficient Statistics

Thm 5.1 Neyman-Fisher Factorization

If we can factor the pdf $p(x;\theta)$ as

$$p(x;\theta) = g(T(x),\theta)\, h(x)$$

where g is a function depending on x only through $T(x)$, and h is a function depending only on x, then $T(x)$ is a sufficient statistic for θ. Conversely, if $T(x)$ is a sufficient statistic for θ, then the pdf can be factored as indicated above.

Proof: Consider the joint pdf $p(x, T(x); \theta)$. We must have $p(x_0, \overset{T_0}{\cancel{T(x_0)}}; \theta) = 0$ unless $T(x_0) = T_0$. With this, the joint pdf becomes $p(x, T(x) = T_0; \theta) = p(x;\theta)\,\delta(T(x) - T_0)$.

We will also need $^{*}p(y) = \int p(x)\,\delta(y - g(x))\,dx$ if $y = g(x)$.

a) Prove $T(x)$ is a sufficient statistic when the factorization holds.

$$p(x \mid T(x) = T_0; \theta) = \frac{p(x, T(x) = T_0; \theta)}{p(T(x) = T_0; \theta)} = \frac{p(x;\theta)\,\delta(T(x) - T_0)}{p(T(x) = T_0; \theta)}$$

using the factorization
$$= \frac{g(T(x) = T_0; \theta)\, h(x)\, \delta(T(x) - T_0)}{p(T(x) = T_0; \theta)} = \frac{g(T(x) = T_0; \theta)\, h(x)\, \delta(T(x) - T_0)}{\int p(x;\theta)\,\delta(T(x) - T_0)\,dx}$$

using $*$

factorization
$$= \frac{g(T(x) = T_0, \theta)\, h(x)\, \delta(T(x) - T_0)}{\int g(T(x) = T_0, \theta)\, h(x)\, \delta(T(x) - T_0)\,dx}$$

$$= \frac{g(T(x)=T_0,\theta)\, h(x)\, \delta(T(x)-T_0)}{g(T(x)=T_0,\theta) \int h(x)\, \delta(T(x)-T_0)\,dx}$$

We could move $g(\cdot)$ out of the integral because the integral is zero except over the surface in $\mathbb{R}^N$ where $T(x)=T_0$. On this surface $g(T(x)=T_0,\theta)$ is constant.

So $p(x \mid T(x)=T_0;\theta) = \dfrac{h(x)\,\delta(T(x)-T_0)}{\int h(x)\,\delta(T(x)-T_0)\,dx}$

which does <u>not</u> depend on θ, so $T(x)$ is a sufficient statistic.

b) Prove that if $T(x)$ is a sufficient statistic, then the factorization holds.

Consider $\overset{**}{p}(x, T(x)=T_0;\theta) = p(x \mid T(x)=T_0;\theta)\, p(T(x)=T_0;\theta)$.

Note that $p(x, T(x)=T_0;\theta) = p(x;\theta)\,\delta(T(x)-T_0)$.

Since $T(x)$ is a sufficient statistic $p(x \mid T(x)=T_0;\theta)$ does <u>not</u> depend on $\theta \Rightarrow p(x \mid T(x)=T_0)$ is adequate. Given $T(x)=T_0$, this conditional pdf is nonzero only on that surface.

Let $p(x \mid T(x)=T_0) = w(x)\,\delta(T(x)-T_0)$, where

$$\int w(x)\,\delta(T(x)-T_0)\,dx = 1. \quad \overset{*}{*}\!*$$

Substituting in $**$

$$p(x;\theta)\,\delta(T(x)-T_0) = w(x)\,\delta(T(x)-T_0)\, p(T(x)=T_0;\theta)$$

and letting $w(x) = \dfrac{h(x)}{\int h(x)\,\delta(T(x)-T_0)\,dx}$ so that $\overset{*}{*}\!*$ is satisfied,

we get $p(x;\theta)\,\delta(T(x)-T_0) = \dfrac{h(x)\,\delta(T(x)-T_0)}{\int h(x)\,\delta(T(x)-T_0)\,dx}\, p(T(x)=T_0;\theta)$

or $p(x;\theta) = g(T(x)=T_0;\theta)\,h(x)$ where

$$g(T(x)=T_0;\theta) = \frac{p(T(x)=T_0;\theta)}{\int h(x)\,\delta(T(x)-T_0)\,dx} \qquad \square.$$

