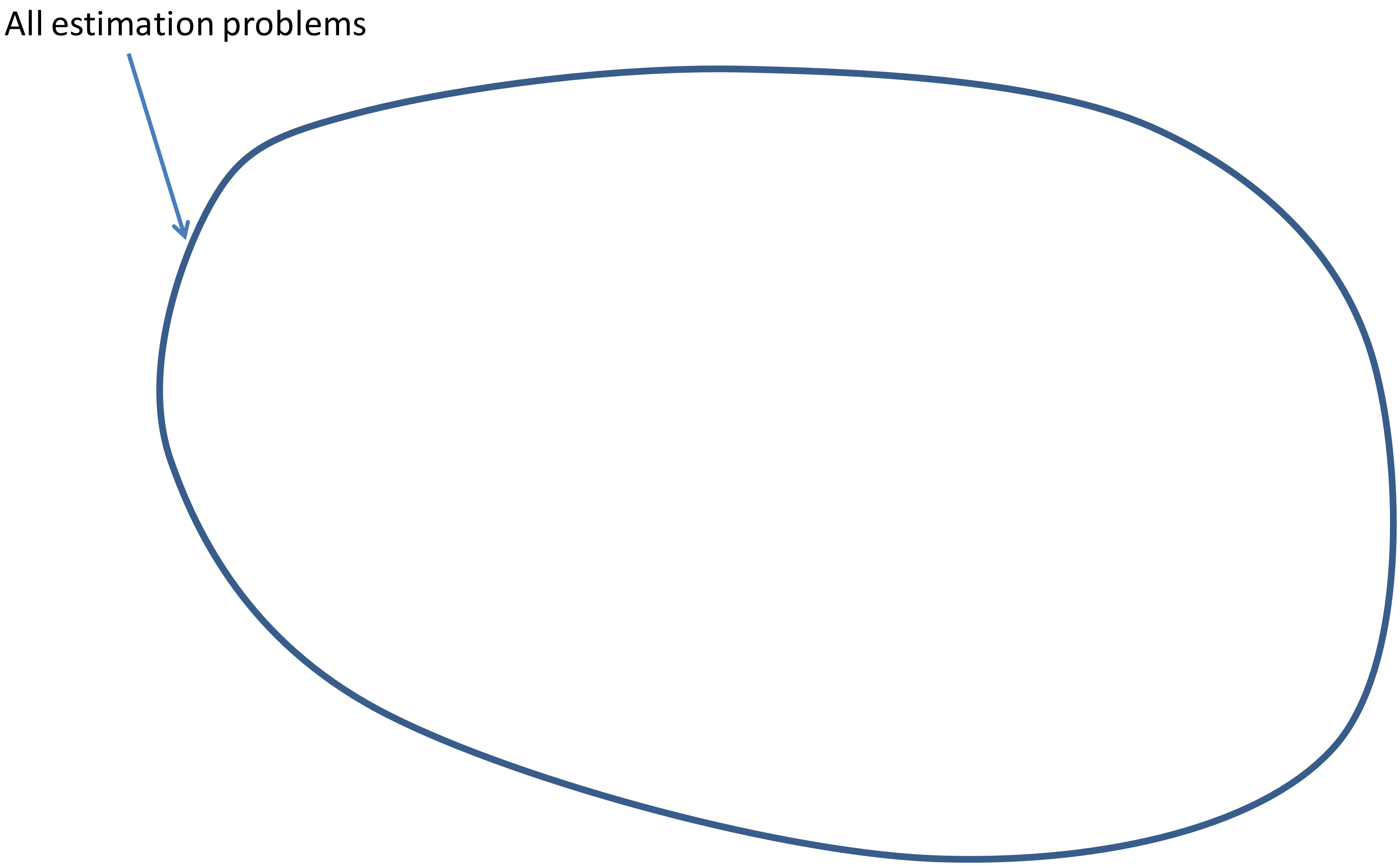
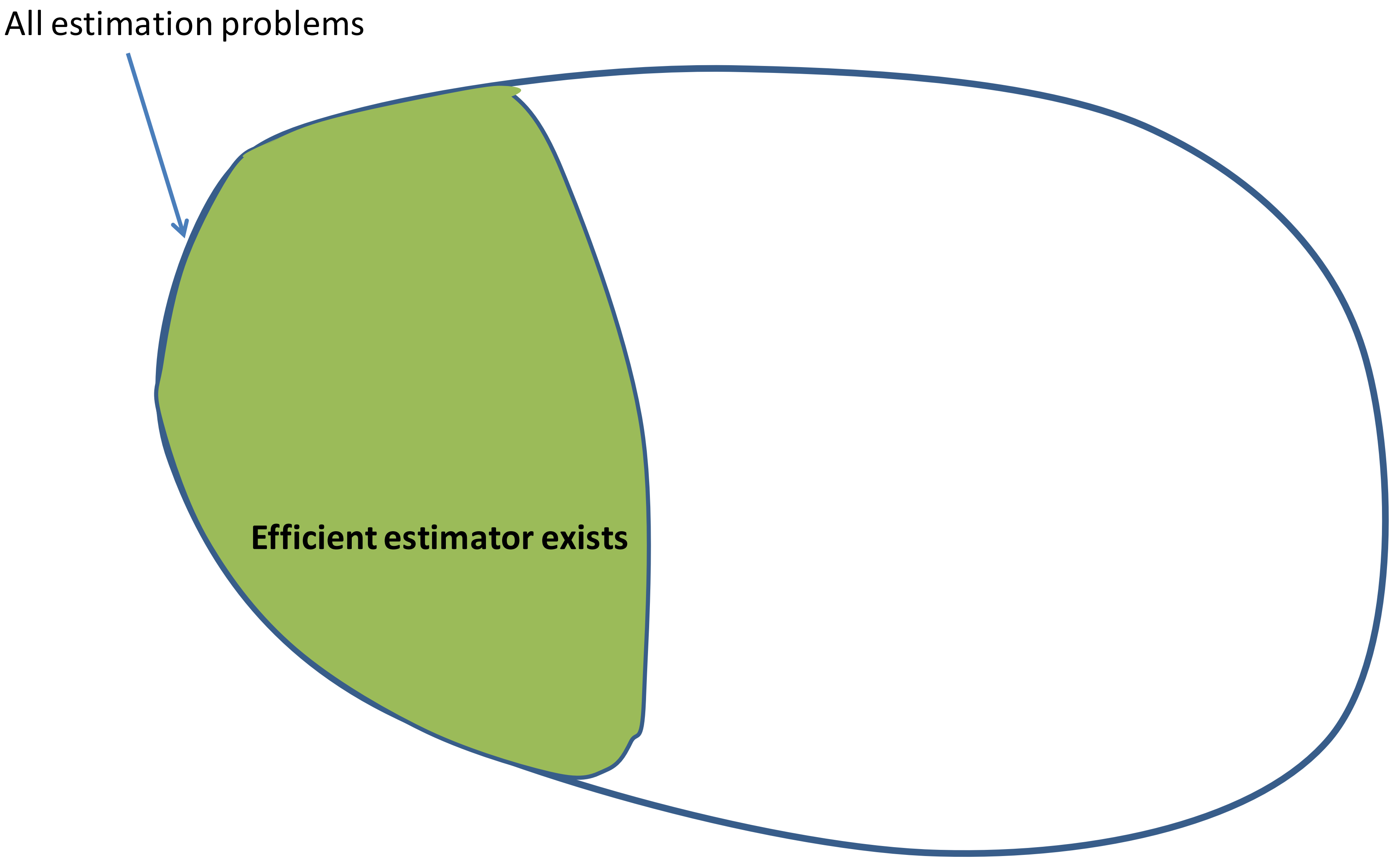
Estimation Theory
Fredrik Rusek

