Estimation Theory
Fredrik Rusek

Chapters 4-5

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

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

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

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

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

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

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

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

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

Chapter 4 – The linear model

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

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

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

Let us now find the MVU estimator

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = \mathbf{b}$$
$$\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$$

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

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

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

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

$$\frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}$$
$$\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$$

Let us now find the MVU estimator

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

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

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

Conclusion 1:

MVU estimator (efficient)

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

Covariance

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

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

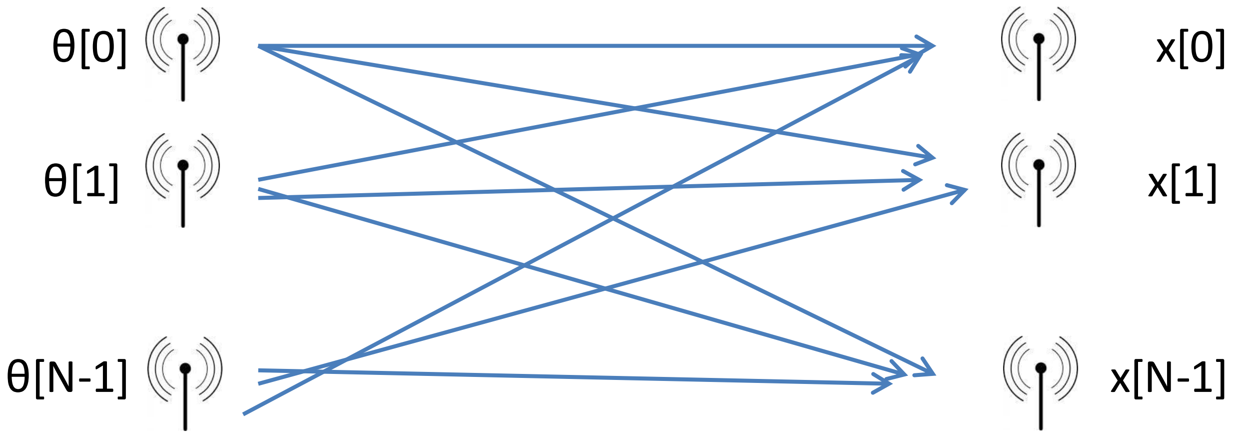
$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}. \tag{4.32}$$

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

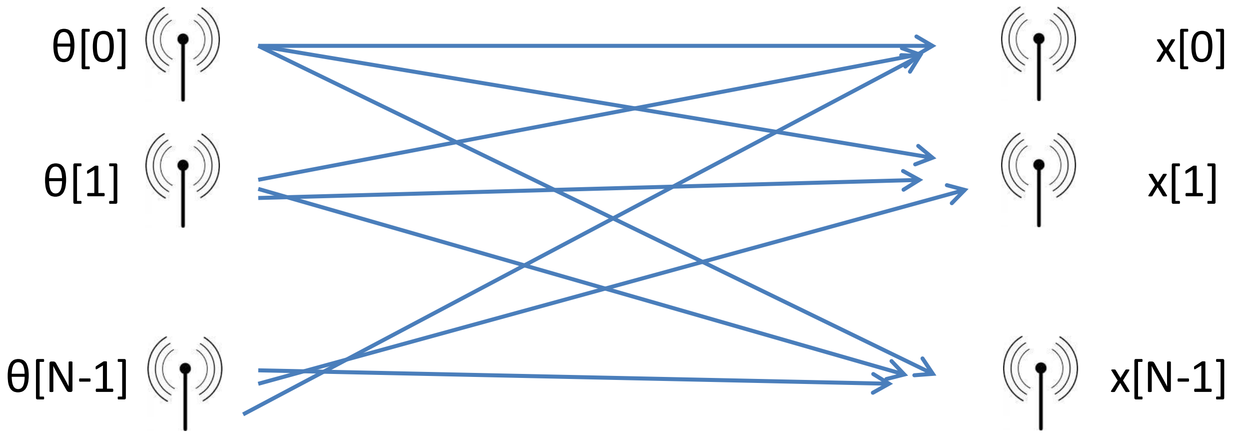


All channels are assumed Different due to the nature of radio propagation

Chapter 4 – The linear model

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

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



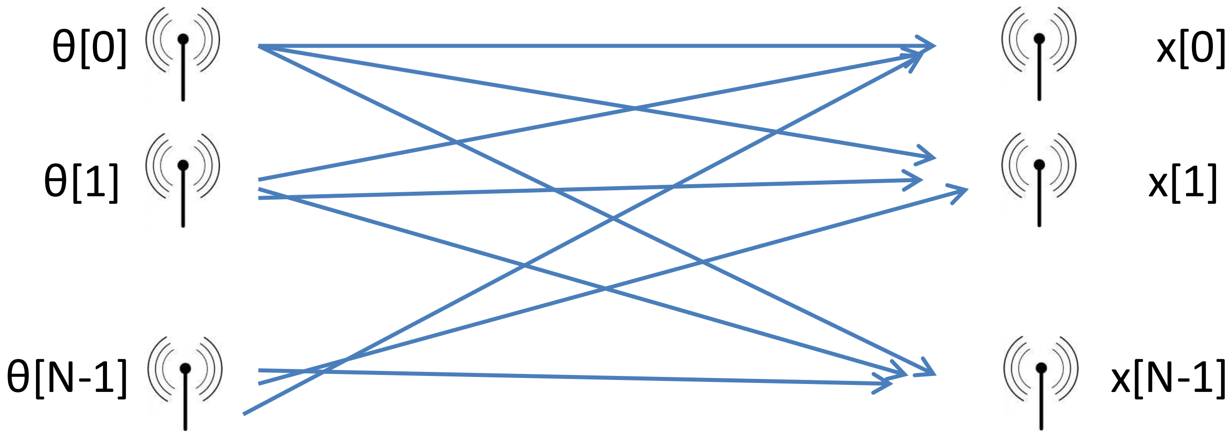
All channels are assumed Different due to the nature of radio propagation

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

Chapter 5– General MVU Estimation

Sufficient statistics

DC level estimation in white noise

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

MVU estimator uses x[0], x[1], …., x[N-1]

But another MVU estimator can be implemented if we are given , e.g.,

$$\begin{aligned} S_1 &= \{x[0], x[1], \ldots, x[N-1]\} \\ S_2 &= \{x[0]+x[1], x[2], x[3], \ldots, x[N-1]\} \\ S_3 &= \left\{\sum_{n=0}^{N-1} x[n]\right\}. \end{aligned}$$

