E3: Cramer-Rao Lower Bound

Estimation accuracy for a parameter will depend on how sensitive the data pdf is to that parameter.

Ex] PDF dependence on unknown parameter

$x[0] = A + w[0]$ where $w[0] \sim \mathcal{N}(0, \sigma^2)$.

We expect to estimate A better when σ^2 is small.

Consider the unbiased estimator $\hat{A} = x[0]$. Its variance is $\text{var}(\hat{A}) = \sigma^2$. Specifically:

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

Defn The pdf of data expressed as a function of unknown parameters is called the <u>likelihood func.</u>

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

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

The curvature increases as σ^2 decreases. For the simple example above, we have

$$\text{var}(\hat{A}) = \frac{1}{\left[-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}\right]}$$

In general the second derivative will depend on data so we will consider the expected curvature:

$$-E\left[\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}\right]$$

of the log-likelihood function. The expectation is w.r.t. $p(x[0];A)$, resulting in a function of A only.

CRLB

Thm 3.1: CRLB for a scalar parameter

Let $p(\mathbf{x};\theta)$ satisfy the regularity condition

$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \quad \forall \theta$ where $E[\cdot]$ is w.r.t. $p(x;\theta)$.

The variance of any unbiased estimator $\hat{\theta}$ must satisfy $\mathrm{var}(\hat{\theta}) \geq \dfrac{1}{-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]}$

where $E[\cdot]$ is w.r.t. $p(x;\theta)$ and $\partial^2/\partial\theta^2$ is evaluated at θ_{true}.

Furthermore, an unbiased estimator that attains the bound for all θ may be found iff

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

for some functions g and I. Specifically, $\hat{\theta} = g(x)$ is a MVU and the minimum variance is $1/I(\theta)$.

The denominator of CRLB is computed using:

$$E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = \int \frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\, p(x;\theta)\,dx$$

Ex | Problem 3.1

If $x[n]$ for $n=0,1,\ldots,N-1$ are iid according to $\mathcal{U}[0,\theta]$, show that the regularity condition in Thm 3.1. does not hold.

Note that $p(x;\theta) = \prod_{n=0}^{N-1} p(x[n];\theta) = \prod_{n=0}^{N-1} \frac{1}{\theta} = \theta^{-N}, \theta \geq 0.$

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

★ Notice that a small perturbation of θ causes $\ln p(x;\theta)$ to change in a non-continuous fashion in the vicinity of θ.

Derivation of Scalar Parameter CRLB

Consider a pdf parameterized by θ and a scalar parameter $\alpha = g(\theta)$. Let $\hat{\alpha}$ be an unbiased estimator of α:

$E[\hat{\alpha}] = \alpha = g(\theta)$, or $\int \hat{\alpha}\, p(x;\theta)\,dx = g(\theta)$ [3A.1]

Examine the regularity condition: $E\left[\frac{\partial \ln p(x;\theta)}{\partial\theta}\right] = 0$

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

∴ The regularity condition will be satisfied if the order of $\partial/\partial\theta$ and $\int\cdot\,dx$ can be interchanged.

order change.

The regularity condition generally holds (i.e. differentiation and integration can be swapped) except when the domain of the pdf for which it is non zero depends on the unknown parameter (recall the example).

Now consider $\int \hat{\alpha} \frac{\partial p(x;\theta)}{\partial\theta} dx = \frac{\partial g(\theta)}{\partial\theta}$ which is obtained by differentiating [3A.1] w.r.t. θ and interchanging the order of integration and differentiation. From this, we get $\int \hat{\alpha} \frac{\partial \ln p(x;\theta)}{\partial\theta} p(x;\theta) dx = \frac{\partial g(\theta)}{\partial\theta}$. Subtracting a term that is equal to zero when the regularity condition holds; we get.

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

Now let $\tilde{w}(x) = p(x;\theta)$, $\tilde{g}(x) = \hat{\alpha}-\alpha$, $\tilde{h}(x) = \frac{\partial \ln p(x;\theta)}{\partial\theta}$ and apply the Cauchy-Schwarz inequality

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

to get $\left(\frac{\partial g(\theta)}{\partial\theta}\right)^2 \le \int (\hat{\alpha}-\alpha)^2 p(x;\theta)dx \int \left(\frac{\partial \ln p(x;\theta)}{\partial\theta}\right)^2 p(x;\theta)dx$

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

Equality holds iff $(\hat{\alpha}-\alpha) = \tilde{c}\frac{\partial \ln p(x;\theta)}{\partial\theta}$ $(\tilde{g}(x) = \tilde{c}\tilde{h}(x))$.

