Minimum Variance Unbiased Estimation

**MVU Estimators**

- **Unbiased Estimators**

  **Unbiased estimator**: on the average, the estimator will yield the true value of the unknown parameter.

  \[ E(\hat{\theta}) = \theta \quad \text{for all } \theta \]

  Let \( \hat{\theta} = g(x) \), where \( x = [x[0] \ x[1] \ldots x[N-1]]^T \)

  Unbiased \( \Rightarrow E(\hat{\theta}) = \int g(x) p(x; \theta) dx = \theta \)

  **Remark**: Generally, we seek unbiased estimators (necessary but not sufficient for good estimator)

  Example: \( x[n] = A + w[n] \quad n = 0, 1, \ldots, N-1 \)

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

  \[
  E(\hat{A}) = E \left[ \frac{1}{N} \sum_{n=0}^{N-1} x[n] \right] \\
  = \frac{1}{N} \sum_{n=0}^{N-1} E(x[n]) \\
  = \frac{1}{N} \sum_{n=0}^{N-1} A \\
  = A \]

  \( \Rightarrow \) unbiased!

  \[ \hat{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n]. \]
\[ E(\hat{A}) = \frac{1}{2} A \]
\[ = A \text{ if } A = 0 \]
\[ \neq A \text{ if } A \neq 0. \implies \text{biased!} \]

**Bias of estimator**

\[ E(\hat{\theta}) - \theta = b(\theta) \]
Minimum Variance Criterion

Need optimality criterion to assess performance

**Mean Square Error (MSE)**

\[ \text{mse}(\hat{\theta}) = E \left[ (\hat{\theta} - \theta)^2 \right] \]

⇒ Good estimator has small MSE.

Unfortunately, adoption of this natural criterion leads to unrealizable estimators.

\[
\text{mse}(\hat{\theta}) = E \left\{ \left[ (\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta) \right]^2 \right\} \\
= \text{var}(\hat{\theta}) + \left[ E(\hat{\theta}) - \theta \right]^2 \\
= \text{var}(\hat{\theta}) + b^2(\theta)
\]

MSE is composed of errors due to the variance of the estimator as well as the bias!!

Example: \( \hat{A} = a \frac{1}{N} \sum_{n=0}^{N-1} x[n] \) for some constant a.

\[ E(\hat{A}) = aA \]

\[ \text{var}(\hat{A}) = a^2 \sigma^2 / N \]

\[ \text{mse}(\hat{A}) = \frac{a^2 \sigma^2}{N} + (a - 1)^2 A^2. \]

\[
\frac{d}{da} \text{mse}(\hat{A}) = \frac{2a \sigma^2}{N} + 2(a - 1)A^2
\]

\[ a_{\text{opt}} = \frac{A^2}{A^2 + \sigma^2 / N}. \]

The optimal value of a depends on the unknown parameter A.

⇒ The estimator is not realizable.
Any criterion that depends on the bias will lead to an unrealizable estimator!!

Generally, we constrain the bias to be zero and find the estimator which minimizes the variance.

⇒ Minimum Variance Unbiased (MVU) estimator.

Existence of MVU Estimator

\[ \hat{\theta} \] must have the smallest variance for all values of \( \theta \).

In general, the MVU estimator does not always exist.

Example: Assume we have two independent observations \( x[0] \) and \( x[1] \).

\[
\begin{align*}
  x[0] & \sim \mathcal{N}(\theta, 1) \\
  x[1] & \sim \begin{cases} 
  \mathcal{N}(\theta, 1) & \text{if } \theta \geq 0 \\
  \mathcal{N}(\theta, 2) & \text{if } \theta < 0.
\end{cases}
\end{align*}
\]

The two estimators

\[
\begin{align*}
  \hat{\theta}_1 & = \frac{1}{2} (x[0] + x[1]) \\
  \hat{\theta}_2 & = \frac{2}{3} x[0] + \frac{1}{3} x[1]
\end{align*}
\]

are unbiased.
\[
\begin{align*}
\text{var}(\hat{\theta}_1) &= \frac{1}{4} \left( \text{var}(x[0]) + \text{var}(x[1]) \right) \\
\text{var}(\hat{\theta}_2) &= \frac{4}{9} \text{var}(x[0]) + \frac{1}{9} \text{var}(x[1])
\end{align*}
\]

\[
\text{var}(\hat{\theta}_1) = \begin{cases} 
\frac{18}{36} & \text{if } \theta \geq 0 \\
\frac{27}{36} & \text{if } \theta < 0 
\end{cases}
\]

\[
\text{var}(\hat{\theta}_2) = \begin{cases} 
\frac{20}{36} & \text{if } \theta \geq 0 \\
\frac{24}{36} & \text{if } \theta < 0.
\end{cases}
\]

Remark: It's possible that there may not exist even a single unbiased estimator.

Finding the MVU Estimator

Approaches:
1. Determine the Cramer-Rao lower bound (CRLB) and check to see if some estimator satisfies it. (Chapters 3 and 4)

Remarks:
1. If an estimator exists whose variance equals the CRLB for each value of \(\theta\), then it must be the MVU estimator.
2. It may happen that no estimator exists whose variance equals the bound.
2. Apply the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem. (Chapter 5)

First find a sufficient statistic
⇒ then find a function of the sufficient statistic which is an unbiased estimator of θ.

3. Further restrict the class of estimators to be not only unbiased but also linear. Then, find the minimum variance estimator within this restricted class. (Chapter 6)

Extension to a Vector Parameter

Assume \( \boldsymbol{\theta} = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_p]^{T} \) is a vector of unknown parameters.

An estimator \( \hat{\boldsymbol{\theta}} = [\hat{\theta}_1 \quad \hat{\theta}_2 \quad \cdots \quad \hat{\theta}_p]^{T} \) is unbiased if

\[
E(\hat{\theta}_i) = \theta_i \quad \alpha_i < \theta_i < \beta_i \quad \text{for } i = 1, 2, \ldots, p.
\]

By define

\[
E(\hat{\boldsymbol{\theta}}) = \begin{bmatrix}
E(\hat{\theta}_1) \\
E(\hat{\theta}_2) \\
\vdots \\
E(\hat{\theta}_p)
\end{bmatrix}
\]

We define an unbiased estimator to have the property \( E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} \).

MVU estimator: \( \text{var}(\hat{\theta}_i), \ i = 1, 2, \ldots, p \) is minimum.
Cramer-Rao Lower Bound (CRLB)

Find a lower bound on the variance of an unbiased estimator.

- If the estimator attains the bound for all values of the unknown parameter \( \Rightarrow \) it’s the MVU estimator.
- It provides a benchmark for the comparison of performance among unbiased estimators.

Estimator Accuracy Considerations

All our information is embodied in the observation data and the underlying pdf for that data.

\( \Rightarrow \) How accurately we can estimate \( \theta \) depends directly on the pdf.

The more the pdf is influenced by the unknown parameter, the better we can estimate the parameter!

Example:

\[
x[0] = A + w[0] \quad \text{where} \quad w[0] \sim \mathcal{N}(0, \sigma^2).
\]

Choose \( \hat{A} = x[0] \)

The estimator accuracy improves as \( \sigma^2 \) decreases.

Alternative Viewpoint:

\[
p_i(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[ -\frac{1}{2\sigma_i^2} (x[0] - A)^2 \right]
\]

The probability of observing \( x[0] \) when \( A \) is the true value.

If \( \sigma = 1/3 \), it is unlikely that \( A > 4 \).

pdf concentration \( \uparrow \) \( \Rightarrow \) parameter accuracy \( \uparrow \)
When the pdf is viewed as a function of the unknown parameter (with $x$ fixed), it is termed the **likelihood function**.

The “sharpness” of the likelihood function determines how accurately we can estimate the unknown parameter.

\[
\ln p(x[0]; A) = -\ln \sqrt{2\pi \sigma^2} - \frac{1}{2\sigma^2} (x[0] - A)^2
\]

\[
\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2} (x[0] - A)
\]

\[- \frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}. \quad \text{the curvature increases as } \sigma \text{ decreases.}
\]

**Note:**

\[
\text{var}(\hat{A}) = \frac{1}{-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}}
\]

In general, $-\frac{\partial^2 \ln p}{\partial \theta^2}$ is random variable.

To measure curvature, we use

\[- E \left[ \frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} \right],
\]

*the average curvature of the log-likelihood function.*
Cramer-Rao Lower Bound

**Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter)** It is assumed that the PDF \( p(x; \theta) \) satisfies the "regularity" condition

\[
E \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta} \right] = 0 \quad \text{for all } \theta
\]

where the expectation is taken with respect to \( p(x; \theta) \). Then, the variance of any unbiased estimator \( \hat{\theta} \) must satisfy

\[
\text{var}(\hat{\theta}) \geq \frac{1}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]}
\]

where the derivative is evaluated at the true value of \( \theta \) and the expectation is taken with respect to \( p(x; \theta) \). Furthermore, an unbiased estimator may be found that attains the bound for all \( \theta \) if and only if

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

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

Remarks:

1. \[ E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = \int \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} p(x; \theta) \, dx \]

2. Regularity Condition is satisfied if the order of differentiation and integration may be interchanged:

\[
\int \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = \int \frac{\partial p(x; \theta)}{\partial \theta} \, dx = \frac{\partial}{\partial \theta} \int p(x, \theta) \, dx = \frac{\partial 1}{\partial \theta} = 0.
\]

This condition is generally true except when the domain of the pdf for which it is nonzero depends on the unknown parameter.

An example that does not satisfy the regularity condition:

\( x[n] \) are iid according to \( U[0, \theta] \), for \( n = 0, 1, ..., N-1 \).

3. \[ E \left[ \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] \]

\[
\Rightarrow \quad \text{var}(\hat{\theta}) \geq \frac{1}{E \left[ \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 \right]}
\]
4. Fisher Information

\[ I(\theta) = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]. \]

- Nonnegative
- Additive for independent observations.

\[ \ln p(x; \theta) = \sum_{n=0}^{N-1} \ln p(x[n]; \theta). \]

\[ -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = - \sum_{n=0}^{N-1} E \left[ \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right] \]

\[ \Rightarrow I(\theta) = Ni(\theta) \]

where \( i(\theta) = -E \left[ \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right] \)

As \( N \to \infty \), CRLB \( \to 0 \) for iid distribution

**Proof of Scalar Parameter CRLB:**

We consider all unbiased estimator \( \hat{\alpha} \) with \( E(\hat{\alpha}) = \alpha = g(\theta) \)

That is, \( \int \hat{\alpha} p(x; \theta) \, dx = g(\theta) \).

\[ \int \hat{\alpha} \frac{\partial p(x; \theta)}{\partial \theta} \, dx = \frac{\partial g(\theta)}{\partial \theta} \]

\[ \int \hat{\alpha} \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = \frac{\partial g(\theta)}{\partial \theta}. \]

Since \( \int \alpha \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = \alpha E \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta} \right] = 0 \),

we have \( \int (\hat{\alpha} - \alpha) \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = \frac{\partial g(\theta)}{\partial \theta} \).

We now apply the Cauchy-Schwarz inequality

\[ \left[ \int w(x) g(x) h(x) \, dx \right]^2 \leq \int w(x) g^2(x) \, dx \int w(x) h^2(x) \, dx \]
\[ w(x) = p(x; \theta) \]

with
\[ g(x) = \dot{\alpha} - \alpha \]
\[ h(x) = \frac{\partial \ln p(x; \theta)}{\partial \theta} \]

\[ \Rightarrow \left( \frac{\partial g(\theta)}{\partial \theta} \right)^2 \leq \int (\dot{\alpha} - \alpha)^2 p(x; \theta) \, dx \int \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 p(x; \theta) \, dx \]

\[ \Rightarrow \text{var}(\dot{\alpha}) \geq \frac{\left( \frac{\partial g(\theta)}{\partial \theta} \right)^2}{E \left[ \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 \right]} \]

On the other hand, we have

\[ E \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta} \right] = 0 \]
\[ \int \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = 0 \]

\[ \frac{\partial}{\partial \theta} \int \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx = 0 \]
\[ \int \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} p(x; \theta) + \frac{\partial \ln p(x; \theta)}{\partial \theta} \frac{\partial p(x; \theta)}{\partial \theta} \right] p(x; \theta) \, dx = 0 \]

\[ -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = \int \frac{\partial \ln p(x; \theta)}{\partial \theta} \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) \, dx \]
\[ = E \left[ \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 \right] \]

\[ \Rightarrow \text{var}(\dot{\alpha}) \geq \frac{\left( \frac{\partial g(\theta)}{\partial \theta} \right)^2}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]} \]

If \( \alpha = g(\theta) = 0 \),

\[ \text{var}(\dot{\theta}) \geq \frac{1}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]} \].
Moreover, the condition for equality is

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{c} (\hat{\theta} - \theta)
\]

where \( c \) can depend on \( \theta \) but not on \( x \).