Any of these sets are *sufficient statistics* for optimal estimation of A

Chapter 5– General MVU Estimation

Sufficient statistics

DC level estimation in white noise

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

MVU estimator uses x[0], x[1], …., x[N-1]

But another MVU estimator can be implemented if we are given , e.g.,

$$\begin{aligned} S_1 &= \{x[0], x[1], \ldots, x[N-1]\} \\ S_2 &= \{x[0]+x[1], x[2], x[3], \ldots, x[N-1]\} \\ S_3 &= \left\{\sum_{n=0}^{N-1} x[n]\right\}. \end{aligned}$$

The data set that contains the least number of elements is minimal sufficient

Chapter 5– General MVU Estimation

Sufficient statistics

There exists a better definition of being minimal sufficient than that in the book

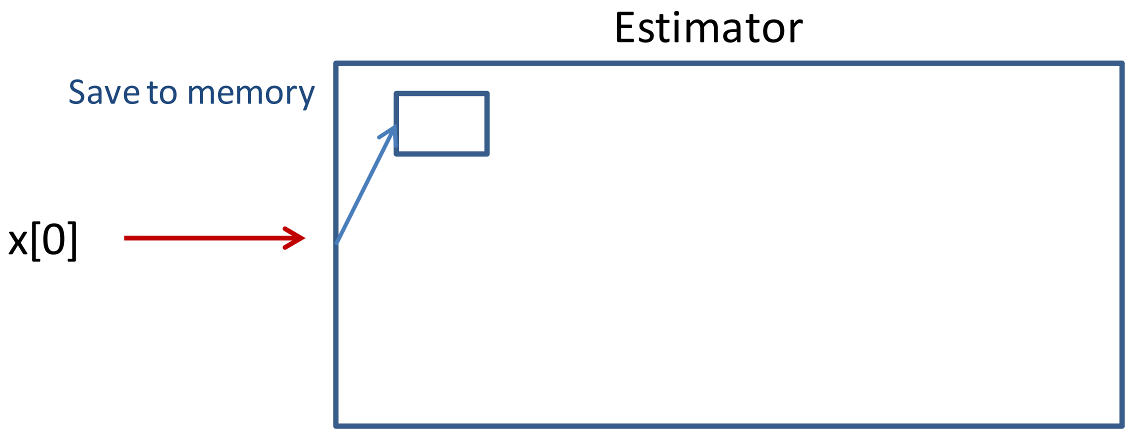
$T_M(\boldsymbol{x})$ is minimal sufficient if and only if

- $T_M(\boldsymbol{x})$ is sufficient
- If $T(\boldsymbol{x})$ is sufficient, then there exist a function q( ), such that $T_M(\boldsymbol{x})=q(T(\boldsymbol{x}))$

Chapter 5– General MVU Estimation

Sufficient statistics

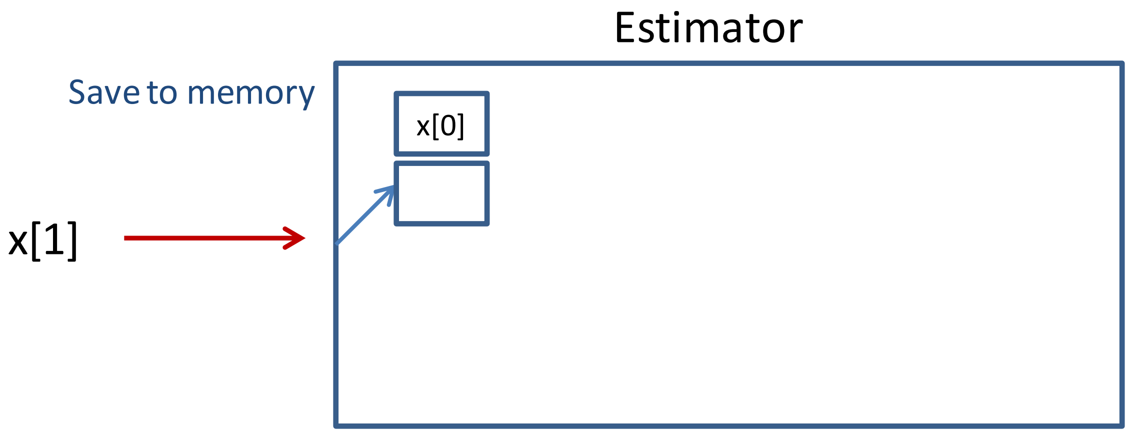
If we do not care about sufficient statistics…



Chapter 5– General MVU Estimation

Sufficient statistics

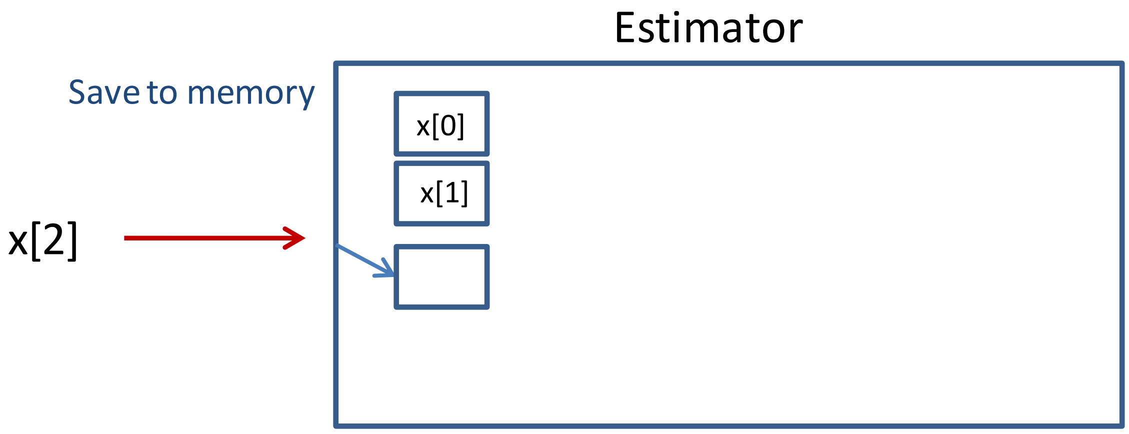
If we do not care about sufficient statistics…



