Estimation Theory

Fredrik Rusek

Chapters 3.5-3.10

Chapter 3 – Cramer-Rao lower bound

Recap

- We deal with unbiased estimators of deterministic parameters
- Performance of an estimator is measured by the variance of the estimate (due to the unbiased condition)
- If an estimator has lower variance for all values of the parameter to estimate, it is the minimum variance estimator (MVU)
- If the estimator is biased, one cannot definte any concept of being an optimal estimator for all values of the parameter to estimate
- The smallest variance possible is determined by the CRLB
- If the CRLB is tight for all values of the parameter, the estimator is efficient
- The CRLB provides us with one method to find the MVU estimator
- Efficient -> MVU. The converse is not true
- To derive the CRLB, one must know the likelihood function

Chapter 3 – Cramer-Rao lower bound

Recap

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

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

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

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

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 $1/I(\theta)$.

Chapter 3 – Cramer-Rao lower bound

Fisher Information

The quantity

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

qualifies as an *information measure*

1. *The more information, the smaller variance*

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

2. *Nonnegative :* $E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]$

3. *Additive for independent observations*

Chapter 3 – Cramer-Rao lower bound

Fisher Information

3. ***Additive for independent observations*** is important:

Chapter 3 – Cramer-Rao lower bound

Fisher Information

3. ***Additive for independent observations*** is important:

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

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

And for IID observations x[n]

$$I(\theta) = N i(\theta)$$

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

Chapter 3 – Cramer-Rao lower bound

Fisher Information

3. ***Additive for independent observations*** is important:

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

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

And for IID observations x[n]

$$I(\theta) = Ni(\theta)$$

where

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

Further verification of "information measure": If x[0]=…x[N-1], then

$$I(\theta) = i(\theta)$$

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

We need to compute the Fisher information $$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]$$

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

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

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

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

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

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

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

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

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

Chapter 3 – Cramer-Rao lower bound

Section 3.5: CRLB for signals in white Gaussian noise

- White Gaussian noise is common. Handy to derive CRLB explicitly for this case

Signal model $$x[n] = s[n;\theta] + w[n] \qquad n = 0, 1, \ldots, N-1$$

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

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

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

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.5: sinusoidal frequency estimation in white noise

Signal model $s[n; f_0] = A\cos(2\pi f_0 n + \phi) \qquad 0 < f_0 < \frac{1}{2}$

Derive CRLB for f_0

Chapter 3 – Cramer-Rao lower bound

Example 3.5: sinusoidal frequency estimation in white noise

Signal model $$s[n; f_0] = A\cos(2\pi f_0 n + \phi) \qquad 0 < f_0 < \frac{1}{2}$$

Derive CRLB for f_0

We have from previous slides that

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

(where $\theta = f_0$)

Chapter 3 – Cramer-Rao lower bound

Example 3.5: sinusoidal frequency estimation in white noise

Signal model $$s[n;f_0] = A\cos(2\pi f_0 n + \phi) \qquad 0 < f_0 < \frac{1}{2}$$

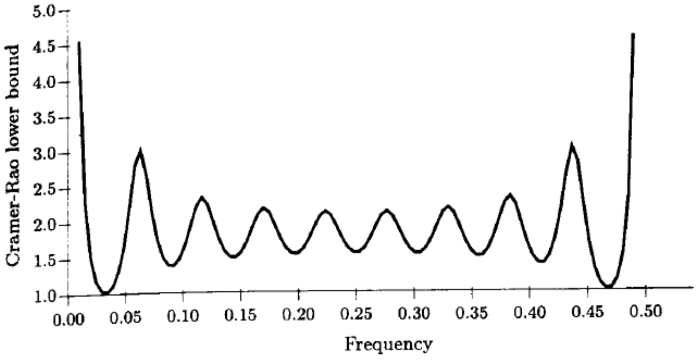
Derive CRLB for f_0

We have from previous slides that

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

(where $\theta = f_0$)

Which yields $$\mathrm{var}(\hat{f}_0) \geq \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1}[2\pi n \sin(2\pi f_0 n + \phi)]^2}$$



Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Now assume that we are interested in estimation of $\alpha = g(\theta)$

We already proved the CRLB for this case

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

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Now assume that we are interested in estimation of $\alpha = g(\theta)$

We already proved the CRLB for this case

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

Question: If we estimate θ first, can we then estimate α as $\hat{\alpha} = g(\hat{\theta})$?

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Works for linear transformations g(x)=ax+b. The estimator for θ is efficient.

Choose the estimator for g(θ) as $\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$

We have:
$$\begin{aligned} E(a\hat{\theta} + b) &= aE(\hat{\theta}) + b = a\theta + b \\ &= g(\theta) \end{aligned}$$

so the estimator is unbiased

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Works for linear transformations g(x)=ax+b. The estimator for θ is efficient.

Choose the estimator for g(θ) as $\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$

We have:
$$\begin{aligned} E(a\hat{\theta}+b) &= aE(\hat{\theta}) + b = a\theta + b \\ &= g(\theta) \end{aligned}$$

so the estimator is unbiased

The CRLB states

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

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Works for linear transformations g(x)=ax+b. The estimator for θ is efficient.