If \( \alpha = g(\theta) = \theta \),

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{c(\theta)} (\hat{\theta} - \theta).
\]

\[
\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} = -\frac{1}{c(\theta)} + \frac{\partial}{\partial \theta} \left( \frac{1}{c(\theta)} \right) (\hat{\theta} - \theta)
\]

\[
-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right] = \frac{1}{c(\theta)}
\]

\[
c(\theta) = \frac{1}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]}
\]

\[
= \frac{1}{I(\theta)} \quad \Rightarrow \quad \frac{\partial \ln p(x; \theta)}{\partial \theta} = I(\theta)(g(x) - \theta)
\]

When the CRLB is attained,

\[
\text{var}(\hat{\theta}) = \frac{1}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]}
\]

Q.E.D.

Example: DC level in white Gaussian noise

\[
x[n] = A + w[n] \quad n = 0, 1, 2, ..., N-1
\]

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

\[
= \frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]
\]
The sample mean estimator attains the bound and must be the MVU estimator.

**Summary:** CRLB theorem gives us a lower bound on variance of any unbiased estimator but also finds estimator that attains bound (if it exists)

Estimators that attains bound are termed "efficient" (good use of data)

efficient $\Rightarrow$ MVU

MVU $\Rightarrow$ efficient
Example: Phase estimation

Assume we wish to estimate the phase $\phi$ of a sinusoid embedded in WGN.

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

The amplitude $A$ and frequency $f_0$ are assumed known.

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

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

$$= -\frac{A}{\sigma^2} \sum_{n=0}^{N-1} [x[n] \sin(2\pi f_0 n + \phi) - A \frac{1}{2} \sin(4\pi f_0 n + 2\phi)]$$

$$\frac{\partial^2 \ln p(x; \phi)}{\partial \phi^2} = -\frac{A}{\sigma^2} \sum_{n=0}^{N-1} [x[n] \cos(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)]$$

$$-E \left[ \frac{\partial^2 \ln p(x; \phi)}{\partial \phi^2} \right] = \frac{A}{\sigma^2} \sum_{n=0}^{N-1} [A \cos^2(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)]$$

$$= \frac{A^2}{\sigma^2} \sum_{n=0}^{N-1} \left[ \frac{1}{2} + \frac{1}{2} \cos(4\pi f_0 n + 2\phi) - \cos(4\pi f_0 n + 2\phi) \right]$$

$$\approx \frac{NA^2}{2\sigma^2}$$

$$\text{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2}$$

Does an efficient estimator exist?

MVU?

How to find it?
General CRLB for Signals in White Gaussian Noise

Assume a deterministic signal with an unknown parameter \( \theta \) is observed in WGN.

\[
x[n] = s[n; \theta] + w[n] \quad n = 0, 1, \ldots, N - 1.
\]

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

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n; \theta]) \frac{\partial s[n; \theta]}{\partial \theta}
\]

\[
\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left\{ (x[n] - s[n; \theta]) \frac{\partial^2 s[n; \theta]}{\partial \theta^2} - \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2 \right\}
\]

\[
E \left( \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right) = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2
\]

\[
\Rightarrow \quad \text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2}
\]

Example: \( s[n; f_0] = A \cos(2\pi f_0 n + \phi) \quad 0 < f_0 < \frac{1}{2} \)

\[
\text{var}(\hat{f}_0) \geq \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1} [2\pi n \sin(2\pi f_0 n + \phi)]^2}
\]
Transformation of Parameters

Assume we wish to estimate $\alpha = g(\theta)$, a function of $\theta$.

\[
\text{var}(\hat{\alpha}) \geq \frac{\left( \frac{\partial g}{\partial \theta} \right)^2}{-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right]}
\]

Example: DC level in WGN $x[n] = A + w[n]$

CRLB for $\alpha = g(A) = A^2$

\[
\text{var}(A^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2 \sigma^2}{N}.
\]

Question: The sample mean estimator is efficient for $A$.

Is $\bar{x}^2$ efficient for $A^2$? \textbf{NO!!}

\[
\bar{x} \sim \mathcal{N}(A, \sigma^2/N),
\]

\[
\text{var}(\bar{x}^2) = E(\bar{x}^4) - E^2(\bar{x}^2)
\]

\[
\text{var}(\bar{x}^2) = \frac{4A^2 \sigma^2}{N} + \frac{2\sigma^4}{N^2} \Rightarrow \text{Not efficient!}
\]

\[
E(\bar{x}^2) = E^2(\bar{x}) + \text{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N} \\
\neq A^2.
\]

$\bar{x}^2$ is not even an unbiased estimator!!

In general, the efficiency of an estimator is destroyed by a nonlinear transformation.
The efficiency can be maintained over a linear (or affine) transformation.

Proof: Assume $\hat{\theta}$ is an efficient estimator for $\theta$ and we want to estimate $g(\theta) = a\theta + b$.

\[
E(a\hat{\theta} + b) = aE(\hat{\theta}) + b = a\theta + b
\]

\[
\text{var}(g(\hat{\theta})) \geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)}
\]

\[
= \left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \text{var}(\hat{\theta})
\]

\[
= a^2 \text{var}(\hat{\theta}).
\]

But we have $\text{var}(g(\hat{\theta})) = \text{var}(a\hat{\theta} + b) = a^2 \text{var}(\hat{\theta}) \Rightarrow$ The CRLB is achieved!

If the data record is large enough, the efficiency of an estimator is approximately maintained over nonlinear transformation!!

Example: DC level in WGN $x[n] = A + w[n]$

\[\alpha = g(A) = A^2\]

\[g(\bar{x}) \approx g(A) + \frac{dg(A)}{dA} (\bar{x} - A)\]

\[E[g(\bar{x})] = g(A) = A^2\] asymptotically unbiased!

\[
\text{var}[g(\bar{x})] = \frac{\left(\frac{dg(A)}{dA}\right)^2}{N} \text{var}(\bar{x})
\]

\[
= \frac{(2A)^2 \sigma^2}{N}
\]

\[
= \frac{4A^2 \sigma^2}{N}\] asymptotically efficient!
Extension to a Vector Parameter

Wish to estimate a vector parameter \( \theta = [\theta_1, \theta_2 \ldots \theta_p]^T \)

Assume \( E(\hat{\theta}) = \theta \)

\[
\text{var}(\hat{\theta}_i) \geq [\text{I}^{-1}(\theta)]_{ii}
\]

where \( [\text{I}(\theta)]_{i,j} = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta_i \partial \theta_j} \right] \quad \text{for} \quad i = 1, 2, \ldots, p; j = 1, 2, \ldots, p. \)

**I(\theta)** Fisher information matrix

Remark:

\[
E \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta_i} \frac{\partial \ln p(x; \theta)}{\partial \theta_j} \right] = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta_i \partial \theta_j} \right]
\]

Example: Line Fitting

\[
x[n] = A + Bn + w[n] \quad \text{w[n]} \text{: WGN}
\]

\[
\theta = \begin{bmatrix} A \\ B \end{bmatrix}
\]

Likelihood function

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

\[
\text{I}(\theta) = \begin{bmatrix} -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial A^2} \right] & -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial A \partial B} \right] \\
-E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial B \partial A} \right] & -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial B^2} \right] \end{bmatrix}
\]

\[
\frac{\partial \ln p(x; \theta)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)
\]

\[
\frac{\partial \ln p(x; \theta)}{\partial B} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)n
\]
\[
\frac{\partial^2 \ln p(x; \theta)}{\partial A^2} = -\frac{N}{\sigma^2}
\]
\[
\frac{\partial^2 \ln p(x; \theta)}{\partial A \partial B} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} n
\]
\[
\frac{\partial^2 \ln p(x; \theta)}{\partial B^2} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^2
\]

\[
I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix}
N & \sum_{n=0}^{N-1} n \\
\sum_{n=0}^{N-1} n & \sum_{n=0}^{N-1} n^2
\end{bmatrix}
\]
\[
= \frac{1}{\sigma^2} \begin{bmatrix}
N & \frac{N(N-1)}{2} \\
\frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6}
\end{bmatrix}
\]

\[
I^{-1}(\theta) = \sigma^2 \begin{bmatrix}
\frac{2(2N-1)}{N(N+1)} & \frac{6}{N(N+1)} \\
\frac{6}{N(N+1)} & \frac{12}{N(N^2-1)}
\end{bmatrix}
\]

\[
\text{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)}
\]
\[
\text{var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}.
\]

Remarks:
1. When \( B \) is known, we have a lower CRLB.
2. \( B \) is easier to estimate for \( N \geq 3 \).

\[
\frac{\text{CRLB}(\hat{A})}{\text{CRLB}(\hat{B})} = \frac{(2N-1)(N-1)}{6} > 1
\]

\textit{CRLB always increases as we estimate more parameters.}
\( \Delta x[n] \approx \frac{\partial x[n]}{\partial A} \Delta A = \Delta A \)
\( \Delta x[n] \approx \frac{\partial x[n]}{\partial B} \Delta B = n \Delta B \)

**Theorem 3.2 (Cramer-Rao Lower Bound - Vector Parameter)** It is assumed that the PDF \( p(x; \theta) \) satisfies the "regularity" conditions

\[
E \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta} \right] = 0 \quad \text{for all } \theta
\]

where the expectation is taken with respect to \( p(x; \theta) \). Then, the covariance matrix of any unbiased estimator \( \hat{\theta} \) satisfies

\[
C_\theta - I^{-1}(\theta) \succeq 0 \quad (3.24)
\]

where \( \succeq 0 \) is interpreted as meaning that the matrix is positive semidefinite. The Fisher information matrix \( I(\theta) \) is given as

\[
[I(\theta)]_{ij} = -E \left[ \frac{\partial^2 \ln p(x; \theta)}{\partial \theta_i \partial \theta_j} \right]
\]

where the derivatives are evaluated at the true value of \( \theta \) and the expectation is taken with respect to \( p(x; \theta) \). Furthermore, an unbiased estimator may be found that attains the bound in that \( C_\theta = I^{-1}(\theta) \) if and only if

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = I(\theta)(g(x) - \theta) \quad (3.25)
\]

for some \( p \)-dimensional function \( g \) and some \( p \times p \) matrix \( I \). That estimator, which is the MVU estimator, is \( \hat{\theta} = g(x) \), and its covariance matrix is \( I^{-1}(\theta) \).

**Remark:**

\[
C_\hat{\theta} - I^{-1}(\theta) \succeq 0 \quad (\text{positive semi-definite})
\]

\[
\Rightarrow \quad [C_\hat{\theta} - I^{-1}(\theta)]_{ii} \geq 0 \Rightarrow \quad \text{var}(\hat{\theta}_i) = [C_\hat{\theta}]_{ii} \geq [I^{-1}(\theta)]_{ii}
\]
Example revisited: Line Fitting

\[ x[n] = A + Bn + w[n] \]

\[ w[n]: \text{WGN} \]

\[ \theta = \begin{bmatrix} A \\ B \end{bmatrix} \]

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial \ln p(x; \theta)}{\partial A} \\ \frac{\partial \ln p(x; \theta)}{\partial B} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn) \\ \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)n \end{bmatrix}
\]

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \begin{bmatrix} \frac{N}{\sigma^2} & \frac{N(N - 1)}{2\sigma^2} \\ \frac{N(N - 1)}{2\sigma^2} & \frac{N(N - 1)(2N - 1)}{6\sigma^2} \end{bmatrix} \begin{bmatrix} \hat{A} - A \\ \hat{B} - B \end{bmatrix}
\]

\[
\hat{A} = \frac{2(2N - 1)}{N(N + 1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N + 1)} \sum_{n=0}^{N-1} nx[n]
\]

\[
\hat{B} = -\frac{6}{N(N + 1)} \sum_{n=0}^{N-1} x[n] + \frac{12}{N(N^2 - 1)} \sum_{n=0}^{N-1} nx[n]
\]

Efficient and thus MVU!!!
Vector Parameter CRLB for Transformations

Wish to estimate $\alpha = g(\theta)$. $g$: r-dimensional function

$$C_\alpha = \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^T}{\partial \theta} \geq 0 \quad \text{or} \quad \text{var}(\hat{\alpha}) \geq \left[ \frac{\partial g}{\partial \theta} I^{-1}(\theta) \frac{\partial g^T}{\partial \theta} \right]_{ii}$$

$$\frac{\partial g(\theta)}{\partial \theta} = \begin{bmatrix}
\frac{\partial g_1(\theta)}{\partial \theta_1} & \frac{\partial g_1(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_1(\theta)}{\partial \theta_p} \\
\frac{\partial g_2(\theta)}{\partial \theta_1} & \frac{\partial g_2(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_2(\theta)}{\partial \theta_p} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial g_r(\theta)}{\partial \theta_1} & \frac{\partial g_r(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_r(\theta)}{\partial \theta_p}
\end{bmatrix} \quad r \times p \text{ Jacobian matrix}$$

Example: $x[n] = A + w[n]$ $\quad w[n]:$ WGN with variance $\sigma^2$

$A$ and $\sigma$ are unknown.