Chapter 5– General MVU Estimation

Sufficient statistics

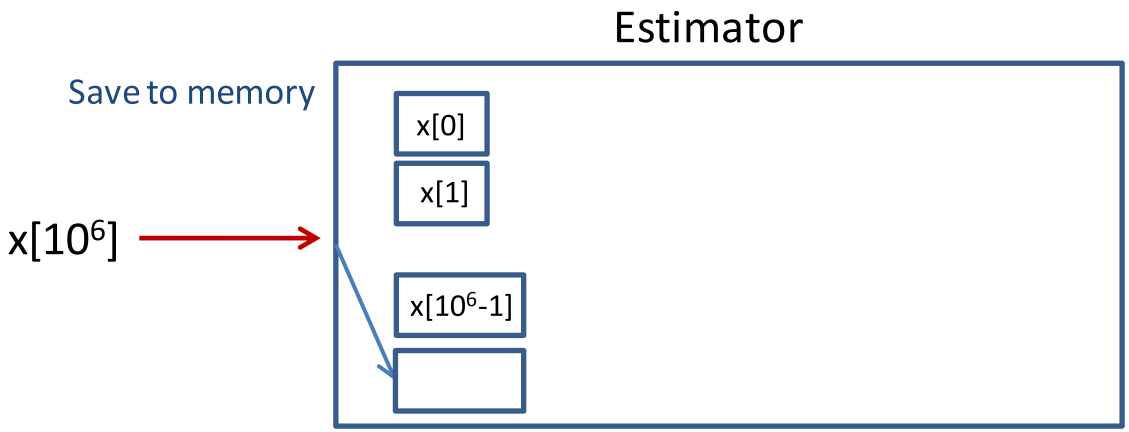
If we do not care about sufficient statistics…



Chapter 5– General MVU Estimation

Sufficient statistics

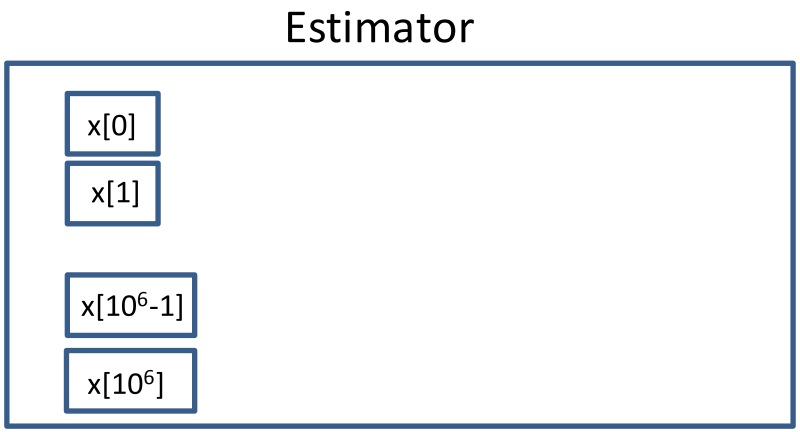
If we do not care about sufficient statistics…



Chapter 5– General MVU Estimation

Sufficient statistics

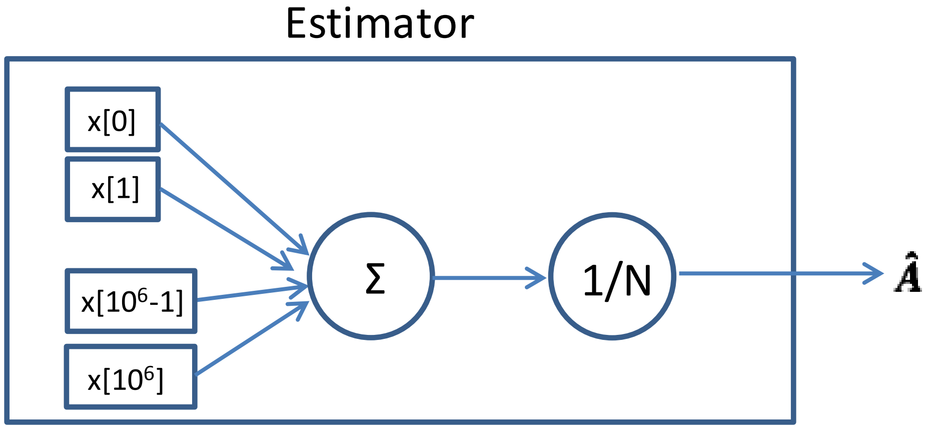
If we do not care about sufficient statistics…



Chapter 5– General MVU Estimation

Sufficient statistics

If we do not care about sufficient statistics…

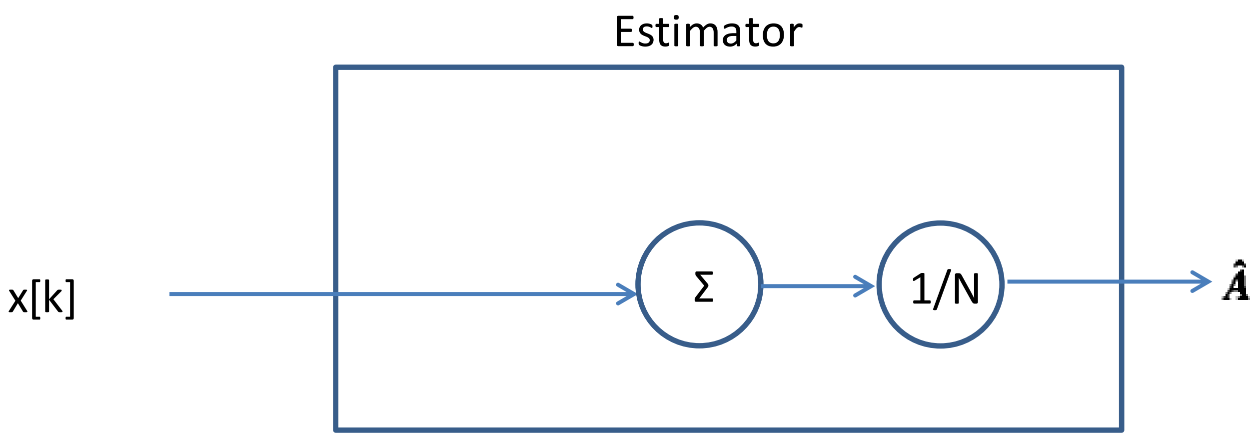


Do estimation

Chapter 5– General MVU Estimation

Sufficient statistics

With sufficient statistics, no memory is needed



Chapter 5– General MVU Estimation

Sufficient statistics

Another unbiased estimator $\check{A} = x[0]$

Estimator can be improved by using also x[1],…

x[0] is not a sufficient statistic for estimation of A

Chapter 5– General MVU Estimation

Sufficient statistics

Another unbiased estimator $\check{A} = x[0]$

Estimator can be improved by using also x[1],…

x[0] is not a sufficient statistic for estimation of A

Question: *How can we (in formula) find if a certain function of the data is sufficent or not?*

