Minimum Variance Unbiased Estimation

Unbiased Estimation

Defⁿ: Let $a < \theta < b$. $\hat{\theta} = g(\theta)$ is an unbiased estimator if $E(\hat{\theta}) = \theta$ for all $a < \theta < b$.

i.e $\quad E(\hat{\theta}) = \int g(x)\, p(x;\theta)\, dx = \theta$

for all $a < \theta < b$

Ex: $\quad x[n] = A + w[n] \quad n = 0, \dots, N-1$

$-\infty < A < \infty \qquad w[n] \hookrightarrow$ white Gaussian

Compare $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ & $\bar{A} = \frac{1}{2N}\sum_{n=0}^{N-1} x[n]$

$E(\hat{A}) = A \quad \forall\ -\infty < A < \infty$

unbiased estimator

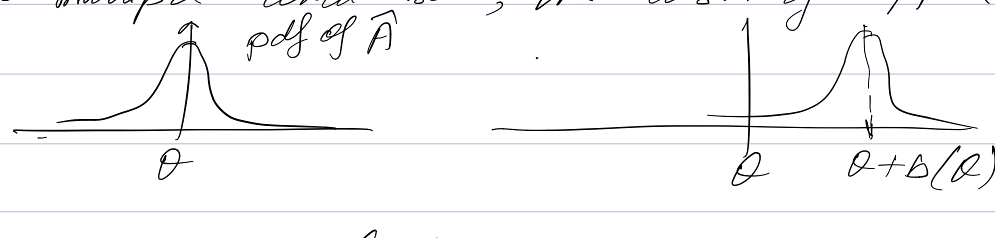
$E(\bar{A}) = A/2$

$E(\bar{A}) = A/2$
$\quad = A$ if $A = 0$
$\quad \neq A$ if $A \neq 0$.

biased estimator

\# Clearly unbiased is better. Why?

(Ans. With multiple data sets, the dist. of $\hat{A}$ & $\bar{A}$)



Minimum Variance Criteria

First we try minimizing mean squared error.

(Not necessarily same as variance)

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

Recall
$$var(\hat{\theta}) = E(\hat{\theta}-E(\hat{\theta}))^2$$

Leads to some problems

$$mse(\hat{\theta}) = E\{(\hat{\theta}-E(\hat{\theta}))+(E(\hat{\theta})-\theta)^2\}$$
$$= E\{(\hat{\theta}-E(\hat{\theta}))^2\} + E\{(E(\hat{\theta})-\theta)^2\}$$
$$+2E(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)$$

$$E[\hat{\theta}E\hat{\theta} - \hat{\theta}\theta - E\hat{\theta}E\hat{\theta} + E\hat{\theta}\theta]$$
$$=(E\hat{\theta})^2 - \theta E\hat{\theta} - E\hat{\theta}^2 + \cancel{\theta E\hat{\theta}}$$
$$=0$$

$$= var\,\hat{\theta} + (E\hat{\theta}-\theta)^2.$$
$$= var\,\hat{\theta} + b^2(\theta) \quad \cdots (1)$$

Let
$b(\theta) = \text{bias}$
$= E\hat{\theta} - \theta$

With suitable assumptions, may not depend on θ

clearly depends on θ (unknown)

Note: Since $mse(\hat{\theta})$ depends on θ, it cannot be optimized easily.

Ex: $x[n] = A + w[n]$ ← same assumption $w[n] \rightarrow WGN \sim N(0,\sigma^2)$

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

Then we saw: $E(\hat{A}) = aA$, $var(\hat{A}) = a^2\frac{\sigma^2}{N}$. Then using (1), $mse(\hat{A}) = \frac{a^2\sigma^2}{N} + (a-1)^2A^2$

Lets forget about unbiased etc & min. mse w.r.t a.