Chapters 6-7

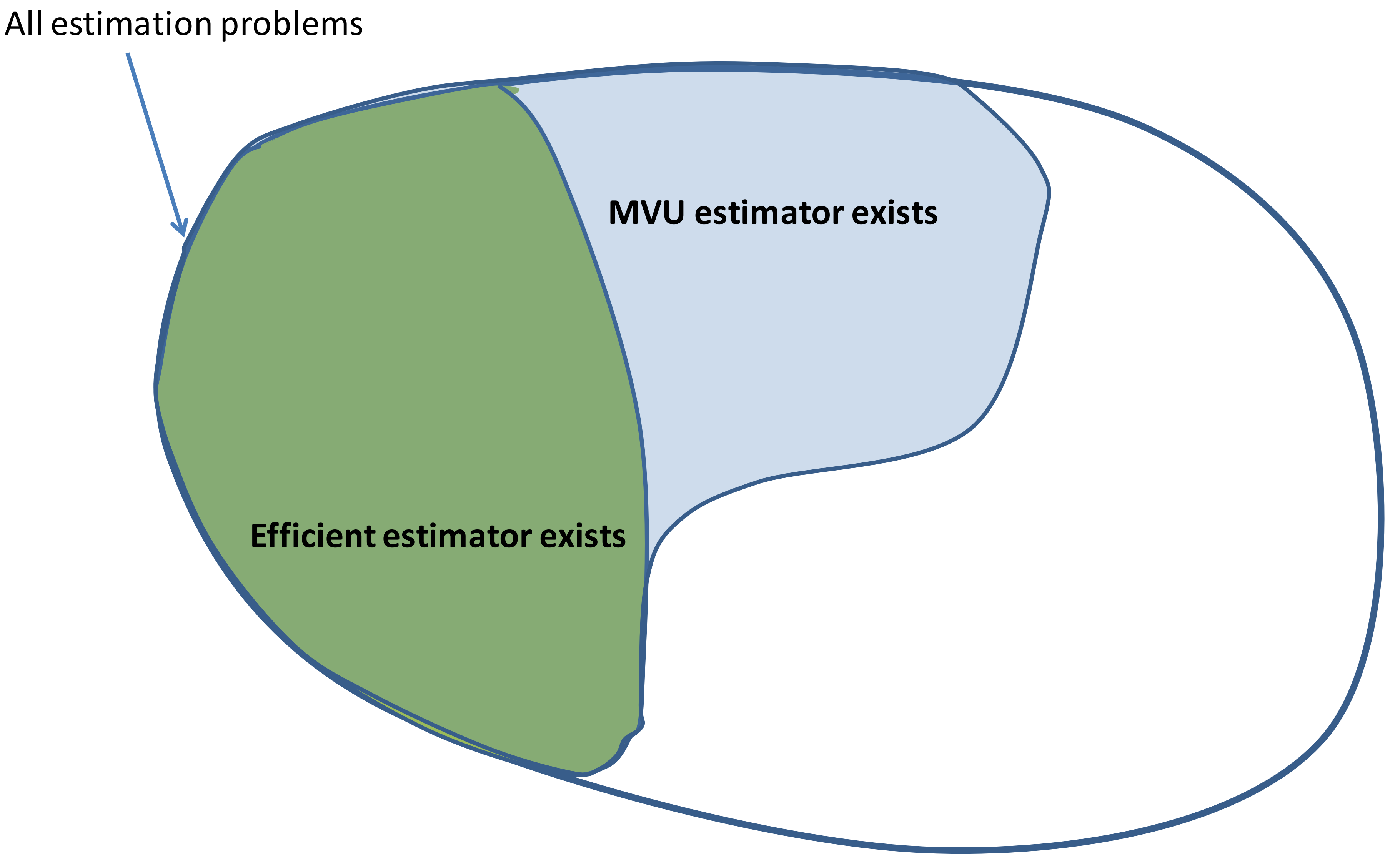
Summary



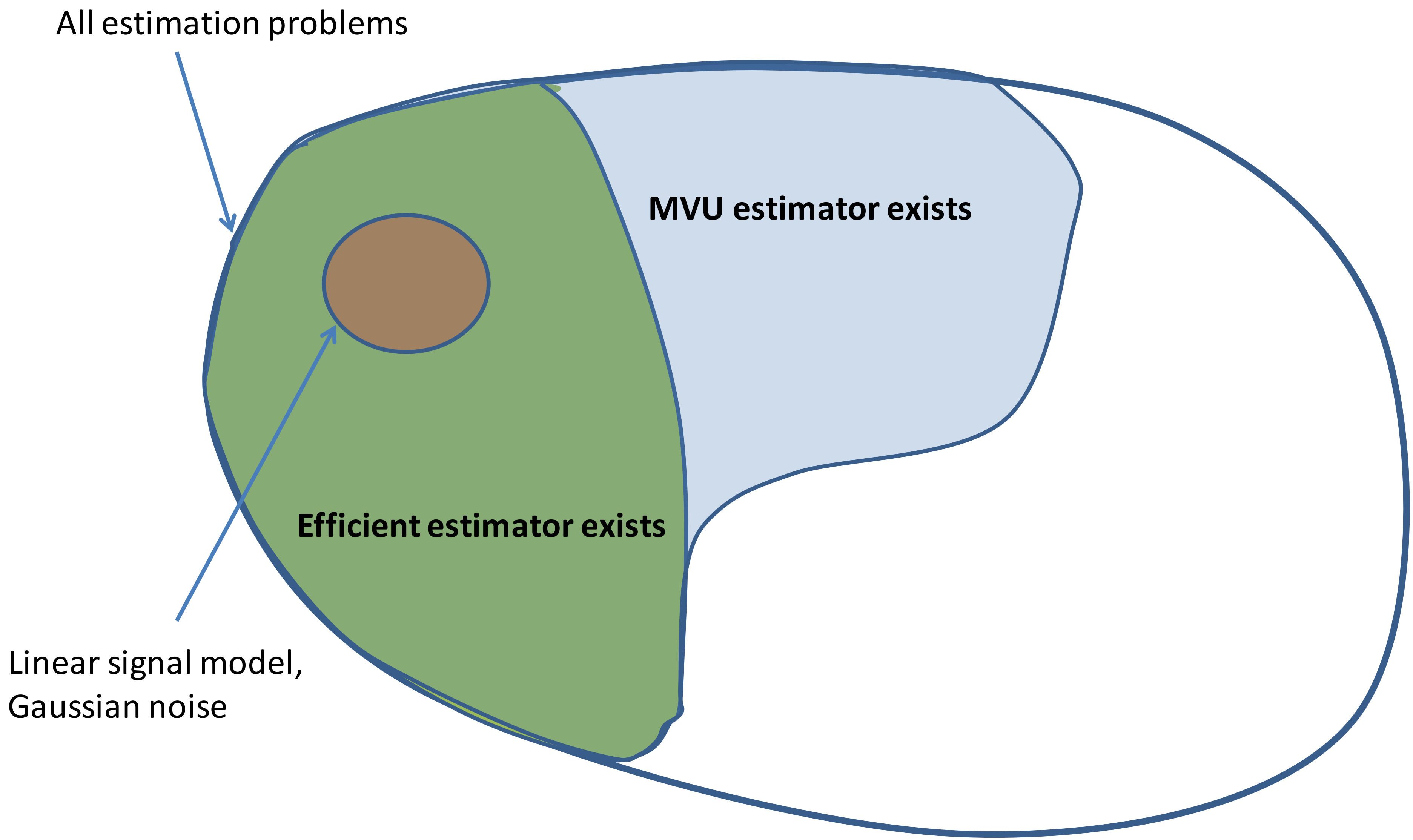
Summary



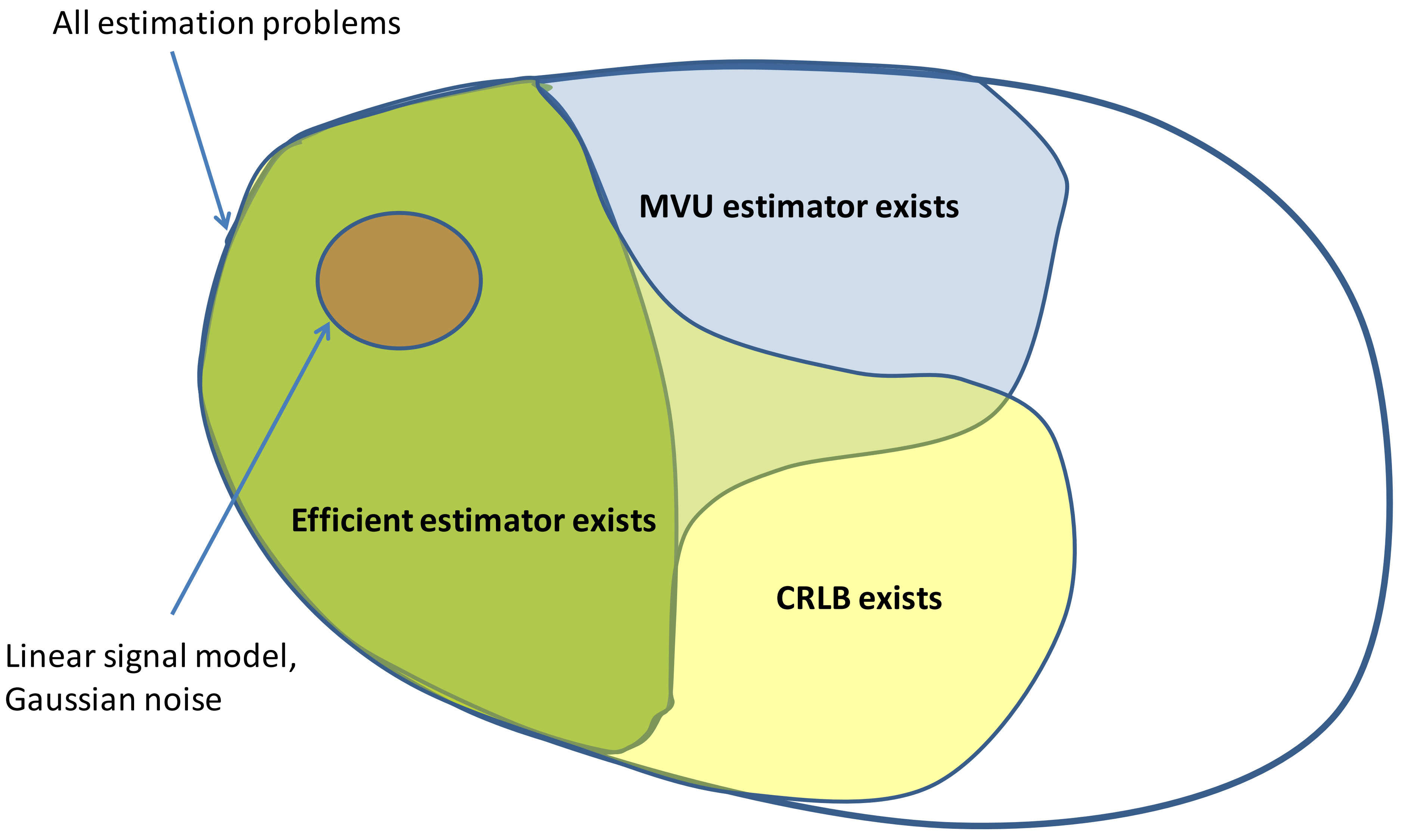
Summary


All estimation problems
MVU estimator exists
Efficient estimator exists

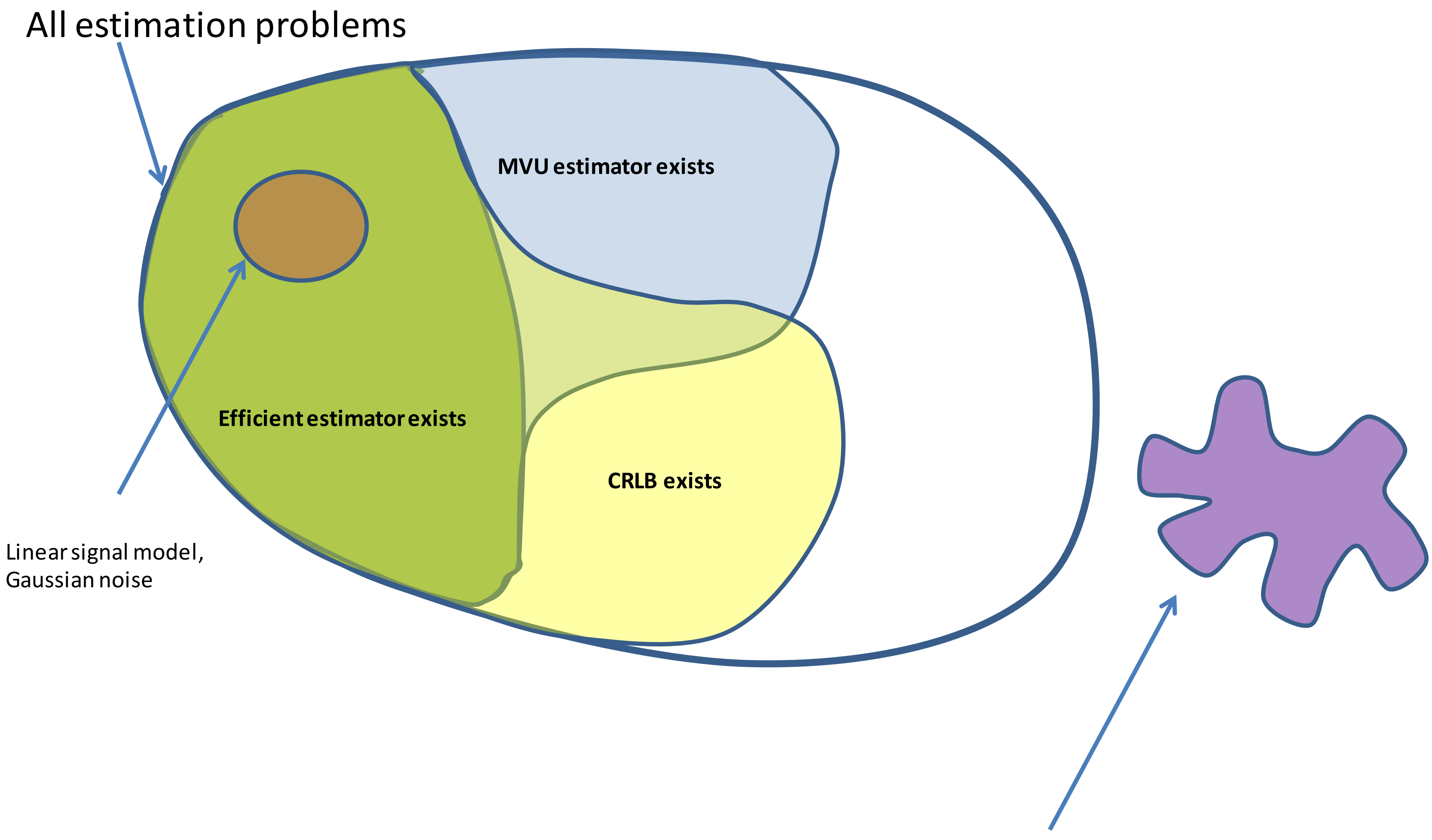
Summary


All estimation problems
MVU estimator exists
Efficient estimator exists
Linear signal model,
Gaussian noise

Summary

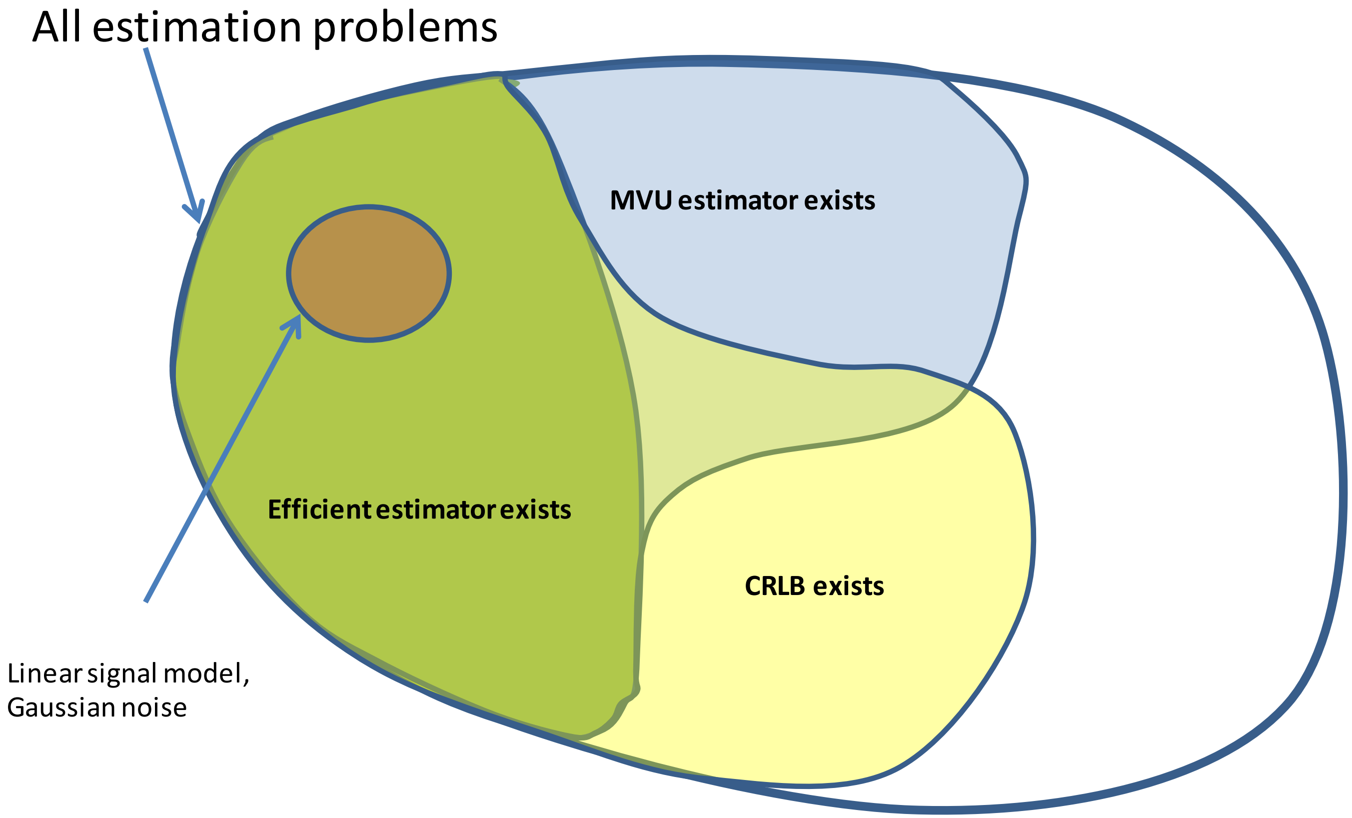

All estimation problems
MVU estimator exists
Efficient estimator exists
CRLB exists
Linear signal model,
Gaussian noise

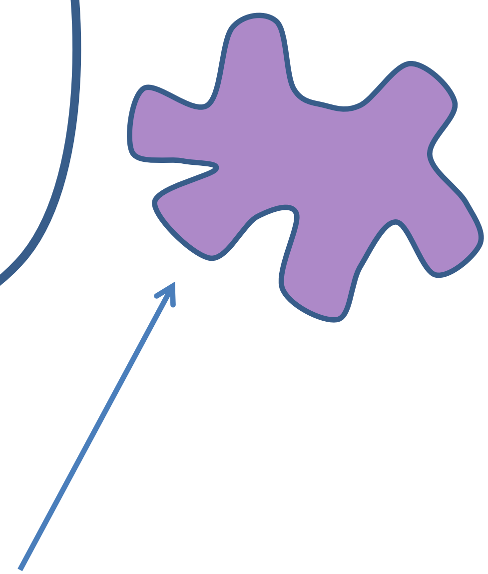
Summary



Today we will place the following piece:
Linear signal in non-Gaussian noise

Summary





Chapter 6 deals with non-Gaussian noise
This is not very well pointed out

Today we will place the following piece:
Linear signal in non-Gaussian noise

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Linear in received data

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

Many possible estimators possible.

The BLUE is the one that is:

- **Unbiased**
- **Smallest possible variance over all $\{a_n\}$**

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Linear in received data

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

Many possible estimators possible.

The BLUE is the one that is:

- **Unbiased**
- **Smallest possible variance over all $\{a_n\}$**

Optimal when

- **MVU is linear in the data**
- **Suboptimal otherwise**

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Linear in received data

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

Examples:

<u>DC level in white noise</u>

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

Many possible estimators possible.

The BLUE is the one that is:

- **Unbiased**
- **Smallest possible variance over all $\{a_n\}$**

Optimal when

- **MVU is linear in the data**
- **Suboptimal otherwise**

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Linear in received data

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

Many possible estimators possible.

The BLUE is the one that is:

- **Unbiased**
- **Smallest possible variance over all $\{a_n\}$**

Optimal when

- **MVU is linear in the data**
- **Suboptimal otherwise**

Examples:

<u>DC level in white noise</u>

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

<u>German tank problem</u>

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

BLUE: (can be shown) $\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$

BLUE ≠ MVU

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Linear in received data

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

Many possible estimators possible.