Choose the estimator for g(θ) as $\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$

We have:
$$\begin{aligned} E(a\hat{\theta} + b) &= aE(\hat{\theta}) + b = a\theta + b \\ &= g(\theta) \end{aligned}$$

so the estimator is unbiased

The CRLB states

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

but

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

so we reach the CRLB

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Works for linear transformations g(x)=ax+b. The estimator for θ is efficient.

Choose the estimator for g(θ) as $\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$

We have: $$\begin{aligned} E(a\hat{\theta}+b) &= aE(\hat{\theta}) + b = a\theta + b \\ &= g(\theta) \end{aligned}$$

so the estimator is unbiased

The CRLB states

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

but

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

so we reach the CRLB

Efficiency is preserved for affine transformations!

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Now move on to non-linear transformations

Recall the DC level in white noise example. Now seek the CRLB for the power A^2

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Now move on to non-linear transformations

Recall the DC level in white noise example. Now seek the CRLB for the power A^2

We have $g(x)=x^2$, so $\quad \operatorname{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Recall: the sample mean estimator is efficient for the DC level estimation

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

Question: Is $\hat{A}^2$ efficient for A^2 ?

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Recall: the sample mean estimator is efficient for the DC level estimation

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

Question: Is $\hat{A}^2$ efficient for A^2 ?

NO! This is not even an unbiased estimator

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

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Recall: the sample mean estimator is efficient for the DC level estimation

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

Question: Is $\hat{A}^2$ efficient for A^2 ?

NO! This is not even an unbiased estimator

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

Efficiency is in general destroyed by non-linear transformations!

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Non-linear transformations with large data records

Take a look at the bias again $$E(\bar{x}^2) = E^2(\bar{x}) + \mathrm{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N}$$

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Non-linear transformations with large data records

Take a look at the bias again $E(\bar{x}^2) \;=\; E^2(\bar{x}) + \mathrm{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N}$

The square of the sample mean is *asymptotically unbiased* or *unbiased as N grows large*

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Non-linear transformations with large data records

Take a look at the bias again $$E(\bar{x}^2) = E^2(\bar{x}) + \mathrm{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N}$$

The square of the sample mean is *asymptotically unbiased* or *unbiased as N grows large*

Further $$\mathrm{var}(\bar{x}^2) = \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}$$

While the CRLB states that $$\mathrm{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Non-linear transformations with large data records

Take a look at the bias again $$E(\bar{x}^2) = E^2(\bar{x}) + \mathrm{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N}$$

The square of the sample mean is *asymptotically unbiased* or *unbiased as N grows large*

Further $$\mathrm{var}(\bar{x}^2) = \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}$$

While the CRLB states that $$\mathrm{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

Hence, the estimator $\hat{A}^2$ is *asymptotically efficient*

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ asymptotically efficient ?

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ **asymptotically efficient ?**

When the data record grows, the data statistic* becomes more concentrated around a stable value. We can linearize g(x) around this value, and as the data record grows large, the non-linear region will seldomly occur

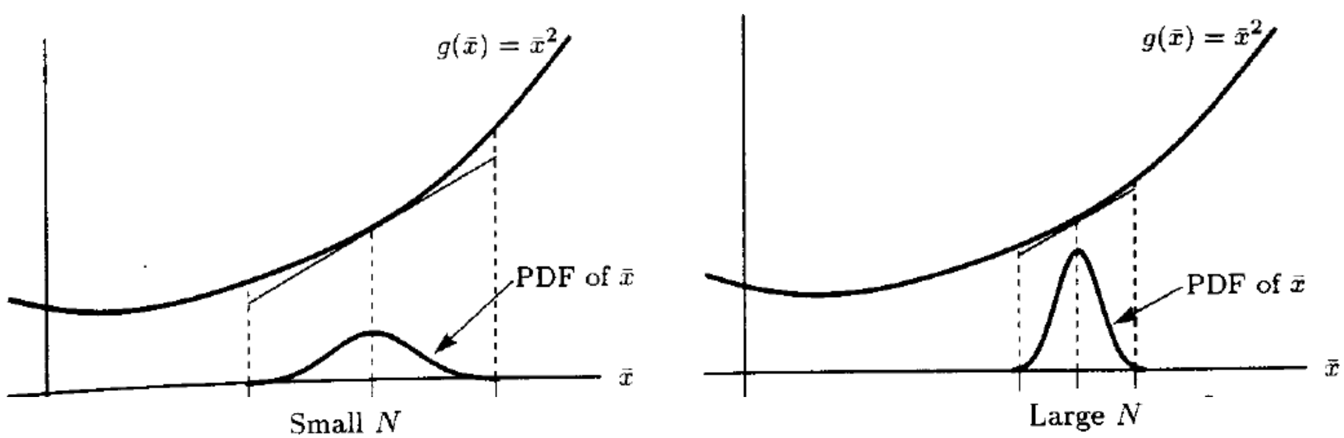
*Definition of statistic: function of data observations used to estimate parameters of interest

Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ asymptotically efficient ?