Wish to estimate the signal-to-noise ratio $\alpha = \frac{A^2}{\sigma^2}$

$$I(\theta) = \begin{bmatrix}
N & 0 \\
0 & \frac{N}{2\sigma^4}
\end{bmatrix}$$

$$\frac{\partial g(\theta)}{\partial \theta} = \begin{bmatrix}
\frac{\partial g(\theta)}{\partial \theta_1} \\
\frac{\partial g(\theta)}{\partial \theta_2}
\end{bmatrix} = \begin{bmatrix}
\frac{\partial g(\theta)}{\partial A} \\
\frac{\partial g(\theta)}{\partial \sigma^2}
\end{bmatrix} = \begin{bmatrix}
\frac{2A}{\sigma^2} \\
\frac{-A^2}{\sigma^4}
\end{bmatrix}$$

$$\frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^T}{\partial \theta} = \begin{bmatrix}
\frac{2A}{\sigma^2} & \frac{-A^2}{\sigma^4} \\
\frac{\sigma^2}{N} & \frac{2\sigma^4}{N}
\end{bmatrix} \begin{bmatrix}
\frac{\sigma^2}{N} & 0 \\
0 & \frac{2\sigma^4}{N}
\end{bmatrix} \begin{bmatrix}
\frac{2A}{\sigma^2} \\
\frac{-A^2}{\sigma^4}
\end{bmatrix} = \frac{4A^2}{N\sigma^2} + \frac{2A^4}{N\sigma^4} = \frac{4\alpha + 2\alpha^2}{N}$$

$$\text{var}(\hat{\alpha}) \geq \frac{4\alpha + 2\alpha^2}{N}.$$
CRLB for the General Gaussian Case

\[ x \sim N(\mu(\theta), C(\theta)) \]

\[
[I(\theta)]_{i,j} = \left[ \frac{\partial \mu(\theta)}{\partial \theta_i} \right]^T C^{-1}(\theta) \left[ \frac{\partial \mu(\theta)}{\partial \theta_j} \right]
+ \frac{1}{2} \text{tr} \left[ C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta_i} C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta_j} \right]
\]

\[
\frac{\partial \mu(\theta)}{\partial \theta_i} = \begin{bmatrix}
\frac{\partial [\mu(\theta)]_1}{\partial \theta_i} \\
\frac{\partial [\mu(\theta)]_2}{\partial \theta_i} \\
\vdots \\
\frac{\partial [\mu(\theta)]_N}{\partial \theta_i}
\end{bmatrix}
\]

\[
\frac{\partial C(\theta)}{\partial \theta_i} = \begin{bmatrix}
\frac{\partial [C(\theta)]_{11}}{\partial \theta_i} & \frac{\partial [C(\theta)]_{12}}{\partial \theta_i} & \cdots & \frac{\partial [C(\theta)]_{1N}}{\partial \theta_i} \\
\frac{\partial [C(\theta)]_{21}}{\partial \theta_i} & \frac{\partial [C(\theta)]_{22}}{\partial \theta_i} & \cdots & \frac{\partial [C(\theta)]_{2N}}{\partial \theta_i} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial [C(\theta)]_{N1}}{\partial \theta_i} & \frac{\partial [C(\theta)]_{N2}}{\partial \theta_i} & \cdots & \frac{\partial [C(\theta)]_{NN}}{\partial \theta_i}
\end{bmatrix}
\]

For the scalar parameter case \( x \sim N(\mu(\theta), C(\theta)) \)

we have

\[
I(\theta) = \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]^T C^{-1}(\theta) \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]
+ \frac{1}{2} \text{tr} \left[ \left( C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta} \right)^2 \right]
\]
Example: $x[n] = s[n; \theta] + w[n]$, $n = 0, 1, \ldots, N-1$. $w[n]$: WGN

$$C = \sigma^2 I$$

$$x[n] \sim N(s[n; \theta], \sigma^2 I)$$

$$I(\theta) = \frac{1}{\sigma^2} \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]^T \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]$$

$$= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial [\mu(\theta)]_n}{\partial \theta} \right)^2$$

$$= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2$$

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial s[n; \theta]}{\partial \theta} \right)^2}$$

Generalizing to a vector signal parameter estimated in the presence of WGN, we have

$$[I(\theta)]_{ij} = \left[ \frac{\partial \mu(\theta)}{\partial \theta_i} \right]^T \frac{1}{\sigma^2} I \left[ \frac{\partial \mu(\theta)}{\partial \theta_j} \right]$$

$$[I(\theta)]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \theta]}{\partial \theta_i} \frac{\partial s[n; \theta]}{\partial \theta_j}$$

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

$w[n]$: WGN with unknown variance $\theta = \sigma^2$

$$C(\sigma^2) = \sigma^2 I$$

$$I(\sigma^2) = \frac{1}{2} \text{tr} \left[ \left( C^{-1}(\sigma^2) \frac{\partial C(\sigma^2)}{\partial \sigma^2} \right)^2 \right]$$

$$= \frac{1}{2} \text{tr} \left[ \left( \frac{1}{\sigma^2} \right)^2 \left( I \right)^2 \right]$$

$$= \frac{1}{2} \text{tr} \left[ \frac{1}{\sigma^4} I \right]$$

$$= \frac{N}{2\sigma^4}$$
Example: $x[n] = A + w[n]$ \quad n = 0, 1, \ldots, N-1

$w[n]$: WGN

A: Gaussian random variable with zero mean and unknown variance $\sigma_A^2$

A is independent of $w[n]$,

$\Rightarrow x = [x[0], x[1], x[2], \ldots, x[N-1]]^T$ is Gaussian with zero mean and an $N \times N$ covariance matrix $C(\sigma_A^2)$ whose $[i,j]$ element is

$$
[C(\sigma_A^2)]_{i,j} = E[x[i-1]x[j-1]]
= E[(A + w[i-1])(A + w[j-1])]
= \sigma_A^2 + \sigma^2 \delta_{ij}.
$$

$C(\sigma_A^2) = \sigma_A^2 I + \sigma^2 I$

Use Woodbury’s identity

$$
C^{-1}(\sigma_A^2) = \frac{1}{\sigma^2} \left( I - \frac{\sigma_A^2}{\sigma^2 + N\sigma_A^2} \cdot 11^T \right)
$$

$$
\frac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = 11^T
$$

$$
C^{-1}(\sigma_A^2) \frac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = \frac{1}{\sigma^2 + N\sigma_A^2} 11^T
$$

$$
I(\sigma_A^2) = \frac{1}{2} \text{tr} \left[ \left( \frac{1}{\sigma^2 + N\sigma_A^2} \right)^2 11^T \cdot 11^T \right]
= \frac{N}{2} \left( \frac{1}{\sigma^2 + N\sigma_A^2} \right)^2 \text{tr}(11^T)
= \frac{1}{2} \left( \frac{N}{\sigma^2 + N\sigma_A^2} \right)^2
$$

$$
\text{var}(\sigma_A^2) \geq 2 \left( \sigma_A^2 + \frac{\sigma^2}{N} \right)^2
$$

As $N \to \infty$, CRLB $\to 2\sigma_A^4$
Asymptotic CRLB for WSS Gaussian Random Processes

At times, it is difficult to analytically compute the CRLB due to the need to invert the covariance matrix.
⇒ An alternative form can be applied to WSS Gaussian processes.

Assume $x[n]$ is a zero-mean Gaussian process with the autocorrelation function $r_{xx}[k] = W[x[n]x[n+k]]$.
Assume the data record length $N$ is much greater than the correlation time of the process.

\[
[I(\theta)]_{ij} = \frac{N}{2} \int_{-F_c}^{F_c} \frac{1}{2} \frac{\partial \ln P_{xx}(f; \theta)}{\partial \theta_i} \frac{\partial \ln P_{xx}(f; \theta)}{\partial \theta_j} df
\]

$P_{xx}(f; \theta) = F\{r_{xx}[k]\}$ PSD (Power Spectral Density) of $x[n]$

Example: Center Frequency of Process

\[
P_{xx}(f; f_c) = Q(f - f_c) + Q(-f - f_c) + \sigma^2
\]

Assume $Q(f)$ and $\sigma^2$ are known. Wish to estimate $f_c$. 

Figure 3.6 Signal PSD for center frequency estimation
The possible center frequencies are constrained to be in the interval \([f_1, 0.5- f_2]\) so that \(Q(f-f_c)\) is always within \([0, 0.5]\).

\[
\text{var}(\hat{f}_c) \geq \frac{1}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{\partial \ln{P_{xx}(f; f_c)}}{\partial f_c} \right)^2 \, df
\]

\[
\frac{\partial \ln{P_{xx}(f; f_c)}}{\partial f_c} = \frac{\partial \ln{[Q(f-f_c) + Q(-f-f_c) + \sigma^2]}}{\partial f_c} = \frac{\partial Q(f-f_c)}{\partial f_c} + \frac{\partial Q(-f-f_c)}{\partial f_c}
\]

\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{\partial \ln{P_{xx}(f; f_c)}}{\partial f_c} \right)^2 \, df = 2 \int_{0}^{\frac{1}{2}} \left( \frac{\partial \ln{P_{xx}(f; f_c)}}{\partial f_c} \right)^2 \, df
\]

\[
\text{var}(\hat{f}_c) \geq \frac{1}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{\partial Q(f)}{\partial f} \right)^2 \left( \frac{Q(f) + \sigma^2}{Q(f)} \right) \, df
\]

For \(Q(f) = \exp \left[-\frac{1}{2} \left( \frac{f}{\sigma_f} \right)^2 \right]\) where \(\sigma_f << 0.5\)

and \(Q(f) >> \sigma^2\),

\[
\text{var}(\hat{f}_c) \geq \frac{1}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{f^2}{\sigma_f^4} \, df = \frac{12\sigma_f^4}{N}
\]

Narrower PSDs ⇒ better accuracy
Signal Processing Examples

Example: Range Estimation

$s(t)$: nonzero over $[0,T_s]$

$x(t) = s(t-\tau_0) + w(t)$

$\tau_0 = 2R/c \quad R$: range, $c$: the speed of propagation

$w(t)$: Gaussian noise

Assume the maximum time delay is $\tau_{0_{\text{max}}}$. 