Chapter 5– General MVU Estimation

Sufficient statistics

Consider an observation ***x*** = x[0],…,x[N-1]

But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x} | \sum_{n=0}^{N-1} x[n] = T_0; A)$

Chapter 5– General MVU Estimation

Sufficient statistics

Consider an observation ***x*** = x[0],…,x[N-1]

But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x}|\sum_{n=0}^{N-1} x[n] = T_0; A)$

Now forget about sufficient statistics for a while

Chapter 5– General MVU Estimation

Sufficient statistics

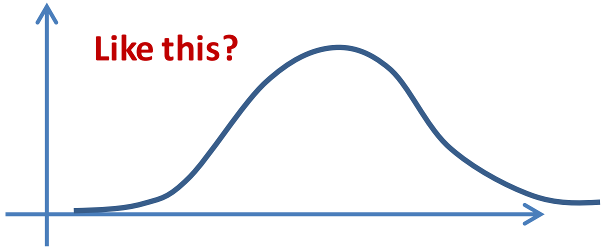
Consider an observation ***x*** = x[0],…,x[N-1]

But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x} | \sum_{n=0}^{N-1} x[n] = T_0; A)$

Now forget about sufficient statistics for a while

If we observe ***x*** but cannot infer the value of an underlying parameter θ, what does the likelihood p(***x***; θ) look like?



Chapter 5– General MVU Estimation

Sufficient statistics

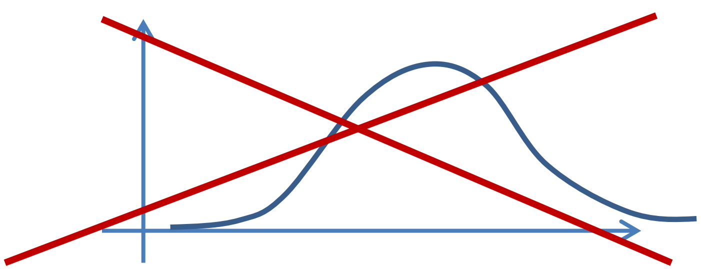
Consider an observation ***x*** = x[0],…,x[N-1]

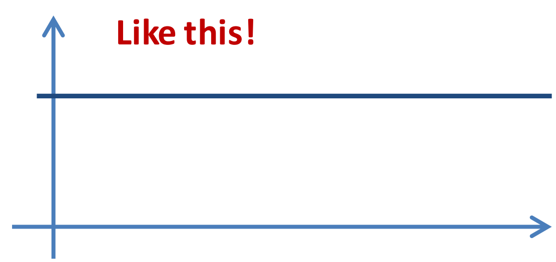
But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x}|\sum_{n=0}^{N-1} x[n] = T_0; A)$

Now forget about sufficient statistics for a while

If we observe ***x*** but cannot infer the value of an underlying parameter θ, what does the likelihood p(***x***; θ) look like?





Chapter 5– General MVU Estimation

Sufficient statistics

Consider an observation ***x*** = x[0],…,x[N-1]

But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x} | \sum_{n=0}^{N-1} x[n] = T_0; A)$

If T(**x**) is sufficient, it is not possible to get information about θ by observing x given T(**x**)

Chapter 5– General MVU Estimation

Sufficient statistics

Consider an observation ***x*** = x[0],…,x[N-1]

But we are also given some statistic T(***x***)

Knowing T(***x***) changes the pdf of ***x*** into $p(\mathbf{x}|\sum_{n=0}^{N-1} x[n] = T_0; A)$

If T(**x**) is sufficient, it is not possible to get information about θ by observing x given T(**x**)

Definition

T(***x***) is sufficient if and only if p(***x***|T(***x***); θ) is independent of θ

Chapter 5– General MVU Estimation

Sufficient statistics

Problem: We guessed that the sample mean was a sufficient statistic. To verify it, we must do

Consider the PDF of (5.1). To prove that $\sum_{n=0}^{N-1} x[n]$ is a sufficient statistic we need to determine $p(\mathbf{x}|T(\mathbf{x}) = T_0; A)$, where $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$. By the definition of the conditional PDF we have

$$p(\mathbf{x}|T(\mathbf{x}) = T_0; A) = \frac{p(\mathbf{x}, T(\mathbf{x}) = T_0; A)}{p(T(\mathbf{x}) = T_0; A)}.$$

But note that $T(\mathbf{x})$ is functionally dependent on $\mathbf{x}$, so that the *joint* PDF $p(\mathbf{x}, T(\mathbf{x}) = T_0; A)$ takes on nonzero values only when $\mathbf{x}$ satisfies $T(\mathbf{x}) = T_0$. The joint PDF is therefore $p(\mathbf{x}; A)\delta(T(\mathbf{x}) - T_0)$, where δ is the Dirac delta function (see also Appendix 5A for a further discussion). Thus, we have that

$$p(\mathbf{x}|T(\mathbf{x}) = T_0; A) = \frac{p(\mathbf{x}; A)\delta(T(\mathbf{x}) - T_0)}{p(T(\mathbf{x}) = T_0; A)}. \quad (5.2)$$

Clearly, $T(\mathbf{x}) \sim \mathcal{N}(NA, N\sigma^2)$, so that

$$\begin{aligned}
& p(\mathbf{x}; A)\delta(T(\mathbf{x}) - T_0) \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right]\delta(T(\mathbf{x}) - T_0) \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n] - 2AT(\mathbf{x}) + NA^2\right)\right]\delta(T(\mathbf{x}) - T_0) \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n] - 2AT_0 + NA^2\right)\right]\delta(T(\mathbf{x}) - T_0).
\end{aligned}$$

From (5.2) we have

$$\begin{aligned}
& p(\mathbf{x}|T(\mathbf{x}) = T_0; A) \\
&= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \exp\left[-\frac{1}{2\sigma^2}(-2AT_0 + NA^2)\right]}{\frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{1}{2N\sigma^2}(T_0 - NA)^2\right]}\delta(T(\mathbf{x}) - T_0) \\
&= \frac{\sqrt{N}}{(2\pi\sigma^2)^{\frac{N-1}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right] \exp\left[\frac{T_0^2}{2N\sigma^2}\right]\delta(T(\mathbf{x}) - T_0)
\end{aligned}$$

which as claimed does not depend on A. Therefore, we can conclude that $\sum_{n=0}^{N-1} x[n]$ is a sufficient statistic for the estimation of A. ◇

Long messy verification

Can we establish that the sample mean is sufficient easier?