This form also demonstrates that the pdf of the sufficient statistic is in the form

$$p(T(x)=T_0;\theta) = g(T(x)=T_0;\theta)\int h(x)\,\delta(T(x)-T_0)\,dx$$

<u>Ex</u> DC level in WGN

$$p(x;A) = \frac{1}{(2\pi\sigma^2)^{N/2}}\, e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2}$$

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

From this, clearly $T(x) = \sum_{n=0}^{N-1} x[n]$ is a sufficient statistic for A.

Note that $T'(x) = 2\sum_{n=0}^{N-1} x[n]$ is also a sufficient statistic for A.

<u>Fact</u>: Sufficient statistics are unique only to within one-to-one transformations.

Ex | Power of WGN

Consider the same pdf as in the previous example with $A=0$ and σ^2 as the unknown parameter. ($x[n]=w[n]$)

$$p(x;\sigma^2) = \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]}}_{g(T(x),\sigma^2)} \cdot \underbrace{1}_{h(x)}$$

Immediately from the factorization $T(x) = \sum_{n=0}^{N-1} x^2[n]$ is a sufficient statistic for σ^2.

Ex | Phase of sinusoid

$x[n] = A\cos(2\pi f_0 n + \phi) + w[n] \quad n = 0, 1, \dots, N-1.$

Here A and f_0 are known. $w[n]$ is WGN with σ^2, and σ^2 is also known. ϕ is to be estimated.

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

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

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

Then

$$p(x;\phi) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1} A^2\cos^2(2\pi f_0 n+\phi) - 2AT_1(x)\cos\phi + 2AT_2(x)\sin\phi\right)} \cdot \underbrace{e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]}}_{h(x)}$$

where $T_1(x) = \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n$, $T_2(x) = \sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n$.

In this example, $p(x;\phi) = g(T_1(x), T_2(x), \phi)\, h(x)$ and we refer to $\begin{pmatrix} T_1(x) \\ T_2(x) \end{pmatrix}$ as jointly sufficient statistics for ϕ.

<u>Defn</u>: If $p(x;\theta) = g(T_1(x), T_2(x), \dots, T_r(x), \theta)\, h(x)$ then $\{T_1(x), \dots, T_r(x)\}$ are sufficient statistics for θ.

In that case $p(x \mid T_1(x), \dots, T_r(x); \theta)$ does <u>not</u> depend on θ. The original data is always a sufficient statistic but usually not minimal.

<u>Using Sufficiency to Find the MVU Estimator</u>

Assuming that we found a sufficient statistic $T(x)$ for θ, we can employ the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem to find the MVU estimator.

<u>Ex</u> DC Level in WGN

$x[n] = A + w[n]$ $n=0,1,\dots,N-1$, $w[n] \sim$ WGN with var σ^2.

A minimal sufficient statistic is $T(x) = \sum_{n=0}^{N-1} x[n]$. There are two methods:

1) Find any unbiased estimator of A, say $\bar{A}$, and determine $\hat{A} = E[\bar{A} \mid T]$. The expectation is w.r.t $p(\bar{A} \mid T)$.

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

Approach 1) Let $\bar{A} = x[0]$. $\hat{A} = E\left[x[0] \middle| \sum_{n=0}^{N-1} x[n]\right]$.

Notice that $x[0]$ and $T(x)$ are jointly Gaussian.

