I225E Statistical Signal Processing

8. Estimation Theory

MAWALIM and UNOKI

candylim@jaist.ac.jp and unoki@jaist.ac.jp

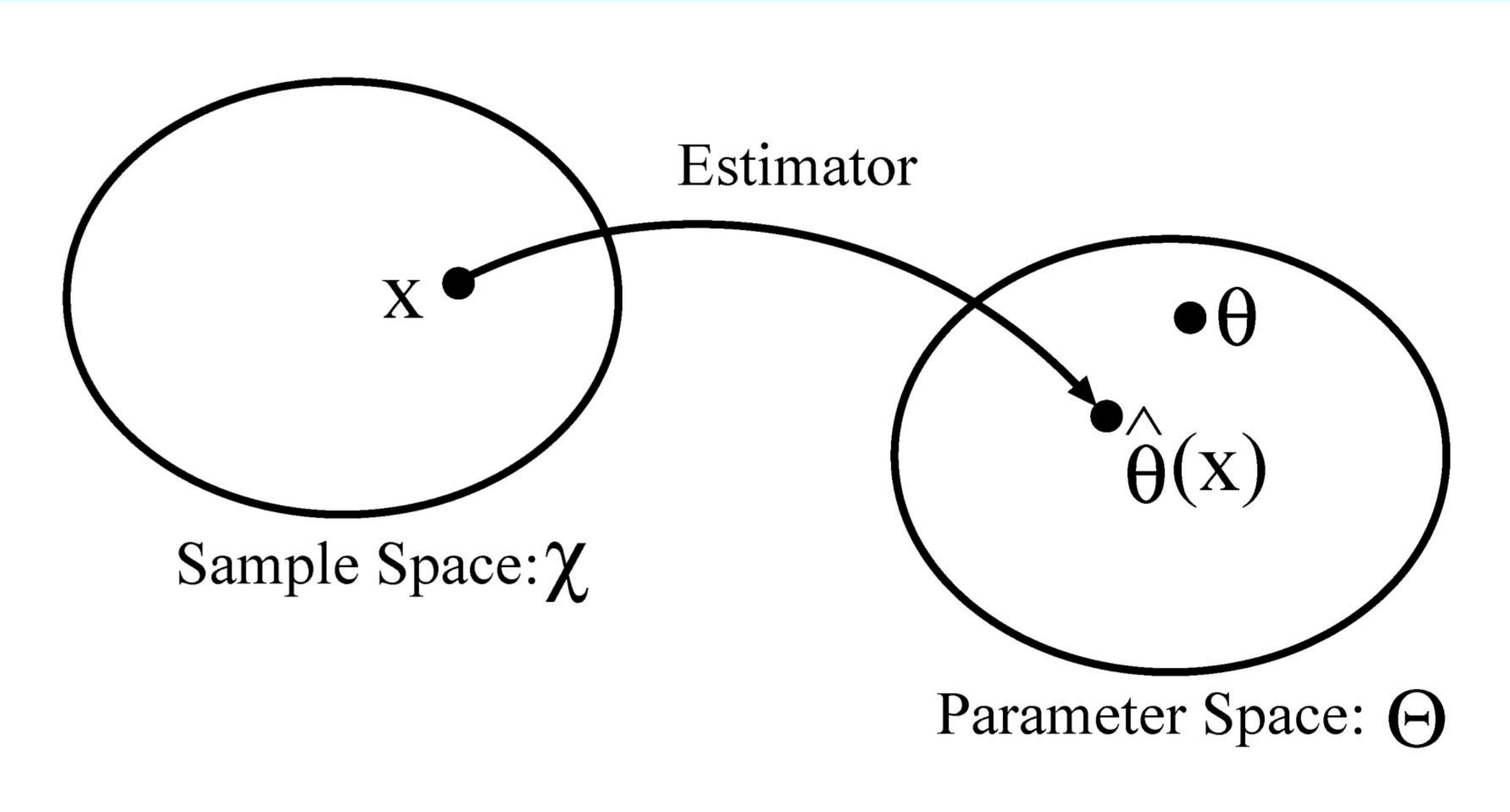
School of Information Science

1. What is (Classical) Estimation?

χ:	**Sample space**
$\boldsymbol{X} \in \chi$:	**Random vector of measurements**
$\boldsymbol{\Theta}$	**Parameter space**
$\theta \in \boldsymbol{\Theta}$:	**Unknown parameter**
$p(x;\theta)$:	**Probability density of $\boldsymbol{X}$**

Problem:

A realization $x \in \chi$ of $\boldsymbol{X}$ is observed. Distribution of $\boldsymbol{X}$ is specified entirely by the unknown parameter θ. Estimate θ using x. Find an estimator $\hat{\theta}: \chi \to \Theta$.


Estimator
x
θ
θ̂(x)
Sample Space: χ
Parameter Space: Θ

Example 1.1

Image restoration.