The BLUE is the one that is:

- **Unbiased**
- **Smallest possible variance over all $\{a_n\}$**

Optimal when

- **MVU is linear in the data**
- **Suboptimal otherwise**

Examples:

<u>DC level in white noise</u>

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

<u>German tank problem</u>

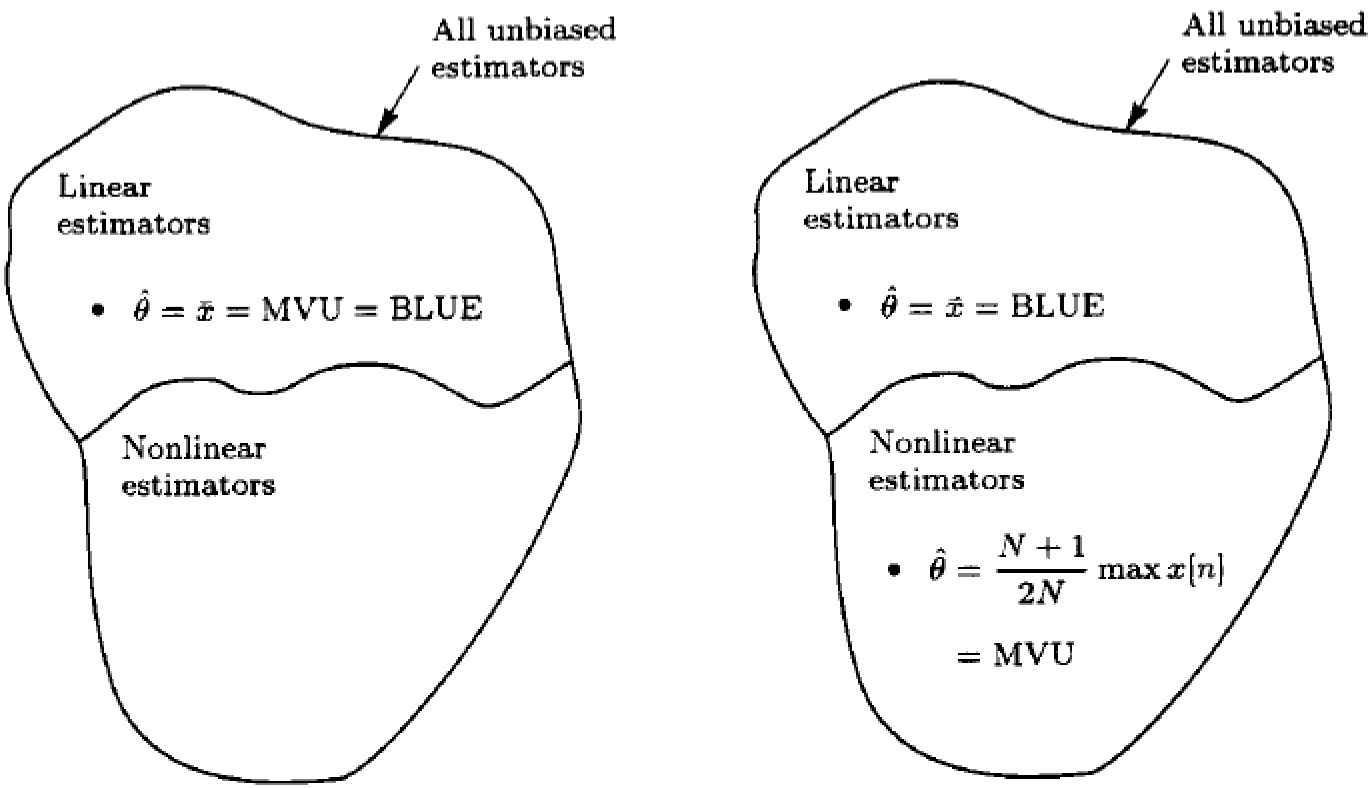
MVU: $\hat{\theta} = \frac{N+1}{2N} \max x[n]$

BLUE: **(can be shown)** $\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$

BLUE ≠ MVU

The MVU is often not linear, hence the name ”best linear”

Chapter 6 – Best linear unbiased estimators



(a) DC level in WGN; BLUE is optimal

(b) Mean of uniform noise; BLUE is suboptimal

The MVU is often not linear, hence the name ”best linear”

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Sometimes LE is very bad

<u>Noise power estimation</u> (x[n]=w[n])

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

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

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

Chapter 6 – Best linear unbiased estimators

Definition of the BLUE

Sometimes LE is very bad

Noise power estimation (x[n]=w[n])

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

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

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

BLUE + ”cleverness”

- Transform data as $y[n] = x^2[n]$
- Apply BLUE to y[n]

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

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

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned}\operatorname{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\ &= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right]\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned} \mathrm{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\ &= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \end{aligned}$$

Use vector notation

$$\boxed{\sum_{n=0}^{N-1} a_n x[n] = \mathbf{a}^T \mathbf{x}}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned}
\mathrm{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\
&= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T E(\mathbf{x})\right)^2\right]
\end{aligned}$$

Use vector notation

$$\boxed{\sum_{n=0}^{N-1} a_n x[n] = \mathbf{a}^T\mathbf{x}}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned}
\mathrm{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\
&= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T E(\mathbf{x})\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))\right)^2\right]
\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned}
\operatorname{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\
&= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T E(\mathbf{x})\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))\right)^2\right] \\
&= E\left[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T\mathbf{a}\right]
\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Variance

$$\begin{aligned}
\operatorname{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\
&= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T E(\mathbf{x})\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))\right)^2\right] \\
&= E\left[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T\mathbf{a}\right] \\
&= \mathbf{a}^T\mathbf{C}\mathbf{a}.
\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Optimization problem

$$\min_{\mathbf{a}} \quad \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

Variance

$$\begin{aligned}
\operatorname{var}(\hat{\theta}) &= E\left[\hat{\theta} - E(\hat{\theta})\right] \\
&= E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left(\sum_{n=0}^{N-1} a_n x[n]\right)\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T \mathbf{x} - \mathbf{a}^T E(\mathbf{x})\right)^2\right] \\
&= E\left[\left(\mathbf{a}^T (\mathbf{x} - E(\mathbf{x}))\right)^2\right] \\
&= E\left[\mathbf{a}^T (\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}\right] \\
&= \mathbf{a}^T \mathbf{C} \mathbf{a}.
\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Constraint

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

Optimization problem

$$\min_{\mathbf{a}} \; \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

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

Optimization problem

$$\min_{\mathbf{a}} \; \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Now write $$x[n] = E(x[n]) + [x[n] - E(x[n])]$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

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

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Now write $$x[n] = E(x[n]) + [x[n] - E(x[n])]$$

Define $$w[n] = [x[n] - E(x[n])]$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

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

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Now write $$x[n] = E(x[n]) + [x[n] - E(x[n])]$$

Define $$w[n] = [x[n] - E(x[n])]$$

Which yields $$\boxed{x[n] = \theta s[n] + w[n]}$$

w[n] not Gaussian, but zero-mean

This is the linear model, but with non-Gaussian noise

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

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

Optimization problem

$$\min_{\mathbf{a}} \quad \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Now write $$x[n] = E(x[n]) + [x[n] - E(x[n])]$$

We have $E\left[\mathbf{a}^T \mathbf{x}\right] = \mathbf{a}^T \mathbf{s}\, \theta$

Define $$w[n] = [x[n] - E(x[n])]$$

Which yields $$\boxed{x[n] = \theta s[n] + w[n]}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

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

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$\mathbf{a}^T \mathbf{s} = 1$$

To satisfy $E\left[\mathbf{a}^T \mathbf{x}\right] = \theta$ we must have $E(x[n]) = s[n]\theta$

Now write $$x[n] = E(x[n]) + [x[n] - E(x[n])]$$

We have $E\left[\mathbf{a}^T \mathbf{x}\right] = \mathbf{a}^T \mathbf{s}\, \theta$

Define $$w[n] = [x[n] - E(x[n])]$$

Which yields $$\boxed{x[n] = \theta s[n] + w[n]}$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T \mathbf{C} \mathbf{a} + \lambda(\mathbf{a}^T \mathbf{s} - 1)$

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T \mathbf{C} \mathbf{a}$$

$$\mathbf{a}^T \mathbf{s} = 1$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s} \qquad \mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}.$$

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{s} = 1$$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

Optimization problem

$$\min_{\mathbf{a}} \ \mathbf{a}^T\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{s} = 1$$

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s} \qquad \mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$$

To find λ, use the constraint function $\mathbf{a}^T\mathbf{s} = 1$ **with** $\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

Optimization problem

$$\min_{\mathbf{a}} \quad \mathbf{a}^T\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{s} = 1$$

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s} \qquad \mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$$

To find λ, use the constraint function $\mathbf{a}^T\mathbf{s} = 1$ with $\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$

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

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

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

Optimization problem

$$\min_{\mathbf{a}} \quad \mathbf{a}^T\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{s} = 1$$

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s} \qquad \mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$$

To find λ, use the constraint function $\mathbf{a}^T\mathbf{s} = 1$ with $\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$

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

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

Plug back: $\mathbf{a}_{\mathrm{opt}} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$

Chapter 6 – Best linear unbiased estimators

Finding the BLUE

Lagrangian $J = \mathbf{a}^T\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{s} - 1)$

Optimization problem

$$\min_{\mathbf{a}} \quad \mathbf{a}^T\mathbf{C}\mathbf{a}$$

$$\mathbf{a}^T\mathbf{s} = 1$$

$$\frac{\partial J}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda\mathbf{s} \qquad \mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$$

To find λ, use the constraint function $\mathbf{a}^T\mathbf{s} = 1$ with $\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$

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

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

Plug back: $\mathbf{a}_{\text{opt}} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}$

Variance
$$\begin{aligned}\operatorname{var}(\hat{\theta}) &= \mathbf{a}_{\text{opt}}^T\mathbf{C}\mathbf{a}_{\text{opt}} \\ &= \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{s}}{(\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s})^2} \\ &= \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}.\end{aligned}$$

Chapter 6 – Best linear unbiased estimators

To compute the BLUE, we need

- Know that $E(\mathbf{x}) = \mathbf{s}\theta$
- Know $\mathbf{s}$
- Know $\mathbf{C}$

Connections to the MVU estimator for the linear model will soon be made

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Constraint

$$E(\hat{\boldsymbol{\theta}}) = \mathbf{A}E(\mathbf{x}) = \boldsymbol{\theta}$$

$$E(\mathbf{x}) = \mathbf{H}\boldsymbol{\theta}$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Constraint

$$E(\hat{\boldsymbol{\theta}}) = \mathbf{A}E(\mathbf{x}) = \boldsymbol{\theta}$$

$$E(\mathbf{x}) = \mathbf{H}\boldsymbol{\theta}$$

Which becomes

$$\boxed{\mathbf{A}\mathbf{H} = \mathbf{I}}$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Constraint

$$E(\hat{\boldsymbol{\theta}}) = \mathbf{A}E(\mathbf{x}) = \boldsymbol{\theta}$$

$$E(\mathbf{x}) = \mathbf{H}\boldsymbol{\theta}$$

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

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \ldots & \mathbf{h}_p \end{bmatrix}$$

