430.714

Estimation Theory

Prof. Songhwai Oh

ECE, SNU

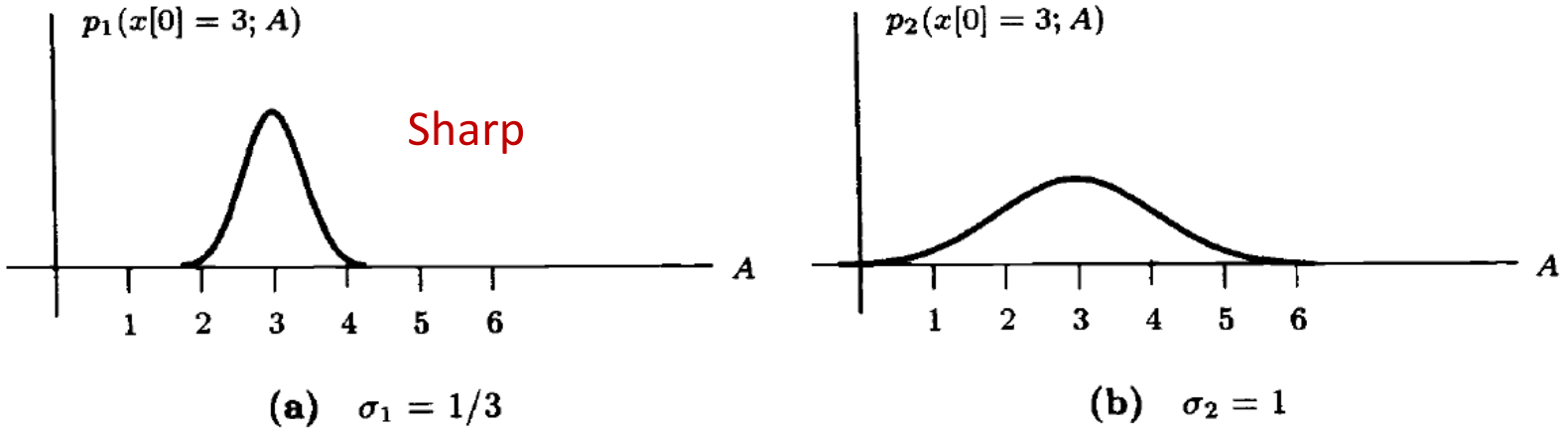
CRAMER-RAO LOWER BOUND (CRLB)

Estimator Accuracy

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

Likelihood – distribution of data given parameters

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



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

As $\mathrm{var}(\hat{A}) \downarrow 0$, 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: Cramer-Rao Lower Bound (CRLB)

Theorem: It is assumed that the probability density function $p(x;\theta)$ satisfies the *regularity* condition

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

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

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

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

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

Fisher Information

Regularity Condition

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

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right) &= \int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta) dx \\
&= \int \frac{1}{p(x;\theta)} \frac{\partial p(x;\theta)}{\partial \theta} p(x;\theta) dx \\
&= \int \frac{\partial p(x;\theta)}{\partial \theta} dx \\
&= \frac{\partial}{\partial \theta} \int p(x;\theta) dx \\
&= 0
\end{aligned}$$

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

$$\begin{aligned}
\int \left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2} p(x;\theta) + \frac{\partial \ln p(x;\theta)}{\partial \theta} \frac{\partial p(x;\theta)}{\partial \theta}\right) dx &= 0 \\
\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right) + \int \frac{\partial \ln p(x;\theta)}{\partial \theta} \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta) dx &= 0
\end{aligned}$$

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

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

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

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

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

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:

$$|\langle x, y\rangle|^2 \leq \|x\|^2 \|y\|^2$$

$$|\langle x, y\rangle| = \|x\|\|y\||\cos\theta| \leq \|x\|\|y\|$$

$$\text{equality if } x = cy.$$

For random variables (r.v.):

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

$$|\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2) \quad \text{; Cauchy-Schwarz Inequality}$$