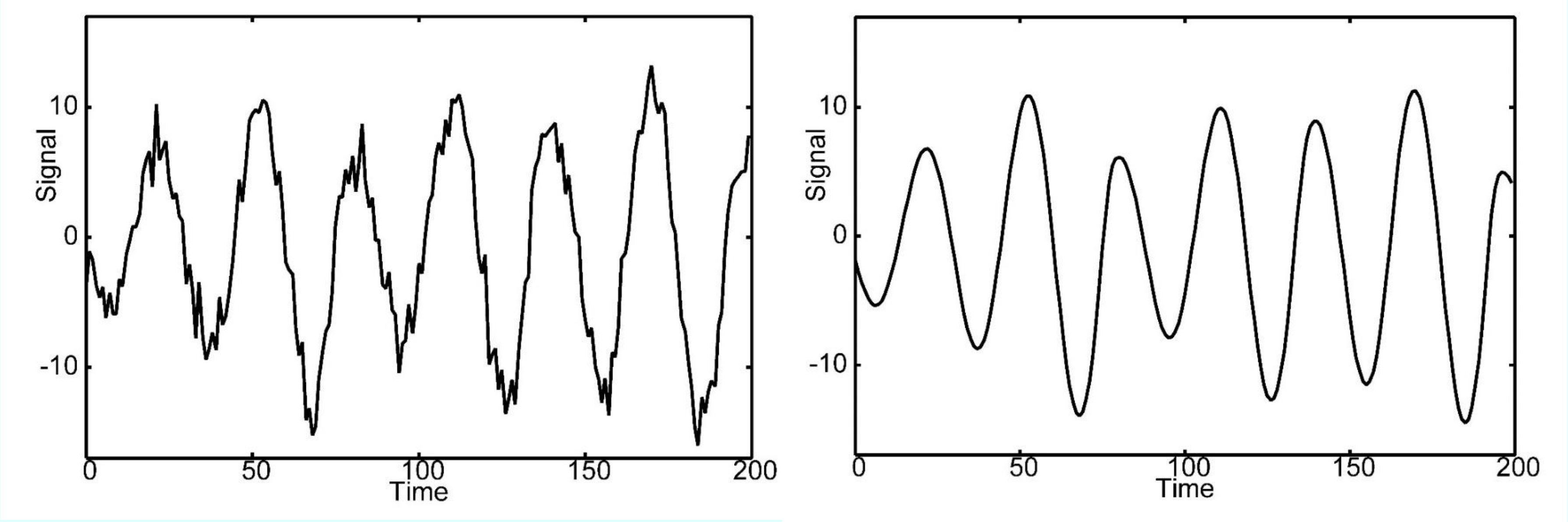
Restore true image (θ) from a picture contaminated with noise.

Example 1.2

Signal denoising.

From signal $x(n) = s(n) + w(n)$,

estimate $\theta = [s(0), \cdots, s(N-1)]^T$.



2. Estimation Categories

1. Optimality Criterion:

A) Classical estimation θ is nonrandom

- Minimum variance unbiased estimation (MVUE)
- Maximum likelihood estimation (MLE)
- Least squared estimation (LSE)

B) Bayesian estimation: θ is random

- Minimum mean squared error (MMSE)

2. Form:

- Linear estimator: $\hat{\theta}(x) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$
- Nonlinear estimator: Other forms than linear

3. Off-line vs. Online:

When we study filtering, we focus on estimators that can efficiently update their estimate as new data comes in.

3. Classical Estimation: Basic Notions

- **Mean Squared Error**

$$\mathrm{MSE}\left(\hat{\theta}\right) = \mathrm{E}\left\{\left\|\hat{\theta} - \theta\right\|^2\right\}$$

- **Variance**

$$\mathrm{Var}\left(\hat{\theta}\right) = \mathrm{E}\left\{\left\|\hat{\theta} - E\{\hat{\theta}\}\right\|^2\right\}$$

- **Bias**

$$\mathrm{Bias}\left(\hat{\theta}\right) = \mathrm{E}\{\hat{\theta}\} - \theta$$

- We say $\hat{\theta}$ is ***unbiased*** if $\mathrm{Bias}\left(\hat{\theta}\right) = 0, \forall \theta \in \Theta$.

- Let $\left\{\hat{\theta}_n\right\}_{n=1}^{\infty}$ be a family of estimators. We say $\left\{\hat{\theta}_n\right\}_{n=1}^{\infty}$ is ***asymptotically unbiased*** if $\mathrm{Bias}\left(\hat{\theta}_n\right) \to 0 \ as \ n \to \infty \ \ \forall \theta \in \Theta$.

- We say $\left\{\hat{\theta}_n\right\}_{n=1}^{\infty}$ is ***consistent*** if $\mathrm{MSE}(\hat{\theta}_n) \to 0$ as $n \to \infty \ \forall \theta \in \Theta$.

Example 3.1

Suppose $\mathbf{X} = [X[0], X[1], \cdots, X[N-1]]^T$, where $X[n] \sim N(\mu, \sigma^2)$, $n = 0, \cdots, N-1$. Consider the estimator of μ given by $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} X[n]$. What is the bias of $\hat{\theta}$?

Note: $\sim$ indicates i.i.d. (independent and identically distributed)

Answer:

$$
\begin{aligned}
\mathrm{E}\{\hat{\theta}\} &= \mathrm{E}\left\{\frac{1}{N}\sum_{n=0}^{N-1} X[n]\right\} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{E}\{X[n]\} = \frac{1}{N}\sum_{n=0}^{N-1} \mu \\
&= \mu.
\end{aligned}
$$

Hence $\hat{\theta}$ is unbiased.

Exercise 3.1

Compute the variance of $\hat{\theta}$ and check whether $\hat{\theta}$ is consistent.

Answer: $\mathrm{Var}(\hat{\theta}) = \mathrm{E}\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\} = \mathrm{E}\left\{\left(\hat{\theta} - \mu\right)^2\right\}$

$$= \mathrm{E}\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1} X[n] - \mu\right)^2\right\} = \mathrm{E}\left\{\sum_{n=0}^{N-1}\left(\frac{X[n]-\mu}{N}\right)^2\right\}$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-1} \mathrm{E}\{(X[n] - \mu)^2\} = \frac{1}{N^2}\sum_{n=0}^{N-1} \sigma^2 = \frac{\sigma^2}{N}.$$

Mean squared error is

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left\{\left(\hat{\theta} - \mu\right)^2\right\} = \mathrm{E}\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\}$$

$$= \mathrm{Var}(\hat{\theta}) = \frac{\sigma^2}{N} \to 0 \text{ as } N \to \infty$$ (Law of large number)

Hence $\hat{\theta}$ is a consistent estimator.

Note

For an estimator $\hat{\theta}$ to be unbiased, its expected value must be the true value for all $\theta \in \Theta$.

For instance, in the previous example, suppose we take $\hat{\theta} = \frac{2}{N}\sum_{n=0}^{N-1} X[n]$.