For jointly Gaussian $\begin{bmatrix} x \\ y \end{bmatrix}$ with mean μ and cov C,

$$E[x|y] = \int_{-\infty}^{\infty} x\, p(x|y)\, dx = \int_{-\infty}^{\infty} x \frac{p(x,y)}{p(y)} dx = E[x] + \frac{\mathrm{cov}(x,y)}{\mathrm{var}(y)}(y - E[y])$$

$$= \mu_1 + \frac{C_{12}}{C_{22}}(y - \mu_2) \quad \text{(Derivation in App 10A)}$$

Then for jointly Gaussian $\begin{bmatrix} x[0] \\ T(x) \end{bmatrix}$, we have $\mu = A\,L\,\mathbf{1} = \begin{bmatrix} A \\ NA \end{bmatrix}$

and $C = \sigma^2 L L^T = \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & N \end{bmatrix}$, where $L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$, since

$\begin{bmatrix} x[0] \\ T(x) \end{bmatrix} = L\,x$ (and x is jointly Gaussian $\sim \mathcal{N}(0, \sigma^2 I)$).

Then $\hat{A} = \mu_1 + \frac{C_{12}}{C_{22}}(T(x) - \mu_2) = A + \frac{1}{N}\left(\sum_{n=0}^{N-1} x[n] - NA\right)$

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

which is the MVU estimator.

Approach 2) Find g such that $\hat{A} = g(T(x))$ is unbiased.

Let $g(z) = z/N$. Then $\hat{A} = \frac{1}{N} T(x)$, which is unbiased.

It is the MVU estimator.

In practice approach 2 is usually easier to employ.

Thm 5.2 (Rao-Blackwell-Lehmann-Scheffe)

If $\bar{\theta}$ is an unbiased estimator of θ and $T(x)$ is a sufficient statistic for θ, then $\hat{\theta} = E[\bar{\theta} \mid T(x)]$ is

1) a valid estimator for θ (not dependent on θ),
2) unbiased,
3) of lesser or equal variance than that of $\bar{\theta}$, $\forall \theta$.

continued

(Thm continued) Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

Proof: (Scalar parameter)

(1) Note that $\hat{\theta} = E[\bar{\theta} \mid T(x)] = \int \bar{\theta}(x)\, p(x \mid T(x); \theta)\, dx$

Since $T(x)$ is a sufficient statistic, the conditional is constant w.r.t θ (does not depend on θ). So,

$\hat{\theta} = \int \bar{\theta}(x)\, p(x \mid T(\mathbf{x}))\, dx$ is only a function of $T(\mathbf{x})$, not dependent on θ.

(2) $\hat{\theta}$ is a function of $T(\mathbf{x})$ only. Therefore

$$
\begin{aligned}
E[\hat{\theta}] &= \iint \bar{\theta}(x)\, p(x \mid T(\mathbf{x}); \theta)\, dx\, p(T(\mathbf{x}); \theta)\, dT \\
&= \int \bar{\theta}(x) \int p(x \mid T(\mathbf{x}); \theta)\, p(T(\mathbf{x}); \theta)\, dT\, dx \\
&= \int \bar{\theta}(x)\, p(x; \theta)\, dx \\
&= E[\bar{\theta}] \\
&= \theta
\end{aligned}
$$

since $\bar{\theta}$ is unbiased

(3) We have $\mathrm{var}(\bar{\theta}) = E\{(\bar{\theta} - E(\bar{\theta}))^2\} = E\{(\bar{\theta} - \hat{\theta} + \hat{\theta} - \theta)^2\}$

$$= E[(\bar{\theta} - \hat{\theta})^2] + 2E[(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta)] + E[(\hat{\theta} - \theta)^2]$$

The cross-term is zero as follows:

$$
\begin{aligned}
E_{T,\bar{\theta}}[(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta)] &= E_T\, E_{\bar{\theta}|T}[(\bar{\theta} - \hat{\theta})(\hat{\theta} - \theta)] \\
&= E_T\left[E_{\bar{\theta}|T}[\bar{\theta} - \hat{\theta}]\,(\hat{\theta} - \theta)\right] = E_T\left\{\big(\underbrace{E_{\bar{\theta}|T}[\bar{\theta} \mid T] - \hat{\theta}}_{=0}\big)(\hat{\theta} - \theta)\right\}
\end{aligned}
$$

Therefore, $\mathrm{var}(\bar{\theta}) = E[(\bar{\theta} - \hat{\theta})^2] + \mathrm{var}(\hat{\theta}) \geq \mathrm{var}(\hat{\theta})$.