When the data record grows, the data statistic* becomes more concentrated around a stable value. We can linearize g(x) around this value, and as the data record grows large, the non-linear region will seldomly occur



*Definition of statistic: function of data observations used to estimate parameters of interest

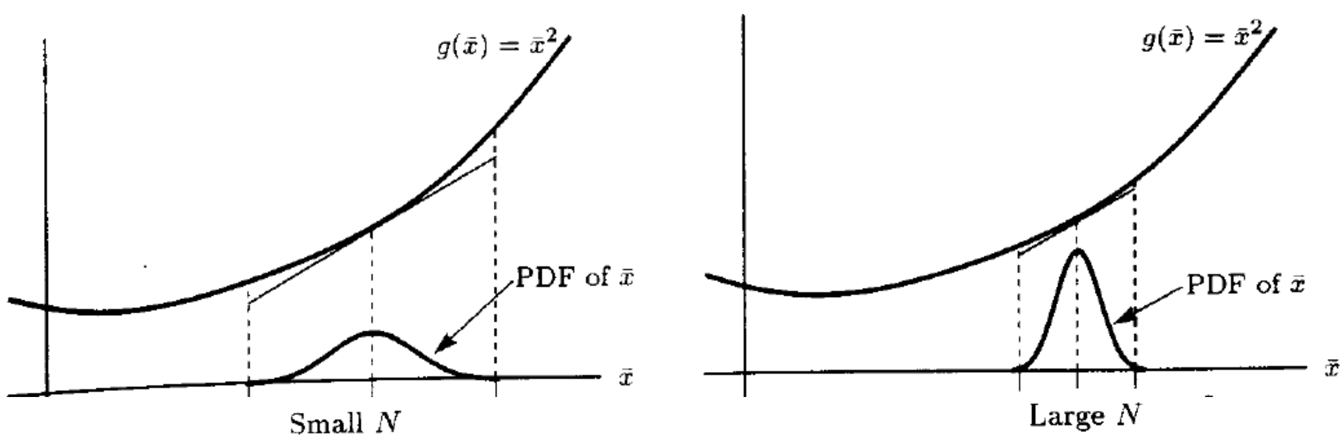
Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ **asymptotically efficient ?**

Linearization at A

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



Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ asymptotically efficient ?

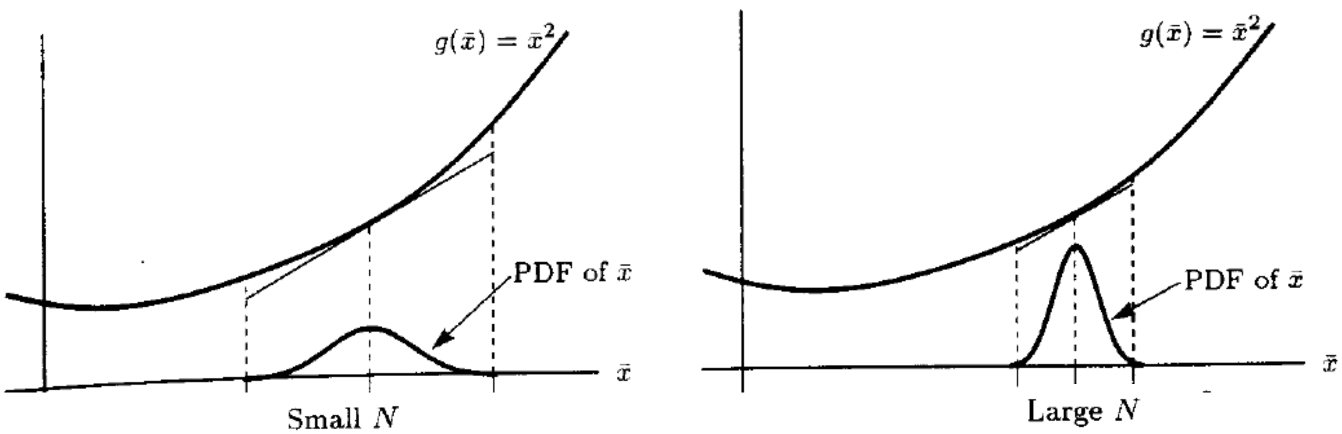
Linearization at A

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

Expectation

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

Unbiased!



Chapter 3 – Cramer-Rao lower bound

Section 3.6: transformation of parameters

Why is the estimator $\hat{\alpha}=g(\hat{\theta})$ **asymptotically efficient ?**