Which becomes

$$\boxed{\mathbf{A}\mathbf{H} = \mathbf{I}}$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Constraint

$$E(\hat{\boldsymbol{\theta}}) = \mathbf{A}E(\mathbf{x}) = \boldsymbol{\theta}$$

$$E(\mathbf{x}) = \mathbf{H}\boldsymbol{\theta}$$

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

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \ldots & \mathbf{h}_p \end{bmatrix}$$

Which becomes

$$\mathbf{a}_i^T \mathbf{h}_j = \delta_{ij} \qquad i = 1, 2, \ldots, p; j = 1, 2, \ldots, p$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1, 2, \ldots, p$$

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{x}$$

Constraint

$$E(\hat{\boldsymbol{\theta}}) = \mathbf{A}E(\mathbf{x}) = \boldsymbol{\theta}$$

$$E(\mathbf{x}) = \mathbf{H}\boldsymbol{\theta}$$

Cost function to optimize

$$\mathrm{var}(\hat{\theta}_i) = \mathbf{a}_i^T \mathbf{C} \mathbf{a}_i$$

Which becomes

$$\mathbf{a}_i^T \mathbf{h}_j = \delta_{ij} \qquad i = 1, 2, \ldots, p; j = 1, 2, \ldots, p$$

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

Optimal solution

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

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

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

Optimal solution

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

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

Summary

- For a model **y=Hθ+n,** a linear estimator has the same form no matter the distribution of the noise (for the same covariance **C** of it)

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

Optimal solution

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

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

Summary

- For a model **y=Hθ+n,** a linear estimator has the same form no matter the distribution of the noise (for the same covariance **C** of it)
- If **n** is Gaussian, then a linear estimator is MVU

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

Optimal solution

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

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

Summary

- For a model **y=Hθ+n,** a linear estimator has the same form no matter the distribution of the noise (for the same covariance **C** of it)
- If **n** is Gaussian, then a linear estimator is MVU
- If **n** is not Gaussian, the BLUE is the best linear estimator, but not MVU

Chapter 6 – Best linear unbiased estimators

BLUE for vector parameter

Optimal solution

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

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

Summary

- For a model **y=Hθ+n,** a linear estimator has the same form no matter the distribution of the noise (for the same covariance **C** of it)
- If **n** is Gaussian, then a linear estimator is MVU
- If **n** is not Gaussian, the BLUE is the best linear estimator, but not MVU
- Performance of the BLUE for non-Gaussian noise is identical to the performance with Gaussian noise

Chapter 6 – Best linear unbiased estimators

Est. Problem with Gaussian noise | **Est. Problem with Non-Gaussian noise**

Variance of linear estimator

Chapter 6 – Best linear unbiased estimators

	Est. Problem with Gaussian noise	Est. Problem with Non-Gaussian noise
Variance of linear estimator	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$

Chapter 6 – Best linear unbiased estimators

	Est. Problem with Gaussian noise	Est. Problem with Non-Gaussian noise
Variance of linear estimator	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$
	Estimator is Optimal	Estimator is not optimal

Chapter 6 – Best linear unbiased estimators

	Est. Problem with Gaussian noise	Est. Problem with Non-Gaussian noise
Variance of linear estimator	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$ Estimator is Optimal	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$ Estimator is not optimal
Variance of non-linear estimator	The same	?? But smaller than $\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$

Therefore, Gaussian noise is the worst noise possible

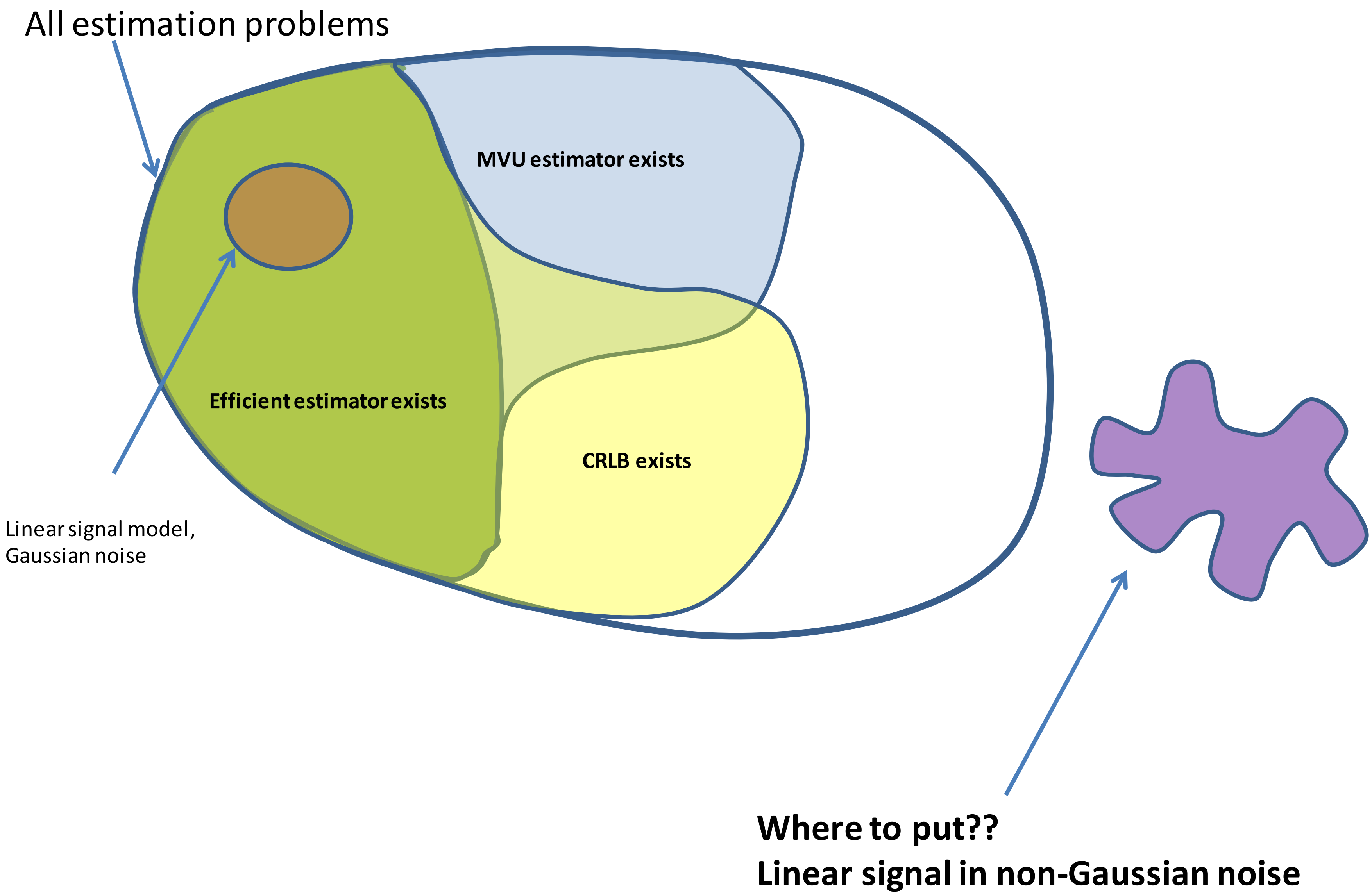
Chapter 6 – Best linear unbiased estimators

	Est. Problem with Gaussian noise	Est. Problem with Non-Gaussian noise
Variance of linear estimator	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$ Estimator is Optimal	$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$ Estimator is not optimal
Variance of non-linear estimator	The same	?? But smaller than $\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}.$

Alternative proofs

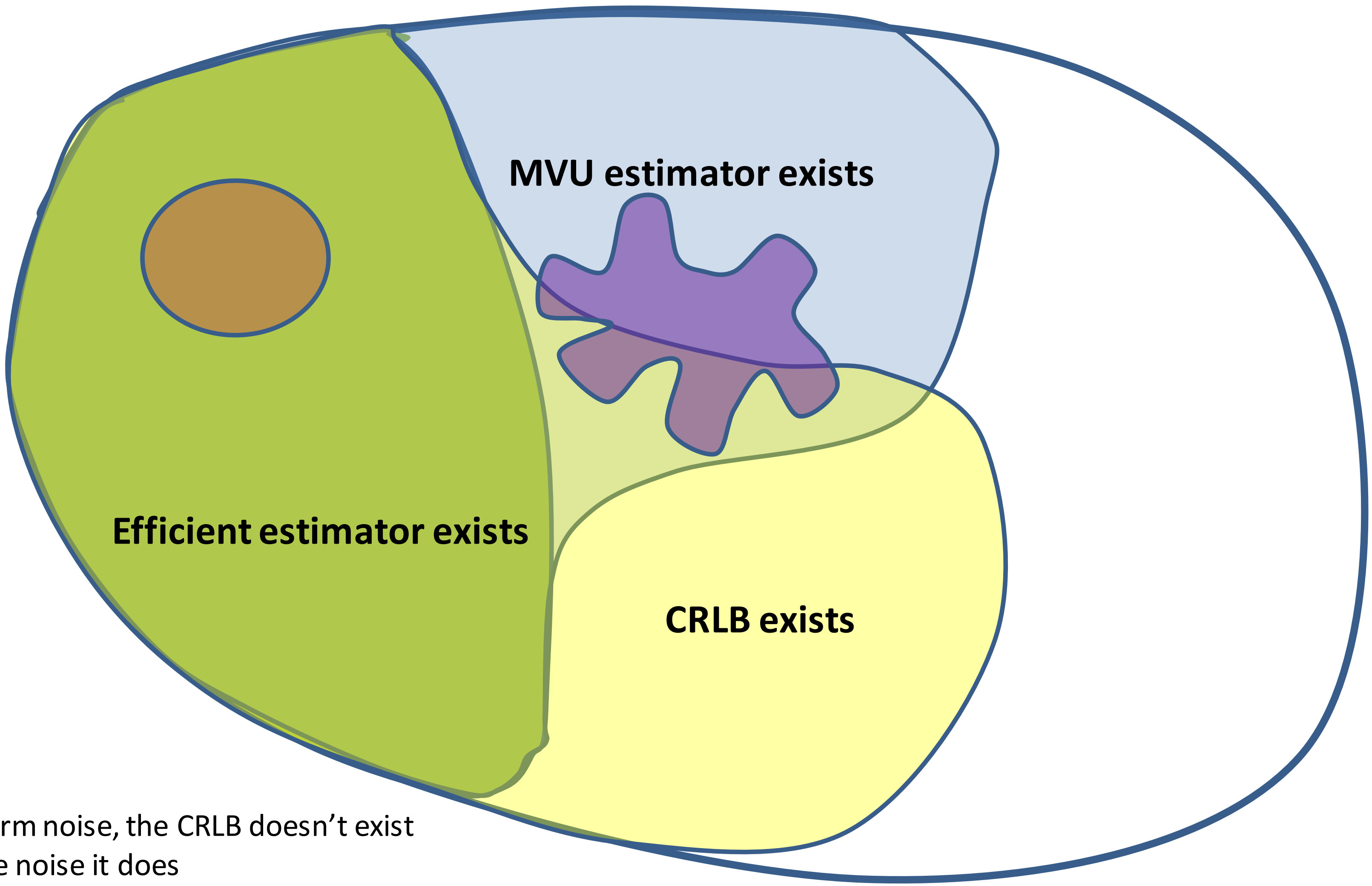
1. Based on calculus of the variations
2. Based on link between Fisher information and Kullback-Leibler divergence (information theory)

