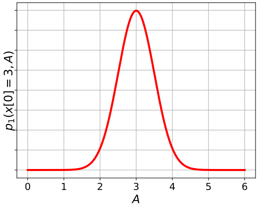
E1 244: Detection and Estimation

Cramer-Rao Lower Bound

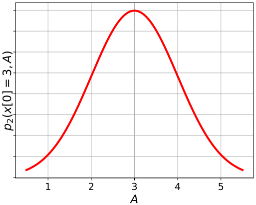
Likelihood function

DC level in white Gaussian noise (WGN)

$$x[n] = A + w[n], \quad n = 0, 1, \ldots, N-1. \quad w[n] \sim \mathcal{N}\left(0, \sigma^2\right)$$



$\sigma^2 = \frac{1}{3}$



$\sigma^2 = \frac{1}{3}$

For a fixed $x[0] = x_0$, the PDF $p(x[0] = x_0; A)$ is a function of the unknown. It is termed as the likelihood function.

For $x[0] = 3$, the values of $A > 4$ are highly unlikely.

The viable values of A are in a much wider interval for large values of σ^2.

Score function

Score function

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

measures the sensitivity of $p(\mathbf{x};\theta)$ to changes in θ.

Curvature

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

measures the sharpness of the log-likelihood function.

Example: $x[0] \sim \mathcal{N}\left(A, \sigma^2\right)$

$$s(\mathbf{x};\theta) = \frac{\partial \ln p(x[0];A)}{\partial A} = \frac{1}{\sigma^2}(x[0]-A) \qquad \text{and} \qquad \mathrm{E}[s(\mathbf{x};\theta)] = 0$$

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

Estimator accuracy and curvature increases as σ^2 decreases.

Theorem: Cramer Rao Lower Bound

Assume that the *regularity condition* holds:

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

The variance of any unbiased estimator $\hat{\theta}$ satisfies

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

An unbiased estimator that attains the bound, i.e., an *efficient estimator* may be found *iff*

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

Then the MVU estimator is $\hat{\theta} = g\left(\mathbf{x}\right)$ has a variance $I^{-1}(\theta)$.

Fisher information has an alternative expression:

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

Regularity condition

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln p\left(\mathbf{x};\theta\right)\right] = \int \frac{\partial}{\partial\theta}\left(\ln p\left(\mathbf{x};\theta\right)\right)p\left(\mathbf{x};\theta\right)d\mathbf{x}$$
$$= \int \frac{1}{p\left(\mathbf{x};\theta\right)}\frac{\partial}{\partial\theta}p\left(\mathbf{x};\theta\right)p\left(\mathbf{x};\theta\right)d\mathbf{x} = \int \frac{\partial}{\partial\theta}p\left(\mathbf{x};\theta\right)d\mathbf{x}$$

If we are allowed to interchange the $\int$ and $\frac{\partial}{\partial\theta}$

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln p\left(\mathbf{x};\theta\right)\right] = \frac{\partial}{\partial\theta}\int p\left(\mathbf{x};\theta\right)d\mathbf{x} = \frac{\partial 1}{\partial\theta} = 0$$

Lebnitz's integration rule: When the limits of the integral is not function of θ, we may swap $\int$ and $\frac{\partial}{\partial\theta}$.

Example:

Suppose $p\left(\mathbf{x};\theta\right) = \mathcal{U}\left(0,\theta\right)$

$$\int_0^\theta \frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)d\mathbf{x} \neq \frac{\partial}{\partial\theta}\int_0^\theta \frac{1}{\theta}d\mathbf{x}.$$

Derivation of CRLB

For an unbiased estimator $\hat{\theta}$

$$\int \left(\hat{\theta} - \theta\right) p\left(\mathbf{x};\theta\right) d\mathbf{x} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial\theta}\int \left(\hat{\theta} - \theta\right) p\left(\mathbf{x};\theta\right) d\mathbf{x} = 0$$

$$\int \left(\hat{\theta} - \theta\right) \frac{\partial}{\partial\theta} p\left(\mathbf{x};\theta\right) d\mathbf{x} = \int p\left(\mathbf{x};\theta\right) d\mathbf{x} = 1$$

Substituting

$$\frac{\partial}{\partial\theta} p\left(\mathbf{x};\theta\right) d\mathbf{x} = \frac{\partial}{\partial\theta} \ln p\left(\mathbf{x};\theta\right) p\left(\mathbf{x};\theta\right)$$

we get

$$\int \left(\hat{\theta} - \theta\right) \frac{\partial}{\partial\theta} \ln p\left(\mathbf{x};\theta\right) p\left(\mathbf{x};\theta\right) d\mathbf{x} = 1$$

$$\int \left(\hat{\theta} - \theta\right) \sqrt{p\left(\mathbf{x};\theta\right)} \frac{\partial}{\partial\theta} \ln p\left(\mathbf{x};\theta\right) \sqrt{p\left(\mathbf{x};\theta\right)} d\mathbf{x} = 1$$