Linearization at A

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

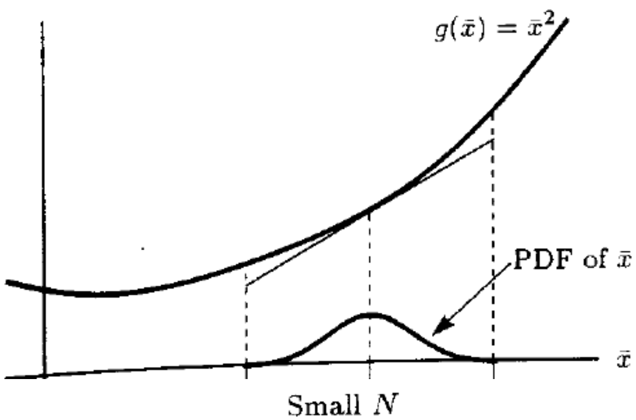
Expectation

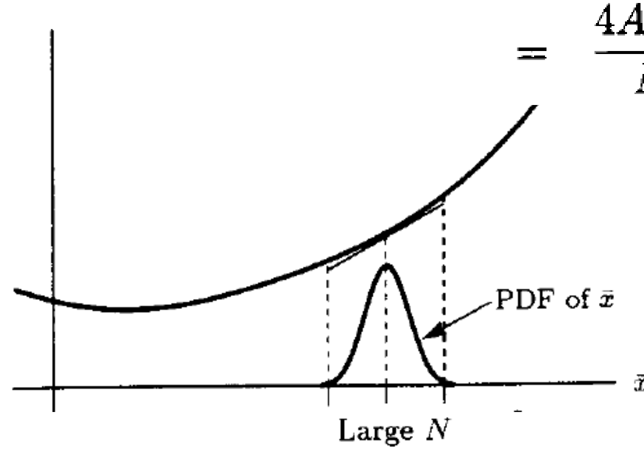
$$E[g(\bar{x})] = g(A) = A^2$$

Unbiased!

Variance

$$\begin{aligned} \mathrm{var}[g(\bar{x})] &= \left[\frac{dg(A)}{dA}\right]^2 \mathrm{var}(\bar{x}) \\ &= \frac{(2A)^2\sigma^2}{N} \\ &= \frac{4A^2\sigma^2}{N} \end{aligned}$$





Efficient!

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB $\boldsymbol{\theta} = [\theta_1\, \theta_2 \ldots \theta_p]^T$

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

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

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 $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB $\boldsymbol{\theta} = [\theta_1\, \theta_2 \ldots \theta_p]^T$

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

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

where the expectation is taken with respect to $p(\mathbf{x}; \boldsymbol{\theta})$.

Same regularity conditions as before

Satisfied (in general) when integration and differentiating can be interchanged

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB $\boldsymbol{\theta} = [\theta_1\, \theta_2 \ldots \theta_p]^T$

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 $p(\mathbf{x}; \boldsymbol{\theta})$.

Covariance of estimator minus Fisher matrix is positive semi-definite

Will soon be a bit simplified

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB $\boldsymbol{\theta} = [\theta_1\, \theta_2 \ldots \theta_p]^T$

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 $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

We can find the MVU estimator also in this case

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB

Let us start with an implication of $\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}$

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB

Let us start with an implication of $\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta) \geq \mathbf{0}$

The diagonal elements of a positive semi-definite matrix are non-negative

Therefore

$$\left[\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta)\right]_{ii} \geq 0$$

Chapter 3 – Cramer-Rao lower bound

Section 3.7: Multi-variable CRLB

Let us start with an implication of $\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta) \geq \mathbf{0}$

The diagonal elements of a positive semi-definite matrix are non-negative

Therefore

$$\left[\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta)\right]_{ii} \geq 0$$

And consequently, $\mathrm{var}(\hat{\theta}_i) = [\mathbf{C}_{\hat{\theta}}]_{ii} \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{ii}$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Start by differentiating the "unbiasedness conditions" $E(\hat{\alpha}_i) = \alpha_i = [\mathbf{g}(\boldsymbol{\theta})]_i \qquad i = 1, 2, \ldots, r.$

$$\frac{\partial}{\partial \theta_i} \int \hat{\alpha}_i \, p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i} \longrightarrow \int \hat{\alpha}_i \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Start by differentiating the "unbiasedness conditions" $E(\hat{\alpha}_i) = \alpha_i = [\mathbf{g}(\boldsymbol{\theta})]_i \qquad i = 1, 2, \ldots, r.$

$$\frac{\partial}{\partial \theta_i} \int \hat{\alpha}_i \, p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i} \longrightarrow \int \hat{\alpha}_i \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

And then add "0" via the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$

to get

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Start by differentiating the "unbiasedness conditions" $E(\hat{\alpha}_i) = \alpha_i = [\mathbf{g}(\boldsymbol{\theta})]_i \qquad i = 1, 2, \ldots, r.$

$$\frac{\partial}{\partial \theta_i} \int \hat{\alpha}_i \, p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathrm{g}(\boldsymbol{\theta})]_i}{\partial \theta_i} \longrightarrow \int \hat{\alpha}_i \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathrm{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

And then add "0" via the regularity condition $E\left[\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$
to get

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathrm{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

For i≠j, we have

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \frac{\partial [\mathrm{g}(\boldsymbol{\theta})]_i}{\partial \theta_j}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Combining into matrix form, we get

$$\int (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial [\mathrm{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_j} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_j}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Combining into matrix form, we get

$$\int \underbrace{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})}_{\text{rx1 vector}} \underbrace{\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T}_{\text{1xp vector}} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

rxp matrix

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_i}$$

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_j} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_j}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Combining into matrix form, we get

$$\int (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Now premultiply with a^T and postmultiply with b, to yield

$$\int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Combining into matrix form, we get

$$\int (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Now premultiply with a^T and postmultiply with b, to yield

$$\int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}$$

Now apply C-S to yield

$$\boxed{\begin{aligned}\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}\right)^2 \leq & \int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T \mathbf{a} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} \\ & \cdot \int \mathbf{b}^T \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x}\end{aligned}}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Combining into matrix form, we get

$$\int (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Now premultiply with a^T and postmultiply with b, to yield

$$\int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} = \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}$$