Chapter 5– General MVU Estimation

Neyman-Fisher Factorization Theorem

Theorem 5.1 (Neyman-Fisher Factorization) *If we can factor the PDF $p(\mathbf{x};\theta)$ as*

$$p(\mathbf{x};\theta) = g(T(\mathbf{x}),\theta)h(\mathbf{x}) \qquad (5.3)$$

where g is a function depending on $\mathbf{x}$ only through $T(\mathbf{x})$ and h is a function depending only on $\mathbf{x}$, then $T(\mathbf{x})$ is a sufficient statistic for θ. Conversely, if $T(\mathbf{x})$ is a sufficient statistic for θ, then the PDF can be factored as in (5.3).

Proof: Not very illuminating. Only manipulations. Read on your own.

Chapter 5– General MVU Estimation

Neyman-Fisher Factorization Theorem

Theorem 5.3 (Neyman-Fisher Factorization Theorem (Vector Parameter)) *If we can factor the PDF* $p(\mathbf{x};\boldsymbol{\theta})$ *as*

$$p(\mathbf{x};\boldsymbol{\theta}) = g(\mathbf{T}(\mathbf{x}),\boldsymbol{\theta})h(\mathbf{x}) \tag{5.11}$$

where g *is a function depending only on* $\mathbf{x}$ *through* $\mathbf{T}(\mathbf{x})$, *an* $r \times 1$ *statistic, and also on* $\boldsymbol{\theta}$, *and* h *is a function depending only on* $\mathbf{x}$, *then* $\mathbf{T}(\mathbf{x})$ *is a sufficient statistic for* $\boldsymbol{\theta}$. *Conversely, if* $\mathbf{T}(\mathbf{x})$ *is a sufficient statistic for* $\boldsymbol{\theta}$, *then the PDF can be factored as in (5.11).*

In many cases, we cannot find a single sufficient statistic

Factorization still holds.

Factorization also gives us the smallest dimension of the sufficient statistic, i.e., the minimal sufficient statistic.

Chapter 5– General MVU Estimation

Interlude: Exponential family (not in book, but rather good to know)

An important class of likelihoods is the exponential family (scalar case presented)

$$f(x;\theta) = h(x)\ \exp(n(\theta)\ T(x) - A(\theta))$$

This is a wide class of pdfs. For example,

$f(x)^{g(\theta)}$ is included since $f(x)^{g(\theta)} = \exp(g(\theta)\ \log f(x))$

Chapter 5– General MVU Estimation

Interlude: Exponential family (not in book, but rather good to know)

Important results for the exponential family $f(x;\theta) = h(x) \exp(n(\theta) \cdot T(x)-A(\theta))$

1. If the likelihood belongs to the exponential family, then $T(x)$ is a sufficient statistic

2. Multivariate case also exists. The number of sufficient statistics equals the number of unknowns

3. With IID observations, the sufficient statistics are the sums of the individual sufficient statistics

4. **Pitman-Darmóis-Koopman Theorem**: If the number of (IID) observations grows asymptotically large, then the number of sufficient statistics is bounded if and only if the pdf belongs to the exponential family. Domain of pdf must not depend on θ.

Chapter 5– General MVU Estimation

Example 5.9

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

$$\boldsymbol{\theta} = [A\; f_0\; \sigma^2]^T$$

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A\cos 2\pi f_0 n)^2\right]$$

Expand the exponent:

$$\sum_{n=0}^{N-1}(x[n]-A\cos 2\pi f_0 n)^2 = \sum_{n=0}^{N-1}x^2[n] - 2A\sum_{n=0}^{N-1}x[n]\cos 2\pi f_0 n + A^2\sum_{n=0}^{N-1}\cos^2 2\pi f_0 n$$

Chapter 5– General MVU Estimation

Example 5.9

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

$$\boldsymbol{\theta} = [A\ f_0\ \sigma^2]^T$$

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2\right]$$

Expand the exponent:

$$\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2 = \sum_{n=0}^{N-1} x^2[n] - 2A\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1}\cos^2 2\pi f_0 n$$

Problem: with f_0 unknown, we cannot express this as g(**T**(***x***),**θ**) except from letting **T**(***x***)=***x***

Hence,

- All of the data is needed for estimation. We cannot compress it
- The pdf is ***not*** belonging to the exponential family (requires extension to multi-variate case)

Chapter 5– General MVU Estimation

Example 5.9 $x[n] = A\cos 2\pi f_0 n + w[n] \qquad n = 0, 1, \ldots, N-1$

$$\boldsymbol{\theta} = [A\ f_0\ \sigma^2]^T$$

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2\right]$$

Expand the exponent:

$$\sum_{n=0}^{N-1}(x[n] - A\cos 2\pi f_0 n)^2 = \sum_{n=0}^{N-1} x^2[n] - 2A\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1}\cos^2 2\pi f_0 n$$

With f_0 known:

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

- **Our second method to find the MVU estimator** (1st was the CRLB)
- Quite difficult to execute in most cases
- Usually refered to as two theorems
  - The Rao-Blackwell Theorem (the first part)
  - The Lehman-Scheffe Theorem (the second part)
- Statement is complicated and looks confusing
- Proof is easy and clean

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If* $\check{\theta}$ *is an unbiased estimator of* θ *and* $T(\mathbf{x})$ *is a sufficient statistic for* θ, *then* $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ *is*

1. *a valid estimator for* θ *(not dependent on* θ*)*

2. *unbiased*

3. *of lesser or equal variance than that of* $\check{\theta}$, *for all* θ.

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ *is the MVU estimator.*

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to interpret this?

$$\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. a valid estimator for θ (not dependent on θ)

2. unbiased

3. of lesser or equal variance than that of $\check{\theta}$, for all θ.

