430.714

Estimation Theory

Prof. Songhwai Oh

ECE, SNU
CRAMER-RAO LOWER BOUND (CRLB)


Estimator Accuracy

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

Likelihood – distribution of data given parameters

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

(a) \( \sigma_1 = 1/3 \)

(b) \( \sigma_2 = 1 \)

Large curvature -> more precise estimation of \( A \) is possible
Curvature of the Log-Likelihood Function

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

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

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

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

As \( \sigma^2 \downarrow 0 \), the curvature \( \uparrow \) (better estimation).

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

As \( \text{var}(\hat{A}) \downarrow 0 \), the curvature \( \uparrow \) (better estimation).

Measure of curvature: \[ -\mathbb{E} \left( \frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} \right) \); Average curvature of log-likelihood
Theorem: It is assumed that the probability density function $p(x; \theta)$ satisfies the regularity condition

$$\mathbb{E} \left( \frac{\partial \ln p(x; \theta)}{\partial \theta} \right) = 0 \quad \text{for all } \theta, \quad (1)$$

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

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E} \left( \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right)}.$$ 

(2)
Additional Facts about the CRLB

1. An unbiased estimator achieves the bound if and only if

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

for some functions \(I\) and \(g\).

2. \(\hat{\theta} = g(x)\) is the minimum variance unbiased estimator.

3. The minimum variance is \(1/I(\theta)\), where

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

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

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

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

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

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

\[ = 0 \]

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

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

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

Hence,

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

Let \( \alpha = g(\theta) \). Consider all unbiased estimators \( \hat{\alpha} \) such that

\[
\mathbb{E}(\hat{\alpha}) = \alpha = g(\theta).
\]

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

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

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

(regularity condition)

\[
\int (\hat{\alpha} - \alpha) \frac{\partial \ln p(x; \theta)}{\partial \theta} p(x; \theta) dx = \frac{\partial g(\theta)}{\partial \theta}
\]
Cauchy-Schwarz Inequality

For vectors:

\[
\left| \langle x, y \rangle \right|^2 \leq \| x \|^2 \| y \|^2
\]

\[
\langle x, y \rangle = \| x \| \| y \| \cos \theta \leq \| x \| \| y \|
\]

equality if \( x = cy \).

For random variables (r.v.):

\[
\mathbb{E}(XY) = \langle X, Y \rangle ; \text{ an inner product between two mean zero r.v.'s}
\]

\[
\left| \mathbb{E}(XY) \right|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2) ; \text{ Cauchy-Schwarz Inequality}
\]

\[
\left| \int X(z)Y(z)p(z)dz \right|^2 \leq \left( \int X^2(z)p(z)dz \right) \cdot \left( \int Y^2(z)p(z)dz \right)
\]
Back to the Proof of CRLB

We have

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

Cauchy-Schwarz:

\[
\left| \int X(z) Y(z) p(z) dz \right|^2 \leq \left( \int X^2(z) p(z) dz \right) \cdot \left( \int Y^2(z) p(z) dz \right)
\]

Let \( X(z) \leftarrow \hat{\alpha} - \alpha, Y(z) \leftarrow \frac{\partial \ln p(x; \theta)}{\partial \theta} \), and \( p(z) \leftarrow p(x; \theta) \).

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

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

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

Let \( \alpha = g(\theta) = \theta \), we get the desired result.
Proof of Additional Results

The Cauchy-Schwarz inequality becomes an equality if $X(z) = cY(z)$ for some constant $c$, not depending on $z$.

In our case, this is when
\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{c} (\hat{\alpha} - \alpha)
\]
\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = \frac{1}{c(\theta)} (\hat{\theta} - \theta) \quad \text{for } \alpha = g(\theta) = \theta
\]

To find $c(\theta)$:
\[
\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left( \frac{1}{c(\theta)} \right) (\hat{\theta} - \theta) - \frac{1}{c(\theta)}
\]
\[
-\mathbb{E} \left( \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \right) = \frac{1}{c(\theta)}
\]

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

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

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

Estimator: \[ \tilde{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \mathbb{E}(\tilde{A}) = A \quad \text{unbiased} \]

\[ \text{var}(\tilde{A}) = \frac{1}{N} \sigma^2 \]

Likelihood of \( x := (x[0], \ldots, x[N-1]) \)

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

\]

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

Regularity condition:
\[
\mathbb{E}\left( \frac{\partial \ln p(x; A)}{\partial A} \right) = 0
\] (checked)