Now apply C-S to yield

$$\boxed{\begin{aligned} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}\right)^2 &\leq \int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T \mathbf{a} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} \\ &\quad \cdot \int \mathbf{b}^T \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} \\ &= \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b} \end{aligned}}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Regularity condition: same story as for single parameter case

Note that

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

but since

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

this is ok

$$\boxed{\begin{aligned}\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{b}\right)^2 &\leq \int \mathbf{a}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T \mathbf{a} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} \\ &\quad \cdot \int \mathbf{b}^T \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x} \\ &= \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b}\end{aligned}}$$

Chapter 3 – Cramer-Rao lower bound

Proof

The vectors a and b are arbitrary. Now select b as $\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}$

$$\begin{aligned}
\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}\right)^2 &\leq \int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T \mathbf{a} p(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x} \\
&\quad \cdot \int \mathbf{b}^T \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x} \\
&= \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b}
\end{aligned}$$

Chapter 3 – Cramer-Rao lower bound

Proof

The vectors a and b are arbitrary. Now select b as $\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}$

Some manipulations yield

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)$$

$$\boxed{\begin{aligned} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 &\leq \int \mathbf{a}^T (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T \mathbf{a} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} \\ &\quad \cdot \int \mathbf{b}^T \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} \\ &= \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b} \end{aligned}}$$

Chapter 3 – Cramer-Rao lower bound

Proof

The vectors a and b are arbitrary. Now select b as $\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}$

Some manipulations yield

$$\boxed{\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)}$$

For later use, we must now prove that the Fisher information matrix I(**θ**) is positive semi-definite.

Chapter 3 – Cramer-Rao lower bound

Interlude (proof that Fisher information matrix is positive semi-definite)

We have

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

So

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]$$

Chapter 3 – Cramer-Rao lower bound

Interlude (proof that Fisher information matrix is positive semi-definite)

We have

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

So

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]$$

Condition for positive semi-definite: $\mathbf{a}^T\,\mathbf{I}(\boldsymbol{\theta})\,\mathbf{a} \geq 0$

Chapter 3 – Cramer-Rao lower bound

Interlude (proof that Fisher information matrix is positive semi-definite)

We have

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

So

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]$$

Condition for positive semi-definite: $\mathbf{a}^T\,\mathbf{I}(\boldsymbol{\theta})\,\mathbf{a} \geq 0$

But $$\mathbf{a}^T\,\mathbf{I}(\boldsymbol{\theta})\,\mathbf{a} = E\left[\mathbf{a}^T\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{a}\right] = E\left[\boldsymbol{d}^T\boldsymbol{d}\right] \geq 0$$

for $\boldsymbol{d} = \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\,\mathbf{a}$

Chapter 3 – Cramer-Rao lower bound

Proof

Now return to the proof, we had

$$\boxed{\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right)}$$

Chapter 3 – Cramer-Rao lower bound

Proof

$$\boxed{\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right)}$$

We have

$\mathbf{I}(\boldsymbol{\theta})$ positive semi-definite

$\mathbf{I}^{-1}(\boldsymbol{\theta})$ positive semi-definite

$\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T$ positive semi-definite

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right) \geq 0$$

Chapter 3 – Cramer-Rao lower bound

Proof

We therefore get

$$\boxed{\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right) \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a}}$$

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\right) \geq 0$$

Chapter 3 – Cramer-Rao lower bound

Proof

and

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

but **a** is arbitrary, so $\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T$ must be positive semi-definite

Chapter 3 – Cramer-Rao lower bound

Proof

Last part:

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*

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

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* $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Chapter 3 – Cramer-Rao lower bound

Proof

Last part:

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*

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

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* $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Condition for equality

$$\begin{aligned}
\mathbf{a}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \\
&= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}.
\end{aligned}$$

$$\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}$$

Chapter 3 – Cramer-Rao lower bound

Proof

Last part:

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*

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

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* $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Condition for equality

$$\begin{aligned} \mathbf{a}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \\ &= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \end{aligned}$$

However, **a** is arbitrary, so it must hold that $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{c}\mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$

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

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

Condition for equality

$$\begin{aligned}\mathbf{a}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \\ &= c\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}.\end{aligned}$$

However, **a** is arbitrary, so it must hold that $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{c}\mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

One more differential (just as in the single-parameter case) yields (via the chain rule)

$$\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \sum_{k=1}^{p} \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(-\delta_{kj}) + \frac{\partial \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \right)}{\partial \theta_j}(\hat{\theta}_k - \theta_k) \right)$$

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\dfrac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

Take expectation

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = E\left[\ \sum_{k=1}^{p}\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(-\delta_{kj}) + \frac{\partial\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}\right)}{\partial \theta_j}(\hat{\theta}_k - \theta_k)\right)\cdot\right]$$

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

Take expectation

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = E\left[\sum_{k=1}^{p}\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(-\delta_{kj}) + \cancel{\frac{\partial\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}\right)}{\partial \theta_j}(\hat{\theta}_k - \theta_k)}\right)\right]$$

Only k=j remains

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

Take expectation

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = E\left[\sum_{k=1}^{p}\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(-\delta_{kj}) + \cancel{\frac{\partial\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}\right)}{\partial \theta_j}(\hat{\theta}_k - \theta_k)}\right)\right]$$

Only k=j remains

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

Chapter 3 – Cramer-Rao lower bound

Proof

Each element of the vector $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ equals

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{p} \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(\hat{\theta}_k - \theta_k)$$

Take expectation

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = E\left[\sum_{k=1}^{p}\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}(-\delta_{kj}) + \cancel{\frac{\partial\left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})}\right)}{\partial \theta_j}(\hat{\theta}_k - \theta_k)}\right)\right]$$

Only k=j remains

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

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

$$c(\boldsymbol{\theta}) = \underline{1}$$

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

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

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

Derivatives are

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

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

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

Derivatives are

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

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

What are the implications of a diagonal Fisher information matrix?