$$\frac{d(mse)}{da} = \frac{2a\sigma^2}{N} + 2(a-1)A^2 = 0$$

$$a_{opt} = \frac{A^2}{A^2 + \sigma^2/N}$$

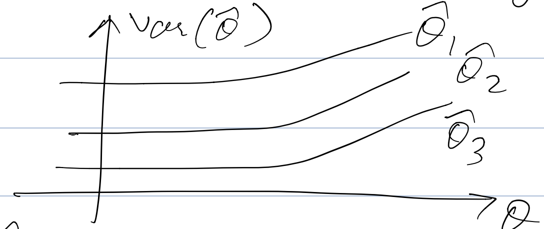
[clearly a_{opt} depends on A → not known so <u>unrealizable</u>]

<u>Strategy</u>: We require bias $= 0$. Then mse $=$ var. We then min. variance, leading to <u>Minimum Variance Unbiased Est</u> (MVUE) Criterion. (<u>Min must hold at each θ</u> → sometimes called uniform MVUE)

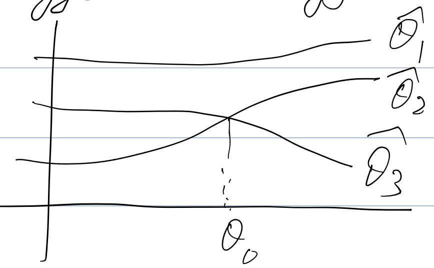
\# Unfortunately MVUE <u>does not</u> always exist.

\# The best estimator might differ at different θ.

Ex:



$\hat{\theta}_3$ is the best among $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ $\forall \theta$



If $\theta < \theta_0$, $\hat{\theta}_2$ is better

if $\theta > \theta_0$, $\hat{\theta}_3$ is better.

Q1) Given an estimator, is it MVUE? - CRLB

Q2) Can you synthesize MVUE? → No.

<u>Cramer - Rao</u> Lower Bound on the variance of <u>any unbiased</u> estimator.

→ we can check if a proposed estimator is MVUE for all θ.

→ we can check how good our estimator is if it is not MVUE.

→ Check if the bound satisfies practical requirements (since we know no better estimator can be found)

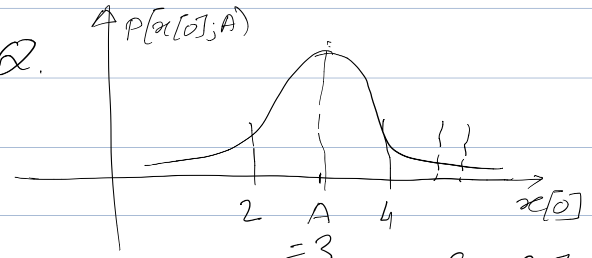
<u>CRLB</u> → intuition → Consider $x[0] = A + w[0]$

$w[0] \sim N(0, \sigma^2)$ (Single observation)

Let $\hat{A} = x[0]$ → unbiased, $var(\hat{A}) = \sigma^2$

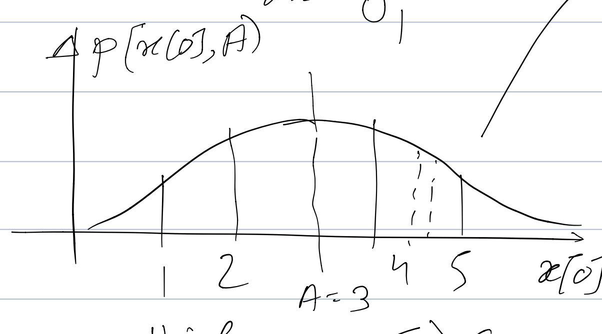
Consider : $p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0]-A)^2\right]$

Q.



Prob of $x[0]$ lying outside $[2, 4]$ negligible

If $A=3$, prob. dist. of $x[0]$ σ_1

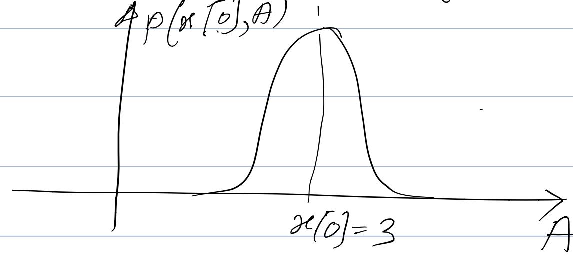


Higher $\sigma_2 > \sigma_1$

→ Prob of $x[0]$ lying outside $[1,5]$ negligible.