Then

$$\mathrm{E}\{\hat{\theta}\} = \mu \text{ if } \mu = 0,$$
$$\mathrm{E}\{\hat{\theta}\} \neq \mu \text{ if } \mu \neq 0,$$

Therefore, $\hat{\theta}$ is biased.

Exercise 3.2

You are a quality control engineer at a manufacturing plant. You take a random sample of n items from a large batch and find Y defective items. You want to estimate p, the true proportion of defective items in the entire batch.

Consider two possible estimators for p:

- The sample proportion: $\hat{p}_1 = \frac{Y}{n}$
- A slightly adjusted estimator: $\hat{p}_2 = \frac{Y+1}{n+2}$ (This is sometimes used as a smoother estimate, especially when you have very few samples or zero successes/failures).

Assume that the number of defective items Y in a sample of size n follows a Binomial distribution: $Y \sim \text{Binomial}(n, p)$.

Exercise 3.2

a) Is the sample proportion $\hat{p}_1$ an unbiased estimator for p? Prove your answer by finding $E[\hat{p}_1]$.
b) Calculate the bias of the estimator $\hat{p}_2$. Is it unbiased?
c) What happens to the bias of $\hat{p}_2$ as the sample size n becomes very large? Is $\hat{p}_2$ asymptotically unbiased?

4. Bias-Variance Tradeoff

- Mean squared error can be decomposed into bias and variance as follows:

$$\mathrm{MSE}(\hat{\theta}) = \left\|\mathrm{Bias}(\hat{\theta})\right\|^2 + \mathrm{Var}(\hat{\theta})$$

When designing an estimator $\hat{\theta}$, decreasing the bias of $\hat{\theta}$ will increase the variance, while increasing the bias will decrease the variance.

[Proof] $\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left\{(\hat{\theta} - \theta)^2\right\}$

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

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

$$+2\mathrm{E}\left\{\left(\hat{\theta} - \mathrm{E}[\hat{\theta}]\right)\left(\mathrm{E}[\hat{\theta}] - \theta\right)\right\} + \mathrm{E}\left\{\left(\mathrm{E}[\hat{\theta}] - \theta\right)^2\right\}$$

Since the second term is

$$\mathrm{E}\left\{\left(\hat{\theta}-\mathrm{E}\left[\hat{\theta}\right]\right)\left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right)\right\} = \left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right)\mathrm{E}\left\{\left(\hat{\theta}-\mathrm{E}\left[\hat{\theta}\right]\right)\right\}$$
$$= \left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right)\left(\mathrm{E}\left[\hat{\theta}\right]-\mathrm{E}\left[\hat{\theta}\right]\right) = 0$$

$$\mathrm{MSE}\left(\hat{\theta}\right) = \mathrm{E}\left\{\left(\hat{\theta}-\mathrm{E}\left[\hat{\theta}\right]\right)^2\right\} + \mathrm{E}\left\{\left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right)^2\right\}$$
$$= \mathrm{E}\left\{\left(\hat{\theta}-\mathrm{E}\left[\hat{\theta}\right]\right)^2\right\} + \left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right)^2$$
$$= \mathrm{Var}\left(\hat{\theta}\right) + \left\|\mathrm{Bias}(\hat{\theta})\right\|^2$$

Example 4.1

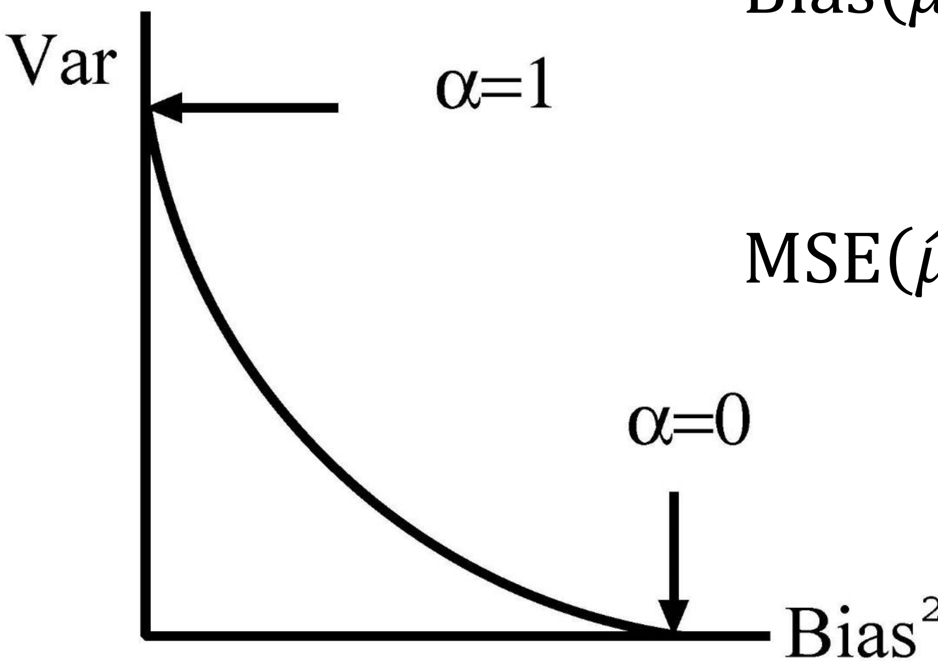
Suppose $\mathbf{X} = [X[0], X[1], \cdots, X[N-1]]^T$, where $X[n] \sim N(\mu, \sigma^2)$, $n = 0, \cdots, N-1$. Consider the class of estimators for μ as $\hat{\mu}_\alpha = \frac{\alpha}{N}\sum_{n=0}^{N-1} X[n]$ $(\alpha \in \Re)$. Let us see how α affects the bias-variance tradeoff.