$\Rightarrow$ observation interval $\quad 0 \leq t \leq T = T_s + \tau_{0_{\text{max}}}$
\[ \Delta = 1/(2B) \]

\[ x(n\Delta) = s(n\Delta - \tau_0) + w(n\Delta) \quad n = 0, 1, \ldots, N - 1. \]

\[ x[n] = s(n\Delta - \tau_0) + w[n]. \]

Note that \( w[n] \) is WGN with variance \( \sigma^2 = r_{ww}(0) = N_0 B \)

\[ x[n] = \begin{cases} 
  w[n] & 0 \leq n \leq n_0 - 1 \\
  s(n\Delta - \tau_0) + w[n] & n_0 \leq n \leq n_0 + M - 1 \\
  w[n] & n_0 + M \leq n \leq N - 1 
\end{cases} \]

\( M: \) length of the sampled signal

\( n_0 = \tau_0/\Delta: \) delay in samples

\[
\text{var}(\hat{\tau}_0) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left( \frac{\partial s[n; \tau_0]}{\partial \tau_0} \right)^2} \\
= \frac{\sigma^2}{\sum_{n=n_0}^{n_0+M-1} \left( \frac{\partial s(n\Delta - \tau_0)}{\partial \tau_0} \right)^2} \\
= \frac{\sigma^2}{\sum_{n=n_0}^{n_0+M-1} \left( \frac{ds(t)}{dt} \bigg|_{t=n\Delta-\tau_0} \right)^2} \\
= \frac{\sigma^2}{\sum_{n=0}^{M-1} \left( \frac{ds(t)}{dt} \bigg|_{t=n\Delta} \right)^2} \\
\text{var}(\hat{\tau}_0) \geq \frac{\sigma^2}{\frac{1}{\Delta} \int_0^T \left( \frac{ds(t)}{dt} \right)^2 \, dt} \\
\text{var}(\hat{\tau}_0) \geq \frac{\frac{N_0}{2}}{\frac{1}{\Delta} \int_0^T \left( \frac{ds(t)}{dt} \right)^2 \, dt} \"]
\[ \mathcal{E} = \int_0^{T_s} s^2(t) \, dt \]

\[ \text{var}(\hat{r}_0) \geq \frac{1}{\mathcal{E}} \frac{\mathcal{E}}{N_0/2} \]

\[ \overline{F^2} = \frac{\int_0^{T_s} \left( \frac{ds(t)}{dt} \right)^2 \, dt}{\int_0^{T_s} s^2(t) \, dt} \]

\[ \overline{F^2} = \frac{\int_{-\infty}^{\infty} (2\pi F)^2 |S(F)|^2 \, dF}{\int_{-\infty}^{\infty} |S(F)|^2 \, dF} \]

mean square bandwidth

\[ S(F) = F \{ s(t) \} \]

If \( s(t) = \exp \left\{-\frac{1}{2} \sigma_F^2 (t - \frac{T_s}{2})^2 \right\} \), \( |S(F)| = \frac{\sigma_F}{\sqrt{2\pi}} \exp \left\{-2\pi^2 F^2 / \sigma_F^2 \right\} \).

\[ \Rightarrow \overline{F^2} = \frac{\sigma_F^2}{2} \]

\[ \text{var}(\hat{R}) \geq \frac{c^2/4}{\mathcal{E}} \frac{\mathcal{E}}{N_0/2} \frac{1}{\overline{F^2}} \]

For good estimation

(1) large \( \overline{F^2} \)

(2) large SNR.
Example: Sinusoidal Parameter Estimation

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

Want to estimate \(A, f_0,\) and \(\phi,\) with \(A > 0\) and \(0 < f_0 < 0.5.\)

\[
[I(\theta)]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial^2 s[n; \theta]}{\partial \theta_i} \frac{\partial^2 s[n; \theta]}{\partial \theta_j} \quad \text{where} \quad \theta = [A \quad f_0 \quad \phi]^T
\]

\[
[I(\theta)]_{11} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \cos^2 \alpha = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left( \frac{1}{2} + \frac{1}{2} \cos 2\alpha \right) \approx \frac{N}{2\sigma^2}
\]

\[
[I(\theta)]_{12} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} A^2 \pi n \cos \alpha \sin \alpha = -\frac{\pi A}{\sigma^2} \sum_{n=0}^{N-1} n \sin 2\alpha \approx 0
\]

\[
[I(\theta)]_{13} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} A \cos \alpha \sin \alpha = -\frac{A}{2\sigma^2} \sum_{n=0}^{N-1} \sin 2\alpha \approx 0
\]

\[
[I(\theta)]_{22} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} A^2 (2\pi n)^2 \sin^2 \alpha = \frac{(2\pi A)^2}{\sigma^2} \sum_{n=0}^{N-1} n^2 \left( \frac{1}{2} - \frac{1}{2} \cos 2\alpha \right) \\
\approx \frac{(2\pi A)^2}{2\sigma^2} \sum_{n=0}^{N-1} n^2
\]

\[
[I(\theta)]_{23} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} A^2 2\pi n \sin^2 \alpha \approx \frac{\pi A^2}{\sigma^2} \sum_{n=0}^{N-1} n
\]

\[
[I(\theta)]_{33} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} A^2 \sin^2 \alpha \approx \frac{NA^2}{2\sigma^2}
\]

\[
I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix}
\frac{N}{2} & 0 & 0 \\
0 & 2A^2\pi^2 \sum_{n=0}^{N-1} n^2 & \pi A^2 \sum_{n=0}^{N-1} n \\
0 & \pi A^2 \sum_{n=0}^{N-1} n & \frac{NA^2}{2}
\end{bmatrix}
\]

\[
\text{var}(\hat{A}) \geq \frac{2\sigma^2}{N}
\]

\[
\text{var}(\hat{f}_0) \geq \frac{12}{(2\pi)^2 \eta N(N^2 - 1)}
\]

\[
\text{var}(\hat{\phi}) \geq \frac{2(2N - 1)}{\eta N(N + 1)}
\]

where \(\eta = A^2/(2\sigma^2)\)
Linear Models

Definition and Properties

Example: Line Fitting

\[ x[n] = A + Bn + w[n] \quad n = 0, 1, \ldots, N - 1 \]

\[ x = \mathbf{H}\theta + w \]

\[
\begin{align*}
x &= [x[0] \ x[1] \ldots \ x[N - 1]]^T \\
w &= [w[0] \ w[1] \ldots \ w[N - 1]]^T \\
\theta &= [A \ B]^T
\end{align*}
\]

\[
\mathbf{H} = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
\vdots & \vdots \\
1 & N - 1
\end{bmatrix}
\]

observation matrix

\[ w \sim \mathcal{N}(0, \sigma^2 I) \]

Linear Model: \[ x = \mathbf{H}\theta + w \]

Assume \[ w \sim \mathcal{N}(0, \sigma^2 I). \Rightarrow x \sim \mathcal{N}(\mathbf{H}\theta, \sigma^2 I) \]

\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left[ -\ln(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2}(x - \mathbf{H}\theta)^T(x - \mathbf{H}\theta) \right]
\]

\[ = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} \left[ x^T x - 2x^T\mathbf{H}\theta + \theta^T\mathbf{H}^T\mathbf{H}\theta \right]. \]

\[
\begin{align*}
\frac{\partial \mathbf{b}^T \theta}{\partial \theta} &= \mathbf{b} \\
\frac{\partial \theta^T A \theta}{\partial \theta} &= 2A\theta \quad (A \text{ is symmetric}) \\
\frac{\partial \ln p(x; \theta)}{\partial \theta} &= \frac{1}{\sigma^2} [\mathbf{H}^T x - \mathbf{H}^T \mathbf{H}\theta]
\end{align*}
\]
Assume $H^T H$ is invertible.

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

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

$$I(\theta) = \frac{H^T H}{\sigma^2}.$$

**MVU estimator**

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

Covariance matrix:

$$C_\hat{\theta} = I^{-1}(\theta) = \sigma^2 (H^T H)^{-1}$$

Remark: $H^T H$ is invertible $\Rightarrow$ the columns of $H$ are linearly independent.

Counterexample:

$$H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}$$

The model parameters will not be identifiable even in the absence of noise.

![Figure 4.1 Nonidentifiability of linear model parameters](image)

In practice, if $H^T H$ is ill-conditioned or $H$ is approximately rank deficient,

$$C_\hat{\theta} = I^{-1}(\theta) = \sigma^2 (H^T H)^{-1} \rightarrow \infty$$

$\Rightarrow$ Cannot estimate $\theta$ reliably.
Theorem 4.1 (Minimum Variance Unbiased Estimator for the Linear Model)

If the data observed can be modeled as

\[ \mathbf{x} = \mathbf{H}\theta + \mathbf{w} \quad (4.8) \]

where \( \mathbf{x} \) is an \( N \times 1 \) vector of observations, \( \mathbf{H} \) is a known \( N \times p \) observation matrix (with \( N > p \)) and rank \( p \), \( \theta \) is a \( p \times 1 \) vector of parameters to be estimated, and \( \mathbf{w} \) is an \( N \times 1 \) noise vector with PDF \( \mathcal{N}(0, \sigma^2 \mathbf{I}) \), then the MVU estimator is

\[ \hat{\theta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \quad (4.9) \]

and the covariance matrix of \( \hat{\theta} \) is

\[ \mathbf{C}_\theta = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}. \quad (4.10) \]

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

\[ \hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}). \]

- Linear Model Examples

Example: Curve Fitting

![Diagram of voltage vs time with points and curve fitting]

\[ x(t_n) = \theta_1 + \theta_2 t_n + \cdots + \theta_p t_n^{p-1} + w(t_n) \quad n = 0, 1, \ldots, N - 1. \]

Assume \( w(t_n) \) are i.i.d. Gaussian random variables with zero mean and variance \( \sigma^2 \).

\[ \mathbf{x} = \mathbf{H}\theta + \mathbf{w} \]

\[ \mathbf{x} = [x(t_0) \ x(t_1) \ldots x(t_{N-1})]^T \]

\[ \theta = [\theta_1 \ \theta_2 \ldots \theta_p]^T \]
\[
\mathbf{H} = \begin{bmatrix}
1 & t_0 & \cdots & t_0^{p-1} \\
1 & t_1 & \cdots & t_1^{p-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & t_{N-1} & \cdots & t_{N-1}^{p-1}
\end{bmatrix} \quad (N \times p)
\]

\[
\hat{\theta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}
\]

\[
\hat{s}(t) = \sum_{i=1}^{p} \hat{\theta}_i e^{j(t-i-1)}
\]

**Example: Fourier Analysis**

\[
x[n] = \sum_{k=1}^{M} a_k \cos \left( \frac{2\pi kn}{N} \right) + \sum_{k=1}^{M} b_k \sin \left( \frac{2\pi kn}{N} \right) + w[n] \quad n = 0, 1, \ldots, N - 1
\]

\(w[n]: \text{WGN}\)

\[
\theta = [a_1 \ a_2 \ \ldots \ a_M \ b_1 \ b_2 \ \ldots \ b_M]^T
\]

\[
\mathbf{H} = \begin{bmatrix}
1 & \cdots & 1 & 0 & \cdots & 0 \\
\cos \left( \frac{2\pi}{N} \right) & \cdots & \cos \left( \frac{2\pi M}{N} \right) & \sin \left( \frac{2\pi}{N} \right) & \cdots & \sin \left( \frac{2\pi M}{N} \right) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\cos \left( \frac{2\pi(N-1)}{N} \right) & \cdots & \cos \left( \frac{2\pi M(N-1)}{N} \right) & \sin \left( \frac{2\pi(N-1)}{N} \right) & \cdots & \sin \left( \frac{2\pi M(N-1)}{N} \right)
\end{bmatrix} \quad N \times 2M \quad (N > 2M)
\]

\[
\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \ldots \ \mathbf{h}_{2M}]
\]

\[
\mathbf{h}_i^T \mathbf{h}_j = 0 \quad \text{for } i \neq j
\]

\[
\mathbf{H}^T \mathbf{H} = \begin{bmatrix}
\mathbf{h}_1^T \\
\vdots \\
\mathbf{h}_{2M}^T
\end{bmatrix}
\begin{bmatrix}
\mathbf{h}_1 & \mathbf{h}_{2M}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathbf{h}_1^T \mathbf{h}_1 & \mathbf{h}_1^T \mathbf{h}_2 & \cdots & \mathbf{h}_1^T \mathbf{h}_{2M} \\
\mathbf{h}_2^T \mathbf{h}_1 & \mathbf{h}_2^T \mathbf{h}_2 & \cdots & \mathbf{h}_2^T \mathbf{h}_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{h}_{2M}^T \mathbf{h}_1 & \mathbf{h}_{2M}^T \mathbf{h}_2 & \cdots & \mathbf{h}_{2M}^T \mathbf{h}_{2M}
\end{bmatrix}
\]
\[
\sum_{n=0}^{N-1} \cos \left( \frac{2\pi in}{N} \right) \cos \left( \frac{2\pi jn}{N} \right) = \frac{N}{2} \delta_{ij}
\]
\[
\sum_{n=0}^{N-1} \sin \left( \frac{2\pi in}{N} \right) \sin \left( \frac{2\pi jn}{N} \right) = \frac{N}{2} \delta_{ij}
\]
\[
\sum_{n=0}^{N-1} \cos \left( \frac{2\pi in}{N} \right) \sin \left( \frac{2\pi jn}{N} \right) = 0 \quad \text{for all } i, j.
\]

\[
H^T H = \begin{bmatrix}
\frac{N}{2} & 0 & \cdots & 0 \\
0 & \frac{N}{2} & \cdots & 0 \\
0 & 0 & \cdots & \frac{N}{2} \\
\end{bmatrix} = \frac{N}{2} I
\]

\[
\hat{\theta} = (H^T H)^{-1} H^T x
\]
\[
= \frac{2}{N} H^T x = \frac{2}{N} \begin{bmatrix} h_1^T \\ \vdots \\ h_{2M}^T \end{bmatrix} x
\]
\[
= \begin{bmatrix}
\frac{2}{N} h_1^T x \\
\vdots \\
\frac{2}{N} h_{2M}^T x
\end{bmatrix}
\sim \text{Gaussian}
\]

\[
\hat{a}_k = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos \left( \frac{2\pi kn}{N} \right)
\]
\[
\hat{b}_k = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin \left( \frac{2\pi kn}{N} \right)
\]

Discrete Fourier Transform Coefficients.

\[
E(\hat{a}_k) = a_k
\]
\[
E(\hat{b}_k) = b_k
\]

\[
C_\theta = \sigma^2 (H^T H)^{-1}
\]
\[
= \sigma^2 \left( \frac{N}{2} I \right)^{-1}
\]
\[
= \frac{2\sigma^2}{N} I.
\]

The amplitude estimates are independent!!
Example: System Identification

A common system model is the tapped delay line (TDL) or finite impulse response (FIR) filter.