→ Accuracy of estimation $\propto \frac{1}{var}$

$\propto$ Sharpness of likelihood fn



Sharpness $\equiv$ -ve of 2nd derivative.
(It is convention & convenient to take log)

Log likelihood fn : $\ln p(x[0];A)$

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

Then, $-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2} = \frac{1}{\sigma^2}$

Here $Var(\hat{A}) = \sigma^2 = \frac{1}{-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}}$

This is the least var since there is no other information about θ in the pdf.

In general, $\frac{\partial^2 \ln p[x[0];A)}{\partial A^2}$ might depend on $x[0]$
Then $E[\cdot]$ is taken.

Cramer – Rao Lower Bound (Scalar Parameter)

Assume that the pdf $p(x;\theta)$ satisfies the "regularity" condition $E\left[\frac{\partial \ln p(x;\theta)}{\partial\theta}\right] = 0 \ \underline{\underline{\forall\theta}}$

where the expectation is w.r.t $p(x;\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy:

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

where derivative is taken at true value of θ & exp. is w.r.t. $p(x;\theta)$. Furthermore, an unbiased estimator may be found that attains the bound for all θ iff

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

for some functions g & I. Then that estimator $\hat{\theta} = g(x)$ is MVU with min. var. $= \frac{1}{I(\theta)}$.

Ex: $x[0] = A + w[0]$ $\quad w[0] \sim N(0,\sigma^2)$

$\hat{A} = x[0]$

From CRLB & our calculations above

$$Var(\hat{A}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right]} = \sigma^2$$

But we know, since $\hat{A} = x[0]$, $Var(\hat{A}) = \sigma^2$

$\Rightarrow$ Since $var(\hat{A})$ attains the CRLB, it is the min. var. unbiased estimator.

Check reg. condition satisfied.

2nd part: Recall,
$$\frac{\partial \ln p(x[0];A)}{\partial A} = \underbrace{\frac{1}{\sigma^2}}_{I(\theta)}(\underbrace{x[0]}_{g(x[0])} - A)$$

Ex: $x[n] = A + w[n]$ $\quad n = 0, 1, \cdots, N-1$

$w[n] \sim N(0, \sigma^2)$

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

To evaluate CRLB,

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

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

$\bar{x}$ ↳ Sample mean

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

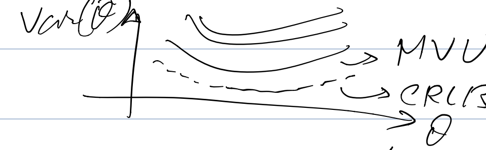
Check: Regularity condition satisfied.

Here CRLB states $\mathrm{Var}(\hat{A}) \geq \frac{\sigma^2}{N}$

On the other hand, we have seen for $\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$, $\mathrm{Var}(\hat{A}) = \frac{\sigma^2}{N} \leftarrow$ MVU.

On the other hand, $\bar{A} = x[0]$ has $\mathrm{Var}(\bar{A}) = \sigma^2 > \mathrm{CRLB}$. So not MVU.

Def$^{\underline{n}}$: An estimator which is unbiased & attains the CRLB is said to be efficient.



Note: An MVU estimator is not necessarily efficient

<u>Proof of CRLB:</u> $E(\hat{\theta}) = \theta$; $\hat{\theta} = g(x)$

i.e. $\int_X \hat{\theta}\, p(x;\theta)\, dx = \theta$

Consider the regularity condition:

$$E\left[\frac{\partial \ln p(x;\theta)}{\partial \theta}\right] = 0 \Leftrightarrow \int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)\, dx = 0$$

$$\int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)\, dx = \int_X \frac{\partial p(x;\theta)}{\partial \theta}\, dx$$

$$= \frac{\partial}{\partial \theta}\int_X p(x;\theta)\, dx = \frac{\partial 1}{\partial \theta} = 0$$

$\hookrightarrow$ If this equality holds then 'regularity' condition is automatically satisfied.

Q. When can this be done? Recall Liebnitz rule.

$\rightarrow$ If the domain of int. X does not involve θ.