Since $\hat{\mu}_\alpha = \alpha\bar{X}, \bar{X} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$,

$$\mathrm{Var}(\hat{\mu}_\alpha) = \mathrm{Var}(\alpha\bar{X}) = \alpha^2\mathrm{Var}(\bar{X}) = \frac{\alpha^2\sigma^2}{N}$$

$$\begin{aligned}\mathrm{Bias}(\hat{\mu}_\alpha) &= \mathrm{E}[\hat{\mu}_\alpha] - \mu = \mathrm{E}[\alpha\bar{X}] - \mu \\ &= \alpha\mu - \mu = (\alpha - 1)\mu\end{aligned}$$

$$\mathrm{MSE}(\hat{\mu}_\alpha) = (\alpha - 1)^2\mu^2 + \frac{\alpha^2\sigma^2}{N}$$



$$\mathrm{Var}(\hat{\mu}_\alpha) = \mathrm{Var}(\alpha\bar{X})$$
$$= \alpha^2\mathrm{Var}(\bar{X}) = \frac{\alpha^2\sigma^2}{N}$$
$$\mathrm{Bias}(\hat{\mu}_\alpha) = \mathrm{E}[\hat{\mu}_\alpha] - \mu$$
$$= \mathrm{E}[\alpha\bar{X}] - \mu$$
$$= \alpha\mu - \mu = (\alpha - 1)\mu$$
$$\mathrm{MSE}(\hat{\mu}_\alpha) = (\alpha - 1)^2\mu^2 + \frac{\alpha^2\sigma^2}{N}$$

Minimum Mean Squared Error

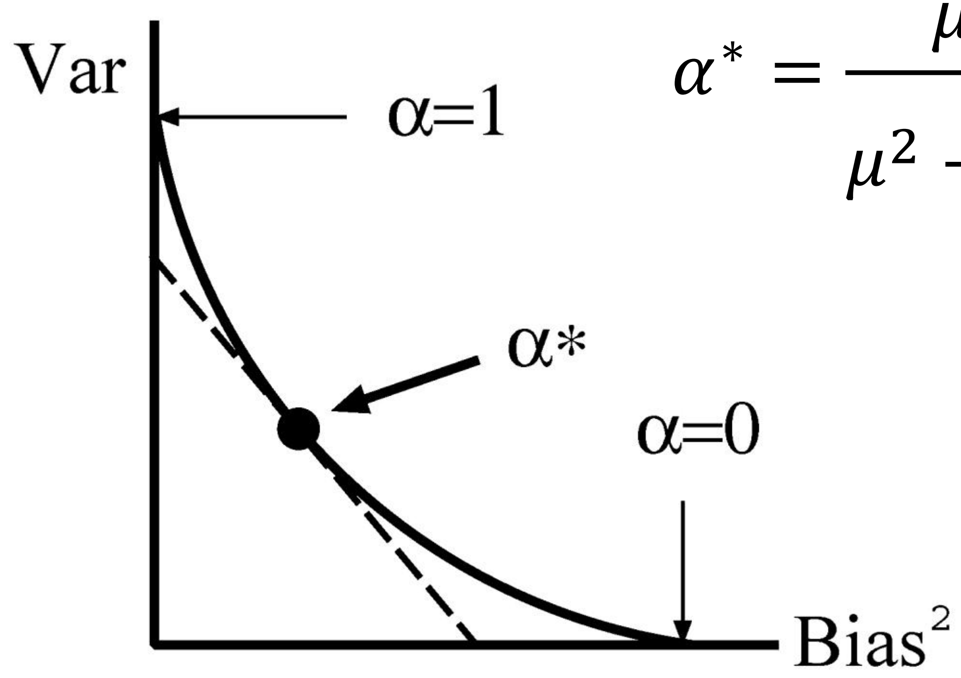
How practical is the mean squared error (MSE) as a criterion to design estimator?

In the previous example, the MSE is minimized when

$$\frac{\partial}{\partial \alpha}\mathrm{MSE}(\hat{\mu}_\alpha) = 2(\alpha - 1)\mu^2 + 2\alpha\frac{\sigma^2}{N} = 0$$

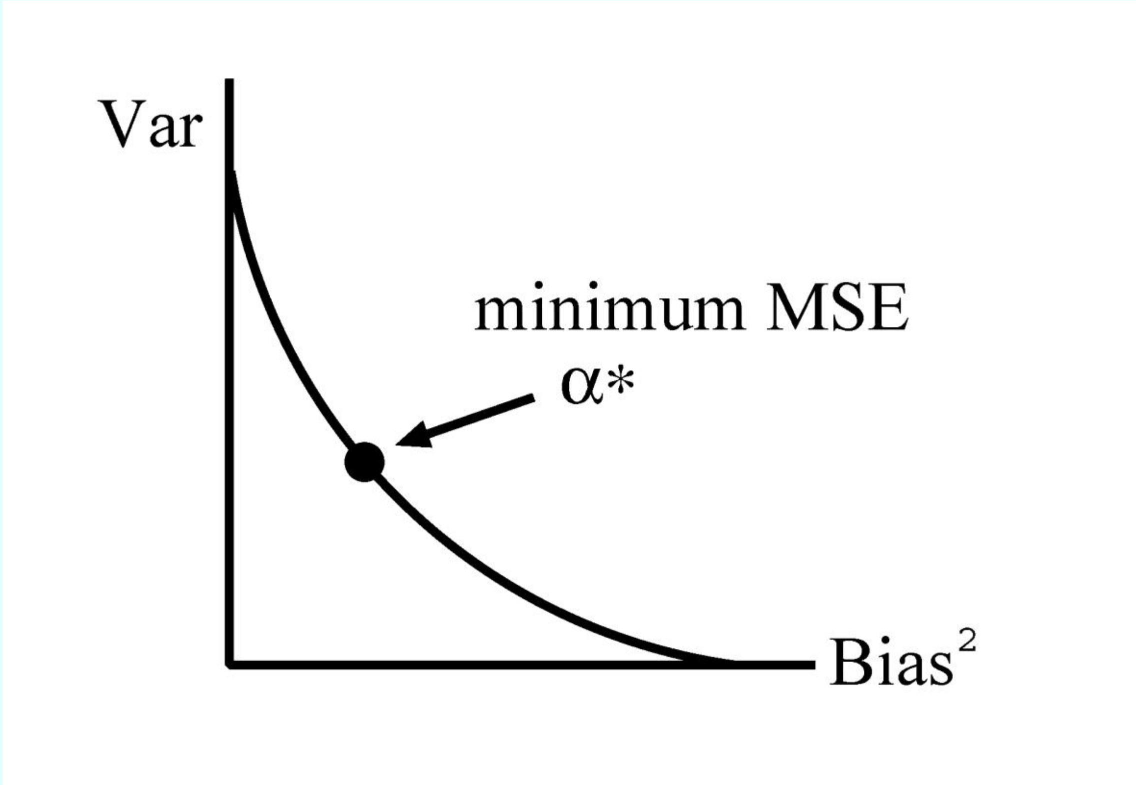
$$\rightarrow \alpha^* = \frac{\mu^2}{\mu^2 + \frac{\sigma^2}{N}}$$