Defn: A statistic is complete if $\exists$ only one function statistic that is unbiased.

Consider all possible unbiased estimators of θ. By determining $E[\bar{\theta} | T(x)]$ we lower the variance of the estimator according to the RBLS theorem. Note that $E[\bar{\theta} | T(x)]$ is solely a function of $T(x)$, since

$$\hat{\theta} = E[\bar{\theta} | T(x)] = \int \bar{\theta}\, p(\bar{\theta} | T(x))\, d\bar{\theta} = g(T(x)).$$

If $T(x)$ is complete, $\exists!$ function of T that is unbiased, so $\hat{\theta}$ is unique independent of the $\bar{\theta}$ we choose. Every $\bar{\theta}$ results in the same $\hat{\theta}$, which is the MVU estimator.

$\therefore$ approach 2 works; by finding the unique $g(\cdot)$ such that $g(T(x))$ is unbiased, we obtain the MVU estimator.

Note that validating the completeness of $T(x)$ is in general difficult.

Ex) Completeness of a Sufficient Statistic.

In the DC level in WGN example, for A, $T(x) = \sum_{n=0}^{N-1} x[n]$ is complete and $\exists!\ g \ni E[g(T(x))] = A$. (Nevertheless) Suppose $\exists$ a second function $h \ni E[h(T(x))] = A$. Then we must have $E[g(T) - h(T)] = A - A = 0\ \forall A$. Since $T \sim \mathcal{N}(NA, N\sigma^2)$

$$\int_{-\infty}^{\infty} v(T) \frac{1}{\sqrt{2\pi N\sigma^2}} e^{-\frac{1}{2N\sigma^2}(T - NA)^2}\, dT = 0 \quad \forall A$$

where $v(T) = g(T) - h(T)$. Letting $\tau = T/N$, $\bar{v}(\tau) = v(N\tau)$

$$\int_{-\infty}^{\infty} \bar{v}(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} e^{-\frac{N}{2\sigma^2}(A-\tau)^2} d\tau = 0 \quad \forall A$$

This is the convolution of $\bar{v}(\tau)$ with a Gaussian $w(\tau)$. For the result to be zero $\forall A$, $\bar{v}(\tau)$ must be identically zero. (From Fourier transforms $\bar{V}(f) W(f) = 0$, we see $\bar{V}(f) = 0 \ \forall f$).

This implies $g = h \ \forall T$. $\therefore$ this proof by contradiction shows g is unique.

Ex) Incomplete Sufficient Statistic

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

A sufficient statistic is $x[0]$; and it is also unbiased. $g(x[0]) = x[0]$ seems to be a viable candidate for the MVU estimator. We need to verify that $x[0]$ is a complete statistic.

Suppose $\exists h \neq g$ s.t. $h(x[0])$ is unbiased. Let $v(T) = g(T) - h(T)$.

$$\int_{-\infty}^{\infty} v(T)\, p(x;A)\, dx = 0 \quad \forall A$$

$\Big\downarrow$ since $T(x) = x[0]$.

$$\int_{-\infty}^{\infty} v(T)\, p(T;A)\, dT = 0 \quad \forall A$$

However, $p(T;A) = \begin{cases} 1 & \text{if } A - \frac{1}{2} \leq T \leq A + \frac{1}{2} \\ 0 & \text{o.w.} \end{cases}$

so the condition becomes

$$\int_{A-\frac{1}{2}}^{A+\frac{1}{2}} v(T)\, dT = 0 \quad \forall A.$$

There are many such $v(T)$.

For instance $v(T) = \sin 2\pi T$. For this choice $h(T) = g(T) - \sin 2\pi T = T - \sin 2\pi T$

$\therefore \hat{A} = x[0] - \sin 2\pi x[0]$ is also based on the sufficient statistic and is unbiased $\therefore$ T is not complete.

Fact: A sufficient statistic T is complete if the condition $\int_{-\infty}^{\infty} v(T)\, p(T;\theta)\, dT = 0 \ \forall \theta$ is satisfied only by the zero function ($v(T)=0 \ \forall T$).