![Tapped delay line diagram]

We provide \( u[n] \) and measure \( x[n] \). Wish to estimate the weights \( h[n] \).

If \( u[n] \) is provided for \( n = 0, 1, .., N-1 \) and \( u[n] = 0 \) for \( n < 0 \), we observe

\[
x[n] = \sum_{k=0}^{p-1} h[k]u[n - k] + w[n] \quad n = 0, 1, \ldots, N - 1
\]

We provide \( u[n] \) and measure \( x[n] \). Wish to estimate the weights \( h[n] \).

If \( u[n] \) is provided for \( n = 0, 1, .., N-1 \) and \( u[n] = 0 \) for \( n < 0 \), we observe

\[
x[n] = \sum_{k=0}^{p-1} h[k]u[n - k] + w[n] \quad n = 0, 1, \ldots, N - 1
\]

\[
x = \begin{bmatrix}
  u[0] & 0 & \ldots & 0 \\
  u[1] & u[0] & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  u[N - 1] & u[N - 2] & \ldots & u[N - p]
\end{bmatrix}
\begin{bmatrix}
  h[0] \\
  h[1] \\
  \vdots \\
  h[p - 1]
\end{bmatrix} + w
\]

If \( w[n] \) is WGN with variance \( \sigma^2 \), \( w \sim N(0, \sigma^2 I) \).

\[
\hat{\theta} = (H^T H)^{-1} H^T x. \quad \text{MVU and efficient}
\]

\[
C_{\hat{\theta}} = \sigma^2 (H^T H)^{-1}
\]
Note that how well we can estimate the tap weights depends on the probing signal $u[n]$.

$\Rightarrow$ The signal should be chosen to be a pseudorandom noise (PRN).

Proof:

$$\text{var}(\hat{\theta}_i) = e_i^T C_\theta e_i \quad e_i = [0 \ldots 0 \ 1 \ldots 0]^T$$

$C_\theta^{-1}$ can be factored as $D^T D$ with $D$ an invertible $p \times p$ matrix.

Note that $1 = (e_i^T D^T D^{-1} e_i)^2$.

Let $\xi_1 = De_i$ and $\xi_2 = D^{-1} e_i$,

$$(\xi_1^T \xi_2)^2 \leq \xi_1^T \xi_1 \xi_2^T \xi_2$$

$$1 \leq (e_i^T D^T De_i)(e_i^T D^{-1} D^{-1} e_i)$$

$$= (e_i^T C_\theta^{-1} e_i)(e_i^T C_\theta^{-1} e_i)$$

$$\text{var}(\hat{\theta}_i) \geq \frac{1}{e_i^T C_\theta^{-1} e_i} = \frac{\sigma^2}{[H^T H]_{ii}}$$

The minimum variance is attained if $De_i = c D^{-1} e_i$

$\Rightarrow D^T De_i = c_i e_i \quad i = 1, 2, \ldots, p.$

$$D^T D = C_\theta^{-1} = \frac{H^T H}{\sigma^2}$$

$$\frac{H^T H}{\sigma^2} e_i = c_i e_i$$

$$H^T H = \sigma^2 \begin{bmatrix} c_1 & 0 & \ldots & 0 \\ 0 & c_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & c_p \end{bmatrix}$$

To minimize the variance of the MVU estimator, $u[n]$ should be chosen to make $H^T H$ diagonal.
Since \( \mathbf{H}_{ij} = u[i - j] \)

\[
[\mathbf{H}^T \mathbf{H}]_{ij} = \sum_{n=1}^{N} u[n-i]u[n-j] \quad i = 1, 2, \ldots, p; \quad j = 1, 2, \ldots, p
\]

For large \( N \),

\[
[\mathbf{H}^T \mathbf{H}]_{ij} \approx \sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|]
\]

\[
\mathbf{H}^T \mathbf{H} = N \begin{bmatrix}
  r_{uu}[0] & r_{uu}[1] & \cdots & r_{uu}[p-1] \\
  r_{uu}[1] & r_{uu}[0] & \cdots & r_{uu}[p-2] \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{uu}[p-1] & r_{uu}[p-2] & \cdots & r_{uu}[0]
\end{bmatrix}
\]

\[
r_{uu}[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} u[n]u[n+k] \sim \text{autocorrelation function of } u[n]
\]

\[
r_{uu}[k] = 0 \quad k \neq 0 \quad \text{white noise.}
\]

Actually, we use pseudo random noise (PRN). Sequence.

\[
\mathbf{H}^T \mathbf{H} = Nr_{uu}[0]I
\]

\[
\operatorname{var}(\hat{h}[i]) = \frac{1}{N r_{uu}[0]/\sigma^2} \quad i = 0, 1, \ldots, p - 1
\]

The TDL weight estimators are independent.

\[
\hat{\theta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}
\]

\[
\hat{h}[i] = \frac{1}{N r_{uu}[0]} \sum_{n=0}^{N-1} u[n-i]x[n]
\]

\[
= \frac{1}{N} \sum_{n=0}^{N-1-i} u[n]x[n+i]
\]

\[
= \frac{r_{uu}[0]}{r_{uu}[0]}
\]
Extension to the Linear Model

A more general linear model allows for noise that is not white.

\[ x = H\theta + w \]

\[ w \sim \mathcal{N}(0, C) \]

To find MVU estimator,

1. Repeat \( \frac{\partial \ln p(x; \theta)}{\partial \theta} \) derivation

2. Whitening approach

Whitening approach

Since \( C \) is positive definite, \( C^{-1} \) is positive definite.

\[ C^{-1} = D^T D \]

where \( D \) is an \( N \times N \) invertible matrix.

\( D \) acts as a whitening transformation.

\[
E \left[ (Dw)(Dw)^T \right] = DC \cdot C^T = DD^{-1}D^{-1} = I.
\]

\[
x' = Dx = DH\theta + Dw = H'\theta + w'
\]

\[
w' = Dw \sim \mathcal{N}(0, I).
\]

\[
\hat{\theta} = \left( H'^T H' \right)^{-1} H'^T x'
\]

\[
= \left( H'^T D^T D H \right)^{-1} H'^T D^T D x
\]

\[
\hat{\theta} = \left( H'^T C^{-1} H \right)^{-1} H'^T C^{-1} x.
\]

\[
C_{\hat{\theta}} = \left( H'^T H' \right)^{-1}
\]

\[
C_{\hat{\theta}} = \left( H'^T C^{-1} H \right)^{-1}
\]

Estimation of signal parameter in colored noise environment.
Example: \( x[n] = A + w[n], \ n = 0, 1, \ldots, N-1. \)

\( w[n] \): colored Gaussian noise with covariance matrix \( C \)

\[
\hat{A} = (H^T C^{-1} H)^{-1} H^T C^{-1} x
\]

\[
= \frac{1^T C^{-1} x}{1^T C^{-1} 1}
\]

\[
\text{var}(\hat{A}) = (H^T C^{-1} H)^{-1}
\]

\[
= \frac{1}{1^T C^{-1} 1}.
\]

Recall \( C^{-1} = D^T D \)

\[
\hat{A} = \frac{1^T D^T D x}{1^T D^T D 1}
\]

\[
= \frac{(D1)^T x'}{1^T D^T D 1}
\]

\[
= \sum_{n=0}^{N-1} d_n x'[n]
\]

\[
d_n = [D1]_n / 1^T D^T D 1.
\]

The data are first prewhitened to form \( x'[n] \) and then averaged using prewhitened averaging weights \( d_n \).

**Another extension to the linear model:** \( x = H\theta + s + w \)

\( s \): known signal

Let \( x' = x - s \).

\[
x' = H\theta + w
\]

\[
\hat{\theta} = (H^T H)^{-1} H^T (x - s)
\]

\[
C_{\theta} = \sigma^2 (H^T H)^{-1}
\]
Example: DC level and Exponential in White Noise
\[ x[n] = A + r^n + w[n], \ n = 0, 1, \ldots, N-1 \]
r: known
w[n]: WGN
A is to be estimated.

\[
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1
\end{bmatrix}
\begin{bmatrix}
x \\
1
\end{bmatrix}
= A + s + w
\]

\[ s = [1 \ r \ \ldots \ r^{N-1}]^T \]

\[
\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - r^n)
\]

\[
\text{var}(\hat{A}) = \frac{\sigma^2}{N}
\]

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

\[
x = H\theta + s + w \tag{4.30}
\]

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

\[
\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} (x - s) \tag{4.31}
\]

and the covariance matrix is

\[
C_{\hat{\theta}} = (H^T C^{-1} H)^{-1}. \tag{4.32}
\]

For the general linear model the MVU estimator is efficient in that it attains the CRLB.
General Minimum Variance Unbiased Estimation

Sufficient Statistics

If an efficient estimator does not exist, it is still of interest to find the MVU estimator.
Recall the problem of estimating a DC level $A$ in WGN.

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

is the MVU estimator with variance $\sigma^2/N$.

Consider $\hat{A} = x[0]$ as an estimator.

$$E[\hat{A}] = A, \ Var[\hat{A}] = \sigma^2$$

The variance is much larger because the data $\{x[1], x[2], \ldots, x[N-1]\}$ are not used.

Which data are important or sufficient for estimation?
Consider

$$S_1 = \{x[0], x[1], \ldots, x[N - 1]\}$$

$$S_2 = \{x[0] + x[1], x[2], x[3], \ldots, x[N - 1]\}$$

$$S_3 = \left\{ \sum_{n=0}^{N-1} x[n] \right\}.$$ 

All sets are sufficient since $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ may be found.

Once the sufficient statistics are known, we no longer need the individual data values. All information has been summarized in the sufficient statistic.

The sufficient data set that contains the least number of elements is called the **minimal** one.

$$S_3 = \left\{ \sum_{n=0}^{N-1} x[n] \right\}$$

is the **minimal sufficient statistic**.
To quantify these ideas, consider

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

and assume \( T(x) = \sum_{n=0}^{N-1} x[n] = T_0 \) has been observed.

\[ p(x | \sum_{n=0}^{N-1} x[n] = T_0; A) \]: pdf of the observations after the sufficient statistic has been observed.

\[ \Rightarrow \quad p(x | \sum_{n=0}^{N-1} x[n] = T_0; A) \] will not depend on \( A \).

Otherwise, the data will provide additional information about \( A \).