Unfortunately, the solution depends on the unknown parameter μ. Therefore the estimator is not realizable.



$$\alpha^* = \frac{\mu^2}{\mu^2 + \frac{\sigma^2}{N}}$$

5. Minimum Variance Unbiased Estimation



In general, the minimum mean squared estimator has nonzero bias and variance. However, in many situations, only the bias depends on the unknown parameter.

In the previous example,

$$\mathrm{Bias}(\hat{\mu}_\alpha) = (\alpha - 1)\mu$$

$$\mathrm{Var}(\hat{\mu}_\alpha) = \frac{\alpha^2\sigma^2}{N}$$

This suggests that we constrain the estimator to be unbiased and minimize the variance. This is equivalent to minimization of the mean squared error (MSE) among all unbiased estimators.

Definition

$\hat{\theta}$ is said to be a (uniform) minimum variance unbiased estimator (MVUE) for θ if

1. $E\{\hat{\theta}\} = \theta,\ \theta \in \Theta$
2. For all unbiased estimators $\hat{\theta}_n$, $\mathrm{Var}(\hat{\theta}) \leq \mathrm{Var}(\hat{\theta}_n)$, $\forall \theta \in \Theta$

Remark:

MVUE criterion requires an estimator to be optimal for all values of θ.

It is a strong requirement for an estimator to have minimal MSE for all θ.

For example, the estimator $\hat{\theta} = 28$ provides the minimum when $\theta = 28$, but it is terrible elsewhere.

Example 5.1

Let $X[0]$ and $X[1]$ be independent random variables, normally distributed as $N(\mu, \sigma^2)$. Consider two estimators for μ as follows:

$$\hat{\theta}_1 = \frac{1}{2}X[0] + \frac{1}{2}X[1]$$

$$\hat{\theta}_2 = \frac{2}{3}X[0] + \frac{1}{3}X[1]$$

$$\mathrm{E}\{\hat{\theta}_1\} = \frac{1}{2}\mathrm{E}\{X[0]\} + \frac{1}{2}\mathrm{E}\{X[1]\} = \mu$$

$$\mathrm{E}\{\hat{\theta}_2\} = \frac{2}{3}\mathrm{E}\{X[0]\} + \frac{1}{3}\mathrm{E}\{X[1]\} = \mu$$

Therefore, $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased.

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\theta}_1\right) &= \frac{1}{4}\mathrm{Var}(X[0]) + \frac{1}{4}\mathrm{Var}(X[1]) \\
&= \frac{1}{2}\sigma^2 \\
\mathrm{Var}\left(\hat{\theta}_2\right) &= \frac{4}{9}\mathrm{Var}(X[0]) + \frac{1}{9}\mathrm{Var}(X[1]) \\
&= \frac{5}{9}\sigma^2
\end{aligned}
$$

Hence, $\mathrm{Var}(\hat{\theta}_1) \leq \mathrm{Var}(\hat{\theta}_2)$.

Existence of MVUE

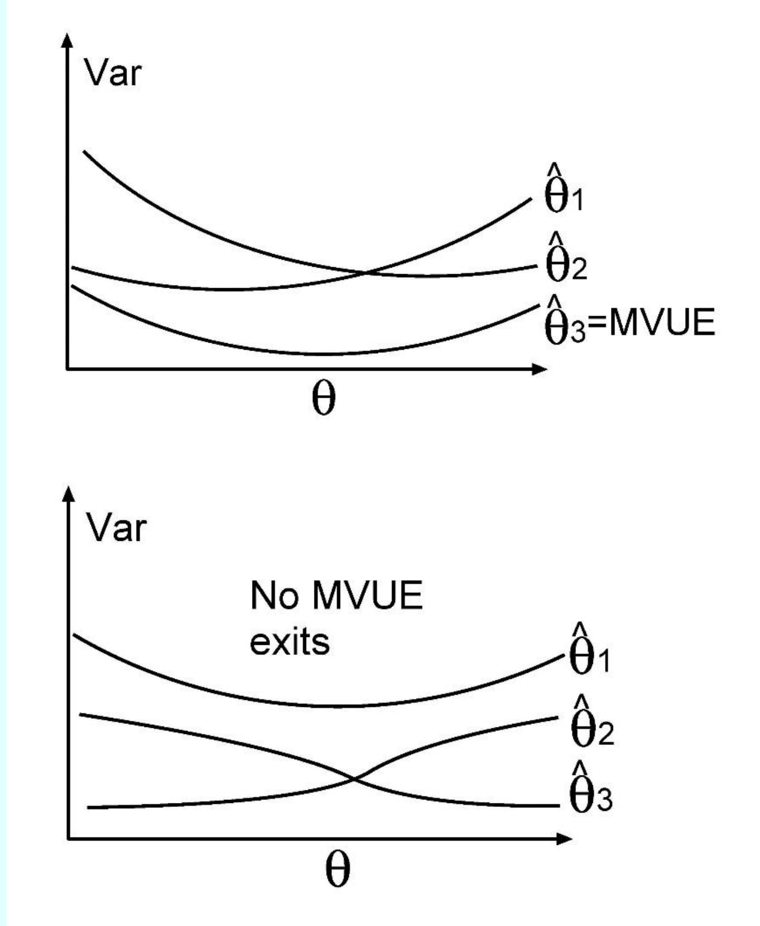
MVUE does not always exist.