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

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 $p(z) \leftarrow p(x;\theta)$.

$$\begin{aligned}
\left(\frac{\partial g(\theta)}{\partial \theta} \right)^2 &= \left(\int (\hat{\alpha} - \alpha) \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta) dx \right)^2 \\
&\leq \underbrace{\left(\int (\hat{\alpha} - \alpha)^2 p(x;\theta) dx \right)}_{\mathrm{var}(\hat{\alpha})} \left(\int \left(\frac{\partial \ln p(x;\theta)}{\partial \theta} \right)^2 p(x;\theta) dx \right)
\end{aligned}$$

Hence,

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

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

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

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = \frac{1}{c(\theta)}(\hat{\theta} - \theta) \qquad \text{for } \alpha = g(\theta) = \theta$$

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,

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

$$\Rightarrow \quad I(\theta) = \mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right).$$

Example 1

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

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

$$\mathbb{E}(\tilde{A}) = A \qquad \text{unbiased}$$

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

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

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

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

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

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

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

- Since $\text{var}(\tilde{A}) = \frac{\sigma^2}{N}$, $\tilde{A}$ is the minumum 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

$$
\begin{aligned}
p(\mathbf{x};\theta) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n;\theta])^2\right) \\
\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n;\theta])\frac{\partial s[n;\theta]}{\partial\theta} \\
\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left((x[n]-s[n;\theta])\frac{\partial^2 s[n;\theta]}{\partial\theta^2} - \left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2\right) \\
\mathbb{E}\left(\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right) &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2 \\
\mathbf{var}(\hat{\theta}) &\geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2}
\end{aligned}
$$

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

Transformation of Parameters

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

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

$$\mathbf{var}(\hat{A}^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

CRLB for $\alpha = g(\theta)$:

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

1. Nonlinear transformation does not preserve efficiency

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

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

$$\begin{aligned} \mathbb{E}(\bar{x}^2) &= \mathbb{E}^2(\bar{x}) + \mathbf{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N} \\ &\neq A^2. \end{aligned}$$

Biased => not efficient

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

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

$$\mathbf{var}(\hat{A}^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

CRLB for $\alpha = g(\theta)$:

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

2. Linear transform preserves efficiency

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

$$\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$$

$$\begin{aligned} \mathbb{E}(a\hat{\theta} + b) &= a\mathbb{E}(\hat{\theta}) + b = a\theta + b \\ &= g(\theta) \end{aligned}$$

$$\mathbf{var}(\widehat{g(\theta)}) = \mathbf{var}(a\hat{\theta} + b) = a^2\mathbf{var}(\hat{\theta})$$

CRLB for $g(\theta)$:

$$\begin{aligned} \mathbf{var}(\widehat{g(\theta)}) &\geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)} \\ &= \left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \mathbf{var}(\hat{\theta}) \\ &= a^2\mathbf{var}(\hat{\theta}) \end{aligned}$$