Summary


All estimation problems
MVU estimator exists
Efficient estimator exists
CRLB exists
Linear signal model,
Gaussian noise
Where to put??
Linear signal in non-Gaussian noise

We do not have explicit info

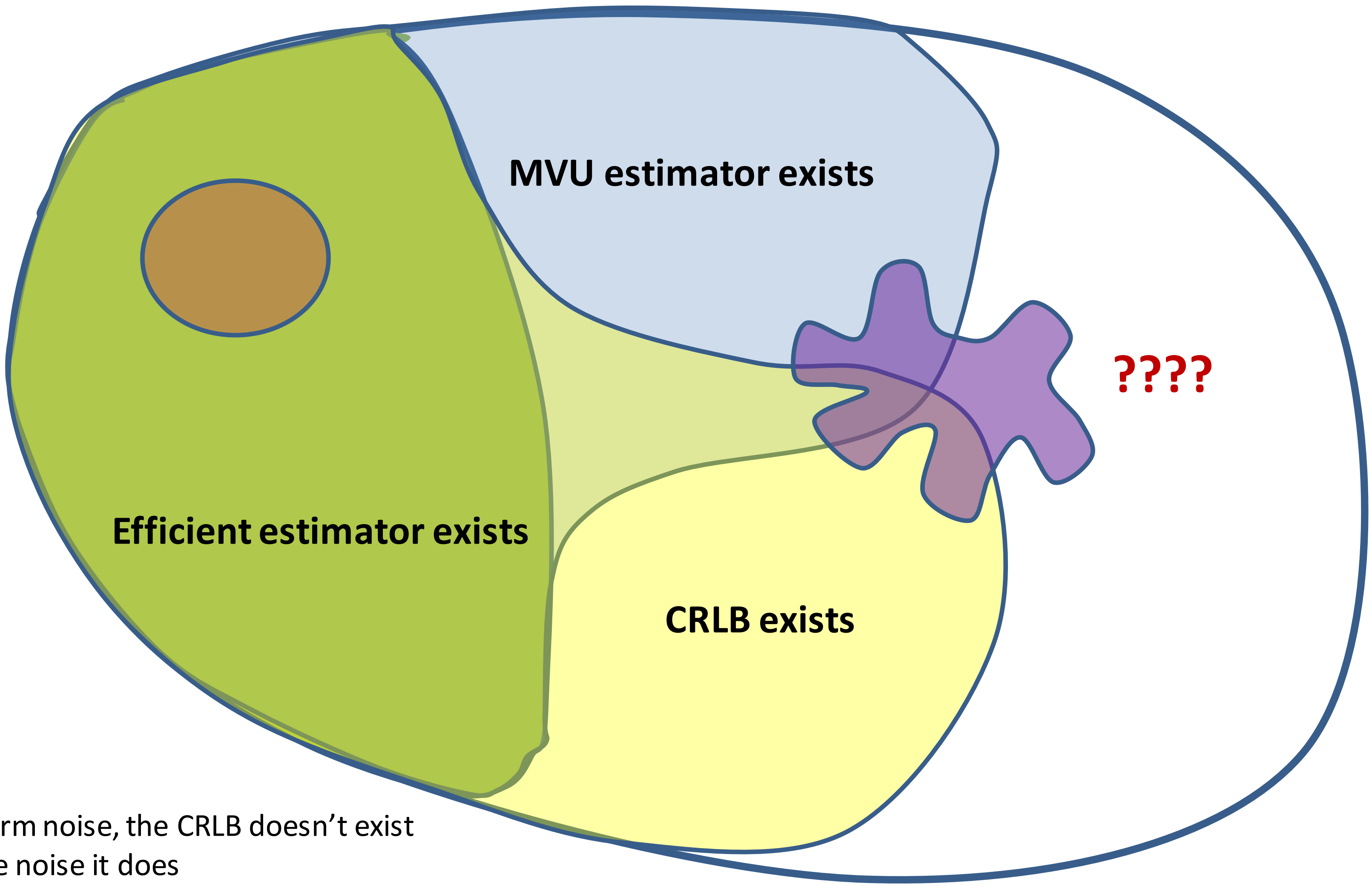
Summary ????



- With uniform noise, the CRLB doesn’t exist
- For Laplace noise it does
- No efficient estimator seems to exist – see my ”Italian proof”
- Perhaps MVU does not always exist….?

We do not have explicit info

Summary ????



- With uniform noise, the CRLB doesn't exist
- For Laplace noise it does
- No efficient estimator seems to exist – see my "Italian proof"
- Perhaps MVU does not always exist….?

Chapter 7 – Maximum Likelihood

$$\hat{\theta} = \underset{\theta}{arg\ max}\ p(\mathbf{x};\theta)$$

Chapter 7 – Maximum Likelihood

$$\hat{\theta} = \underset{\theta}{arg\ max}\ p(\mathbf{x};\theta)$$

Theorem. If an efficient estimator exists, then it is given by the MLE

Chapter 7 – Maximum Likelihood

$$\hat{\theta} = \underset{\theta}{arg\ max}\ p(\mathbf{x};\theta)$$

Theorem. If an efficient estimator exists, then it is given by the MLE

Proof.

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

Chapter 7 – Maximum Likelihood

$$\hat{\theta} = \underset{\theta}{arg\ max}\ p(\mathbf{x};\theta)$$

Theorem. If an efficient estimator exists, then it is given by the MLE

Proof.

$$0 = \left.\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right|_{\theta_{ML}} = I(\theta)(g(\mathbf{x}) - \theta)\Big|_{\theta_{ML}}$$

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

Check point 1: Try the CRLB $\quad \dfrac{\partial \ln p(\mathbf{x}; A)}{\partial A} = I(A)(\hat{A} - A)$

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

Check point 2: Try Neyman-Fisher

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{(2\pi A)^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\left(\frac{1}{A}\sum_{n=0}^{N-1}x^2[n] + NA\right)\right]}_{g\left(\sum_{n=0}^{N-1}x^2[n],\, A\right)} \underbrace{\exp(N\bar{x})}_{h(\mathbf{x})}$$

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

Check point 2: Try Neyman-Fisher

If $T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$ *is complete, then an unbiased function of T(**x**) is MVU*

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

Check point 2: Try Neyman-Fisher

If $T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$ *is complete, then an unbiased function of T(**x**) is MVU*

$$\begin{aligned} E\left[\sum_{n=0}^{N-1} x^2[n]\right] &= NE[x^2[n]] \\ &= N\left[\operatorname{var}(x[n]) + E^2(x[n])\right] \\ &= N(A + A^2) \end{aligned}$$

Not clear what function to choose

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

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

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

No more strategies to find the MVU – proceed to MLE

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x^2[n] + \frac{1}{4}}$$

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

Possible to show:

- As N →∞, $E(\hat{A}) \rightarrow A$
- As N →∞, $\mathrm{var}(\hat{A}) \rightarrow \mathrm{CRLB}$

Asymptotically efficient

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

Possible to show:

- As N →∞, $E(\hat{A}) \rightarrow A$
- As N →∞, $\mathrm{var}(\hat{A}) \rightarrow \mathrm{CRLB}$

Asymptotically efficient

We saw this notation before, when we claimed that as N grows, we can estimate a transformation of a variable as the transformation of the estimate

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

Possible to show:

- As N →∞, $E(\hat{A}) \rightarrow A$
- As N →∞, $\mathrm{var}(\hat{A}) \rightarrow \mathrm{CRLB}$

Asymptotically efficient

- By the CLT, we have that $\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$ is Gaussian as N grows.

  Asymptotically, one can show that the MLE is a linear transformation of this Gaussian variable. Thus, the estimator is Gaussian distributed

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}$$

Linearity is not easy to see, but possible…..

Chapter 7 – Maximum Likelihood

Example 7.3. DC level in white Gaussian noise *with variance = DC-level*

Possible to show:

- As N →∞, $E(\hat{A}) \rightarrow A$
- As N →∞, $\text{var}(\hat{A}) \rightarrow \text{CRLB}$

Asymptotically efficient

- By the CLT, we have that $\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]$ is Gaussian as N grows.