Suppose there are three unbiased estimators, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$.

There are two possibilities:

- One estimator has uniformly smaller variance.
- No estimator has uniformly smaller variance.

There may not exist even a single unbiased estimator.



Example 5.2

Suppose we observe a single scalar realization of $X \sim U\left(0, \frac{1}{\theta}\right)$ $(0 < \theta)$. Show that an unbiased estimator for θ does not exist.

Density of $\boldsymbol{X}$ is

$$p(x,\theta) = \theta \quad \left(0 \le x \le \frac{1}{\theta}\right)$$
$$= 0 \quad \text{otherwise}$$

If $\hat{\theta}$ is unbiased then

For all $\theta > 0$,

$$\theta = \mathrm{E}\{\hat{\theta}\} = \int_{-\infty}^{\infty} \hat{\theta}(x) p(x,\theta) dx$$
$$= \int_{0}^{\frac{1}{\theta}} \hat{\theta}(x) \theta dx$$

Hence, $$\int_0^{\frac{1}{\theta}} \hat{\theta}(x)dx = 1.$$

Differentiation by $\frac{1}{\theta}\ (> 0)$,

For all $\theta > 0$, $$\hat{\theta}\left(\frac{1}{\theta}\right) = 0.$$

This is a contradiction.

Finding of the MVUE

No systematic procedure exists.

- Calculate the Cramér-Rao lower bound and see if some estimator achieves the bound.
- Apply the Rao-Blackwell theorem with a complete sufficient statistics.
- Restrict the class of possible estimators to be linear.

6. Cramér-Rao lower bound

Cremér-Rao lower bound (CRLB) is a lower bound on the variance of any unbiased estimator of a parameter θ.

- If $\hat{\theta}$ achieves the CRLB for all θ, $\hat{\theta}$ is an MVUE.
- CRLB provides a benchmark against which we can compare the performance of any unbiased estimator. We are doing well if our estimator is close to the CRLB.
- CRLB allows us to rule out impossible estimators. It is impossible to find an estimator better than the CRLB. This is useful in feasibility studies.
- Theory behind CRLB tells precisely when the bound is achievable.

Notions

- Likelihood function: $p(x;\theta)$

- Log likelihood function: $\ln p(x;\theta)$

- Fisher's information:

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

Theorem

Cramér-Rao Lower Bound – Scalar Parameter

Consider $X \sim p(x;\theta)$, where θ is fixed but unknown. Assume that $p(x;\theta)$ satisfies the "regularity" condition: $E\left\{\frac{\partial \ln p(x;\theta)}{\partial \theta}\right\} = 0$, where the expectation is with respect to $p(x;\theta)$. Then the variance of any unbiased estimator $\hat{\theta}$ satisfies

$$Var\left(\hat{\theta}\right) \geq \frac{1}{I(\theta)}.$$

The bound holds with equality if

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = I(\theta)(g(x) - \theta), \quad \forall x \in \chi.$$

- Estimator $\hat{\theta} = g(x)$, which satisfies the equality condition, provides the MVU estimator.

- The corresponding minimum variance is

$$Var(\hat{\theta}) = \frac{1}{I(\theta)}.$$

- Unbiased estimator $\hat{\theta}$ that gives CRLB is said to be ***efficient***.

Cauchy-Schwarz inequality

- Cauchy-Schwarz inequality

$$\mathrm{Var}(X)\mathrm{Var}(Y) \geq \mathrm{Cov}(X,Y)^2$$

- [Proof]

$$0 \leq \mathrm{Var}(X+tY) = \mathrm{Var}(X) + 2t\mathrm{Cov}(X,Y) + t^2\mathrm{Var}(Y)$$

$$= \mathrm{Var}(Y)\left\{t^2 + \frac{2\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}t + \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)^2}\right\} + \mathrm{Var}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}$$

$$= \mathrm{Var}(Y)\left\{t + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\right\}^2 + \frac{\mathrm{Var}(X)\mathrm{Var}(Y) - \mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}$$

Sketchy proof of Cramer-Rao theorem

- Cramer-Rao lower bound (CRLB)

$$\mathrm{Var}\left(\hat{\theta}\right) \geq \frac{1}{I(\theta)}.$$

- [Proof] Let us introduce an unbiased estimator $\hat{\theta} = g(X)$ and $V(X,\theta) = {}^{\partial}/_{\partial\theta} \ln p(X;\theta)$. Then using

$$\mathrm{Var}\{V(X,\theta)\} = \mathrm{E}\left\{\left(\frac{\partial \ln p(X;\theta)}{\partial\theta}\right)^2\right\} = I(\theta),$$

$$\mathrm{Cov}\left(\hat{\theta}, V\right) = \mathrm{E}\left\{g(X)\frac{\partial \ln p(X;\theta)}{\partial\theta}\right\} = 1,$$

and the Cauchy-Schwarz inequality, we find

$$I(\theta)\cdot\mathrm{Var}\left(\hat{\theta}\right) \geq 1.$$

Therefore, the CRLB holds.

Example 6.1

Suppose that random variable X is normally distributed as $N(\mu, \sigma^2)$. Variance σ^2 is known, but mean μ is to be estimated. Calculate CRLB of μ.

First, we need to check the regularity condition.