Efficient

3. Efficiency is approximately maintained over a nonlinear transformation for a **large dataset**.

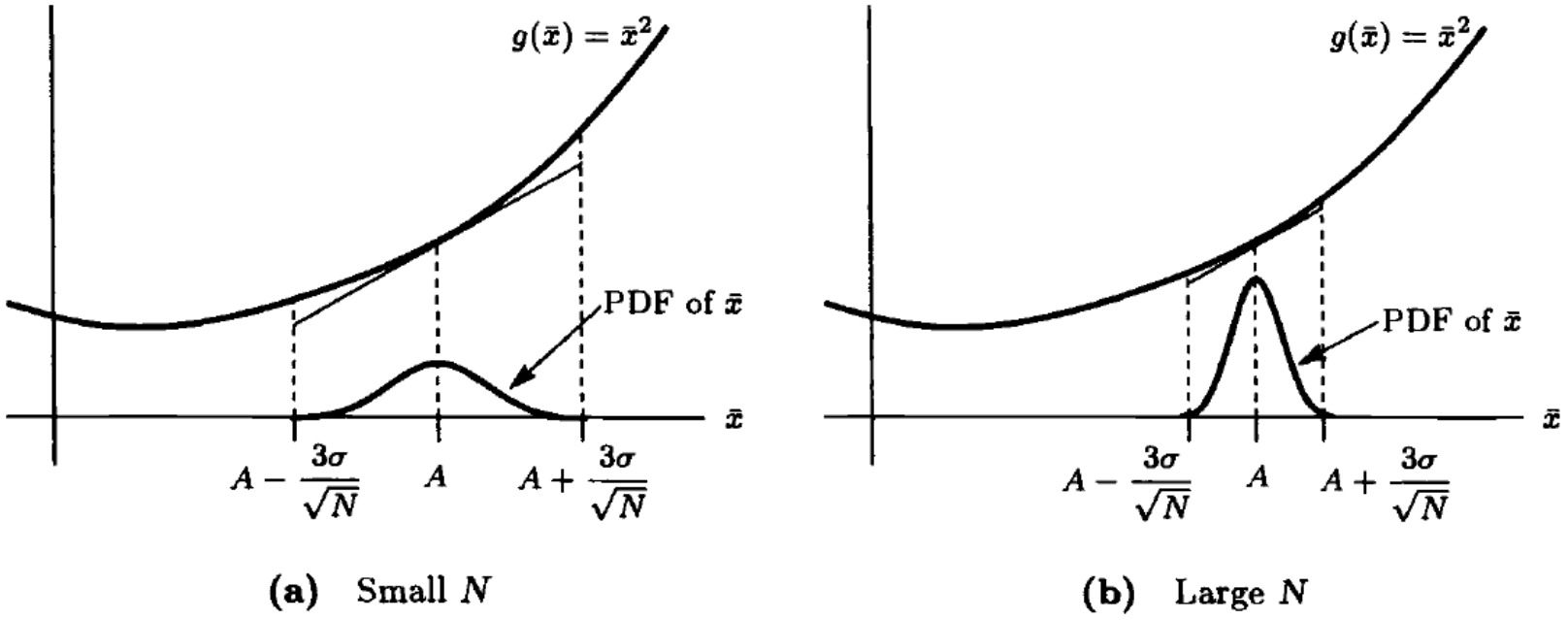
$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n], \qquad \bar{x} \sim \mathcal{N}\left(A, \frac{\sigma^2}{N}\right)$$

$\bar{x}^2$ is approximately unbiased

CRLB for A^2:

$$\mathbf{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

$$\mathbf{var}(\bar{x}^2) = E(\bar{x}^4) - E^2(\bar{x}^2)$$
$$= \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2} \quad \text{as } N \to \infty,\ \mathrm{var}(\bar{x}^2) \to \tfrac{4A^2\sigma^2}{N}.$$



(a) Small N **(b)** Large N

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

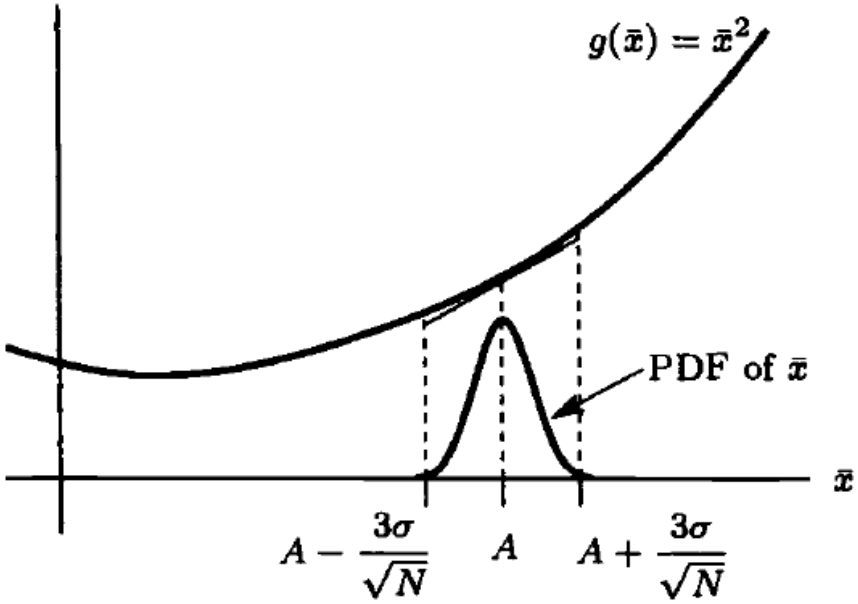
CRLB for A^2:

$$\text{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

$$g(\bar{x}) = \bar{x}^2$$

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

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

$\hat{\theta}$, an unbiased estimate of θ.

CRLB:

$$\mathbf{var}(\hat{\boldsymbol{\theta}}_i) \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{ii}$$

$I(\theta)$, Fisher information matrix

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

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

Unknowns: A and σ^2

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

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

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

$$\mathbf{var}(\hat{A}) \geq \frac{\sigma^2}{N}$$

$$\mathbf{var}(\hat{\sigma^2}) \geq \frac{2\sigma^4}{N}$$

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

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

$$\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2} = -\frac{N}{\sigma^2}$$

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

$$\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\sigma^{2\,2}} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{n=0}^{N-1}(x[n]-A)^2.$$

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

$$\boldsymbol{\theta} = [A\ B]^T$$

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

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

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

$$\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2} = -\frac{N}{\sigma^2}$$

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

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

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix} N & \sum_{n=0}^{N-1} n \\ \sum_{n=0}^{N-1} n & \sum_{n=0}^{N-1} n^2 \end{bmatrix}$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} N & \frac{N(N-1)}{2} \\ \frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

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

$$\mathrm{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)}$$

$$\mathrm{var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}.$$

1. CRLB always increases as we estimate more parameters

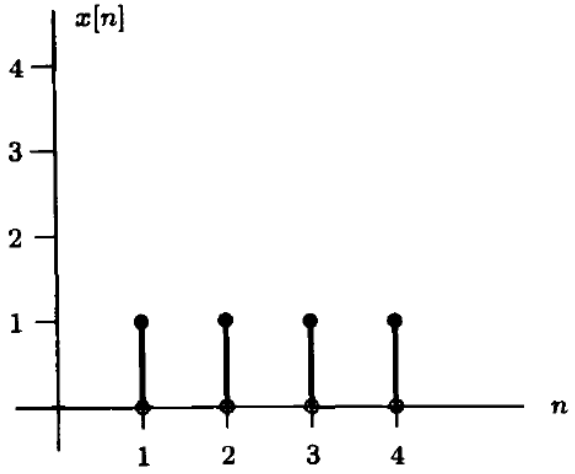
- When only A is unknown, $\mathbf{var}(\hat{A}) \geq \frac{\sigma^2}{N}$.
- When both A and B are unknown, $\mathbf{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)}$.
- For $N \geq 2$, $\frac{2(2N-1)\sigma^2}{N(N+1)} > \frac{\sigma^2}{N}$.

2. Some parameters are more sensitive than others

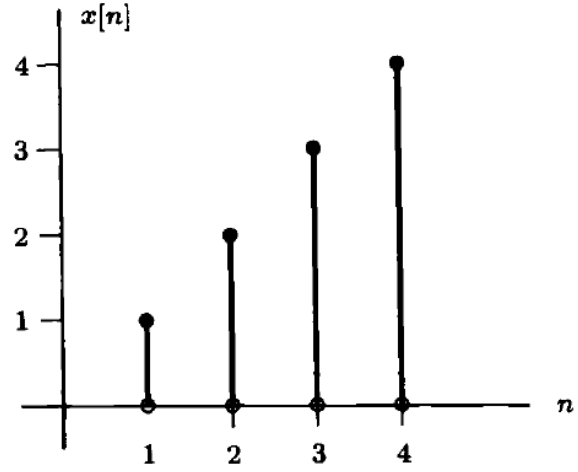
$$\frac{\mathrm{CRLB}(\hat{A})}{\mathrm{CRLB}(\hat{B})} = \frac{(2N-1)(N-1)}{6} > 1 \quad \text{for } N \geq 3.$$