To show that \( S_3 = \left\{ \sum_{n=0}^{N-1} x[n] \right\} \) is a sufficient statistics, we compute the conditional pdf:

\[ p(x | T(x) = T_0; A) = \frac{p(x, T(x) = T_0; A)}{p(T(x) = T_0; A)} \quad \text{where} \quad T(x) = \sum_{n=0}^{N-1} x[n] \]

\[ p(x | T(x) = T_0; A) = \frac{p(x; A) \delta(T(x) - T_0)}{p(T(x) = T_0; A)} \]

\[ T(x) \sim \mathcal{N}(NA, N\sigma^2) \]
\[ p(x; A) \delta(T(x) - T_0) \]

\[ = \frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \delta(T(x) - T_0) \]

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

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

\[ p(x|T(x) = T_0; A) \]

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

\[ = \frac{\sqrt{N}}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x[n] \right] \exp \left[ \frac{T_0^2}{2\sigma^2} \right] \delta(T(x) - T_0) \]

\[ \Rightarrow p(x|T(x) = T_0; A) \text{ doesn’t depend on } A. \]

\section*{Finding Sufficient Statistics}

\textbf{Theorem 5.1 (Neyman-Fisher Factorization)} If we can factor the PDF \( p(x; \theta) \) as

\[ p(x; \theta) = g(T(x), \theta)h(x) \] (5.3)

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 \( \theta \). Conversely, if \( T(x) \) is a sufficient statistic for \( \theta \), then the PDF can be factored as in (5.3).

Proof:

\((\Rightarrow)\) Assume the factorization holds.

\[ p(x|T(x) = T_0; \theta) = \frac{p(x, T(x) = T_0; \theta)}{p(T(x) = T_0; \theta)} \]

\[ = \frac{p(x; \theta) \delta(T(x) - T_0)}{p(T(x) = T_0; \theta)}. \]
\[ p(x|T(x) = T_0; \theta) = \frac{g(T(x) = T_0, \theta)h(x)\delta(T(x) - T_0)}{p(T(x) = T_0; \theta)} \]

\[ p(T(x) = T_0; \theta) = \int p(x; \theta)\delta(T(x) - T_0) \, dx. \]

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

\[ p(x|T(x) = T_0; \theta) = \frac{h(x)\delta(T(x) - T_0)}{\int h(x)\delta(T(x) - T_0) \, dx} \text{ does not depend on } \theta. \]

⇒ T(x) is a sufficient statistic.

(⇐) Assume T(x) is a sufficient statistic.

Consider the joint pdf

\[ p(x, T(x) = T_0; \theta) = p(x|T(x) = T_0; \theta)p(T(x) = T_0; \theta) \]

\[ p(x, T(x) = T_0; \theta) = p(x; \theta)\delta(T(x) - T_0) \]

Since T(x) is sufficient statistic,

\[ p(x|T(x) = T_0; \theta) = p(x|T(x) = T_0) \]

Moreover, T(x) = T_0 defines a surface in \( \mathbb{R}^N \). The conditional pdf is nonzero only on that surface.

\[ p(x|T(x) = T_0) = w(x)\delta(T(x) - T_0) \]

where

\[ \int w(x)\delta(T(x) - T_0) \, dx = 1 \]

\[ p(x; \theta)\delta(T(x) - T_0) = w(x)\delta(T(x) - T_0)p(T(x) = T_0; \theta) \]

We can let

\[ w(x) = \frac{h(x)}{\int h(x)\delta(T(x) - T_0) \, dx} \]

\[ p(x; \theta)\delta(T(x) - T_0) = \frac{h(x)\delta(T(x) - T_0)}{\int h(x)\delta(T(x) - T_0) \, dx}p(T(x) = T_0; \theta) \]
where \[ g(T(x) = T_0; \theta) = \frac{p(T(x) = T_0; \theta)}{\int h(x) \delta(T(x) - T_0) \, dx}. \]

Moreover, the pdf of the sufficient statistic can be found based on the factorization.

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

Example: DC Level in WGN

\[
\sum_{n=0}^{N-1} (x[n] - A)^2 = \sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] + NA^2
\]

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

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

Example: Power of WGN

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

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

Example: Phase of Sinusoid

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

A, \( f_0 \): known. \( w[n] \): WGN

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

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

It doesn’t appear that the pdf is factorable as required by the Neyman-Fisher theorem. No single sufficient statistic exists. However, the pdf can be factored as

\[
p(x; \phi) = \frac{1}{(2\pi \sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi) - 2AT_1(x) \cos \phi + 2AT_2(x) \sin \phi \right] \right\} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right\} \cdot g(T_1(x), T_2(x), \phi) \cdot h(x) \]

\[ T_1(x) = \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \]

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

\( T_1(x) \) and \( T_2(x) \) are jointly sufficient statistics for the estimation of \( \phi \).

The r statistics \( T_1(x), T_2(x), \ldots, T_r(x) \) are jointly sufficient statistics if the conditional pdf \( p(x \mid T_1(x), T_2(x), \ldots, T_r(x), \theta)h(x) \) does not depend on \( \theta \).

\[
p(x; \theta) = g(T_1(x), T_2(x), \ldots, T_r(x), \theta)h(x) \]

Remark: The original data are always sufficient statistics

\[ T_{n+1}(x) = x[n] \quad n = 0, 1, \ldots, N - 1 \]

\[ g = p(x; \theta) \]

\[ h = 1 \]
Using Sufficiency to Find the MVU Estimator

Example: DC Level in WGN

Sufficient statistic: \( T(x) = \sum_{n=0}^{N-1} x[n] \)

Based on \( T(x) \), there are two ways to find the MVU estimator.

1. Find any unbiased estimator of \( A \), say \( \tilde{A} = x[0] \), and determine
   \[ \hat{A} = E[\tilde{A} | T] \]

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

Approach 1: \( \hat{A} = E\{x[0]\} \sum_{n=0}^{N-1} x[n] \}

For a Gaussian random vector \([x y]^T\) with mean vector \( \mu = [E(x) E(y)]^T \) and covariance matrix

\[
C = \begin{bmatrix}
\text{var}(x) & \text{cov}(x, y) \\
\text{cov}(y, x) & \text{var}(y)
\end{bmatrix}
\]

\[
E(x|y) = \int_{-\infty}^{\infty} x p(x|y) \, dx
\]

\[
= \int_{-\infty}^{\infty} x \frac{p(x, y)}{p(y)} \, dx
\]

\[
= E(x) + \frac{\text{cov}(x, y)}{\text{var}(y)} (y - E(y))
\]

Let \( x = x[0] \) and \( y = \sum_{n=0}^{N-1} x[n] \).

\[
\begin{bmatrix}
x \\
y
\end{bmatrix} = \begin{bmatrix}
x[0] \\
\sum_{n=0}^{N-1} x[n]
\end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix}
x[0] \\
x[1] \\
\vdots \\
x[N-1]
\end{bmatrix}
\]

Since \([x y]^T\) is a linear transformation of a Gaussian random vector, the pdf of \([x y]^T\) is \( N(\mu, C) \).
\[
\begin{align*}
\mu &= LE(x) = LA1 = \left[ \begin{array}{c} A \\ NA \end{array} \right] \\
C &= \sigma^2 LL^T = \sigma^2 \left[ \begin{array}{cc} 1 & 0 \\ 1 & N \end{array} \right].
\end{align*}
\]

\[
\hat{A} = E(x|y) = A + \frac{\sigma^2}{N\sigma^2} \left( \sum_{n=0}^{N-1} x[n] - NA \right)
\]

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

Approach 2: \[\hat{A} = g\left(\sum_{n=0}^{N-1} x[n]\right)\]

Choose a function \(g\) so that \(\hat{A}\) is unbiased.

By inspection, \(g(x) = x/N\).

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

Much easier to apply!!

**Theorem 5.2** *(Rao-Blackwell-Lehmann-Scheffe)* If \(\hat{\theta}\) is an unbiased estimator of \(\theta\) and \(T(x)\) is a sufficient statistic for \(\theta\), then \(\hat{\theta} = E(\hat{\theta}|T(x))\) is

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

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

Remark: A statistic is complete if there is only one function of the statistic that is unbiased.

Assume \(T(x)\) is complete.

1. \(\hat{\theta}\) is unique
2. All \(\bar{\theta}\) produce the same \(\hat{\theta}\)
3. But variance must be decreased.

\[\Rightarrow \hat{\theta}\text{ is the MVU estimator.}\]
Proof of RBLS Theorem:

1. \( \hat{\theta} \) is a valid estimator of \( \theta \) (not a function of \( \theta \))

\[
\hat{\theta} = E(\tilde{\theta}|T(x)) = \int \tilde{\theta}(x)p(x|T(x); \theta) \, dx.
\]

Since \( p(x|T(x); \theta) \) does not depend on \( \theta \), \( \hat{\theta} \) is solely a function of \( T(x) \).

2. \( \hat{\theta} \) is unbiased.

\[
E(\hat{\theta}) = \int \int \tilde{\theta}(x)p(x|T(x); \theta) \, dx \, p(T(x); \theta) \, dT
= \int \tilde{\theta}(x) \int p(x|T(x); \theta)p(T(x); \theta) \, dT \, dx
= \int \tilde{\theta}(x)p(x; \theta) \, dx
= E(\tilde{\theta}).
\]

3. \( \text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta}) \)

\[
\text{var}(\hat{\theta}) = E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right]
= E \left[ (\hat{\theta} - \hat{\theta} + \hat{\theta} - \theta)^2 \right]
= E[(\hat{\theta} - \hat{\theta})^2] + 2E[(\hat{\theta} - \hat{\theta})(\hat{\theta} - \theta)] + E[(\hat{\theta} - \theta)^2]
\]

\[
E_{T,\hat{\theta}}[(\hat{\theta} - \hat{\theta})(\hat{\theta} - \theta)] = E_T E_{\hat{\theta}|T}[(\hat{\theta} - \hat{\theta})(\hat{\theta} - \theta)]
\]

\[
E_{\hat{\theta}|T}[(\tilde{\theta} - \hat{\theta})(\tilde{\theta} - \theta)] = E_{\hat{\theta}|T}[\tilde{\theta} - \hat{\theta}](\tilde{\theta} - \theta)
\]

\[
E_{\hat{\theta}|T}[\hat{\theta} - \tilde{\theta}] = E_{\hat{\theta}|T}(\tilde{\theta}|T) - \tilde{\theta} = \hat{\theta} - \tilde{\theta} = 0
\]

\[
\text{var}(\hat{\theta}) = E[(\hat{\theta} - \tilde{\theta})^2] + \text{var}(\tilde{\theta})
\geq \text{var}(\hat{\theta}).
\]
Remarks:
1. In general, it is difficult to validate the completeness of a sufficient statistic.
2. For the exponential family of PDFs, sufficient statistics are complete.

Example: Completeness of a Sufficient Statistic

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

Wish to show that if \( E[g(T)] = A \) for all \( A \), then there is only one solution for \( g \).
Suppose there exists a second function \( h \) with \( E[h(T)] = A, \forall A \).
\[
E[g(T) - h(T)] = A - A = 0 \quad \text{for all } A
\]

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

where \( v(T) = g(T) - h(T) \).

Letting \( \tau = T/N \) and \( v'(\tau) = v(N\tau) \), we have
\[
\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp \left[ -\frac{N}{2\sigma^2} (A - \tau)^2 \right] d\tau = 0 \quad \text{for all } A
\]
which can be recognized as the convolution of \( v'(\tau) \) with a Gaussian pulse \( w(\tau) \).

(a) Integral equals zero

(b) Integral does not equal zero
But \( W(f) = F\{\text{Gaussian pulse}\} > 0 \quad \forall f \)
\[ \Rightarrow v'(f) = 0 \quad \forall f \]
\[ \Rightarrow v'(t) = 0 \quad \forall t \]
\[ \Rightarrow g = h. \]

Example: Incomplete Sufficient Statistic
\[ x[0] = A + w[0], \text{ where } x[0] \sim U[-0.5, 0.5] \]
x[0] is a sufficient statistic
x[0] is an unbiased estimator of A.
g(x[0]) = x[0].
Assume we have another function h for which \( E\{h(x[0])\} = A \).
Let \( v(T) = g(T) - h(T) \), where \( T = x[0] \).
\[ \int_{-\infty}^{\infty} v(T)p(T; A) \, dT = 0 \quad \text{for all } A \]

\[ p(T; A) = \begin{cases} 
1 & A - \frac{1}{2} \leq T \leq A + \frac{1}{2} \\
0 & \text{otherwise}
\end{cases} \]
\[ \int_{A - \frac{1}{2}}^{A + \frac{1}{2}} v(T) \, dT = 0 \quad \text{for all } A \]

The nonzero function \( v(T) = \sin(2\pi T) \) will satisfy this condition.