Derivation of CRLB

From the *Cauchy-Schwartz inequality:*

$$\int f^2(x)\,dx \int g^2(x)\,dx \geq \left(\int f(x)\,g(x)\,dx\right)^2$$

we have

$$\int \left(\hat{\theta}-\theta\right)^2 p(\mathbf{x};\theta)\,d\mathbf{x} \int \left(\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right)^2 p(\mathbf{x};\theta)\,d\mathbf{x} \geq 1$$

Since $\int \left(\hat{\theta}-\theta\right)^2 p(\mathbf{x};\theta)\,d\mathbf{x} = \mathrm{var}(\hat{\theta})$

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

Fisher Information

To show that the Fisher Information

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

Let us use the regularity condition

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = 0 \quad \Rightarrow \quad \frac{\partial}{\partial\theta}\int\left(\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right)p(\mathbf{x};\theta)\,d\mathbf{x} = 0$$

$$\int\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)p(\mathbf{x};\theta) + \frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\ \frac{1}{p(\mathbf{x};\theta)}\frac{\partial}{\partial\theta}p(\mathbf{x};\theta)p(\mathbf{x};\theta)\right]d\mathbf{x} = 0$$

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

Properties of Fisher information

- *Non-negativity*

$$I(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right)^2\right] \geq 0$$

- *Additivity* for independent observations

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

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

Efficiency

Suppose the score function admits the factorization

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

we want to show that $\mathrm{E}\left[\hat{\theta}\right] = \theta$ and $\mathrm{var}\left(\hat{\theta}\right) = \frac{1}{I(\theta)}$.

Unbiasedness:

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = \mathrm{E}\left[I(\theta)\left(\hat{\theta}-\theta\right)\right] = I(\theta)\left(\mathrm{E}\left[\hat{\theta}\right]-\theta\right) = 0$$

Efficiency:

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

Taking the negative expected value, $-\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right] = I(\theta)$.

Since

$$\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right)^2\right] = I(\theta)^2\,\mathrm{E}\left[\left(\hat{\theta}-\theta\right)^2\right] \quad\Rightarrow\quad \mathrm{var}\left(\hat{\theta}\right) = \frac{1}{I(\theta)}$$

Example: nonlinear model in additive Gaussian noise

Suppose we are given

$$\mathbf{x} = \mathbf{h}(\theta) + \mathbf{w}, \quad \mathbf{x} \in \mathbb{R}^M, \ \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}), \mathbf{C} : M \times M$$

The log likelihood function

$$\begin{aligned}
\ln p(\mathbf{x};\theta) &= \text{const.} - \frac{1}{2}(\mathbf{x} - \mathbf{h}(\theta))^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{h}(\theta)) \\
\frac{\partial}{\partial\theta} \ln p(\mathbf{x};\theta) &= \frac{\partial}{\partial\theta} \mathbf{h}(\theta)^{\mathrm{T}} \mathbf{C}^{-1} [\mathbf{x} - \mathbf{h}(\theta)] \\
\frac{\partial^2}{\partial\theta^2} \ln p(\mathbf{x};\theta) &= \frac{\partial^2}{\partial\theta^2} \mathbf{h}(\theta)^{\mathrm{T}} \mathbf{C}^{-1} [\mathbf{x} - \mathbf{h}(\theta)] - \frac{\partial}{\partial\theta} \mathbf{h}(\theta)^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{h}(\theta) \\
\Rightarrow \mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} \ln p(\mathbf{x};\theta)\right] &= -\frac{\partial}{\partial\theta} \mathbf{h}(\theta)^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{h}(\theta)
\end{aligned}$$

CRLB depends on θ for a non-linear model. The more $\mathbf{h}(\theta)$ depends on θ, smaller will be the CRLB.