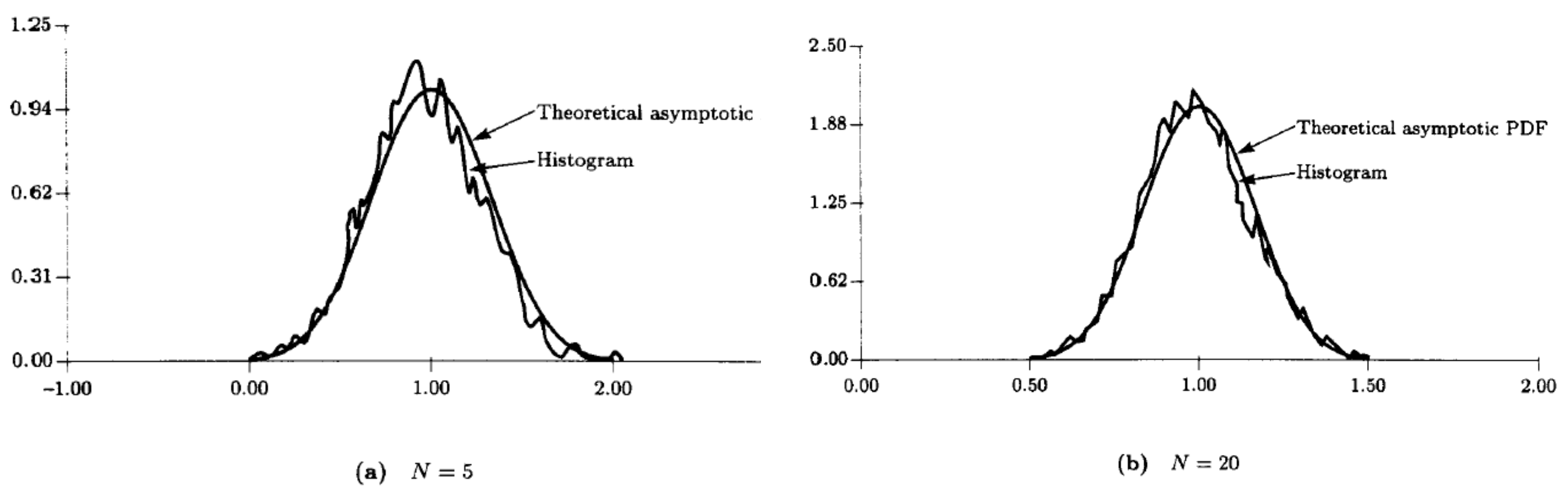
Asymptotically, one can show that the MLE is a linear transformation of this Gaussian variable. Thus, the estimator is Gaussian distributed

But since the variance of the estimator is given by the CRLB (asympt.),

$$\boxed{\hat{\theta} \overset{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta))}$$

Chapter 7 – Maximum Likelihood

”Effect of N”



$$\hat{\theta} \overset{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta))$$

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

1. The first-order and second-order derivatives of the log-likelihood function are well defined.
2. $$E\left[\frac{\partial \ln p(x[n];\theta)}{\partial \theta}\right] = 0.$$

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

Proof of mean (conistency):

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

Let θ_0 be the true value of θ

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

Proof of mean (conistency):

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

Let θ_0 be the true value of θ
By the CLT, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\ln p(x[n];\theta) \rightarrow \int \ln p(x[n];\theta)p(x[n];\theta_0)\,dx[n]$$

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

Proof of mean (conistency):

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

Let θ_0 be the true value of θ
By the CLT, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\ln p(x[n];\theta) \rightarrow \int \ln p(x[n];\theta)p(x[n];\theta_0)\,dx[n]$$

$$\boxed{\int \ln\left[p(x[n];\theta_1)\right]p(x[n];\theta_1)\,dx[n] \geq \int \ln\left[p(x[n];\theta_2)\right]p(x[n];\theta_1)\,dx[n]}$$

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

Proof of mean (conistency):

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

Let θ_0 be the true value of θ

By the CLT, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\ln p(x[n];\theta) \rightarrow \overbrace{\int \ln p(x[n];\theta)p(x[n];\theta_0)\,dx[n]}^{\textbf{Maximized for } \theta=\theta_0}$$

$$\boxed{\int \ln\left[p(x[n];\theta_1)\right] p(x[n];\theta_1)\,dx[n] \geq \int \ln\left[p(x[n];\theta_2)\right] p(x[n];\theta_1)\,dx[n]}$$

Chapter 7 – Maximum Likelihood

Theorem 7.1 (Asymptotic Properties of the MLE) *If the PDF $p(\mathbf{x};\theta)$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter θ is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \tag{7.11}$$

where $I(\theta)$ is the Fisher information evaluated at the true value of the unknown parameter.

Proof of mean (conistency):

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

As N grows, the MLE is $\boldsymbol{\theta_0}$

Let θ_0 be the true value of θ

By the CLT, we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\ln p(x[n];\theta) \rightarrow \overbrace{\int \ln p(x[n];\theta)p(x[n];\theta_0)\,dx[n]}^{\textbf{Maximized for } \boldsymbol{\theta = \theta_0}}$$

$$\boxed{\int \ln\left[p(x[n];\theta_1)\right]p(x[n];\theta_1)\,dx[n] \geq \int \ln\left[p(x[n];\theta_2)\right]p(x[n];\theta_1)\,dx[n]}$$

Chapter 7 – Maximum Likelihood

Neyman-Fisher factorization

$$\boxed{p(\mathbf{x};\theta) = g(T(\mathbf{x}),\theta)h(\mathbf{x})}$$

MLE

$$\boxed{\max_{\theta} \; p(\mathbf{x};\theta)}$$

Chapter 7 – Maximum Likelihood

Neyman-Fisher factorization

$$\boxed{p(\mathbf{x};\theta) = g(T(\mathbf{x}),\theta)h(\mathbf{x})}$$

MLE

$$\boxed{\max_{\theta}\; p(\mathbf{x};\theta)}$$

To find the ML, it is sufficient to optimize the function g(T(x),θ)

Hence, MLE can be made on the basis of the sufficient statistic(s) only

Chapter 7 – Maximum Likelihood

Transformation of parameters

Theorem 7.2 (Invariance Property of the MLE) *The MLE of the parameter* $\alpha = g(\theta)$, *where the PDF* $p(\mathbf{x};\theta)$ *is parameterized by* θ, *is given by*

$$\hat{\alpha} = g(\hat{\theta})$$

where $\hat{\theta}$ *is the MLE of* θ. *The MLE of* $\hat{\theta}$ *is obtained by maximizing* $p(\mathbf{x};\theta)$. *If* g *is not a one-to-one function, then* $\hat{\alpha}$ *maximizes the modified likelihood function* $\bar{p}_T(\mathbf{x};\alpha)$, *defined as*

$$\bar{p}_T(\mathbf{x};\alpha) = \max_{\{\theta:\alpha=g(\theta)\}} p(\mathbf{x};\theta).$$

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists
- **From before:** The (non-linear) transformed estimate of an efficient estimator does not preserve efficiency

$$\hat{\alpha} = g(\hat{\theta})$$

efficient

Not efficient

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists
- **From before:** The (non-linear) transformed estimate of an efficient estimator does not preserve efficiency

$$\hat{\alpha} = g(\hat{\theta})$$

efficient

Not efficient

This argument didn't say that an efficient estimator for α does not exist

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists
- **From before:** The (non-linear) transformed estimate of an efficient estimator does not preserve efficiency

$$\hat{\alpha} = g(\hat{\theta})$$

Not efficient

efficient

- **We can deduce:** A non-linear function of a parameter that can be efficiently estimated ***cannot be efficiently estimated***

$$\hat{\alpha} = g(\hat{\theta})$$

Not efficient

(but would have been if an efficient existed since it is the transformed MLE)

efficient

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists
- **From before:** The (non-linear) transformed estimate of an efficient estimator does not preserve efficiency

$$\underbrace{\hat{\alpha}}_{\text{Not efficient}} = g(\underbrace{\hat{\theta}}_{\text{efficient}})$$

- **We can deduce:** A non-linear function of a parameter that can be efficiently estimated ***<u>cannot be efficiently estimated</u>***
- Due to previous linearization argument, the transformed estimate is asymptotically efficient.

Chapter 7 – Maximum Likelihood

Transformation of parameters

Let us tie things together

- The MLE of a transformed parameter is the transformed MLE of the parameter
- The MLE is efficient if an efficient estimator exists
- **From before:** The (non-linear) transformed estimate of an efficient estimator does not preserve efficiency

$$\hat{\alpha} = g(\hat{\theta})$$

efficient

Not efficient

- **We can deduce:** A non-linear function of a parameter that can be efficiently estimated ***cannot be efficiently estimated***
- Due to previous linearization argument, the transformed estimate is asymptotically efficient.
- This could also have been realized by the observation that

  transformed estimate = MLE = asymptotically efficient

Chapter 7 – Maximum Likelihood

Transformation of parameters

Theorem 7.2 (Invariance Property of the MLE) *The MLE of the parameter* $\alpha = g(\theta)$, *where the PDF* $p(\mathbf{x};\theta)$ *is parameterized by* θ, *is given by*

$$\hat{\alpha} = g(\hat{\theta})$$

where $\hat{\theta}$ *is the MLE of* θ. *The MLE of* $\hat{\theta}$ *is obtained by maximizing* $p(\mathbf{x};\theta)$. *If* g *is not a one-to-one function, then* $\hat{\alpha}$ *maximizes the modified likelihood function* $\bar{p}_T(\mathbf{x};\alpha)$, *defined as*

$$\bar{p}_T(\mathbf{x};\alpha) = \boxed{\max_{\{\theta:\alpha=g(\theta)\}}} p(\mathbf{x};\theta).$$

What is this?