Now let's focus on the denominator. From the regularity condition: $\int \frac{\partial \ln p(x;\theta)}{\partial\theta} p(x;\theta)\,dx = 0$.

So $\frac{\partial}{\partial\theta}\int \frac{\partial \ln p(x;\theta)}{\partial\theta} p(x;\theta)\,dx = \frac{\partial}{\partial\theta} 0 = 0$

$$\hookrightarrow \int \left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2} p(x;\theta) + \frac{\partial \ln p(x;\theta)}{\partial\theta} \cdot \frac{\partial p(x;\theta)}{\partial\theta} \right] dx = 0$$

$$\hookrightarrow -E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = \int \frac{\partial \ln p(x;\theta)}{\partial\theta} \cdot \frac{\partial \ln p(x;\theta)}{\partial\theta} p(x;\theta)\,dx$$

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

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

Returning to the case when equality is achieved...

$$\frac{\partial \ln p(x;\theta)}{\partial\theta} = \frac{1}{c(\theta)}(\hat{\alpha} - \alpha)$$ where c can depend on θ but not X.

If $\alpha = g(\theta) = \theta$ then $\hookrightarrow = \frac{1}{c(\theta)}(\hat{\theta} - \theta)$

In the latter case $\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left[\frac{1}{c(\theta)}(\hat{\theta} - \theta)\right]$

$$= \frac{-1}{c(\theta)} + \frac{\partial(1/c(\theta))}{\partial\theta}(\hat{\theta} - \theta)$$

$$-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = -\int p(x;\theta)\left[\frac{-1}{c(\theta)} + \cancel{\frac{\partial(1/c(\theta))}{\partial\theta}(\hat{\theta} - \theta)}\right] dx$$

0 unbiased

$$= \frac{1}{c(\theta)}$$

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

$\square$

Ex DC level in AWGN: $x[n] = A + w[n] \quad n = 0, 1, \ldots, N-1$
where $A \in \mathbb{R}$ and $w[n]$ is WGN with variance σ^2.

$$p(\mathbf{x}; A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x[n]-A)^2}$$

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

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

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

where $\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ is the sample average.

$$\frac{\partial^2 \ln p(x;A)}{\partial A^2} = -\frac{N}{\sigma^2} \Rightarrow \boxed{\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{N}}$$ for any unbiased $\hat{A}$.

Regularity condition: $\int p(\mathbf{x};A)\,\frac{N}{\sigma^2}(\bar{x}-A)\,dx = E_x\left[\frac{N}{\sigma^2}\left(\frac{\mathbf{1}^T}{N}x - A\right)\right] = 0.$

When the CRLB is attained $\mathrm{var}(\hat{\theta}) = \frac{1}{I(\theta)}$ where
$I(\theta) = -E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right]$. From Thm 3.1 and the derivation on page E3-5, $\frac{\partial \ln p(x;\theta)}{\partial\theta} = I(\theta)(\hat{\theta}-\theta)$. Differentiating this:

$$\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2} = \frac{\partial I(\theta)}{\partial\theta}(\hat{\theta}-\theta) - I(\theta).$$ Taking $-E[\cdot]$ of this:

$$-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = -\frac{\partial I(\theta)}{\partial\theta}\left[\underbrace{\cancel{E[\hat{\theta}]-\theta}}_{0 \text{ unbiased}}\right] + I(\theta) = I(\theta)$$

$\therefore\ \mathrm{var}(\hat{\theta}) = \frac{1}{I(\theta)}$ when the CRLB is achieved.

Ex) Phase Estimation $x[n] = A\cos(2\pi f_0 n + \phi) + w[n]$

$n = 0, 1, \ldots, N-1$

A and f_0 are known, $w[n]$ is WGN.

$$p(x;\phi) = (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos(2\pi f_0 n + \phi))^2}$$

$$\frac{\partial \ln p(x;\phi)}{\partial \phi} = \frac{-1}{\sigma^2}\sum_{n=0}^{N-1}(x[n] - A\cos(2\pi f_0 n + \phi))A\sin(2\pi f_0 n + \phi)$$

some trigonometry

$$= \frac{-A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\sin(2\pi f_0 n + \phi) - \frac{A}{2}\sin(4\pi f_0 n + 2\phi)\right]$$

$$\frac{\partial^2 \ln p(x;\phi)}{\partial \phi^2} = \frac{-A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\cos(2\pi f_0 n + \phi) - A\cos(4\pi f_0 n + 2\phi)\right]$$

$$-E\left[\frac{\partial^2 \ln p(x;\phi)}{\partial \phi^2}\right] = \frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[A\cos^2(2\pi f_0 n + \phi) - A\cos(4\pi f_0 n + 2\phi)\right]$$

some trig.