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to interpret this?

$$\hat{\theta} \;=\; E(\check{\theta}|T(\mathbf{x})) = \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to interpret this?

$$\hat{\theta} = E(\check{\theta}|T(\mathbf{x})) = \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta} = g(T(\mathbf{x}))$$

The new estimator is a function only of the sufficient statistic T(***x***)!

However, from this derivation, it may possibly depend on θ, but we show next that it does not (1.)

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\begin{aligned}\hat{\theta} &= E(\check{\theta}|T(\mathbf{x})) \\ &= \int \check{\theta}(\mathbf{x}) p(\mathbf{x}|T(\mathbf{x});\theta)\, d\mathbf{x}.\end{aligned}$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\begin{aligned} \hat{\theta} &= E(\check{\theta}|T(\mathbf{x})) \\ &= \int \check{\theta}(\mathbf{x}) p(\mathbf{x}|T(\mathbf{x});\theta)\, d\mathbf{x}. \end{aligned}$$

But since T(***x***) is sufficient

$$p(\mathbf{x}|T(\mathbf{x});\theta) = p(\mathbf{x}|T(\mathbf{x}))$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\begin{aligned}\hat{\theta} &= E(\check{\theta}|T(\mathbf{x})) \\ &= \int \check{\theta}(\mathbf{x})p(\mathbf{x}|T(\mathbf{x});\theta)\,d\mathbf{x}.\end{aligned}$$

But since T(***x***) is sufficient

$$p(\mathbf{x}|T(\mathbf{x});\theta) = p(\mathbf{x}|T(\mathbf{x}))$$

$$\hat{\theta} = E(\check{\theta}|T(\mathbf{x})) = \int \check{\theta}(\mathbf{x})p(\mathbf{x}|T(\mathbf{x}))d\mathbf{x}$$

After integrating over ***x***, only the value of T(***x***) remains, therefore not a function of θ

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\hat{\theta} = \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}$$

$$E(\hat{\theta}) = \int \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}\; p(T(\mathbf{x});\theta)\, dT$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\hat{\theta} = \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}$$

$$E(\hat{\theta}) = \int \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}\, p(T(\mathbf{x});\theta)\, dT$$

$$= \int \check{\theta} \int p(\check{\theta}|T(\mathbf{x}))\, p(T(\mathbf{x});\theta)\, dT\, d\check{\theta}$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. a valid estimator for θ (not dependent on θ)

2. unbiased

3. of lesser or equal variance than that of $\check{\theta}$, for all θ.

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

How to prove this?

$$\hat{\theta} = \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}$$

$$E(\hat{\theta}) = \int \int \check{\theta} p(\check{\theta}|T(\mathbf{x}))\, d\check{\theta}\, p(T(\mathbf{x});\theta)\, dT$$

$$= \int \check{\theta} \int p(\check{\theta}|T(\mathbf{x})) p(T(\mathbf{x});\theta)\, dT\, d\check{\theta} = \int \check{\theta} \int p(\check{\theta} T(\mathbf{x});\theta)\, dT\, d\check{\theta} = \int \check{\theta}\, p(\check{\theta};\theta)\, d\check{\theta}$$

$$= E(\check{\theta}) = \theta$$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

Read on your own. Simple

This means that if we can

- Find one unbiased estimator
- Find one sufficient statistic

We can improve the first estimator

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

Read on your own. Simple

This means that if we can

- Find one unbiased estimator
- Find one sufficient statistic

We can improve the first estimator

Important remark:
We cannot improve the improved estimator
By conditioning it on the same sufficient statistic

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Theorem 5.2 (Rao-Blackwell-Lehmann-Scheffe) *If $\check{\theta}$ is an unbiased estimator of θ and $T(\mathbf{x})$ is a sufficient statistic for θ, then $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$ is*

1. *a valid estimator for θ (not dependent on θ)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\theta}$, for all θ.*

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

This is the Lehmann-Scheffe Theorem

- A complete statistic is a fairly complicated thing
- The property of a complete sufficient statistic needed to prove the L-S thm is

"There is only one function of the statistic that is unbiased"

$$\int_{-\infty}^{\infty} v(T)p(T;\theta)\,dT = 0 \qquad \text{for all } \theta$$

Implies v(T)=0, all T

Remark:Exponential family is complete (also multivariate case)

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then θ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then θ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$
- Condition on some sufficient statistic T(***x***) to produce $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then $\hat{\theta}$ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$
- Condition on some sufficient statistic T(***x***) to produce $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$
- Three properties now hold
  - $\hat{\theta}$ is unbiased
  - $\hat{\theta}$ has lower variance than $\check{\theta}$
  - $\hat{\theta}$ *is only a function of* T(***x***)

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then θ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$
- Condition on some sufficient statistic T(***x***) to produce $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$
- Three properties now hold
  - $\hat{\theta}$ is unbiased
  - $\hat{\theta}$ has lower variance than $\check{\theta}$
  - $\hat{\theta}$ *is only a function of* T(***x***)
- However, there is *only one* function of T(***x***) that is unbiased if T(***x***) is complete

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then θ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$
- Condition on some sufficient statistic T(***x***) to produce $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$
- Three properties now hold
  - $\hat{\theta}$ is unbiased
  - $\hat{\theta}$ has lower variance than $\check{\theta}$
  - $\hat{\theta}$ *is only a function of* T(***x***)
- However, there is *only one* function of T(***x***) that is unbiased if T(***x***) is complete
- Hence, no matter from which $\check{\theta}$ we start, we reach the same (unique) $\hat{\theta}$

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Additionally, if the sufficient statistic is complete, then θ is the MVU estimator.

- Start with some unbiased estimator $\check{\theta}$
- Condition on some sufficient statistic T(***x***) to produce $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$
- Three properties now hold
  - $\hat{\theta}$ is unbiased
  - $\hat{\theta}$ has lower variance than $\check{\theta}$
  - $\hat{\theta}$ *is only a function of* T(***x***)
- However, there is *only one* function of T(***x***) that is unbiased if T(***x***) is complete
- Hence, no matter from which $\check{\theta}$ we start, we reach the same (unique) $\hat{\theta}$
- Since $\hat{\theta}$ has lower variance, this must be the MVU estimator

Chapter 5– General MVU Estimation

Rao-Blackwell-Lehman-Scheffe Theorem

Steps to find MVU estimator

1. Find a single sufficient statistic for θ, that is, $T(\mathbf{x})$, by using the Neyman-Fisher factorization theorem.
2. Determine if the sufficient statistic is complete and, if so, proceed; if not, this approach cannot be used.
3. Find a function g of the sufficient statistic that yields an unbiased estimator $\hat{\theta} = g(T(\mathbf{x}))$. The MVU estimator is then $\hat{\theta}$.