Second derivative:
\[
\frac{\partial^2 \ln p(x; A)}{\partial A^2} = -\frac{N}{\sigma^2}
\]

CRLB:
\[
\text{var}(\hat{A}) \geq \frac{\sigma^2}{N} \quad \text{for any unbiased estimator } \hat{A}
\]

- Since \( \text{var}(\hat{A}) = \frac{\sigma^2}{N} \), \( \hat{A} \) is the minimum variance unbiased estimator.
- If an unbiased estimator achieves the CRLB, it is said to be efficient.
Example 2

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

- Likelihood of $\mathbf{x} := (x[0], \ldots, x[N - 1])$ is

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

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

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

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

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

Special case: $s[n; \theta] = \theta$. Then CRLB $= \frac{\sigma^2}{N}$. 

Prof. Songhwai Oh

Estimation Theory
Transformation of Parameters
\[ x[n] = A + w[n], \quad \text{where } w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \]

Estimation of \( A^2 \) (\( \alpha = g(A) = A^2 \))

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

1. Nonlinear transformation does not preserve efficiency

\[
\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n], \quad \bar{x} \sim \mathcal{N}(A, \frac{\sigma^2}{N})
\]

Is \( \bar{x}^2 \) efficient for \( A^2 \)?

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

\[
\neq A^2.
\]

Biased => not efficient
\[ x[n] = A + w[n], \quad \text{where } w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \]

Estimation of \( A^2 \) (\( \alpha = g(A) = A^2 \))

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

**2. Linear transform preserves efficiency**

Assume: \( \hat{\theta} \) efficient for \( \theta \), \( g(\theta) = a\theta + b \) (affine transform)

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

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

\[
\text{var}(\hat{g}(\theta)) = \text{var}(a\hat{\theta} + b) = a^2\text{var}(\hat{\theta})
\]

**CRLB for \( g(\theta) \):**

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

Efficient

\[
= \left( \frac{\partial g(\theta)}{\partial \theta} \right)^2 \text{var}(\hat{\theta}) = a^2\text{var}(\hat{\theta})
\]
3. Efficiency is approximately maintained over a nonlinear transformation for a large dataset.

\[
\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n], \quad \bar{x} \sim \mathcal{N} \left( \frac{\sigma^2}{N} \right)
\]

\(\bar{x}^2\) is approximately unbiased

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

\[
= \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}
\]

as \(N \to \infty\), \(\text{var}(\bar{x}^2) \to \frac{4A^2\sigma^2}{N}\).

CRLB for \(A^2\):

\[
\text{var}(\bar{A}^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}
\]
\[ \bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n], \quad \bar{x} \sim \mathcal{N} \left( A, \frac{\sigma^2}{N} \right) \]

\[ g(\bar{x}) = \bar{x}^2 \]

Linearize \( g \) about \( A \):

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

Under this approximation:

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

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

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

\[ = \frac{4A^2 \sigma^2}{N} \]

Asymptotically Efficient
CRLB for Vector Parameters
Vector parameter

\[
\theta = \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\vdots \\
\theta_p 
\end{bmatrix}
\]

CRLB:

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

\[
I(\theta), \text{ Fisher information matrix}
\]

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

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

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

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

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

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

\[
\text{var}(\hat{\sigma}^2) \geq \frac{2\sigma^4}{N}
\]

Unknowns: \( A \) and \( \sigma^2 \)

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

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

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

\[
\frac{\partial^2 \ln p(x; \theta)}{\partial A \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x[n] - A)
\]

\[
\frac{\partial^2 \ln p(x; \theta)}{\partial \sigma^2^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=0}^{N-1} (x[n] - A)^2.
\]
\[ x[n] = A + Bn + w[n] \quad n = 0, 1, \ldots, N - 1 \]

\[ \theta = [A \ B]^T \]

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

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

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

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

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

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

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

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

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

\[ \text{var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}. \]
1. CRLB always increases as we estimate more parameters

- When only $A$ is unknown, $\text{var}(\hat{A}) \geq \frac{\sigma^2}{N}$.
- When both $A$ and $B$ are unknown, $\text{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)}$.
- For $N \geq 2$, $\frac{2(2N-1)\sigma^2}{N(N+1)} > \frac{\sigma^2}{N}$.