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

What are the implications of a diagonal Fisher information matrix?

$$\begin{aligned} \operatorname{var}(\hat{A}) &\geq \mathrm{I}^{-1}(\boldsymbol{\theta})_{11} = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A^2}\right] = \frac{N}{\sigma^2} \\ \operatorname{var}(\hat{\sigma^2}) &\geq \mathrm{I}^{-1}(\boldsymbol{\theta})_{22} = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial {\sigma^2}^2}\right] = \frac{2\sigma^4}{N} \end{aligned}$$

But this is precisely what one would have obtained if only one parameter was unknown

Chapter 3 – Cramer-Rao lower bound

Example 3.6

DC level in white Gaussian noise with unknown density: estimate both A and σ^2

Fisher Information matrix is

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

A diagonal Fisher Information matrix means that the parameters to estimate are ”independent” and that the quality of the estimate are not degraded when the other parameters are unknown.

Chapter 3 – Cramer-Rao lower bound

Example 3.7

Line fitting problem: Estimate A and B from the data x[n]

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

Fisher Information matrix

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.7

Line fitting problem: Estimate A and B from the data x[n]

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

Fisher Information matrix

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

Chapter 3 – Cramer-Rao lower bound

Example 3.7

Line fitting problem: Estimate A and B from the data x[n]

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

Fisher Information matrix

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

$$\begin{aligned} \operatorname{var}(\hat{A}) &\geq \frac{2(2N-1)\sigma^2}{N(N+1)} \\ \operatorname{var}(\hat{B}) &\geq \frac{12\sigma^2}{N(N^2-1)}. \end{aligned}$$

In this case, the quality of A is degraded if B is unknown (4 times if N is large)

Chapter 3 – Cramer-Rao lower bound

Example 3.7

How to find the MVU estimator?

Chapter 3 – Cramer-Rao lower bound

Example 3.7

How to find the MVU estimator?

Use the CRLB!

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* $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Chapter 3 – Cramer-Rao lower bound

Example 3.7

Not easy to see this….

Find the MVU estimator.

We have

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

$$= \begin{bmatrix} \dfrac{N}{\sigma^2} & \dfrac{N(N-1)}{2\sigma^2} \\ \dfrac{N(N-1)}{2\sigma^2} & \dfrac{N(N-1)(2N-1)}{6\sigma^2} \end{bmatrix} \begin{bmatrix} \hat{A}-A \\ \hat{B}-B \end{bmatrix}$$

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.7

Not easy to see this….

Find the MVU estimator.

We have

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

$$= \begin{bmatrix} \dfrac{N}{\sigma^2} & \dfrac{N(N-1)}{2\sigma^2} \\ \dfrac{N(N-1)}{2\sigma^2} & \dfrac{N(N-1)(2N-1)}{6\sigma^2} \end{bmatrix} \begin{bmatrix} \hat{A}-A \\ \hat{B}-B \end{bmatrix}$$

No dependency on parameters to estimate

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

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

No dependency on x[n]

Chapter 3 – Cramer-Rao lower bound

Example 3.8

DC level in Gaussian noise with unknown density.

Estimate SNR $\alpha = \frac{A^2}{\sigma^2}$

Chapter 3 – Cramer-Rao lower bound

Example 3.8

DC level in Gaussian noise with unknown density.

Estimate SNR $\alpha = \frac{A^2}{\sigma^2}$

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.8

DC level in Gaussian noise with unknown density.

Estimate SNR $\alpha = \frac{A^2}{\sigma^2}$

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

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.8

DC level in Gaussian noise with unknown density.

Estimate SNR $\alpha = \frac{A^2}{\sigma^2}$

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

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

So

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

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

Chapter 3 – Cramer-Rao lower bound

Section 3.9. Derivation of Fisher matrix for Gaussian signals

A common case is that the received signal is Gaussian

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

Convenient to derive formulas for this case (see appendix for proof)

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

When the covariance is not dependent on **θ**, the second term vanishes, and one can reach

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

Chapter 3 – Cramer-Rao lower bound

Example 3.14

Consider the estimation of A, f_0, and ϕ

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

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

Chapter 3 – Cramer-Rao lower bound

Example 3.14

Consider the estimation of A, f_0, and ϕ

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

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

Interpretation of the structure?