Example: linear model in Gaussian noise

For the observations model $\mathbf{x} = \mathbf{h}\theta + \mathbf{w}$, we have $\operatorname{var}\left(\hat{\theta}\right) \geq \frac{1}{\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{h}}$ and

$$\frac{\partial}{\partial\theta} p\left(\mathbf{x};\theta\right) = \mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\left[\mathbf{x} - \mathbf{h}\theta\right] = \mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x} - \mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{h}\theta$$

$$= \underbrace{\left(\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{h}\right)}_{I(\theta)}[\underbrace{(\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{h})^{-1}\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}}_{\hat{\theta}} - \theta]$$

For the IID Case of $\mathbf{x} = A\mathbf{1} + \mathbf{w}$ with $\mathbf{h} = \mathbf{1}$

$\mathbf{C} = \sigma^2\mathbf{I}$, where $\mathbf{I} : M \times M$ identity matrix

$$\operatorname{var}\left(\hat{\theta}\right) \geq \frac{1}{\mathbf{1}^{\mathrm{T}}(\sigma^2\mathbf{I})^{-1}\mathbf{1}} = \frac{\sigma^2}{N}$$

and

$$\hat{\theta} = \left(\mathbf{h}^{\mathrm{T}}C^{-1}\mathbf{h}\right)^{-1}\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x} = \frac{1}{N}\mathbf{1}^{\mathrm{T}}\mathbf{x} = \tfrac{1}{N}\sum_{n=1}^{N-1} x[n]$$

Example: Poisson distribution

Suppose $\mathbf{x} = [x_0, x_1, \ldots, x_{N-1}]^{\mathrm{T}}$ denote observations of size N from a Poisson distribution i.e., $x_0, x_1, \ldots, x_{N-1}$ are IID observations from a Poisson(θ) distribution with marginal pdf

$$p(x_i; \theta) = \frac{\theta^{x_i}}{x_i!} e^{-\theta}$$

and $\mathrm{E}[x_i] = \theta$. Then,

1. Calculate CRLB for the parameter θ,
2. Find the MVU estimator for θ.

Since the observations are i.i.d., we have

$$p(\mathbf{x}; \theta) = \frac{\theta^K}{\prod_{i=0}^{N-1} x_i!} e^{-N\theta}, \text{ where } K = \sum_{i=0}^{N-1} x_i,$$

and hence we have the score function $\frac{\partial}{\partial \theta} \ln p(\mathbf{x}; \theta) = -N + \frac{1}{\theta} \sum_{i=0}^{N-1} x_i$.

Example: Poisson distribution

Further, $\frac{\partial^2}{\partial \theta^2} \ln p(\mathbf{x}; \theta) = \frac{1}{\theta^2} \sum_{i=0}^{N-1} x_i$ and since $\mathrm{E}[x_i] = \theta$, we have

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

Hence, from the CRLB $\mathrm{var}\left(\hat{\theta}\right) \geq \dfrac{\theta}{N}$. Further, writing the score function as $I(\theta)(g(\mathbf{x}) - \theta)$:

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln p(\mathbf{x}; \theta) &= \frac{1}{\theta} \sum_{i=1}^{N-1} x_i - N, \\ &= \underbrace{\frac{N}{\theta}}_{I(\theta)} [\underbrace{\tfrac{1}{N} \textstyle\sum_{i=0}^{N-1} x_i}_{g(\mathbf{x})} - \theta]. \end{aligned}$$

Transformation of parameters

The CRLB of a transformed parameter $\alpha = g(\theta)$ is

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

Example:
For the DC in WGN model, $x[n] = A + w[n]$, the CRLB for $\alpha = g\left(A\right) = A^2$ (power of the signal) in terms of the CRLB for A:

$$\mathrm{var}\left(\hat{A}^2\right) \geq \frac{(2A)^2}{N/\sigma^2} = 4A^2\sigma^2/N$$

Transformation of parameters

Given that $\hat{A} = \bar{x} = \frac{1}{N}\sum_{i=0}^{N-1}$ is an efficient estimator of A, is $\bar{x}^2$ an efficient estimator of A^2?

Note that $\bar{x} \sim \mathcal{N}\left(A, \sigma^2/N\right)$

- Biased: $\mathrm{E}\left[\bar{x}^2\right] = A^2 + \sigma^2/N \neq A^2$
- Does not attain CRLB: $\mathrm{var}(\bar{x}^2) = \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}$

Efficiency is NOT maintained under non-linear transformations

However, as $N \to \infty$

- Uniased: $\mathrm{E}\left[\bar{x}^2\right] \xrightarrow{N\uparrow} A^2$
- Attains CRLB: $\mathrm{var}(\bar{x}^2) \xrightarrow{N\uparrow} \frac{4A^2\sigma^2}{N}$

Thus $\bar{x}^2$ is an asymptotically efficient estimator of A^2.