Hence, a solution is \( v(T) = g(T) - h(T) = \sin(2\pi T) \) or \( h(T) = T - \sin(2\pi T) \).
\[ \Rightarrow \hat{A} = x[0] - \sin 2\pi x[0] \] is also based on the sufficient statistic and is unbiased for A.
\[ \Rightarrow \text{The RBLS theorem no longer holds. It is not possible to assert that} \]
\[ \hat{A} = x[0] \text{ is the MVU estimator.} \]
A sufficient statistic is complete if the condition
\[
\int_{-\infty}^{\infty} v(T)p(T; \theta) \,dT = 0 \quad \text{for all } \theta
\]
is satisfied only by \( v(T) = 0 \) for all \( T \).

**Procedure for finding MVU estimator.**

1. Find a single sufficient statistic for \( \theta \) by using the Neyman-Fisher factorization theorem.
2. Determine if the sufficient statistic is complete. If so, proceed; otherwise, this approach cannot be used.
3. Find a function \( g \) of the sufficient statistic that yields an unbiased estimator \( \hat{\theta} = g(T(x)) \). The MVU estimator is then \( \hat{\theta} \).

(Alternative implementation)

3'. Evaluate \( \hat{\theta} = E(\bar{\theta} | T(x)) \), where \( \bar{\theta} \) is any unbiased estimator.
Example: Mean of Uniform Noise
\[ x[n] = w[n], \ n = 0, 1, \ldots, N-1 \]
where \( w[n] \) is i.i.d. noise with pdf \( u[0, \beta] \) for \( \beta > 0 \).
Wish to find the MVU estimator for the mean \( \theta = \beta/2 \).
Since the pdf doesn’t satisfy the regularity condition, we cannot adopt
the CRLB approach.
\[
\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \text{ is an unbiased estimator.}
\]
\[
\text{var}(\hat{\theta}) = \frac{1}{N} \text{var}(x[n])
= \frac{\beta^2}{12N}.
\]
\[
p(x[n]; \theta) = \frac{1}{\beta} [u(x[n]) - u(x[n] - \beta)]
\]
where \( u(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x < 0. \end{cases} \)
\[
p(x; \theta) = \frac{1}{\beta^N} \prod_{n=0}^{N-1} [u(x[n]) - u(x[n] - \beta)]
\]
\[
p(x; \theta) = \begin{cases} \frac{1}{\beta^N} & 0 < x[n] < \beta \quad n = 0, 1, \ldots, N - 1 \\ 0 & \text{otherwise.} \end{cases}
\]
\[
p(x; \theta) = \begin{cases} \frac{1}{\beta^N} \max x[n] < \beta, \min x[n] > 0 \\ 0 & \text{otherwise} \end{cases}
\]
\[
p(x; \theta) = \frac{1}{\beta^N} u(\beta - \max x[n]) u(\min x[n])
\]
\[
g(T(x), \theta) \quad h(x)
\Rightarrow T(x) = \max(x[n]) \text{ is a sufficient statistic for } \theta.
\]
It can also be shown that \( T(x) = \max(x[n]) \) is complete.
Next, we must determine a function $g$ so that $g(\max \{x[n]\})$ is unbiased.

\[
\Pr\{T \leq \xi\} = \prod_{n=0}^{N-1} \Pr\{x[n] \leq \xi\} = \Pr\{x[n] \leq \xi\}^N
\]

\[
p_T(\xi) = \frac{d\Pr\{T \leq \xi\}}{d\xi} = N \Pr\{x[n] \leq \xi\}^{N-1} \frac{d\Pr\{x[n] \leq \xi\}}{d\xi}
\]

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

\[
\Pr\{x[n] \leq \xi\} = \begin{cases} 
0 & \xi < 0 \\
\frac{\xi}{\beta} & 0 < \xi < \beta \\
1 & \xi > \beta
\end{cases}
\]

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

\[
E(T) = \int_{-\infty}^{\infty} \xi p_T(\xi) \, d\xi
\]

\[
= \int_{0}^{\beta} \xi N \left(\frac{\xi}{\beta}\right)^{N-1} \frac{1}{\beta} \, d\xi
\]

\[
= \frac{N}{N+1} \beta
\]

\[
= \frac{2N}{N+1} \theta.
\]

⇒ The MVU estimator is

\[
\hat{\theta} = \frac{N + 1}{2N} \max x[n].
\]
\[
\text{var}(\hat{\theta}) = \left(\frac{N + 1}{2N}\right)^2 \text{var}(T)
\]

\[
\text{var}(T) = \int_{0}^{1} x^2 \frac{N x^{N-1}}{\beta^N} \, dx = \left(\frac{N \beta}{N + 1}\right)^2
\]

\[
= \frac{N \beta^2}{(N + 1)^2(N + 2)}
\]

\[
\text{var}(\hat{\theta}) = \frac{\beta^2}{4N(N + 2)}.
\]

Remarks:

1. The sample mean is not the MVU estimator of the mean for uniformly distributed noise. For sample mean, $\text{var}(\hat{\theta}) = \frac{\beta^2}{12N}$

2. In the phase estimation example, we require two sufficient statistics, $T_1(x)$ and $T_2(x)$. We need to evaluate $\hat{\phi} = E(\phi|T_1, T_2)$ or find $g(T_1, T_2)$ that is unbiased. $\Rightarrow$ difficult!!

Extension to a Vector Parameter

Assume $\theta$ is a $p \times 1$ vector. A vector statistic $T(x) = [T_1(x), T_2(x), \ldots, T_r(x)]^T$ is sufficient for the estimation of $\theta$ if $p(x|T(x); \theta)$ does not depend on $\theta$.

It is possible that

(1) $r > p$ more sufficient statistics than parameters.

(2) $r = p$ equal sufficient statistics as parameters

(3) $r < p$ less sufficient statistics than parameters.

$r = p \Rightarrow$ allows us to determine the MVU estimator by transforming the sufficient statistics to be unbiased.
Example: Sinusoidal Parameter Estimation

A sinusoidal signal in WGN

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

If \( A, f_0, \) and \( \sigma^2 \) are unknown,
\[
\theta = [A f_0 \sigma^2]^T
\]

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

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

Cannot reduce to the form of \( p(x; \theta) = g(T(x), \theta)h(x) \).

If \( f_0 \) is known but \( A \) and \( \sigma^2 \) are unknown, \( \theta = [A \sigma^2]^T \)
\[
p(x; \theta) = \frac{1}{(2\pi \sigma^2)^{\frac{N}{2}}} \exp \left[ - \frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} x[n]^2 - 2A \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right) \right] \cdot \frac{1}{g(T(x), \theta)}
\]

\[
T(x) = \begin{bmatrix}
\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \\
\sum_{n=0}^{N-1} x[n]^2
\end{bmatrix}
\]
Remark: In the vector parameter case, completeness means that for \( v(T) \), an arbitrary \( r \times 1 \) function of \( T \), if

\[
E(v(T)) = \int v(T)p(T; \theta) dT = 0 \quad \text{for all } \theta
\]

then \( v(T) = 0 \) for all \( T \).

Example: DC Level in White Noise with Unknown Noise Power
\[
x[n] = A + w[n], \quad n = 0, 1, \ldots, N-1
\]
\( w[n] \): WGN with variance \( \sigma^2 \)
\( A, \sigma^2 \): unknown.
\( \theta = [A \sigma^2]^T \)

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

\[
T(x) = \begin{bmatrix}
\sum_{n=0}^{N-1} x[n] \\
\sum_{n=0}^{N-1} x^2[n]
\end{bmatrix}
\]
jointly sufficient
\( r = p = 2 \Rightarrow \) we can find MVU estimator.

\[
E(T(x)) = \begin{bmatrix}
NA \\
NE(x^2[n])
\end{bmatrix} = \begin{bmatrix}
NA \\
N(\sigma^2 + A^2)
\end{bmatrix}
\]
\[ g(T(x)) = \begin{bmatrix} \frac{1}{N} T_1(x) \\ \frac{1}{N} T_2(x) - \left[ \frac{1}{N} T_1(x) \right]^2 \end{bmatrix} \]

\[ = \begin{bmatrix} \bar{x} \\ \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \end{bmatrix} \]

\[ E(\bar{x}) = A \]

\[ E\left( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \right) = \sigma^2 + A^2 - E(\bar{x}^2) \]

Since \( \bar{x} \sim \mathcal{N}(A, \sigma^2/N) \),

\[ E(\bar{x}^2) = A^2 + \sigma^2/N \]

\[ E\left( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \right) = \sigma^2 (1 - \frac{1}{N}) = \frac{N-1}{N} \sigma^2. \]

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

\[ = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2 \end{bmatrix} \]

\[ \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 = \sum_{n=0}^{N-1} x^2[n] - 2 \sum_{n=0}^{N-1} x[n] \bar{x} + N\bar{x}^2 \]

\[ = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2, \]

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

Since this pdf belongs to the vector exponential family of PDFs, \( \Rightarrow \) complete.
\( \Rightarrow \) MVU estimator

However, this MVU estimator is not efficient!
Pf: \( \tilde{x} \sim \mathcal{N} \left( A, \frac{\sigma^2}{N} \right) \)

\[
\frac{(N-1)\sigma^2}{\sigma^2} \sim \chi^2_{N-1}.
\]

\[
C_\delta = \begin{bmatrix}
\frac{\sigma^2}{N} & 0 \\
0 & \frac{2\sigma^4}{N-1}
\end{bmatrix}
\]

However, it has been shown that \( \Gamma^{-1}(\theta) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\
0 & \frac{2\sigma^4}{N} \end{bmatrix} \)

Alternate approach:

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

\[
\sum_{n=0}^{N-1} (x[n] - A)^2 = \sum_{n=0}^{N-1} (x[n] - \bar{x} + \bar{x} - A)^2
\]

\[
= \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 + 2(\bar{x} - \bar{x}) \sum_{n=0}^{N-1} (x[n] - \bar{x}) + N(\bar{x} - A)^2
\]

\[
p(x; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 + N(\bar{x} - A)^2 \right] \right\} \cdot \frac{1}{g(T'(x), \theta)} \cdot h(x)
\]

where

\[
T'(x) = \begin{bmatrix}
\bar{x} \\
\sum_{n=0}^{N-1} (x[n] - \bar{x})^2
\end{bmatrix}
\]

NCTU EE  P.61
**Best Linear Unbiased Estimators (BLUE)**

- **Definition**

  Restrict the estimator to be linear in the data

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

  and find MVU estimator within this class.

  MVU estimator = BLUE if the MVU estimator happens to be linear.

- **Example: DC Level in WGN**

  MVU: \( \hat{\theta} = \bar{x} = \sum_{n=0}^{N-1} \frac{1}{N} x[n] \)

  BLUE = MVU

- **Example: Mean of Uniformly Distributed Noise**

  MVU: \( \hat{\theta} = \frac{N + 1}{2N} \max x[n] \)

  BLUE is suboptimal

---

(a) **DC level in WGN; BLUE is optimal**

(b) **Mean of uniform noise; BLUE is suboptimal**
For some estimation problems, the use of a BLUE can be totally inappropriate. Example: estimation of the power of WGN.

\[
\text{MVU: } \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]
\]

If we force the estimator to be linear,

\[
\hat{\sigma}^2 = \sum_{n=0}^{N-1} a_n x[n]
\]

\[
E(\hat{\sigma}^2) = \sum_{n=0}^{N-1} a_n E(x[n]) = 0
\]

We cannot even find a single linear estimator that is unbiased!

Instead, we may transform the data into \( y[n] = x[n]^2 \)

\[
\hat{\sigma}^2 = \sum_{n=0}^{N-1} a_n y[n] = \sum_{n=0}^{N-1} a_n x^2[n]
\]

\[
E(\hat{\sigma}^2) = \sum_{n=0}^{N-1} a_n \sigma^2 = \sigma^2.
\]

\[\Rightarrow\] The BLUE may still be used.