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

Chapter 3 – Cramer-Rao lower bound

Example 3.14

Consider the estimation of A, f_0, and ϕ

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

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

$$\begin{aligned} \operatorname{var}(\hat{A}) &\geq \frac{2\sigma^2}{N} \\ \operatorname{var}(\hat{f}_0) &\geq \frac{12}{(2\pi)^2\eta N(N^2-1)} \\ \operatorname{var}(\hat{\phi}) &\geq \frac{2(2N-1)}{\eta N(N+1)} \end{aligned}$$

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

$$\eta = A^2/(2\sigma^2)$$

Frequency estimation decays as $1/N^3$

Chapter 3 – Cramer-Rao lower bound

Section 3.10: Asymptotic CRLB for Gaussian WSS processes

For Gaussian WSS processes (first and second order statistics are constant) over time
The elements of the Fisher matrix can be found easy

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

Where P_{xx} is the PSD of the process and N (observation length) grows unbounded

This is widely used in e.g. ISI problems

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

This is the linear model, note that in this book, the noise is white Gaussian

Chapter 4 – The linear model

Definition $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

$$\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$$
$$\mathbf{w} = [w[0]\, w[1] \ldots w[N-1]]^T$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

Let us now find the MVU estimator….How to proceed?

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned}
\mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\
\mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T
\end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

Let us now find the MVU estimator

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

Chapter 4 – The linear model

Definition $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

$$\begin{aligned}\mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T\end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

Let us now find the MVU estimator

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

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

Chapter 4 – The linear model

Definition $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

$$\begin{aligned}\mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T\end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

Let us now find the MVU estimator

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

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\ln(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right] \\ &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right]\end{aligned}$$

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\boxed{\begin{aligned} \frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= \mathbf{b} \\ \frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= 2\mathbf{A}\boldsymbol{\theta} \end{aligned}}$$

Let us now find the MVU estimator

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

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\ln(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\right] \\ &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right] \end{aligned}$$

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\begin{aligned} \frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= \mathbf{b} \\ \frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= 2\mathbf{A}\boldsymbol{\theta} \end{aligned}$$

Let us now find the MVU estimator

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

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

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\begin{aligned} \frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} &= \mathbf{b} \\ \frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} &= 2\mathbf{A}\boldsymbol{\theta} \end{aligned}$$

Let us now find the MVU estimator

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

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

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

Chapter 4 – The linear model

Definition

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x[0]\, x[1] \ldots x[N-1]]^T \\ \mathbf{w} &= [w[0]\, w[1] \ldots w[N-1]]^T \end{aligned}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\boxed{\begin{aligned} \frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= \mathbf{b} \\ \frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= 2\mathbf{A}\boldsymbol{\theta} \end{aligned}}$$

Let us now find the MVU estimator

$$\boxed{\begin{aligned} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}) \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{1}{\sigma^2}[\mathbf{H}^T\mathbf{x} - \mathbf{H}^T\mathbf{H}\boldsymbol{\theta}] \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2}[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \boldsymbol{\theta}] \end{aligned}}$$

Conclusion 1:

MVU estimator (efficient)

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

Covariance

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$$

Conclusion 2:

Statistical performance

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2(\mathbf{H}^T\mathbf{H})^{-1})$$

Chapter 4 – The linear model

Example 4.1: curve fitting

Task is to fit data samples with a second order polynomial

$$x(t_n) = \theta_1 + \theta_2 t_n + \theta_3 t_n^2 + w(t_n) \qquad n = 0, 1, \ldots, N-1$$

We can write this as

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$$\begin{aligned} \mathbf{x} &= [x(t_0)\, x(t_1) \ldots x(t_{N-1})]^T \\ \boldsymbol{\theta} &= [\theta_1\, \theta_2\, \theta_3]^T \end{aligned}$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}.$$

and the (MVU) estimator is

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

Chapter 4 – The linear model

Section 4.5: Extended linear model

Now assume that the noise is not white, so

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

Further assume that the data contains a known part **s**, so that we have

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w}$$

We can transfer this back to the linear model by applying the following transformation:

$$\mathbf{x}' = \mathbf{D}(\mathbf{x} - \mathbf{s})$$

where

$$\mathbf{C}^{-1} = \mathbf{D}^{T}\mathbf{D}$$

Chapter 4 – The linear model

Section 4.5: Extended linear model

In general we have

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

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w} \tag{4.30}$$

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

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{s}) \tag{4.31}$$

and the covariance matrix is

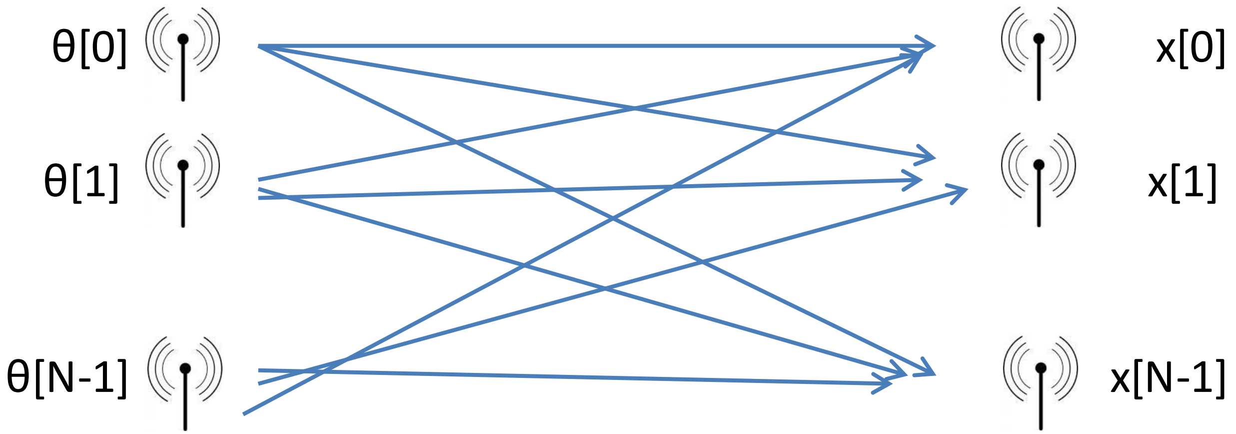
$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}. \tag{4.32}$$