$$= \frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left[\frac{1}{2} + \frac{1}{2}\cos(4\pi f_0 n + 2\phi) - \cos(4\pi f_0 n + 2\phi)\right]$$

$$\approx \frac{NA^2}{2\sigma^2} \quad \text{since } \frac{1}{N}\sum_{n=0}^{N-1}\cos(4\pi f_0 n + 2\phi) \approx 0$$

for large N and f_0 not near 0 or $1/2$.

For unbiased $\hat{\phi}$, $\mathrm{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2}$ (approximate bound).

The first order derivative does not have the form in Thm 3.1 for the bound to be attained, so an unbiased phase estimator that achieves CRLB does not exist.

At this point, we do not know how the find the MVU phase estimator.

<u>Defn</u>: An estimator $\hat{\theta}$ of θ is said to be <u>efficient</u> iff $E\{\hat{\theta}\} = \theta$ and $var(\hat{\theta}) = \frac{1}{I(\theta)}$.
($\hat{\theta}$ is unbiased) ($\hat{\theta}$ attains CRLB)

<u>Fact</u>: $\hat{\theta}$ is efficient $\Rightarrow$ $\hat{\theta}$ is a MVU estimator

<u>Defn</u>: $I(\theta) = -E\left\{ \frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2} \right\}$ is referred to as the Fisher information for data x.

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

sometimes easier to calculate.

<u>Fact</u>: $I(\theta) \geq 0$

$I(\theta)$ is additive for independent observations.

The latter property of $I(\theta) \Rightarrow$ CRLB for N iid observations is $\frac{1}{N}$ times the CRLB for one observation.

To see this consider $\ln p(\mathbf{x};\theta) = \sum_{n=0}^{N-1} \ln p(x[n];\theta)$.

$$I_N(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} \right] = -\sum_{n=0}^{N-1} E\left[\frac{\partial^2 \ln p(x[n];\theta)}{\partial \theta^2} \right] = N I_1(\theta)$$

Fisher information for N iid samples   Fisher information for one sample

(somewhat)
For dependent N samples $I_N(\theta) \leq N I_1(\theta)$ and for completely dependent ~~N~~ samples, $I_N(\theta) = I_1(\theta)$. In the latter case, additional measurements will not reduce CRLB and the variance of the estimator.

General CRLB for Signals in WGN

$x[n] = s[n;\theta] + w[n]$ $n = 0, 1, \ldots, N-1$ where $w[n]$ is WGN and $s[n;\theta]$ is a deterministic signal with an unknown parameter θ.

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

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

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

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

$$\Rightarrow \operatorname{var}(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2}$$ * if the regularity condition is satisfied.

(verify that the regularity condition is satisfied.)

Ex Sinusoidal Frequency Estimation

Assume that $s[n;f_0] = A\cos(2\pi f_0 n + \phi)$ $0 < f_0 < \frac{1}{2}$ where A and ϕ are known. Using the expression above we determine that

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

Note that $\lim_{\substack{f_0\to 0^+ \\ f_0 \to \frac{1}{2}^-}} \operatorname{var}(\hat{f}_0) = \infty$. Also considering A^2/σ^2 as SNR, as $\text{SNR}\to\infty$ $\operatorname{var}(\hat{f}_0)$ decreases (possibly $\to 0$).

Transformation of Parameters

In many cases we are interested in estimating a parameter ~~of~~ that is a function of a more fundamental model parameter. If $\alpha = g(\theta)$, then

$$\mathrm{var}(\hat{\alpha}) \geq \frac{(\partial g(\theta)/\partial\theta)^2}{-E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right]}$$ (as we have shown before)

(Note that this is like linearizing $g(\cdot)$ around the true θ and relating the var of $\hat{\alpha}$ to that of $\hat{\theta}$.)

<u>Fact</u>: In general, the efficiency of an estimator does not transition ~~to parameters~~ through nonlinear parameter transformations. In fact, before one considers variance, even the unbiased-ness property will not even transfer to nonlinear parameter redefinitions.

Ex) Let $\hat{\theta}$ be an efficient estimator for θ. Assume we want to estimate $g(\theta) = a\theta + b$. Consider $\widehat{g(\theta)} = a\hat{\theta} + b$.