Procedure: Finding the MVU estimator using the RBLS thm.

1) Find a single sufficient statistic $T(x)$ for θ using the Neyman-Fisher Factorization Theorem
2) Determine if $T(x)$ is complete. If so, proceed; if not this approach cannot be used.
3) Find a function g s.t. $\hat{\theta} = g(T(x))$ is unbiased. $\hat{\theta}$ is then the MVU estimator.

or 3) Evaluate $\hat{\theta} = E[\bar{\theta} \mid T(x)]$ for any unbiased $\bar{\theta}$.

Ex) Mean of Uniform Noise

We observe $x[n] = w[n]$ $n=0,1,\ldots,N-1$ where $w[n]$ is IID noise with pdf $\mathcal{U}[0,\beta]$ for $\beta>0$. Find the MVU estimator for $\theta = \beta/2$ (the mean). In this problem the CRLB-based method cannot be employed since the regularity condition is not satisfied.

Consider the sample ~~mean~~ average estimator $\bar{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$. Clearly $\bar{\theta}$ is unbiased and $\mathrm{var}(\bar{\theta}) = \frac{1}{N}\mathrm{var}(x[n]) = \frac{\beta^2}{12N}$. Is $\bar{\theta}$ the MVU estimator? We have

$$p(x[n];\theta) = \frac{1}{\beta}\left[u(x[n]) - u(x[n]-\beta)\right] \quad \text{where } \beta = 2\theta$$

$u(x) = \begin{cases} 1 & ; x>0 \\ 0 & ; x<0 \end{cases}$ is the unit step function.

The joint data pdf is

$$p(x;\theta) = \frac{1}{\beta^N} \prod_{n=0}^{N-1} \left(u(x[n]) - u(x[n]-\beta) \right)$$

$$= \begin{cases} \frac{1}{\beta^N}, & 0 < x[n] < \beta \quad n=0,1,\dots,N-1 \\ 0, & \text{o.w.} \end{cases}$$

Alternatively $p(x;\theta) = \begin{cases} \frac{1}{\beta^N}, & \max x[n] < \beta, \min x[n] > 0 \\ 0, & \text{o.w.} \end{cases}$

So $p(x;\theta) = \frac{1}{\beta^N} \underbrace{u(\beta - \max x[n])}_{g(T(x);\theta)} \underbrace{u(\min x[n])}_{h(x)}$ order statistic

By Neyman-Fisher factorization $T(x) = \max_n x[n]$ is a sufficient statistic for θ. $T(x)$ is also complete.

* (As exercise show that $T(x)$ is a complete statistic.)

We now need to identify g s.t. $g(T(x))$ is unbiased.

The cdf of T is

$$\Pr\{T \le \xi\} = \Pr\{x[0] \le \xi, x[1] \le \xi, \dots, x[N-1] \le \xi\}$$

IID samples $= \prod_{n=0}^{N-1} \Pr\{x[n] \le \xi\} = \Pr\{x[n] \le \xi\}^N$

The pdf of T is $p_T(\xi) = \frac{d \Pr\{T \le \xi\}}{d\xi} = N \Pr\{x[n] \le \xi\}^{N-1} \frac{d \Pr\{x[n] \le \xi\}}{d\xi}$

This is the cdf of $x[n]$ This is the pdf of $x[n]$.

$$p_{x[n]}(\xi;\theta) = \begin{cases} 1/\beta & 0 < \xi < \beta \\ 0 & \text{o.w.} \end{cases}$$

$$\Pr\{x[n] \le \xi\} = \begin{cases} 0 & \xi < 0 \\ \xi/\beta & 0 \le \xi \le \beta \\ 1 & \xi > \beta \end{cases}$$

Substituting ...

$$p_T(\xi) = \begin{cases} 0 & \xi < 0 \\ N\left(\frac{\xi}{\beta}\right)^{N-1}\frac{1}{\beta} & 0 \le \xi \le \beta \\ 0 & \xi > \beta \end{cases}$$

$$E[T] = \int_{-\infty}^{\infty} \xi\, p_T(\xi)\, d\xi = \int_0^{\beta} \xi N \left(\frac{\xi}{\beta}\right)^{N-1} \frac{1}{\beta}\, d\xi = \frac{N}{N+1}\beta$$

$$= \frac{2N}{N+1}\theta$$

To make this unbiased, let $\hat{\theta} = \frac{N+1}{2N} T(x)$ ...