As an alternative implementation of step 3 we may

3.′ Evaluate $\hat{\theta} = E(\check{\theta}|T(\mathbf{x}))$, where $\check{\theta}$ is any unbiased estimator.

Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.

Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.

Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?

Chapter 5– General MVU Estimation

Example 5.7

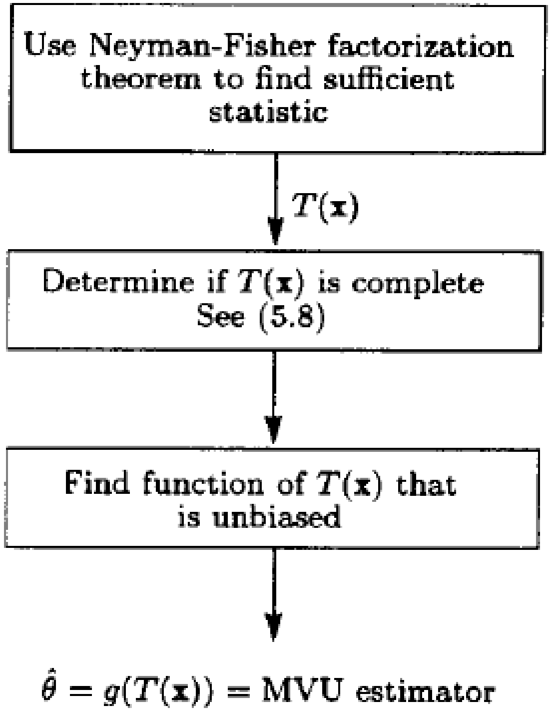
Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.

Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?



Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

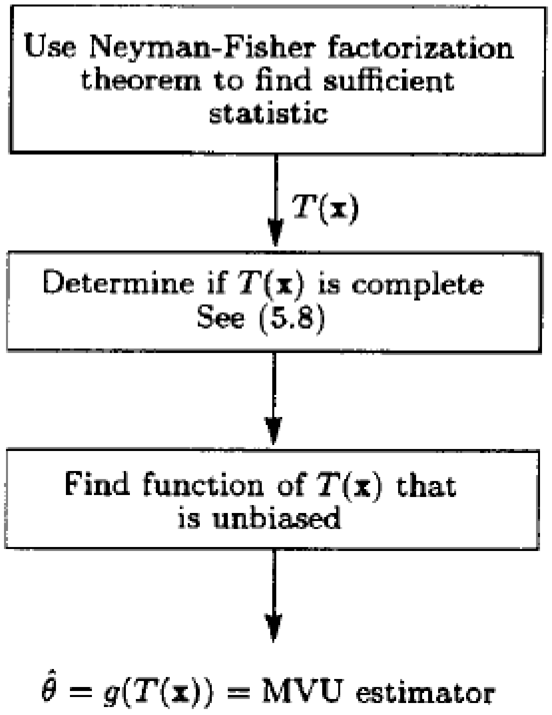
where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.

Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?

Step 1: Find sufficient statistic.

x[0] is clearly sufficient since it is the data



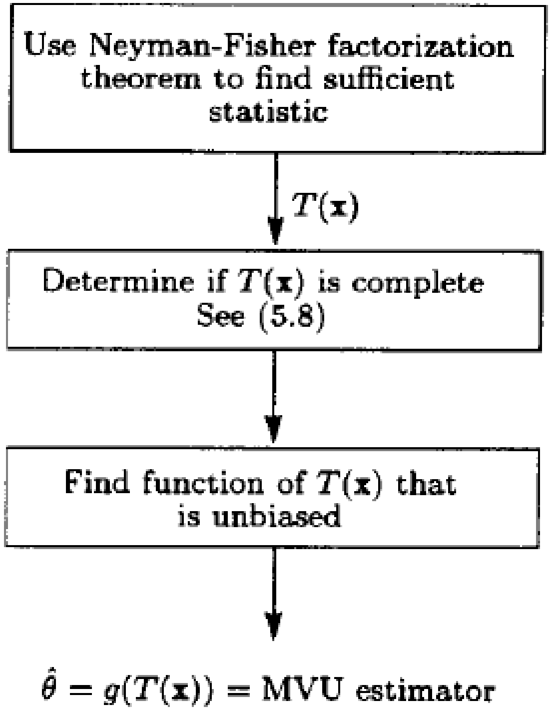
Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.



Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?

Step 2: check if T(*x*)=x[0] is complete

”There should only be one unbiased function of x[0]”

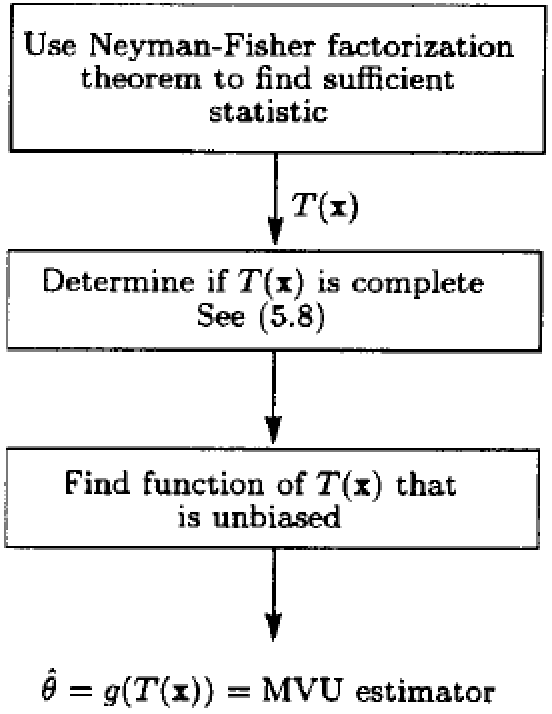
Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.



Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?