$E[a\hat{\theta}+b] = a\theta + b = g(\theta)$ {unbiased ✓}

(CRLB) $\mathrm{var}(\widehat{g(\theta)}) \geq \frac{(\partial g/\partial\theta)^2}{I(\theta)} \underset{\text{since } \hat{\theta} \text{ is efficient}}{=} \left(\frac{\partial g(\theta)}{\partial\theta}\right)^2 \mathrm{var}(\hat{\theta}) = a^2 \mathrm{var}(\hat{\theta})$

Exact variance: $\mathrm{var}(a\hat{\theta}+b) = a^2 \mathrm{var}(\hat{\theta}) = \mathrm{CRLB}$ so $\widehat{g(\theta)}$ is efficient.

Fact: Efficiency is preserved over linear parameter transforms. It is approximately preserved through nonlinear transformations if $N \to \infty$. (verify this by considering a linearization of $g(\theta)$ around θ_T and $\text{var}(\hat{\theta}) \to \text{CRLB}(\theta)$ as $N \to \infty$.)

Defn: An unbiased estimator $\hat{\theta}$ of parameter θ is asymptotically efficient iff $\lim_{N\to\infty} \text{var}(\hat{\theta}) = \text{CRLB}(\hat{\theta})$ $\forall \theta$.

Defn: An estimator $\hat{\theta}$ of θ is asymptotically unbiased iff $\lim_{N\to\infty} E[\hat{\theta}] = \theta$ $\forall \theta$.

Extension to a Vector Parameter

Let $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^T$ and $\hat{\theta}$ be an unbiased estimator.

CRLB: $\quad \text{Cov}(\hat{\theta}) - I^{-1}(\theta) \geq 0$ ← positive semidefinite

OR $\quad \text{Cov}(\hat{\theta}) \geq I^{-1}(\theta)$ ← (more commonly).

Corollary: Since the diagonal entries of a positive semidefinite matrix are nonnegative, we have

$$\text{var}(\hat{\theta}_i) \geq [I^{-1}(\theta)]_{ii}$$

Here, the $p \times p$ Fisher Information Matrix is

$$I(\theta) = -E\left[\underbrace{\nabla_\theta^T \nabla_\theta}_{\text{Hessian operator}} \ln p(x;\theta)\right]$$

More formally (for $E[\cdot]$ over $p(x;\theta)$ and ∇_θ at θ_{true})

<u>Thm 3.2</u> ~~If~~ If $p(x;\theta)$ satisfies the regularity conditions $E\left[\nabla_\theta^T \ln p(x;\theta)\right] = 0 \quad \forall \theta$,

then $C_{\hat{\theta}} - I^{-1}(\theta) \geq 0$, where $I(\theta) = -E\left[\nabla_\theta^T \nabla_\theta \ln p(x;\theta)\right]$

covariance matrix of $\hat{\theta}$

Furthermore, an unbiased estimator may be found that attains the bound ($C_{\hat{\theta}} = I^{-1}(\theta)$) iff $\nabla_\theta^T \ln p(x;\theta) = I(\theta)(g(x) - \theta)$, for some $g(\cdot)$, and its covariance is $I^{-1}(\theta)$.

<u>Proof</u>: See Appendix 3B.

Suppose we are interested in estimating $\alpha = g(\theta)$ for $g: \mathbb{R}^p \to \mathbb{R}^r$. Then (as in App 3B)

$$C_{\hat{\alpha}} - J_g(\theta)\, I^{-1}(\theta)\, J_g^T(\theta) \geq 0$$

where $J_g(\theta) = \frac{\partial g(\theta)}{\partial \theta}$ is the $r \times p$ Jacobian matrix of g.

<u>Ex</u> CRLB for the general Gaussian case

Let $x \sim \mathcal{N}(\mu(\theta), C(\theta))$ be a multivariate Gaussian random variable with mean and covariance parameterized by θ.

With some effort, we can show that (see App 3C)

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

Ex) DC level in WGN

$x[n] = A + w[n]$, $n = 0, 1, \ldots, N-1$ where $w[n]$ is WGN with variance σ^2. $\theta = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$.