i.e., B is easier to estimate

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



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



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

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

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

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

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

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

$$\text{var}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} \tag{3.6}$$

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

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{3.7}$$

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

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

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

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

$$\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0} \tag{3.24}$$

where $\geq \mathbf{0}$ is interpreted as meaning that the matrix is positive semidefinite. The Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is given as

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

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

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}) \tag{3.25}$$

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

Vector Parameter CRLB for Transformations

$\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ r-dimensional function

$$\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \geq \mathbf{0}$$

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

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

Jacobian matrix (r x p):

$$\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}$$

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

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

$$\text{var}(\hat{\alpha}) \geq \frac{4\alpha + 2\alpha^2}{N}$$

CRLB for the General Gaussian Case

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta})\right)$$

$$\begin{aligned}[\mathbf{I}(\boldsymbol{\theta})]_{ij} &= \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}\right]^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}\right] \\ &+ \frac{1}{2}\mathrm{tr}\left[\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right]\end{aligned}$$

Appendix 3C

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_1}{\partial \theta_i} \\ \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_2}{\partial \theta_i} \\ \vdots \\ \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_N}{\partial \theta_i} \end{bmatrix} \qquad \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{11}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{12}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{1N}}{\partial \theta_i} \\ \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{21}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{22}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{2N}}{\partial \theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{N1}}{\partial \theta_i} & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{N2}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{NN}}{\partial \theta_i} \end{bmatrix}$$

$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\theta), \mathbf{C}(\theta))$ (scalar parameter)

$$\begin{aligned}I(\theta) &= \left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta}\right]^T \mathbf{C}^{-1}(\theta) \left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta}\right] \\ &+ \frac{1}{2}\mathrm{tr}\left[\left(\mathbf{C}^{-1}(\theta)\frac{\partial \mathbf{C}(\theta)}{\partial \theta}\right)^2\right]\end{aligned}$$

$x[n] = A + w[n]$, where

- $w[n] \sim \mathcal{N}(0, \sigma^2)$
- $A \sim \mathcal{N}(0, \sigma_A^2)$

$$\begin{aligned}\left[\mathbf{C}(\sigma_A^2)\right]_{ij} &= E\left[x[i-1]x[j-1]\right] \\ &= E\left[(A + w[i-1])(A + w[j-1])\right] \\ &= \sigma_A^2 + \sigma^2\delta_{ij}.\end{aligned}$$

$$\mathbf{C}(\sigma_A^2) = \sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I}$$

Matrix inversion lemma:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$$

Woodbury identity:

$$(\mathbf{A} + \mathbf{uu}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{uu}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}$$

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

$$\frac{\partial \mathbf{C}(\sigma_A^2)}{\partial \sigma_A^2} = \mathbf{1}\mathbf{1}^T$$

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

$$\begin{aligned}I(\theta) &= \left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta}\right]^T \mathbf{C}^{-1}(\theta)\left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta}\right] \\ &\quad + \frac{1}{2}\mathrm{tr}\left[\left(\mathbf{C}^{-1}(\theta)\frac{\partial \mathbf{C}(\theta)}{\partial \theta}\right)^2\right]\end{aligned}$$

$$\begin{aligned}I(\sigma_A^2) &= \frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma^2 + N\sigma_A^2}\right)^2 \mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\right] \\ &= \frac{N}{2}\left(\frac{1}{\sigma^2 + N\sigma_A^2}\right)^2 \mathrm{tr}(\mathbf{1}\mathbf{1}^T) \\ &= \frac{1}{2}\left(\frac{N}{\sigma^2 + N\sigma_A^2}\right)^2\end{aligned}$$

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