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

$$\ln p(x;\mu) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial \ln p(x;\mu)}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu)$$

$$E\left\{\frac{\partial \ln p(x;\mu)}{\partial \mu}\right\} = \frac{1}{\sigma^2}(E\{x\}-\mu) = 0.$$

Hence, regularity condition holds. Then we compute CRLB as

$$\frac{\partial^2}{\partial \mu^2} \ln p(x; \mu) = \frac{\partial}{\partial \mu} \left\{ \frac{1}{\sigma^2} (x - \mu) \right\} = -\frac{1}{\sigma^2},$$

$$I(\mu) = -E\left\{ \frac{\partial^2 \ln p(x;\mu)}{\partial \mu^2} \right\} = -E\left\{ -\frac{1}{\sigma^2} \right\} = \frac{1}{\sigma^2}.$$

Hence, any unbiased estimator $\hat{\mu}$ satisfies

$$Var(\hat{\mu}) \geq \sigma^2.$$

Therefore, $\hat{\mu} = X$ is efficient and MVUE.

Example 6.2

Suppose $X = [X[0], X[1], \cdots, X[N-1]]^T$, where $X[n] \sim N(\mu, \sigma^2)$, $n = 0, \cdots, N-1$. Variance σ^2 is known, but mean μ is to be estimated. Calculate CRLB of μ.

First, we check the regularity condition.

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

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

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

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

Hence, regularity condition holds. Then we compute CRLB as

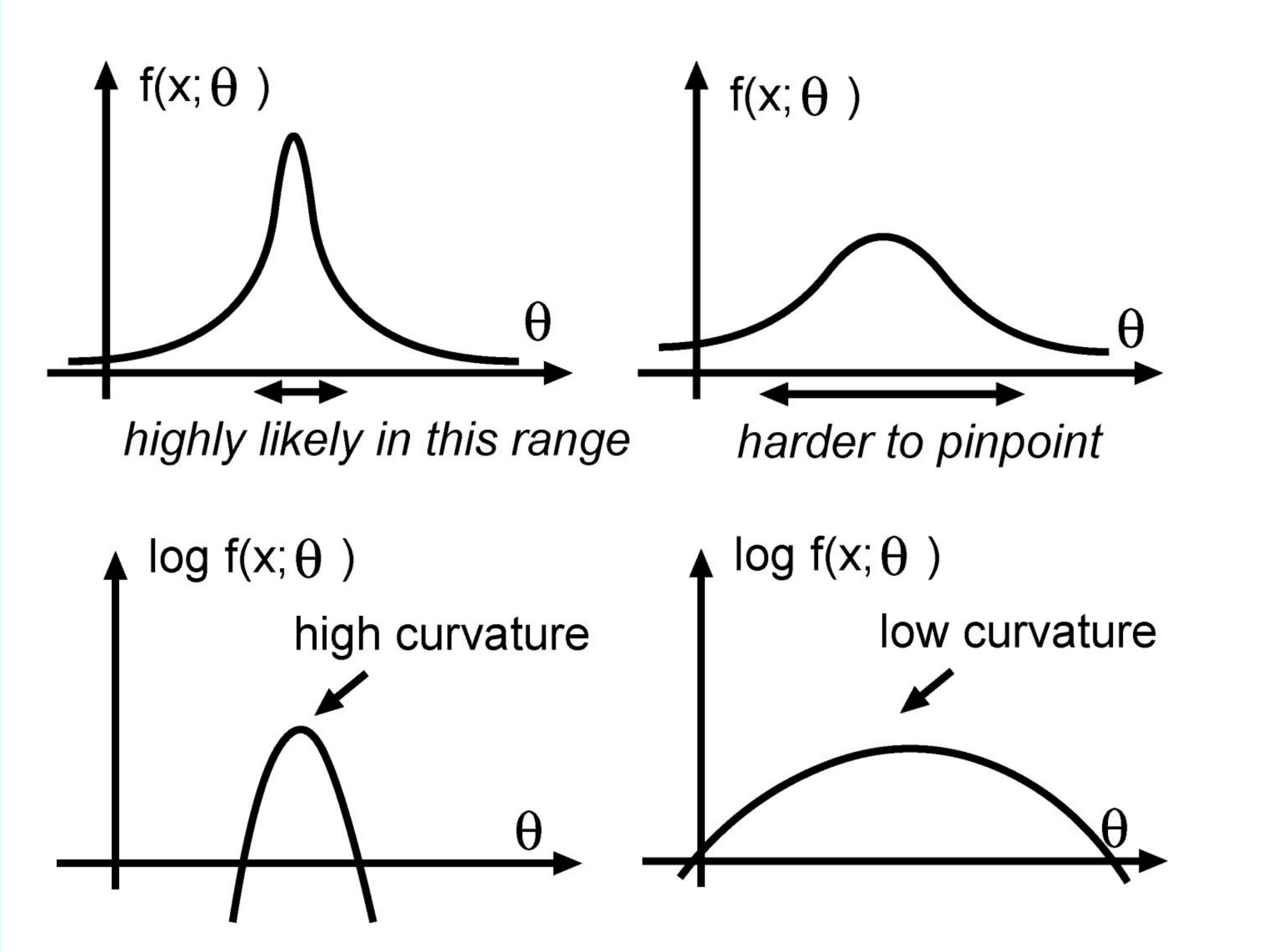
$$\frac{\partial^2}{\partial \mu^2} \ln p(\boldsymbol{x};\mu) = \frac{\partial}{\partial \mu}\left\{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-\mu)\right\} = -\frac{N}{\sigma^2},$$

$$\boldsymbol{I}(\mu) = -E\left\{\frac{\partial^2 \ln p(\boldsymbol{x};\mu)}{\partial \mu^2}\right\} = -E\left\{-\frac{N}{\sigma^2}\right\} = \frac{N}{\sigma^2}.$$

Hence, any unbiased estimator $\hat{\mu}$ satisfies $Var(\hat{\mu}) \geq \frac{\sigma^2}{N}$.