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

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

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

$$\frac{\partial \ln p(x;\theta)}{\partial \sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}(x[n]-A)^2$$

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

$$\frac{\partial^2 \ln p(x;\theta)}{\partial A \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{n=0}^{N-1}(x[n]-A)$$

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

$$\Rightarrow \mathbf{I}(\theta) = \begin{bmatrix} N/\sigma^2 & 0 \\ 0 & N/(2\sigma^4) \end{bmatrix} \Rightarrow \begin{array}{l} \mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{N} \\ \mathrm{var}(\hat{\sigma^2}) \geq \frac{2\sigma^4}{N} \end{array}$$

Ex] CRLB for SNR for DC level in WGN

$x[n] = A + w[n]$ $n = 0, 1, \ldots, N-1$, $w[n] \sim$ WGN with var σ^2

A, σ^2 are unknown and we want to estimate $\alpha = \frac{A^2}{\sigma^2}$

$\theta = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$, $\alpha = g(\theta) = \frac{\theta_1^2}{\theta_2}$. We found in the previous example that $I(\theta) = \begin{bmatrix} N/\sigma^2 & 0 \\ 0 & N/(2\sigma^4) \end{bmatrix}$. The Jacobian

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

$$\Rightarrow \text{var}(\hat{\alpha}) \geq \frac{\partial g(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial g(\theta)}{\partial \theta}^T = \begin{bmatrix} \frac{2A}{\sigma^2} & -\frac{A^2}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix} \begin{bmatrix} 2A/\sigma^2 \\ -A^2/\sigma^4 \end{bmatrix}$$

$$= \frac{4A^2}{N\sigma^2} + \frac{2A^4}{N\sigma^4} = \frac{4\alpha + 2\alpha^2}{N}$$

Efficiency over general linear transformations:

Let $\alpha = A\theta + b$ where A is an $r \times p$ matrix.

Consider $\hat{\alpha} = A\hat{\theta} + b$ where $\hat{\theta}$ is efficient.

$$E[\hat{\alpha}] = E[A\hat{\theta} + b] = A\theta + b \quad \text{(unbiased)}$$

$$C_{\hat{\alpha}} = A C_{\hat{\theta}} A^T \underset{\uparrow \text{ since } \hat{\theta} \text{ is efficient}}{=} \overbrace{A I^{-1}(\theta) A^T}^{\text{covariance}} = \underbrace{J_g(\theta) I^{-1}(\theta) J_g^T(\theta)}_{\text{CRLB}}$$

So $\hat{\alpha}$ is efficient.

Ex] Random DC level in WGN

$x[n] = A + w[n]$ $n = 0, 1, \ldots, N-1$ with $w[n] \sim$ WGN with var σ^2 and A is random, independent from $w[n]$, ~~and has~~ and has a zero-mean σ_A^2-variance Gaussian distribution. Here σ_A^2 is unknown. Let $x = [x[0]\ x[1]\ \ldots\ x[N-1]]^T$. x is Gaussian with zero mean and $C = \sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2 I$.

Using the matrix inversion lemma we get

$$C^{-1}(\sigma_A^2) = \frac{1}{\sigma^2}\left[I - \frac{\sigma_A^2}{\sigma^2+N\sigma_A^2} 1 1^T \right]$$

Recall the CRLB for the general Gaussian case. When θ is a scalar, it reduces to

$$I(\theta) = \frac{\partial \mu(\theta)}{\partial \theta}^T C^{-1}(\theta) \frac{\partial \mu(\theta)}{\partial \theta} + \frac{1}{2} tr\left[\left(C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta} \right)^2 \right]$$

In this example $\dfrac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = 1 1^T$, and $C^{-1}(\sigma_A^2)\dfrac{\partial C(\sigma_A^2)}{\partial \sigma_A^2} = \dfrac{1 1^T}{\sigma^2+N\sigma_A^2}$.

Substituting everything in $I(\theta)$, we get

$$I(\sigma_A^2) = \frac{1}{2} tr\left[\left(\frac{1}{\sigma^2+N\sigma_A^2} \right)^2 1 1^T 1 1^T \right] = \frac{1}{2}\left(\frac{N}{\sigma^2+N\sigma_A^2} \right)^2$$

so $var(\sigma_A^2) \geq 2\left(\sigma_A^2 + \frac{\sigma^2}{N} \right)^2$

Ex] Asymptotic CRLB for WSS Gaussian random processes

Due to the Wolfe decomposition of random processes, we know that almost any WSS Gaussian random process $x[n] = h[n] * u[n]$ where $u[n]$ is WGN.

$$x[n] = \sum_{k=0}^{\infty} h[k] u[n-k] \quad \text{(with causal LTI } h)$$

Here $h[0] = 1$. For this to hold, from Szegö's theorem, we need $\int_{-1/2}^{1/2} \ln P_{xx}(f)\, df > -\infty$. From the Einstein-Wiener-Kintchine theorem, $P_{xx}(f) = |H(f)|^2 \sigma_u^2$ where σ_u^2 is the variance of $u[n]$ (and the PSD of that WGN).