Step 2: check if T(*x*)=x[0] is complete

”There should only be one unbiased function of x[0]”

g(x)=x, g(x[0])=x[0], is clearly unbiased in this case

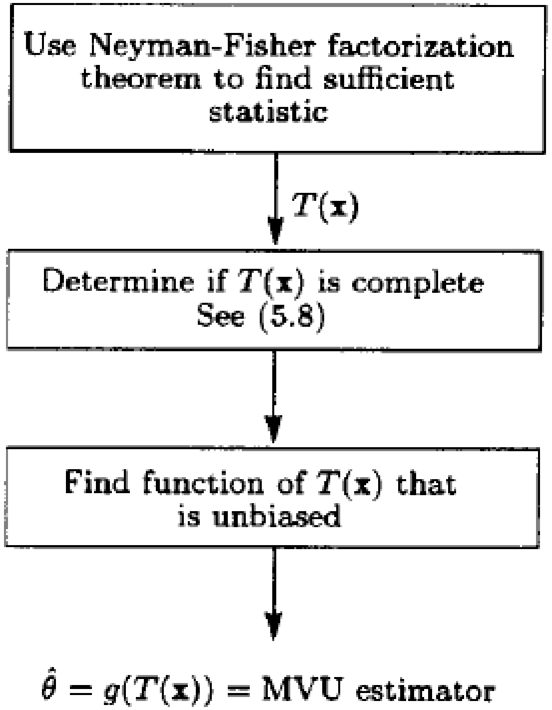
Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.



Clearly, x[0] is unbiased, since E(w[0])=0

Is it the MVU estimator? How to check?

Step 2: check if T(*x*)=x[0] is complete

"There should only be one unbiased function of x[0]"

g(x)=x, g(x[0])=x[0], is clearly unbiased in this case

But so is g(x) = x-sin(2πx), g(x[0]) = x[0]-sin(2πx[0])

Chapter 5– General MVU Estimation

Example 5.7

Consider the estimation of A for the datum

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$.

Clearly, x[0] is unbiased, since E(w[0])=0

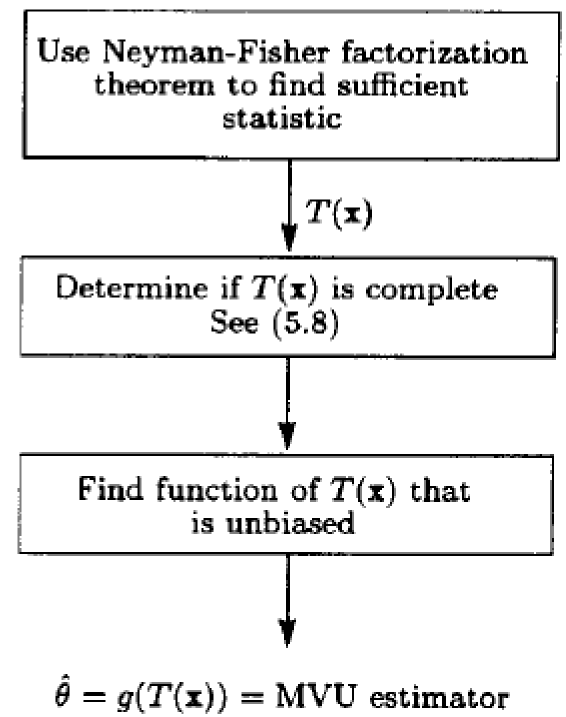
Is it the MVU estimator? How to check?

Step 2: check if T(*x*)=x[0] is complete

”There should only be one unbiased function of x[0]”

g(x)=x, g(x[0])=x[0], is clearly unbiased in this case

But so is g(x) = x-sin(2πx), g(x[0]) = x[0]-sin(2πx[0])



We cannot say that x[0] is the MVU estimator

Method fails

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Procedure:

1. Find complete sufficient statistic

2. Find unbiased function of the complete sufficient statistic

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Interlude

We may guess that the result is $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Procedure:

1. Find complete sufficient statistic. **Apply Neyman-Fisher factorization to the likelihood**

2. Find unbiased function of the complete sufficient statistic

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Likelihood function

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\beta^N} & 0 < x[n] < \beta \quad n = 0, 1, \ldots, N-1 \\ 0 & \text{otherwise.} \end{cases}$$

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Likelihood function

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\beta^N} & 0 < x[n] < \beta \quad n = 0, 1, \ldots, N-1 \\ 0 & \text{otherwise.} \end{cases}$$

Neyman-Fisher

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

$$p(\mathbf{x}; \theta) = \frac{1}{\beta^N} u(\beta - \max x[n])\, u(\min x[n]) .$$

$$u(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Likelihood function

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\beta^N} & 0 < x[n] < \beta \quad n = 0, 1, \ldots, N-1 \\ 0 & \text{otherwise.} \end{cases}$$

$$p(\mathbf{x}; \theta) = \underbrace{\frac{1}{\beta^N} u(\beta - \max x[n])}_{g(T(\mathbf{x}), \theta)} \underbrace{u(\min x[n])}_{h(\mathbf{x})}.$$

max x[n] is sufficient for the estimation of θ

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Is max x[n] complete?

$$\int_{-\infty}^{\infty} v(T)p(T;\theta)\,dT = 0 \qquad \text{for all } \theta$$

Implies v(T)=0, all T

Book skips the proof, but it is simple.....

$$p_T(\xi) = \begin{cases} 0 & \xi < 0 \\ N\left(\frac{\xi}{\beta}\right)^{N-1}\frac{1}{\beta} & 0 < \xi < \beta \\ 0 & \xi > \beta. \end{cases}$$

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Complete statistic $T(\mathbf{x}) = \max x[n]$

We need to find one unbiased function of max x[n]

Chapter 5– General MVU Estimation

Example 5.8 (Basically the german tank problem)

We observe the data

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

where $w[n]$ is IID noise with PDF $\mathcal{U}[0, \beta]$ for $\beta > 0$. We wish to find the MVU estimator for the mean $\theta = \beta/2$.

Guess: Unbiased estimator $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

Complete statistic $T(\mathbf{x}) = \max x[n]$

We need to find one unbiased function of max x[n]

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} \xi p_T(\xi)\, d\xi \\ &= \int_0^{\beta} \xi N \left(\frac{\xi}{\beta}\right)^{N-1} \frac{1}{\beta}\, d\xi \\ &= \frac{N}{N+1}\beta \\ &= \frac{2N}{N+1}\theta. \end{aligned}$$

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

Natural guess is wrong. Sample mean is not a sufficient statistic