Now consider $\int \hat{\theta}\, p(x,\theta)\, dx = \theta$. Diff. wrt. θ,

$$\frac{\partial}{\partial \theta}\int \hat{\theta}\, p(x,\theta)\, dx = 1 \Leftrightarrow \int \hat{\theta}\, \frac{\partial p(x;\theta)}{\partial \theta}\, dx = 1$$

$\uparrow$ Due to regularity condition.

$$\Leftrightarrow \int \hat{\theta}\, \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)\, dx = 1 .$$

Now subtract the regularity condition (after mult. by θ)

$$\int (\hat{\theta} - \theta) \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)\, dx = (1 - 0) \quad --- \; (*)$$

Now apply the Cauchy-Schwarz inequality:

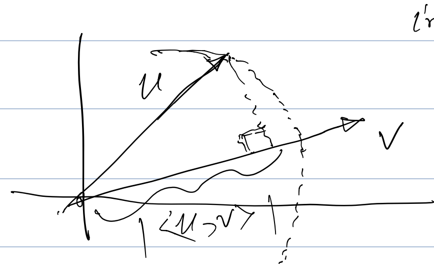
$$\left[\int w(x)\, g(x)\, h(x)\, dx\right]^2 \le \int w(x)\, g^2(x)\, dx \int w(x)\, h^2(x)\, dx$$

Usual Statement: $|\langle u, v\rangle|^2 \le \langle u, u\rangle \cdot \langle v, v\rangle$

inner product

or

$|\langle u, v\rangle| \le \|u\| \cdot \|v\|$



Exercise: Check $\int \phi(x)\, g(x)\, h(x)\, dx$ is an inner product between $\langle g, h\rangle$. For all ϕ?

Equality holds if $g(x) = c\,h(x)$. g, & h are arbitrary but $w(x) \ge 0$.

We identify: $w(x) = p(x;\theta)$, $g(x) = \hat{\theta} - \theta$

$$h(x) = \frac{\partial \ln p(x;\theta)}{\partial \theta}$$

From (*),
$$(1)^2 \le \underbrace{\int (\hat{\theta} - \theta)^2 p(x;\theta)\, dx}_{Var(\hat{\theta})} \int \left[\frac{\partial \ln p(x;\theta)}{\partial \theta}\right]^2 p(x;\theta)\, dx$$

$$\Leftrightarrow Var(\hat{\theta}) \ge \frac{1}{E\left[\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right]}$$

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

Proof: Exercise

Proof of 2nd part:

When equality holds: MVU has been found.

$$g(x) = c\,h(x)$$

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

$c(\theta)$ → can depend on θ but not on x.

For calculating $c(\theta)$, diff. again.

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

Take expectation, $-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right] = \frac{1}{c(\theta)} = I(\theta)$

→ Hence MVU ~~v~~ variance must satisfy

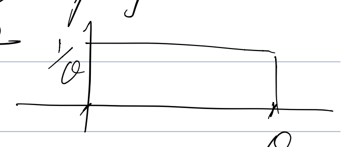
$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = \frac{1}{I(\theta)}(\hat{\theta} - \theta)$$

$\frac{1}{I(\theta)}$ → MVU variance.

Ex. Regularity Condition does not hold

Let $x[0], \dots, x[N-1]$ be iid with pdf

$$p(x[i];\theta) = \frac{1}{\theta}, \quad 0 < x < \theta$$



Joint pdf $= f(x[1], \dots, x[N-1]; \theta) = \frac{1}{\theta^n}$

Regularity condition: $E\left[\frac{\partial \ln p(x;\theta)}{\partial \theta}\right] = -\frac{n}{\theta} \neq 0 \quad \forall \theta$

So, not valid (CRLB should not be used)