Affine transformations

Efficiency of an estimator is maintained under an affine transformation

If $\hat{\theta}$ is an estimator of θ and $\alpha = a\theta + b$, the estimator

$$\hat{\alpha} = a\hat{\theta} + b$$

is efficient

- Unbiased: $\mathrm{E}\left[\hat{\alpha}\right] = a\theta + b = \alpha$
- CRLB: $a^2/I\left(\theta\right) = \mathrm{var}\left(\hat{\alpha}\right) = \mathrm{var}(a\hat{\theta} + b) = a^2 \mathrm{var}\left(\hat{\theta}\right)$

Vector parameters

Assume that the pdf of the observation $\mathbf{x}$ parametrized by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{p-1}]^{\mathrm{T}}$ satisfies the following regularity constraint

$$\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta})\right] = 0, \ \forall\, \boldsymbol{\theta}.$$

Then the covariance matrix of any unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{I}(\boldsymbol{\theta}) \geq \mathbf{0}, \ \forall\, \boldsymbol{\theta},$$

where $\geq \mathbf{0}$ means that the matrix is positive semi-definite. The Fisher matrix $\mathbf{I}(\boldsymbol{\theta})$ is given by

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

Since the diagonal elements of a positive semi-definite matrix are non-negative $\mathrm{var}\left(\hat{\theta}_i\right) \geq [\mathbf{I}(\boldsymbol{\theta})]_{i,i}$.
Further, an unbiased estimator that attains the CRLB can be found if and only if

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

Transformations of vector parameters

Suppose we want to estimate a function

$$\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}), \ \boldsymbol{\alpha} \in \mathbb{R}^r$$

The covariance matrix satisfies the following condition

$$\underbrace{\mathbf{C}_{\hat{\alpha}}}_{\text{rxr}} - \underbrace{\left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]}_{\text{rxp}} \underbrace{\mathbf{I}^{-1}(\boldsymbol{\theta})}_{\text{pxp}} \underbrace{\left[\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^{\mathrm{T}}}_{\text{pxr}} \geq \mathbf{0}.$$

Affine transformation: Suppose $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} + \mathbf{b}$ and the estimator $\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b}$, then

$$\mathrm{E}[\hat{\boldsymbol{\alpha}}] = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} = \boldsymbol{\alpha}$$

$$\mathbf{C}_{\hat{\boldsymbol{\alpha}}} = \mathbf{A}\mathbf{C}_{\hat{\boldsymbol{\theta}}}\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{I}^{-1}(\boldsymbol{\theta})\,\mathbf{A}^{\mathrm{T}} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{I}^{-1}(\boldsymbol{\theta})\,\frac{\partial \mathbf{g}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Example

Estimate parameters A, σ^2 given observations

$$x[n] = A + w[n],\ w[n] \sim \mathcal{N}\left(0, \sigma^2\right),\ n = 1, 2, \dots, N-1,$$

then $\mathbf{x} \sim \mathcal{N}\left(A\mathbf{1}, \sigma^2\mathbf{I}\right)$, and $\boldsymbol{\theta} = \left[A, \sigma^2\right]^{\mathrm{T}}$, $p = 2$.

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

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

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

Hence we have that $\mathrm{var}\left(\hat{A}\right) \geq \sigma^2/N$ and $\mathrm{var}\left(\hat{\sigma^2}\right) \geq 2\sigma^2/N$. Knowing A does not influence the estimator for σ^2

Example

Now suppose we want to estimate

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

from the same observations. Then we have $\boldsymbol{\theta} = \left[A, \sigma^2\right]^{\mathrm{T}}$ and $\alpha = g\left(\boldsymbol{\theta}\right) = \frac{\theta_1^2}{\theta_2}$. Compute $\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ for this model as

$$\frac{\partial}{\partial \boldsymbol{\theta}} g\left(\boldsymbol{\theta}\right) = \left[\frac{\partial g\left(\boldsymbol{\theta}\right)}{\partial A}, \frac{\partial g\left(\boldsymbol{\theta}\right)}{\partial \sigma^2}\right]^{\mathrm{T}} = \left[\frac{2A}{\sigma^2}, \frac{A}{\sigma^4}\right]^{\mathrm{T}}$$

Use this to get the CRLB for covariance of the estimate as

$$\left[\frac{\partial}{\partial \boldsymbol{\theta}} g\left(\boldsymbol{\theta}\right)\right] \mathbf{I}^{-1}\left(\boldsymbol{\theta}\right) \left[\frac{\partial}{\partial \boldsymbol{\theta}} g^{\mathrm{T}}\left(\boldsymbol{\theta}\right)\right] = \frac{4A + 2\sigma^2}{N}.$$

Linear model with vector parameters

Suppose we have the observations

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},\ \theta : \text{px1},\ \mathbf{H} : \text{Nxp and } \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right).$$

Then we have

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

Using $\frac{\partial \mathbf{b}^{\mathrm{T}}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}$ and $\frac{\partial \boldsymbol{\theta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$, we have

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

Further

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

so that the Fisher matrix

$$\mathbf{I}\left(\boldsymbol{\theta}\right) = \frac{\mathbf{H}^{\mathrm{T}}\mathbf{H}}{\sigma^2}$$