$H(f) = \sum_{k=0}^{\infty} h[k] e^{-j2\pi f k}$ is the Fourier transform.

If we observe $\{x[0], x[1], \ldots, x[N-1]\}$ where N is large, then the representation is approximated well with a truncated convolution (equivalent to setting $u[n]=0 \ \forall n<0$)

$$x[n] = \sum_{k=0}^{n} h[k]u[n-k] + \sum_{k=n+1}^{\infty} h[k]u[n-k]$$

$$\approx \sum_{k=0}^{n} h[k]u[n-k]$$

Clearly, early samples (small n) will be poorly approximated but as $n \to \infty$ the approximation for $x[n]$ will improve in accuracy.

$$r_{xx}[k] = \sigma_u^2 \sum_{n=0}^{\infty} h[n]h[n+k]$$

From this we conclude that the correlation time of $x[n]$ is the same as the impulse response length. The asymptotic CRLB will be a good approximation when $N >>$ correlation time.

Let $u = [u[0], u[1], \ldots, u[N-1]]^T$. Then

$$\mathbf{x} = \begin{bmatrix} h[0] & 0 & \cdots & 0 \\ h[1] & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ h[N-1] & \cdots & h[1] & h[0] \end{bmatrix} u = Hu$$

Note that $\det H = (h[0])^N = 1$ so H^{-1} exists.

Since $u \sim \mathcal{N}(0, \sigma_u^2 I)$, $x \sim \mathcal{N}(0, \sigma_u^2 HH^T)$.

After some simplification $p(x;\theta) = \frac{1}{(2\pi\sigma_u^2)^{N/2}} e^{-\frac{u^T u}{2\sigma_u^2}}$.

With $X(f) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi fn}$ and $U(f) = \sum_{n=0}^{N-1} u[n] e^{-j2\pi fn}$

we (approximately) have $X(f) = H(f)U(f)$.

By Parseval's theorem

$$\frac{1}{\sigma_u^2} u^T u = \frac{1}{\sigma_u^2} \sum_{n=0}^{N-1} u^2[n] = \frac{1}{\sigma_u^2} \int_{-1/2}^{1/2} |u(f)|^2 df$$

$$\approx \int_{-1/2}^{1/2} \frac{|X(f)|^2}{\sigma_u^2 |H(f)|^2} df = \int_{-1/2}^{1/2} \frac{|X(f)|^2}{P_{xx}(f)} df$$

$$\ln \sigma_u^2 = \int_{-1/2}^{1/2} \ln \sigma_u^2 df = \int_{-1/2}^{1/2} \ln \left(\frac{P_{xx}(f)}{|H(f)|^2} \right) df$$

$$= \int_{-1/2}^{1/2} \ln P_{xx}(f) df - \int_{-1/2}^{1/2} \ln |H(f)|^2 df$$

$$\int_{-1/2}^{1/2} \ln |H(f)|^2 df = \int_{-1/2}^{1/2} [\ln H(f) + \ln H^*(f)] df$$

$$= 2 \operatorname{Re} \int_{-1/2}^{1/2} \ln H(f) df = 2 \operatorname{Re} \oint_C \ln \mathcal{H}(z) \frac{dz}{2\pi j z}$$

$$= 2 \operatorname{Re} \left[\mathcal{Z}^{-1} \{ \ln \mathcal{H}(z) \} \big|_{n=0} \right]$$

where C is the unit circle in the z-plane and $\mathcal{H}(z)$ corresponds to the system function (of a causal filter).

By the initial value theorem for $\mathcal{Z}$-transform

$$\mathcal{Z}^{-1} \{ \ln \mathcal{H}(z) \} \big|_{n=0} = \lim_{z \to \infty} \ln \mathcal{H}(z) = \ln \lim_{z \to \infty} \mathcal{H}(z) = \underbrace{\ln h[0]}_{=0}$$

$$\Rightarrow \int_{-1/2}^{1/2} \ln |H(f)|^2 df = 0$$

$$\Rightarrow \ln \sigma_u^2 = \int_{-1/2}^{1/2} \ln P_{xx}(f) df$$

Then the $\ln$-pdf of x becomes (asymptotically)