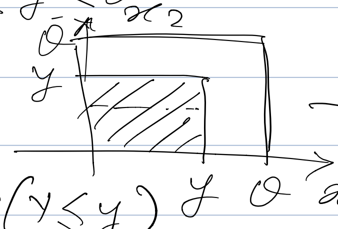
→ Lets try anyway:

$$E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = -\frac{n}{\theta^2} \Rightarrow \text{ Using CRLB, } \mathrm{Var}(\hat{\theta}) \geq \frac{\theta^2}{n}$$

any unbiased estimator

But, consider $Y = \max(X_1, \dots, X_n)$ as estimator

pdf of Y: $p(Y;\theta) = \dfrac{n y^{n-1}}{\theta^n}$; $0 < y \leq \theta$



Exercise

$$\left[P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) = P(Y \leq y) \right]$$

diff. this

Then $E(Y) = \displaystyle\int_0^\theta \frac{n y^n}{\theta^n}\, dy = \frac{n}{n+1}\theta$

$\Rightarrow \left(\frac{n+1}{n}\right) Y$ is an unbiased estimator of θ.

$$\mathrm{Var}\left(\frac{n+1}{n} Y\right) = \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(Y)$$

$$= \left(\frac{n+1}{n}\right)^2 \left[EY^2 - (EY)^2 \right] = \left(\frac{n+1}{n}\right)^2 \left[\left(\frac{n}{n+2}\right)\theta^2 - \left(\frac{n\theta}{n+1}\right)^2 \right]$$

$$= \frac{1}{n(n+2)}\theta^2 < \frac{\theta^2}{n} \rightarrow \text{CRLB (invalid)}$$

uniformly smaller at each θ

Ex: Signal dependence on parameter

$$x[n] = s[n;\theta] + w[n] \quad n = 0, \dots, N-1$$

$\hookrightarrow$ WGN

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

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

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

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

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

. If $s[n;\theta]$ changes fast with θ, it is possible to have a better estimate.

CRLB: Vector Parameter: $\theta = [\theta_1 \cdots \theta_p]^T$

$\hat{\theta} = g(x[0], \cdots, x[N-1])$ is unbiased if

$$\begin{bmatrix}\hat{\theta}_1 \\ \vdots \\ \hat{\theta}_p\end{bmatrix} = \begin{bmatrix} g_1(x[0], \cdots, x[N-1]) \\ \vdots \\ g_p(x[0], \cdots, x[N-1])\end{bmatrix}$$

$E\hat{\theta}_i = \theta_i$; $a_i < \theta_i < b_i$

OR

$E\hat{\theta} = \theta$

$\hat{\theta}$ is MVU if for any unbiased estimator $\bar{\theta} = \begin{bmatrix}\bar{\theta}_1 \\ \vdots \\ \bar{\theta}_p\end{bmatrix}$

$\mathrm{var}(\hat{\theta}_i)$ $(i = 1, \cdots, p)$ satisfies

$$\mathrm{Var}(\hat{\theta}_i) \leq \mathrm{var}(\bar{\theta}_i) \quad i = 1, \cdots, p$$

CRLB: Assume $p(x;\theta)$ satisfies "regularity" condition

$$E\left[\frac{\partial \ln p(x;\theta)}{\partial \theta}\right] = 0 \quad \forall \theta$$

→ Recall $\theta = [\theta_1 \cdots \theta_p]^T$

where expectation is w.r.t. $p(x;\theta)$. Then the covariance matrix of any unbiased estimator $\hat{\theta}$ satisfies:

$$\underbrace{E\left(\left[\hat{\theta}-E\hat{\theta}\right]\left[\hat{\theta}-E\hat{\theta}\right]^T\right)}_{} - I^{-1}(\theta) \geq 0$$

$$\mathrm{Cov}(\hat{\theta}) - I^{-1}(\theta) \geq 0 \rightarrow \underline{\underline{PSD}}$$

The Fisher information matrix $I(\theta)$ is given as

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

1) derivative at true value of θ
2) Exp. w.r.t $p(x,\theta)$

Furthermore, an unbiased estimator may be found that attains the bound i.e. $C(\hat{\theta}) = I^{-1}(\theta)$ iff

$$\underset{p\times 1}{\frac{\partial \ln p(x;\theta)}{\partial\theta}} = \underset{p\times p}{I(\theta)}\left(\underset{p\times 1}{g(x)} - \underset{p\times 1}{\theta}\right)$$

The MVU estimator is $\hat{\theta} = g(x)$ with covariance $\underline{I^{-1}(\theta)}$.