Chapter 7 – Maximum Likelihood

Transformation of parameters

Consider the case $\alpha = A^2$

For some value of α, there are two values of A that produces α, $A = \pm\sqrt{\alpha}$

The likelihood of α, is the largest of the two likelihoods

$$\boxed{\max_{\{\theta:\alpha=g(\theta)\}} p(\mathbf{x};\theta)}$$

Chapter 7 – Maximum Likelihood

Example 7.10 - Power of WGN in dB

We observe N samples of WGN with variance σ^2 whose power in dB is to be estimated. To do so we first find the MLE of σ^2. Then, we use the invariance principle to find the power P in dB, which is defined as

$$P = 10\log_{10}\sigma^2.$$

The PDF is given by

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

Differentiating the log-likelihood function produces

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\sigma^2)}{\partial\sigma^2} &= \frac{\partial}{\partial\sigma^2}\left[-\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \\ &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}x^2[n]\end{aligned}$$

Chapter 7 – Maximum Likelihood

Example 7.10 - Power of WGN in dB

We observe N samples of WGN with variance σ^2 whose power in dB is to be estimated. To do so we first find the MLE of σ^2. Then, we use the invariance principle to find the power P in dB, which is defined as

$$P = 10\log_{10}\sigma^2.$$

The PDF is given by

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

Differentiating the log-likelihood function produces

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\sigma^2)}{\partial\sigma^2} &= \frac{\partial}{\partial\sigma^2}\left[-\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]\\ &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}x^2[n]\end{aligned}$$

MLE (linear)

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

Chapter 7 – Maximum Likelihood

Example 7.10 - Power of WGN in dB

We observe N samples of WGN with variance σ^2 whose power in dB is to be estimated. To do so we first find the MLE of σ^2. Then, we use the invariance principle to find the power P in dB, which is defined as

$$P = 10\log_{10}\sigma^2.$$

The PDF is given by

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

Differentiating the log-likelihood function produces

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\sigma^2)}{\partial\sigma^2} &= \frac{\partial}{\partial\sigma^2}\left[-\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \\ &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}x^2[n]\end{aligned}$$

MLE (linear)

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

MLE (dB) due to invariance

$$\boxed{\begin{aligned}\hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1}x^2[n].\end{aligned}}$$

Chapter 7 – Maximum Likelihood

Efficient ?
Meaningful to compare with CRLB ?

Efficient (Var = $2\sigma^4/N$=CRLB)

MLE (linear)

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

MLE (dB) due to invariance

$$\boxed{\begin{aligned} \hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]. \end{aligned}}$$

Chapter 7 – Maximum Likelihood

Efficient ? No! Asymptotically: Yes!
Meaningful to compare with CRLB ? No! Not even unbiased
Meets CRLB asymptotically

Efficient (Var = $2\sigma^4/N$=CRLB)

MLE (linear)

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

MLE (dB) due to invariance

$$\boxed{\begin{aligned} \hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]. \end{aligned}}$$

Chapter 7 – Maximum Likelihood

Set N=10.
Let us anyway check the CRLB for the dB case: 1.8861

Measured variance (Matlab): ≈ 4.17

Var=2.21xCRLB

Efficient ? **No! Asymptotically: Yes!**
Meaningful to compare with CRLB ? **No! Not even unbiased**
Meets CRLB asymptotically

Efficient (Var = $2\sigma^4/N$=CRLB)

MLE (linear)

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

MLE (dB) due to invariance

$$\begin{aligned} \hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n]. \end{aligned}$$

Chapter 7 – Maximum Likelihood

Set N=50.
Let us anyway check the CRLB for the dB case: 0.3771

Measured variance (Matlab): ≈ 0.77

Var=2.04xCRLB

Efficient ? **No! Asymptotically: Yes!**
Meaningful to compare with CRLB ? **No! Not even unbiased**
Meets CRLB asymptotically

Efficient (Var = 2σ^4/N=CRLB)

MLE (linear)

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

MLE (dB) due to invariance

$$\begin{aligned}\hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n].\end{aligned}$$

Chapter 7 – Maximum Likelihood

Set N=150.
Let us anyway check the CRLB for the dB case: 0.126

Measured variance (Matlab): ≈ 0.25

Var=2xCRLB

Efficient ? No! Asymptotically: Yes!
Meaningful to compare with CRLB ? No! Not even unbiased
Meets CRLB asymptotically

Efficient (Var = 2σ⁴/N=CRLB)

MLE (linear)

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

MLE (dB) due to invariance

$$\begin{aligned}\hat{P} &= 10\log_{10}\hat{\sigma^2} \\ &= 10\log_{10}\frac{1}{N}\sum_{n=0}^{N-1} x^2[n].\end{aligned}$$

Chapter 7 – Maximum Likelihood

Extension to vector parameter: Straightforward

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

Theorem 7.3 (Asymptotic Properties of the MLE (Vector Parameter)) *If the PDF $p(\mathbf{x};\boldsymbol{\theta})$ of the data $\mathbf{x}$ satisfies some "regularity" conditions, then the MLE of the unknown parameter $\boldsymbol{\theta}$ is asymptotically distributed according to*

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\theta}, \mathbf{I}^{-1}(\boldsymbol{\theta})) \tag{7.36}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix evaluated at the true value of the unknown parameter.

Theorem 7.4 (Invariance Property of MLE (Vector Parameter)) *The MLE of the parameter $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$, where $\mathbf{g}$ is an r-dimensional function of the $p \times 1$ parameter $\boldsymbol{\theta}$, and the PDF $p(\mathbf{x};\boldsymbol{\theta})$ is parameterized by $\boldsymbol{\theta}$, is given by*

$$\hat{\boldsymbol{\alpha}} = \mathbf{g}(\hat{\boldsymbol{\theta}})$$

for $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$. If $\mathbf{g}$ is not an invertible function, then $\hat{\boldsymbol{\alpha}}$ maximizes the modified likelihood function $\bar{p}_T(\mathbf{x};\boldsymbol{\alpha})$, defined as

$$\bar{p}_T(\mathbf{x};\boldsymbol{\alpha}) = \max_{\{\boldsymbol{\theta}:\boldsymbol{\alpha}=\mathbf{g}(\boldsymbol{\theta})\}} p(\mathbf{x};\boldsymbol{\theta}).$$

Chapter 7 – Maximum Likelihood

MLE for linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ **in Gaussian noise**

With no derivations: How should we argue in order to establish the MLE ?

Chapter 7 – Maximum Likelihood

MLE for linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ in Gaussian noise

With no derivations: How should we argue in order to establish the MLE ?

- In this case a linear estimator was optimal (Chapter 4)

Chapter 7 – Maximum Likelihood

MLE for linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ in Gaussian noise

With no derivations: How should we argue in order to establish the MLE ?

- In this case a linear estimator was optimal (Chapter 4)
- This linear estimator was also efficient

Chapter 7 – Maximum Likelihood

MLE for linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ in Gaussian noise

With no derivations: How should we argue in order to establish the MLE ?

- In this case a linear estimator was optimal (Chapter 4)
- This linear estimator was also efficient
- An efficient estimator is always the MLE

Chapter 7 – Maximum Likelihood

MLE for linear model $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ in Gaussian noise

With no derivations: How should we argue in order to establish the MLE ?

- In this case a linear estimator was optimal (Chapter 4)
- This linear estimator was also efficient
- An efficient estimator is always the MLE
- **MLE is the linear estimator from Chapter 4** $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$

Chapter 7 – Maximum Likelihood

Asymptotic MLE: some intuition

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}(\boldsymbol{\theta}))} \exp\left[-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{x}\right]$$

With a stationary Gaussian process, the log-likelihood becomes

$$\ln p(\mathbf{x};\boldsymbol{\theta}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\ln P_{xx}(f) + \frac{I(f)}{P_{xx}(f)}\right] df$$

$$I(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi f n)\right|^2$$

Appears strange, but is not if one knows his linear algebra + Szegö's Theorem

Chapter 7 – Maximum Likelihood

Asymptotic MLE: some intuition

- Covariance matrix is Toeplitz
- Elements of C are autocorrelation values of the process (of course)
- Asymptotic Toeplitz matrix admits an Eigenvalue factorization as C=QSQ* (Q=DFT)
- PSD is the Fourier transform of the autocorrelation sequence (e.g. from the Proakis course)

- **not exactly Szegö's Thm, but a consequence thereof:**
  Asymptotically and very loosely speaking, the Eigenvalues of C, i.e. S, converges
  to Fourier transform of the autocorrelation sequence.

Chapter 7 – Maximum Likelihood

Asymptotic MLE: some intuition

- Covariance matrix is Toeplitz
- Elements of C are autocorrelation values of the process (of course)
- Asymptotic Toeplitz matrix admits an Eigenvalue factorization as C=QSQ* (Q=DFT)
- PSD is the Fourier transform of the autocorrelation sequence (e.g. from the Proakis course)

- **not exactly Szegö's Thm, but a consequence thereof:**
  Asymptotically and very loosely speaking, the Eigenvalues of C, i.e. S, converges to Fourier transform of the autocorrelation sequence.

- Consider now log det(C)
  - det(C) is the product of eigenvalues
  - So, log det(C) is the sum of the logarithm of the eigenvalues
  - But: eigenvalues = Fourier transform of autocorrelation = PSD

Chapter 7 – Maximum Likelihood

Asymptotic MLE: some intuition

- Covariance matrix is Toeplitz
- Elements of C are autocorrelation values of the process (of course)
- Asymptotic Toeplitz matrix admits an Eigenvalue factorization as C=QSQ* (Q=DFT)
- PSD is the Fourier transform of the autocorrelation sequence (e.g. from the Proakis course)

- **not exactly Szegö's Thm, but a consequence thereof:**
  Asymptotically and very loosely speaking, the Eigenvalues of C, i.e. S, converges to Fourier transform of the autocorrelation sequence.

- Consider now log det(C)
  - det(C) is the product of eigenvalues
  - So, log det(C) is the sum of the logarithm of the eigenvalues
  - But: eigenvalues = Fourier transform of autocorrelation = PSD

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}(\boldsymbol{\theta}))}$$

$$\ln p(\mathbf{x};\boldsymbol{\theta}) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f)\, df$$

Chapter 7 – Maximum Likelihood

Asymptotic MLE: some intuition

- Covariance matrix is Toeplitz
- Elements of C are autocorrelation values of the process (of course)
- Asymptotic Toeplitz matrix admits an Eigenvalue factorization as C=QSQ* (Q=DFT)
- PSD is the Fourier transform of the autocorrelation sequence (e.g. from the Proakis course)

- **not exactly Szegö's Thm, but a consequence thereof:**
  Asymptotically and very loosely speaking, the Eigenvalues of C, i.e. S, converges to Fourier transform of the autocorrelation sequence.

- Consider now $-x^T C^{-1} x$
  - $x^T C^{-1} x = x^T Q S^{-1} Q^* x$
  - Q^*x is the Fourier transform of x
  - S^{-1} is "one divided with the PSD"
  - $x^T Q$ together with Q^*x is the periodogram of x

$$p(\mathbf{x};\boldsymbol{\theta}) = \exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{x}\right] \qquad \ln p(\mathbf{x};\boldsymbol{\theta}) = -\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{P_{xx}(f)}\,df$$