$\hat{\theta} = \frac{N+1}{2N} \max_n x[n]$ is the MVU estimator.

Consider $\mathrm{var}(\hat{\theta}) = \left(\frac{N+1}{2N}\right)^2 \mathrm{var}(T)$.

$$\mathrm{var}(T) = \int_0^{\beta} \xi^2 \frac{N \xi^{N-1}}{\beta^N}\, d\xi - \left(\frac{N\beta}{N+1}\right)^2 = \frac{N\beta^2}{(N+1)^2(N+2)}$$

$$\Rightarrow \mathrm{var}(\hat{\theta}) = \frac{\beta^2}{4N(N+2)}$$

Note that $\mathrm{var}(\hat{\theta}) \le \mathrm{var}(\bar{\theta})$ for $N \ge 1$ and the equality holds only for $N=1$, in which case the sample average and the MVU estimator based on the order statistic become identical.

Also note that as $N \to \infty$ $\mathrm{var}(\bar{\theta}) \to 0$ at a rate of $\frac{1}{N}$ while $\mathrm{var}(\hat{\theta}) \to 0$ at a rate of $\frac{1}{N^2}$. This is a significant difference in convergence speed.

Extension to a Vector Parameter

We search for a $p\times 1$ dimensional MVU estimator. The number of sufficient statistics may be more than the number of parameters $(r>p)$, exactly the same number $(r=p)$ or fewer $(r<p)$. The $r=p$ case would be desirable. In this case we could use the second approach (to find a function to make the sufficient statistics unbiased).

<u>Defn</u>: A vector statistic $\mathbf{T}(x)=[T_1(x), T_2(x), \ldots, T_r(x)]^T$ is said to be sufficient for the estimation of θ if the pdf of the data conditioned on the statistic or $p(x|T(x);\theta)$ does not depend on θ, i.e. $p(x|T(x);\theta)=p(x|T(x))$.

In general, many such $T(x)$ exist (including x). We are interested in the minimal sufficient statistic, or $T(x)$ with minimum dimension.

<u>Thm 5.3</u> Neyman-Fisher Factorization Theorem (Vector Parameter)

If we can factor the data pdf as $p(x;\theta)=g(\mathbf{T}(x),\theta)h(x)$ where g ~~only~~ depends on x only through $T(x)$, an $r\times 1$ statistic, and on θ; and h depends only on x, then $T(x)$ is a sufficient statistic for θ. Conversely, if $T(x)$ is a sufficient statistic for θ, then the pdf can be factored as indicated.

Ex] Sinusoidal Parameter Estimation

$x[n] = A\cos(2\pi f_0 n) + w[n]$ $n = 0, \ldots, N-1$; $w[n] \sim WGN$ with σ^2 var.

A, f_0, σ^2 are unknown $\Rightarrow$ $\theta = [A, f_0, \sigma^2]^T$.

$$p(x;\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos(2\pi f_0 n))^2}$$

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

Since f_0 is unknown, we cannot reduce the pdf to the form in Thm 5-3.

If f_0 is known, $\theta = [A, \sigma^2]^T$ and the pdf would be

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

where $T(x) = \left[\sum_{n=0}^{N-1} x[n]\cos(2\pi f_0 n), \sum_{n=0}^{N-1} x^2[n]\right]^T$ with $r=2$.

$\therefore$ If f_0 is known, $T(x)$ is a sufficient statistic for $\theta = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$.

Ex] DC Level in White Noise with unknown Noise Power

$x[n] = A + w[n]$ $n = 0, \ldots, N-1$ with $w[n] \sim WGN$ having var σ^2.