<u>Note:</u> $\mathrm{Var}(\hat{\theta}_i) = [C(\hat{\theta})]_{ii} \geq [I^{-1}(\theta)]_{ii}$ (Why?)

<u>Ex:</u> $x[n] = A + Bn + w[n]$ $\quad n = 0, \dots, N-1$
$w[n] \rightarrow$ WGN

Here $\theta = [A \ \ B]^T$. $I(\theta) = -E\begin{bmatrix} \frac{\partial^2 \ln p(x;\theta)}{\partial A^2} & \frac{\partial^2 \ln p(x;\theta)}{\partial A \partial B} \\ \frac{\partial^2 \ln p(x;\theta)}{\partial B \partial A} & \frac{\partial^2 \ln p(x;\theta)}{\partial B^2} \end{bmatrix}$

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

Calculations — Exercise.

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

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

Hence $\mathrm{Var}(\hat{A}) \geq \underline{\frac{2(2N-1)\sigma^2}{N(N+1)}}$ $\quad \mathrm{Var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}$

Lets check the 2nd part of the Thm. here.

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

Exercise

$$\downarrow$$

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

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

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

Q. Is $\hat{A}, \hat{B}$ unbiased $\rightarrow$ Here & in general ?

Transformation of Parameters : Sometimes we want to estimate $\alpha = h(\theta)$ $\rightarrow$ One can of course solve θ in term of α if possible $\theta = h^{-1}(\alpha)$ & then use CRLB on θ. OR

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

For vector case, $\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{bmatrix} = \begin{bmatrix} g_1(\theta_1, \cdots, \theta_p) \\ \vdots \\ g_r(\theta_1, \cdots, \theta_p) \end{bmatrix}$

$$\alpha = g(\theta)$$

CRLB :
$$\underbrace{C(\hat{\alpha})}_{r \times r} - \underbrace{\frac{\partial g(\theta)}{\partial \theta}}_{r \times p} \underbrace{I^{-1}(\theta)}_{p \times p} \underbrace{\left[\frac{\partial g(\theta)}{\partial \theta}\right]^T}_{p \times r} \geq 0$$

The above model is called a linear model & is useful in many contexts.

$\rightarrow$ MVU can be found for CRLB directly.

In general : $x = H\theta + w$ $\rightarrow$ linear model

$$x = \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix} \quad w = \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix} \quad \theta = \begin{bmatrix} A \\ B \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 \\ \cdot & 1 \\ \vdots & \\ 1 & N-1 \end{bmatrix}$$

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

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

$$= \frac{1}{\sigma^2}\left[H^T x - H^T H\theta\right] = \frac{H^T H}{\sigma^2}\left[(H^T H)^{-1} H^T x - \theta\right]$$

↳ When is this invertible? Exercise

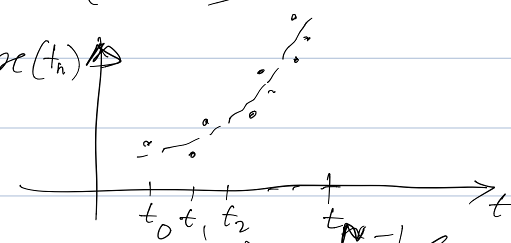
Using the 2nd part of CRLB:

$$\hat{\theta} = (H^T H)^{-1} H^T x \quad \& \quad C(\hat{\theta}) = \sigma^2 (H^T H)^{-1}.$$

Additionally, it is easy to calculate (linear transformation of x) that

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

Ex: Curve Fitting:



Assume: $x(t_n)$, $n = 0, \dots, N-1$ is observed.

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

Surprisingly this is an example of linear model. $\quad x = H\theta + w$

$$x = \begin{bmatrix} x(t_0) \\ \vdots \\ x(t_{N-1}) \end{bmatrix}^T \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \quad H = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}$$

↳ Vandermonde matrix

$\hat{\theta} = (H^T H)^{-1} H^T x \rightarrow$ estimated / fitted curve

$$\hat{x}(t) = \sum_{i=1}^{3} \hat{\theta}_i t^{i-1}$$

Check invertibility of $H^T H$!