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

$$= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \int_{-1/2}^{1/2} \left[\ln P_{xx}(f) + \frac{(1/N)|X(f)|^2}{P_{xx}(f)} \right] df$$

$$\frac{\partial \ln p(x;\theta)}{\partial \theta_i} = -\frac{N}{2}\int_{-1/2}^{1/2}\left(\frac{1}{P_{xx}(f)} - \frac{(1/N)\,|X(f)|^2}{P_{xx}^2(f)}\right)\frac{\partial P_{xx}(f)}{\partial \theta_i}\,df$$

$$\frac{\partial^2 \ln p(x;\theta)}{\partial \theta_i \partial \theta_j} = \frac{-N}{2}\int_{-1/2}^{1/2}\left[\left(\frac{1}{P_{xx}(f)} - \frac{(1/N)\,|X(f)|^2}{P_{xx}^2(f)}\right)\frac{\partial P_{xx}(f)}{\partial \theta_i} + \left(\frac{-1}{P_{xx}^2(f)} + \frac{(2/N)\,|X(f)|^2}{P_{xx}^3(f)}\right)\frac{\partial P_{xx}(f)}{\partial \theta_i}\frac{\partial P_{xx}(f)}{\partial \theta_j}\right]df$$

For the $E[\cdot]$ in Fisher information, we encounter $E[|X(f)|^2/N]$. For large N, this converges to $P_{xx}(f)$ (since it is the periodogram estimator of PSD).

Expectation of the first term in the Hessian of $\ln p(x;\theta)$ becomes asymptotically zero due to this property of the periodogram estimator. Then

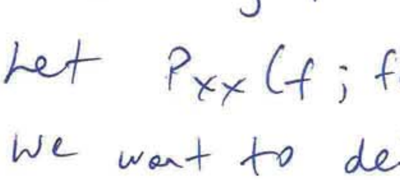
$$I(\theta) = \frac{N}{2}\int_{-1/2}^{1/2} \nabla_\theta^T \ln P_{xx}(f)\, \nabla_\theta \ln P_{xx}(f)\, df.$$

Ex] Estimating the center frequency of a random process.

Let $P_{xx}(f;f_c) = Q(f-f_c) + Q(-f-f_c) + \sigma^2$.

We want to determine f_c. $Q(f)$ and σ^2 are known.

Here x is a random process embedded in white noise.



$f_c \in [f_1, \frac{1}{2} - f_2]$, $\theta = f_c$.

From the previous example

$$\mathrm{var}(\hat{f}_c) \geq \frac{1}{\left(\frac{N}{2}\right)\int_{-1/2}^{1/2}\left(\frac{\partial \ln P_{xx}(f;f_c)}{\partial f_c}\right)^2 df}$$

But $\frac{\partial \ln P_{xx}(f;f_c)}{\partial f_c} = \frac{\partial \ln [Q(f-f_c)+Q(-f-f_c)+\sigma^2]}{\partial f_c}$

$$= \frac{\left(\frac{\partial Q(f-f_c)}{\partial f_c} + \frac{\partial Q(-f-f_c)}{\partial f_c}\right)}{\left(Q(f-f_c)+Q(-f-f_c)+\sigma^2\right)}$$

This is an odd function, so

$$\int_{-1/2}^{1/2}\left(\frac{\partial \ln P_{xx}(f;f_c)}{\partial f_c}\right)^2 df = 2\int_0^{1/2}\left(\frac{\partial \ln P_{xx}(f;f_c)}{\partial f_c}\right)^2 df$$

For $f \geq 0$, $Q(-f-f_c)=0$ so $\left.\frac{\partial Q(-f-f_c)}{\partial f_c}\right|_{f\geq 0} = 0$

With these observations

$$\text{var}(\hat{f}_c) \geq \frac{1}{N\int_0^{1/2}\left(\frac{\partial Q(f-f_c)/\partial f_c}{Q(f-f_c)+\sigma^2}\right)^2 df}$$

with change of variable $f' = f - f_c$

$$= N\int_{-f_c}^{\frac{1}{2}-f_c}\left(\frac{\partial Q(f')/\partial f'}{Q(f')+\sigma^2}\right)^2 df'$$

Since $\frac{1}{2}-f_c \geq \frac{1}{2} f_{c_{max}} = f_2$ and $-f_c \leq -f_{c_{min}} = -f_1$ we may change the integration limits to $[-1/2, 1/2)$:

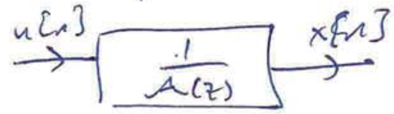
$$\text{var}(\hat{f}_c) \geq \frac{1}{N\int_{-1/2}^{1/2}\left(\frac{\partial Q(f)/\partial f}{Q(f)+\sigma^2}\right)^2 df} = \frac{1}{N\int_{-1/2}^{1/2}\left(\frac{\partial \ln(Q(f)+\sigma^2)}{\partial f}\right)^2 df}$$

For example, if $Q(f) = e^{-\frac{f^2}{2\sigma_f^2}}$, then

$$\text{var}(\hat{f}_c) \geq \frac{1}{N\int_{-1/2}^{1/2}\frac{f^2}{\sigma_f^4}df} = \frac{12\sigma_f^4}{N}$$

$\sigma_f^2 \downarrow$ (narrower bandwidth) yield lower bounds for f_c.

Ex) AR parameter estimation



$x[n]$ is the output of an all-pole (AR) filter with WGN input $u[n]$ where the PSD of WGN u is σ_u^2.

$$A(z) = 1 + \sum_{m=1}^{p} a[m] z^{-m} \quad \text{and} \quad x = [x[0], x[1], \dots, x[N-1]]^T.$$

$$\hat{P}_{xx}(f) = \frac{\hat{\sigma}_u^2}{\left| 1 + \sum_{m=1}^{p} \hat{a}[m] e^{-j2\pi f m} \right|^2} \qquad P_{xx}(f) = \frac{\sigma_u^2}{|A(f)|^2}$$

$$\theta = [a[1], \dots, a[p], \sigma_u^2]^T, \quad A(f) = 1 + \sum_{m=1}^{p} a[m] e^{-j2\pi f m}$$

$$\frac{\partial \ln P_{xx}(f;\theta)}{\partial a[k]} = -\frac{\partial \ln |A(f)|^2}{\partial a[k]} = -\frac{1}{|A(f)|}\left[A(f) e^{j2\pi f k} + A^*(f) e^{-j2\pi f k}\right]$$

$$\frac{\partial \ln P_{xx}(f;\theta)}{\partial \sigma_u^2} = \frac{1}{\sigma_u^2}$$

For $k, l \in \{1, \dots, p\}$, we have (from ~~the~~ a previous example)

$$[I(\theta)]_{kl} = \frac{N}{2} \int_{-1/2}^{1/2} \frac{1}{|A(f)|^4} \left[A(f) e^{j2\pi f k} + A^*(f) e^{-j2\pi f k}\right] \left[A(f) e^{j2\pi f l} + A^*(f) e^{-j2\pi f l}\right] df$$

after some work

$$= N \int_{-1/2}^{1/2} \left[\frac{1}{|A(f)|^2} e^{j2\pi f(k-l)} + \frac{1}{A^*(f)^2} e^{j2\pi f(k+l)} \right] df$$

inverse Fourier transform

The second term in the integral is zero for $n = k+l > 0$.

~~since the~~ Therefore $[I(\theta)]_{kl} = \frac{N}{\sigma_u^2} r_{xx}[k-l]$.

For $k \in \{1, \dots, p\}$, $l = p+1$

$$[I(\theta)]_{kl} = \frac{-N}{2} \int_{-1/2}^{1/2} \frac{1}{\sigma_u^2} \frac{1}{|A(f)|^2} [A(f) e^{j2\pi f k} + A^*(f) e^{-j2\pi f k}] \, df$$

$$= \frac{-N}{\sigma_u^2} \int_{-1/2}^{1/2} \frac{1}{A^*(f)} e^{j2\pi f k} \, df = 0$$

(similarly).

For $k, l \in \{p+1\}$

$$[I(\theta)]_{kl} = \frac{N}{2} \int_{-1/2}^{1/2} \frac{1}{\sigma_u^4} \, df = \frac{N}{2\sigma_u^4}$$

$$\Rightarrow I(\theta) = \begin{bmatrix} \frac{N}{\sigma_u^2} R_{xx} & 0 \\ 0^T & \frac{N}{2\sigma_u^4} \end{bmatrix}$$

where $[R_{xx}]_{ij} = r_{xx}[i-j]$ is the Toeplitz autocorr. matrix of x. From this, we get

$$\mathrm{var}(\hat{a}[k]) \geq \frac{\sigma_u^2}{N} [R_{xx}^{-1}]_{kk} \qquad k = 1, \dots, p$$

$$\mathrm{var}(\hat{\sigma_u^2}) \geq \frac{2\sigma_u^4}{N}$$

Suggested Problems: 1, 4, 10, 12, 13, 15, 16, 20