Sample average $\bar{X} = \frac{1}{N}\sum_{n=0}^{N-1} X[n]$ satisfies $E\{\bar{X}\} = \mu$, $Var\,(\bar{X}) = \frac{\sigma^2}{N}$. Hence, $\bar{X}$ is efficient and MVUE.

Fisher Information and Average Curvature



Fisher Information and Average Curvature

$I(\theta)$ reflects average curvature of log-likelihood $\ln p(x;\theta)$, since operator $\frac{\partial^2}{\partial\theta^2}$ measure curvature.

θ is easy to estimate

$\leftrightarrow p(x;\theta)$ is "peaky" near θ (on average)

$\leftrightarrow \ln p(x;\theta)$ has high curvature of θ (on average)

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

$\leftrightarrow$ CRLB is small

Exercise 6.1

$$\boldsymbol{X} = \begin{bmatrix} X[0] \\ X[1] \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

ρ is known but μ is to be estimated. Calculate CRLB of μ. Does efficient estimator exist?

Hint: $\ln p(\boldsymbol{x};\mu) = \frac{-1}{1+\rho}\{\mu^2 - \mu(x[0] + x[1])\} + C$.

Answer:

$$\frac{\partial \ln p(x;\mu)}{\partial \mu} = -\frac{2\mu - (x[0]+x[1])}{1+\rho},$$

$$E\left\{-\frac{2\mu - (x[0]+x[1])}{1+\rho}\right\} = -\frac{2\mu - (\mu+\mu)}{1+\rho} = 0$$

$$(E\{x[0]\} = E\{x[1]\} = \mu)$$

Hence, regularity condition holds:

$$\frac{\partial^2}{\partial \mu^2} \ln p(\boldsymbol{x};\mu) = -\frac{2}{1+\rho} \Rightarrow I(\mu) = -E\left\{-\frac{2}{1+\rho}\right\} = \frac{2}{1+\rho}$$

Hence, any unbiased estimator $\hat{\mu}$ satisfies

$$Var(\hat{\mu}) \geq \frac{1+\rho}{2}.$$

$$\frac{\partial \ln p(\boldsymbol{x};\mu)}{\partial \mu} = \frac{2}{1+\rho}\left(\frac{x[0]+x[1]}{2} - \mu\right) = I(\mu)(\hat{\mu} - \mu).$$

Therefore, $\hat{\mu} = \frac{x[0]+x[1]}{2}$ is efficient and MVUE.

7. Extension of CRLB to Vector Parameter

- **Theorem**

  (Cramér-Rao Lower Bound – Vector Parameter)

Consider $X \sim p(x;\theta)$, where $\theta \in \Theta \subseteq \Re^p$ is fixed but unknown. Assume that $p(x;\theta)$ satisfies the "regularity" condition:

$E\left\{\frac{\partial \ln p(x;\theta)}{\partial \theta}\right\} = 0$ for all θ, where the expectation is with respect to $p(x;\theta)$. Then the covariance matrix of any unbiased estimator $\hat{\theta}$ satisfies

$$Cov(\hat{\theta}) \geq I(\theta)^{-1}$$

The bound holds with equality if

$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = I(\theta)(g(x) - \theta), \quad \forall x \in \chi.$$

- Estimator $\hat{\theta} = g(x)$, which satisfies the equality condition, provides the MVU estimator.

- The corresponding minimum variance is

$$Var(\hat{\theta}_i) \geq |I(\theta)|_{ii}^{-1}.$$

Meaning of CRLB for Vector Parameter

- **Fisher information matrix:**

$$I(\theta) = \begin{bmatrix} -E\left\{\frac{\partial^2 \ln p(x;\theta)}{\partial \theta_1^2}\right\} & \cdots & -E\left\{\frac{\partial^2 \ln p(x;\theta)}{\partial \theta_1 \ \partial \theta_p}\right\} \\ \vdots & \ddots & \vdots \\ -E\left\{\frac{\partial^2 \ln p(x;\theta)}{\partial \theta_p \ \partial \theta_1}\right\} & \cdots & -E\left\{\frac{\partial^2 \ln p(x;\theta)}{\partial \theta_p^2}\right\} \end{bmatrix}$$

Covariance matrix:

$$Cov(\hat{\theta}) = \begin{bmatrix} Var(\hat{\theta}_1) & \cdots & Cov(\hat{\theta}_1, \hat{\theta}_p) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_p \hat{\theta}_1) & \cdots & Var(\hat{\theta}_p) \end{bmatrix}$$

Inequality "$Cov(\hat{\theta}) \geq I(\theta)^{-1}$" means "$Cov\left(\hat{\theta}\right) - I(\theta)^{-1}$ is positive semi definite." Each diagonal element satisfies

$$Var\left(\hat{\theta}_i\right) \geq |I(\theta)|_{ii}^{-1}$$

Vector CRLB implies that the lower bound exists for each element $\hat{\theta}_i$ of the estimator.

Example 7.1

Suppose $\boldsymbol{X} = [X[0], X[1], \cdots, X[N-1]]^T$, where $X[n] \sim N(\mu, \sigma^2)$, $n = 0, \cdots, N-1$. Calculate CRLB of $\theta = [\mu, \sigma]^T$.

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

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

$$\frac{\partial^2 \ln p(\boldsymbol{x};\mu)}{\partial\mu^2} = -\frac{N}{\sigma^2},$$

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

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

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

Therefore,

$$I(\theta) = E\begin{bmatrix} -\frac{N}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{n=0}^{N-1}(x[n]-\mu) \\ -\frac{1}{\sigma^4}\sum_{n=0}^{N-1}(x[n]-\mu) & \frac{N}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{n=0}^{N-1}(x[n]-\mu)^2 \end{bmatrix} = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

Any unbiased estimator $\hat{\theta} = [\hat{\mu}, \hat{\sigma}]^T$ satisfies

$$Var(\hat{\mu}) \geq \frac{\sigma^2}{N}$$

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