**Finding the BLUE**

Unbiased constraint: \( E(\hat{\theta}) = \sum_{n=0}^{N-1} a_n E(x[n]) = \theta \)

\[
\text{var}(\hat{\theta}) = E \left[ \left( \sum_{n=0}^{N-1} a_n x[n] - E \left( \sum_{n=0}^{N-1} a_n x[n] \right) \right)^2 \right]
\]

Let \( \mathbf{a} = [a_0 \ a_1 \ \ldots \ \ a_{N-1}]^T \)
\[
\text{var}(\hat{\theta}) = E \left[ (a^T \mathbf{x} - a^T E(\mathbf{x}))^2 \right] \\
= E \left[ (a^T(\mathbf{x} - E(\mathbf{x})))^2 \right] \\
= E \left[ a^T(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T a \right] \\
= a^T \mathbf{C} a.
\]

In order to satisfy the unbiased constraint, \(E\{x[n]\}\) must be linear in \(\theta\).
\[
E(x[n]) = s[n]\theta
\]
where \(s[n]\) are known.
\[
x[n] = E(x[n]) + [x[n] - E(x[n])] = \theta s[n] + w[n].
\]

**BLUE is applicable to amplitude estimation of signals in noise.**

**Summary:**

We minimize the variance \(\text{var}(\hat{\theta}) = a^T \mathbf{C} a\) subject to the constraint
\[
\sum_{n=0}^{N-1} a_n E(x[n]) = \theta \\
\sum_{n=0}^{N-1} a_n s[n] \theta = \theta \\
\sum_{n=0}^{N-1} a_n s[n] = 1
\]
or
\[
a^T s = 1
\]
where \(s = [s[0] \ s[1] \ldots s[N-1]]^T\)

**Solution:** Use the method of Lagrangian multipliers:
\[
J = a^T \mathbf{C} a + \lambda (a^T s - 1)
\]
\[
\frac{\partial J}{\partial a} = 2\mathbf{C}a + \lambda s
\]
\[
a = -\frac{\lambda}{2} \mathbf{C}^{-1} s
\]
The Lagrangian multiplier $\lambda$ is found using the constraint

$$a^T s = -\frac{\lambda}{2} s^T C^{-1} s = 1$$

$$-\frac{\lambda}{2} = \frac{1}{s^T C^{-1} s}$$

$$a_{opt} = \frac{C^{-1} s}{s^T C^{-1} s}$$

$$\hat{\theta} = \frac{s^T C^{-1} x}{s^T C^{-1} s}$$

$$\text{var}(\hat{\theta}) = a_{opt}^T C a_{opt}$$

$$= \frac{s^T C^{-1} C s}{(s^T C^{-1} s)^2}$$

$$= \frac{1}{s^T C^{-1} s}.$$  

To verify that $\hat{\theta}$ is unbiased,

$$E(\hat{\theta}) = \frac{s^T C^{-1} E(x)}{s^T C^{-1} s}$$

$$= \frac{s^T C^{-1} \theta s}{s^T C^{-1} s}$$

$$= \theta.$$  

To verify that $a_{opt}$ is indeed the global minimum, we check

$$G = (a - a_{opt})^T C (a - a_{opt})$$

$$G = a^T Ca - 2a_{opt}^T Ca + a_{opt}^T C a_{opt}$$

$$= a^T Ca - 2 \frac{s^T C^{-1} Ca}{s^T C^{-1} s} + \frac{1}{s^T C^{-1} s}$$

$$= a^T Ca - \frac{1}{s^T C^{-1} s}$$

$$a^T Ca = (a - a_{opt})^T C (a - a_{opt}) + \frac{1}{s^T C^{-1} s}.$$  

$\triangleright$
To determine the BLUE, we only require knowledge of
1. $s$, the scaled mean
2. $C$, the covariance
We don’t need the entire pdf!!

Example: DC Level in White Noise

$x[n] = A + w[n]$, $n = 0, 1, ..., N-1$

$w[n]$: white noise of unspecified pdf with variance $\sigma^2$

Since $E(x[n]) = A$, we have $s[n] = 1$ and thus $s = 1$.

\[
\hat{A} = \frac{1^T \frac{1}{\sigma^2} \mathbf{1} x}{1^T \frac{1}{\sigma^2} \mathbf{1}}
\]

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

\[
\text{var}(\hat{A}) = \frac{1}{1^T \frac{1}{\sigma^2} \mathbf{1}}
\]

\[
= \frac{\sigma^2}{N}.
\]

Remarks:
1. The sample mean is the BLUE independent of the pdf of the data.
2. For Gaussian data, the sample mean is the MVU estimator.

Example: DC Level in Uncorrelated Noise

$x[n] = A + w[n]$, $n = 0, 1, ..., N-1$

$w[n]$: zero mean uncorrelated noise with $\text{car}(w[n]) = \sigma_n^2$.

\[
\hat{A} = \frac{1^T C^{-1} x}{1^T C^{-1} 1}
\]

\[
\text{var}(\hat{A}) = \frac{1}{1^T C^{-1} 1}.
\]

\[
C = \begin{bmatrix}
\sigma_0^2 & 0 & \ldots & 0 \\
0 & \sigma_1^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_{N-1}^2
\end{bmatrix}
\]
The BLUE weights those samples most heavily with smallest variances.

Remark: Our model \( E(x[n]) = s[n] \theta \) is actually the general linear model
\[
x[n] = E(x[n]) + [x[n] - E(x[n])] = \theta s[n] + w[n].
\]

For Gaussian noise, BLUE = MVU estimator.

○ Extension to a Vector Parameter

\[ \theta = [\theta_1 \ \theta_2 \ \ldots \ \theta_p]^T \]

Linear estimator: \[
\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n]
\]

\[
\hat{\theta} = A x
\]

To be unbiased,
\[
E(\hat{\theta}_i) = \sum_{n=0}^{N-1} a_{in} E(x[n]) = \theta_i \quad i = 1, 2, \ldots, p
\]

\[
E(\hat{\theta}) = A E(x) = \theta
\]
To satisfy the unbiased constraint,

\[ E(x) = \mathbf{H}\theta \]

\[ \mathbf{A} \mathbf{H} = \mathbf{I}. \]

Let \( \mathbf{a}_i = [a_{i0} \ a_{i1} \ldots a_{i(N-1)}]^{T} \) so that \( \hat{\theta}_i = \mathbf{a}_i^{T} \mathbf{x} \)

\[ \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^{T} \\ \mathbf{a}_2^{T} \\ \vdots \\ \mathbf{a}_p^{T} \end{bmatrix} \]

Assume \( \mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \ldots & \mathbf{h}_p \end{bmatrix} \)

\[ \mathbf{a}_i^{T} \mathbf{h}_j = \delta_{ij} \quad i = 1, 2, \ldots, p; j = 1, 2, \ldots, p. \]

\[ \text{var}(\hat{\theta}_i) = \mathbf{a}_i^{T} \mathbf{C} \mathbf{a}_i \]

The BLUE estimator is

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

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

Remark: The form of the BLUE is identical to the MVU estimator for the general linear model: \( \mathbf{x} = \mathbf{H}\theta + \mathbf{w} \), where \( \mathbf{w} \sim N(0, \mathbf{C}) \)

**Theorem 6.1 (Gauss-Markov Theorem)** If the data are of the general linear model form

\[ \mathbf{x} = \mathbf{H}\theta + \mathbf{w} \]

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

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

and the minimum variance of \( \hat{\theta}_i \) is

\[ \text{var}(\hat{\theta}_i) = [(\mathbf{H}^{T} \mathbf{C}^{-1} \mathbf{H})^{-1}]_{ii} \cdot \]

In addition, the covariance matrix of \( \hat{\theta} \) is

\[ \mathbf{C}_\hat{\theta} = (\mathbf{H}^{T} \mathbf{C}^{-1} \mathbf{H})^{-1}. \]
Signal Processing Example

Example: Source Localization
Determine the position of a source based on the emitted signal of the source.

If $R_1 = R_2 = R_3 \Rightarrow t_2 - t_1 = t_3 - t_2 = 0$.

Assume N antennas have been placed at known locations and the time of arrival measurements $t_i$ for $i = 0, 1, \ldots, N-1$ are available. The arrival times are corrupted by noise with zero mean and known covariance but otherwise unknown pdf.

For signal emitted by the source at $T_0$,

$$t_i = T_0 + \frac{R_i}{c} + \epsilon_i \quad i = 0, 1, \ldots, N-1$$

$\epsilon_i$'s: measurement noise zero mean, uncorrelated, variance $\sigma^2$

c: propagation speed.

$(x_i, y_i)^T$ position of the $i$th antenna
$(x_s, y_s)^T$ unknown position

$$R_i = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}.$$
The source is near the nominal position \((x_n, y_n)^T\), which has been obtained from previous measurements.

\[
R_i \approx R_{ni} + \frac{x_n - x_i}{R_{ni}} \delta x_s + \frac{y_n - y_i}{R_{ni}} \delta y_s
\]

\[
t_i = T_0 + \frac{R_{ni}}{c} + \frac{x_n - x_i}{R_{ni} c} \delta x_s + \frac{y_n - y_i}{R_{ni} c} \delta y_s + \epsilon_i
\]

\[
\frac{x_n - x_i}{R_{ni}} = \cos \alpha_i
\]

\[
\frac{y_n - y_i}{R_{ni}} = \sin \alpha_i
\]

\[
t_i = T_0 + \frac{R_{ni}}{c} + \frac{\cos \alpha_i}{c} \delta x_s + \frac{\sin \alpha_i}{c} \delta y_s + \epsilon_i
\]

Let \(\tau_i = t_i - \frac{R_{ni}}{c}\), we have

\[
\tau_i = T_0 + \frac{\cos \alpha_i}{c} \delta x_s + \frac{\sin \alpha_i}{c} \delta y_s + \epsilon_i
\]

Unknown parameters: \(T_0, \delta x_s, \delta y_s\).

To get rid of \(T_0\), we consider time difference of arrivals.

\[
\xi_1 = \tau_1 - \tau_0
\]

\[
\xi_2 = \tau_2 - \tau_1
\]

\[
\vdots
\]

\[
\xi_{N-1} = \tau_{N-1} - \tau_{N-2}
\]

\[
\xi_i = \frac{1}{c} (\cos \alpha_i - \cos \alpha_{i-1}) \delta x_s + \frac{1}{c} (\sin \alpha_i - \sin \alpha_{i-1}) \delta y_s + \epsilon_i - \epsilon_{i-1}
\]
\[\theta = \begin{bmatrix} \delta x_s & \delta y_s \end{bmatrix}^T\]

\[H = \frac{1}{c} \begin{bmatrix}
\cos \alpha_1 - \cos \alpha_0 & \sin \alpha_1 - \sin \alpha_0 \\
\cos \alpha_2 - \cos \alpha_1 & \sin \alpha_2 - \sin \alpha_1 \\
\vdots & \vdots \\
\cos \alpha_{N-1} - \cos \alpha_{N-2} & \sin \alpha_{N-1} - \sin \alpha_{N-2}
\end{bmatrix}\]

\[w = \begin{bmatrix}
\epsilon_1 - \epsilon_0 \\
\epsilon_2 - \epsilon_1 \\
\vdots \\
\epsilon_{N-1} - \epsilon_{N-2}
\end{bmatrix}\]

\(w\): zero mean, but no longer uncorrelated.

\[w = \begin{bmatrix}
-1 & 1 & 0 & 0 & \ldots & 0 \\
0 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & -1 & 1
\end{bmatrix} \begin{bmatrix}
\epsilon_0 \\
\epsilon_1 \\
\vdots \\
\epsilon_{N-1}
\end{bmatrix}\]

\[C = E[A\epsilon \epsilon^T A^T] = \sigma^2 AA^T\]

\[\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} \xi\]

\[\hat{\theta} = [H^T (AA^T)^{-1} H]^{-1} H^T (AA^T)^{-1} \xi\]

\[\text{var}(\hat{\theta}_i) = \sigma^2 \left[H^T (AA^T)^{-1} H\right]^{-1}_ii\]

\[C_\delta = \sigma^2 \left[H^T (AA^T)^{-1} H\right]^{-1}\]

e.g., \(N = 3\)
\[
H = \frac{1}{c} \begin{bmatrix}
-\cos \alpha & 1 - \sin \alpha \\
-\cos \alpha & -(1 - \sin \alpha)
\end{bmatrix}
\]

\[
A = \begin{bmatrix}
-1 & 1 & 0 \\
0 & -1 & 1
\end{bmatrix}.
\]

\[
C_\theta = \sigma^2 c^2 \begin{bmatrix}
\frac{1}{2 \cos^2 \alpha} & 0 & 0 \\
0 & 3/2 & \frac{3/2}{(1 - \sin \alpha)^2}
\end{bmatrix}
\]