2. Some parameters are more sensitive than others

$$\frac{\text{CRLB}(\hat{A})}{\text{CRLB}(\hat{B})} = \frac{(2N-1)(N-1)}{6} > 1 \quad \text{for } N \geq 3.$$ i.e., $B$ is easier to estimate

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

$$\Delta x[n] \approx \frac{\partial x[n]}{\partial A}\Delta A = \Delta A$$

$$\Delta x[n] \approx \frac{\partial x[n]}{\partial B}\Delta B = n\Delta B$$

(a) $A = 0, B = 0$ to $A = 1, B = 0$

(b) $A = 0, B = 0$ to $A = 0, B = 1$
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 $1/I(\theta)$. 
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
\[
C_\hat{\theta} - I^{-1}(\theta) \geq 0
\] (3.24)
where $\geq 0$ 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 $C_\hat{\theta} = I^{-1}(\theta)$ if and only if
\[
\frac{\partial \ln p(x; \theta)}{\partial \theta} = I(\theta)(g(x) - \theta)
\] (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 $I^{-1}(\theta)$. 
Vector Parameter CRLB for Transformations

\[ \alpha = g(\theta) \quad \text{r-dimensional function} \]

\[ C_\alpha - \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^T}{\partial \theta} \geq 0 \]

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

- Unknowns: A and \( \sigma^2 \)
- Estimate \( \alpha = \frac{A^2}{\sigma^2} \) (SNR)

\[ \frac{\partial g(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial g(\theta)}{\partial \theta_1} \\ \frac{\partial g(\theta)}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial g(\theta)}{\partial A} \\ \frac{\partial g(\theta)}{\partial \sigma^2} \end{bmatrix} \]

\[ = \begin{bmatrix} 2A/\sigma^2 \\ -A^2/\sigma^4 \end{bmatrix} \]

\[ \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)^T}{\partial \theta} = \begin{bmatrix} 2A/\sigma^2 \\ -A^2/\sigma^4 \end{bmatrix} \begin{bmatrix} \sigma^2/N \\ 0 \end{bmatrix} \begin{bmatrix} 2A/\sigma^2 \\ -A^2/\sigma^4 \end{bmatrix} \]

\[ = \frac{4A^2}{N\sigma^2} + \frac{2A^4}{N\sigma^4} = \frac{4\alpha + 2\alpha^2}{N}. \]

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

\[ \mathbf{x} \sim \mathcal{N}(\mu(\theta), \mathbf{C}(\theta)) \]

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

Appendix 3C

\[
\frac{\partial \mu(\theta)}{\partial \theta_i} = \left[ \begin{array}{c}
\frac{\partial [\mu(\theta)]_1}{\partial \theta_i} \\
\frac{\partial [\mu(\theta)]_2}{\partial \theta_i} \\
\vdots \\
\frac{\partial [\mu(\theta)]_N}{\partial \theta_i}
\end{array} \right] \\
\frac{\partial \mathbf{C}(\theta)}{\partial \theta_i} = \left[ \begin{array}{cccc}
\frac{\partial [\mathbf{C}(\theta)]_{11}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\theta)]_{12}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\theta)]_{1N}}{\partial \theta_i} \\
\frac{\partial [\mathbf{C}(\theta)]_{21}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\theta)]_{22}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\theta)]_{2N}}{\partial \theta_i} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial [\mathbf{C}(\theta)]_{N1}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\theta)]_{N2}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\theta)]_{NN}}{\partial \theta_i}
\end{array} \right]
\]

\[ \mathbf{x} \sim \mathcal{N}(\mu(\theta), \mathbf{C}(\theta)) \quad \text{(scalar parameter)} \]

\[
I(\theta) = \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right]^T \mathbf{C}^{-1}(\theta) \left[ \frac{\partial \mu(\theta)}{\partial \theta} \right] \\
+ \frac{1}{2} \text{tr} \left[ \left( \mathbf{C}^{-1}(\theta) \frac{\partial \mathbf{C}(\theta)}{\partial \theta} \right)^2 \right]
\]
\[ x[n] = A + w[n], \text{ where} \]
- \[ w[n] \sim \mathcal{N}(0, \sigma^2) \]
- \[ A \sim \mathcal{N}(0, \sigma_A^2) \]

\[
C(\sigma_A^2) = \sigma_A^2 11^T + \sigma^2 I
\]

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

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

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

Matrix inversion lemma:
\[
(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}
\]

Woodbury identity:
\[
(A + uu^T)^{-1} = A^{-1} - A^{-1}uu^TA^{-1}\frac{1}{1 + u^TA^{-1}u}
\]

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

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

\[
\frac{1}{2} \text{tr} \left[ \left( C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta} \right)^2 \right]
\]

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