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

Chapter 4 – The linear model

Example: Signal transmitted over multiple antennas and received by multiple antennas

Assume that an unknown signal **θ** is transmitted and received over equally many antennas

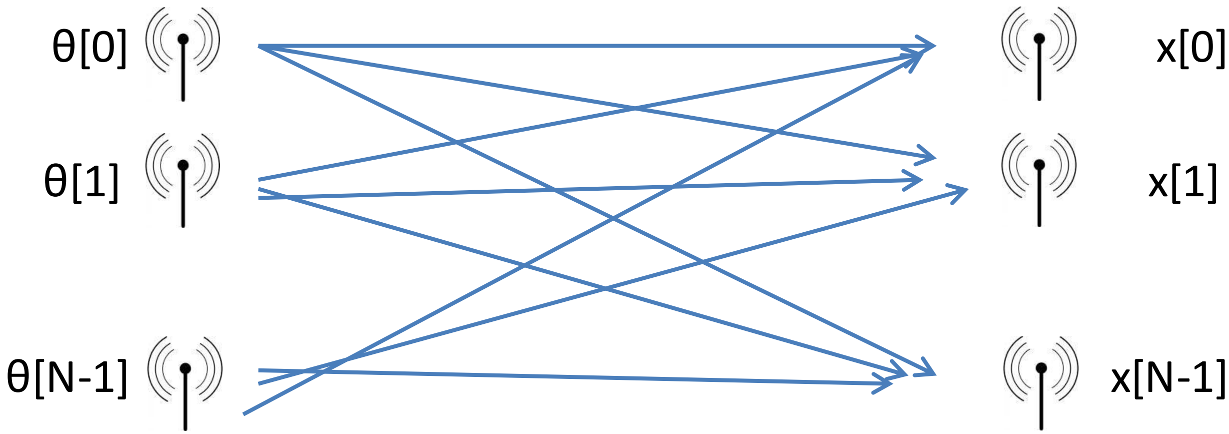


All channels are assumed Different due to the nature of radio propagation

Chapter 4 – The linear model

Example: Signal transmitted over multiple antennas and received by multiple antennas

Assume that an unknown signal $\boldsymbol{\theta}$ is transmitted and received over equally many antennas



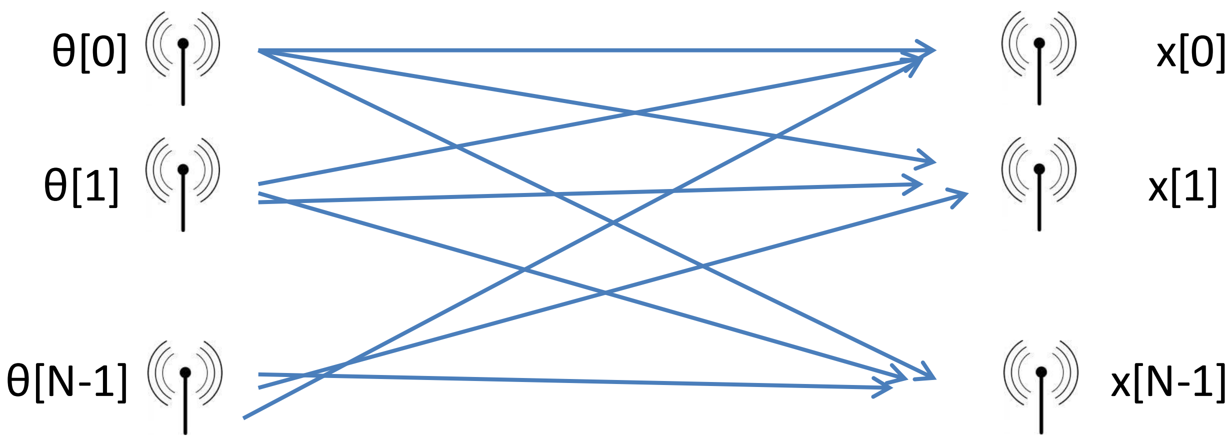
All channels are assumed Different due to the nature of radio propagation

The linear model applies $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

Chapter 4 – The linear model

Example: Signal transmitted over multiple antennas and received by multiple antennas

Assume that an unknown signal $\boldsymbol{\theta}$ is transmitted and received over equally many antennas



All channels are assumed Different due to the nature of radio propagation

The linear model applies $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

So, the best estimator (MVU) is $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ which is the ZF equalizer in MIMO