A and σ^2 are unknown. $\theta = [A, \sigma^2]^T$. In the previous example, let $f_0 = 0$ (to get a DC signal).

$$T(x) = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}$$ is a sufficient statistic for $\theta = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$.

Note: In this case $T_1(x)$ is sufficient for A and $T_2(x)$ is sufficient for σ^2. They are jointly sufficient for θ. This is not the case in general.

Thm 5.4 RBLS (Vector Parameter)

If $\bar{\theta}$ is an unbiased estimator of θ and $T(x)$ is an $r\times 1$ sufficient statistic for θ, then $\hat{\theta} = E[\bar{\theta} \mid T(x)]$ is

1) a valid estimator for θ (not dependent on θ)
2) unbiased
3) of lesser or equal variance than that of $\bar{\theta}$
   ($\mathrm{var}(\hat{\theta}_i) \le \mathrm{var}(\bar{\theta}_i) \quad \forall i$)

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

Defn. $T(x)$ is a complete statistic (with $r\times 1$ $\mathbf{T}$)

if $\exists!\ g$ such that $E[g(T(x))] = \theta$.

(There is only one function that makes T unbiased.)

Ex) DC Level in WGN with Unknown Noise Power

$$T(x) = \begin{bmatrix} T_1(x) \\ T_2(x) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$ Taking the expected value:

$$E[T(x)] = \begin{bmatrix} NA \\ N(\sigma^2+A^2) \end{bmatrix}.$$ Consider $$g(T(x)) = \begin{bmatrix} T_1(x)/N \\ \frac{T_2(x)}{N} - \frac{T_1^2(x)}{N^2} \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \end{bmatrix}$$

we have $E[\bar{x}] = A$. $E\left[\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \bar{x}^2\right]$

$= \sigma^2 + A^2 - E[\bar{x}^2]$

Since $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$, $E[\bar{x}^2] = A^2 + \sigma^2/N$.

So $E\left[\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] - \bar{x}^2\right] = \sigma^2\left(1 - \frac{1}{N}\right) = \frac{N-1}{N}\sigma^2$ This is not unbiased for σ^2.

Let $g(T(x)) = \begin{bmatrix} \frac{1}{N} T_1(x) \\ \frac{1}{N-1}\left(T_2(x) - \frac{1}{N} T_1^2(x)\right) \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1}\left(\sum_{n=0}^{N-1} x^2[n] - \bar{x}^2\right) \end{bmatrix}$

$\sum_{n=0}^{N-1} (x[n]-\bar{x})^2 = \dots = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2$, so $E[g(T(x))] = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$.

(unbiased)

$\therefore \quad \hat{\theta} = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2 \end{bmatrix} = \begin{bmatrix} \hat{A} \\ \hat{\sigma^2} \end{bmatrix}$

Here $\hat{A}$ and $\hat{\sigma^2}$ are independent and $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$,

$\frac{(N-1)\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{N-1}$, so $C_{\hat{\theta}} = \begin{bmatrix} \sigma^2/N & 0 \\ 0 & \frac{2\sigma^4}{(N-1)} \end{bmatrix}$. The CRLB

is $I^{-1}(\theta) = \begin{bmatrix} \sigma^2/N & 0 \\ 0 & 2\sigma^4/N \end{bmatrix}$. Clearly $\mathrm{var}(\hat{\sigma^2}) > \mathrm{CRLB}_{\hat{\sigma^2}}$

so the MVU estimator is <u>not</u> efficient and could not have been found using the CRLB.

The MVU estimator could have been found using $p(x;\theta) = g(T'(x), \theta)\, h(x)$ where $h(x) = 1$ and

$T'(x) = \begin{bmatrix} \bar{x} \\ \sum_{n=0}^{N-1}(x[n]-\bar{x})^2 \end{bmatrix}$ $\begin{matrix} \leftarrow \text{sample mean} \\ \leftarrow \propto \text{sample variance} \end{matrix}$

To make the second entry unbiased we need to divide by $(N-1)$ (the g function is simple to find). This gives the MVU estimator $\hat{\theta}$ above. Completeness can be verified.

Suggested Problems: 3, 5, 7, 12, 15, 16, 18, 19