现代信号处理

Lecture 17

唐晓颖
电子与电气工程系
南方科技大学

December 2, 2025

The CRLB may also be expressed in a slightly different form. Since

$$E\left[\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]$$ (see Appendix 3A)

we have

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{E\left[\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right]}$$ (see Problem 3.8)

$I(\theta) = -E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]$ is referred to as the Fisher information $I(\theta)$ for the data x

When the CRLB is attained, the variance is the reciprocal of the Fisher information. Intuitively, the more information, the lower the bound. It has the essential properties of an *information measure* in that it is:

1. Nonnegative $E\left[\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]$
2. Additive for independent observations

As such, the CRLB for N IID observations is $1/N$ times that for one observation.
To verify this, note that for independent observations

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

This results in

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

And thus

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

where

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

is the Fisher information for one sample.

For non-independent samples we might expect that the information will be less than $Ni(\theta)$.
For completely dependent samples, as for example, $x[0]=x[1]=...=x[N-1]$, we will have $I(\theta)=i(\theta)$
Therefore, additional observations carry no information, and the CRLB will not decrease with increasing data record length.

Example – Sinusoidal Frequency Estimation

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

Assume that the signal is sinusoidal and is represented as

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

The amplitude A and phase ϕ are assumed known.

From above, the CRLB becomes

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

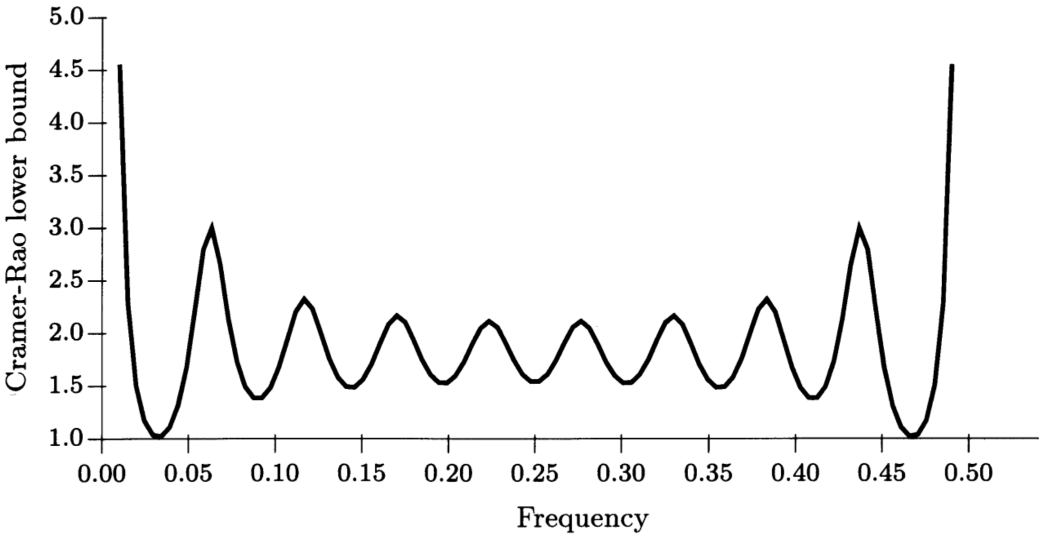


Figure 3.3 Cramer-Rao lower bound for sinusoidal frequency estimation

An SNR of $A^2/\sigma^2 = 1$

A data record length of $N = 10$

A phase of $\phi = 0$

There appear to be preferred frequencies. Also, as $f_0 \to 0$, the CRLB goes to infinity.

3.6 Transformation of Parameters

It frequently occurs in practice that the parameter we wish to estimate is a function of some more fundamental parameter.

If it is desired to estimate $\alpha = g(\theta)$, then the CRLB is

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

(see Appendix 3A)

Example – DC Level in White Gaussian Noise

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

In this example, the sample mean estimator $\bar{x}$ was efficient for A. It might be supposed that $\bar{x}^2$ is efficient for A^2?

We first show that $\overline{x}^2$ is not even an unbiased estimator.

Since $\bar{x} \sim N(A, \sigma^2/N)$

$$E(\overline{x}^2) = E^2(\overline{x}) + \mathrm{var}(\overline{x}) = A^2 + \frac{\sigma^2}{N}$$

$$\neq A^2$$

If X and Y are independent random variables that are normally distributed (and therefore also jointly so), then their sum is also normally distributed. i.e., if

$$X \sim N(\mu_X, \sigma_X^2)$$
$$Y \sim N(\mu_Y, \sigma_Y^2)$$
$$Z = X + Y,$$

then

$$Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Hence, the efficiency of an estimator is destroyed by a nonlinear transformation. It is maintained by linear (actually affine) transformations.

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + (\mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + (\mathrm{E}[X])^2 = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$$

Let $g(\theta) = a\theta + b,\ \hat{g}(\theta) = a\hat{\theta} + b$

$$E(a\hat{\theta} + b) = aE(\hat{\theta}) + b = a\theta + b$$
$$= g(\theta)$$

so that $\hat{g}(\theta)$ is unbiased.

The CRLB for $g(\theta)$ is $\mathrm{var}(\hat{g}(\theta)) \geq \dfrac{\left(\dfrac{\partial g}{\partial \theta}\right)^2}{I(\theta)} = \left(\dfrac{\partial g}{\partial \theta}\right)^2 \mathrm{var}(\hat{\theta})$ and $\mathrm{var}(\hat{g}(\theta)) = \mathrm{var}(a\hat{\theta} + b) = a^2\,\mathrm{var}(\hat{\theta})$

$$= a^2\,\mathrm{var}(\hat{\theta})$$

so that the CRLB is achieved.

Although efficiency is reserved only over linear transformations, it is *approximately* maintained over nonlinear transformations *if the data record is large enough*.

$$\bar{x} \sim N(A, \sigma^2/N)$$

$$E(\overline{x}^2) = E^2(\overline{x}) + \mathrm{var}(\overline{x}) = A^2 + \frac{\sigma^2}{N}$$

$$E(\overline{x}^2) \to A^2, \text{ when } N \to \infty \qquad \mathrm{var}(\overline{x}^2) = E(\overline{x}^4) - E^2(\overline{x}^2)$$

If $\xi \sim N(\mu, \sigma^2)$, then

$$E(\xi^2) = \mu^2 + \sigma^2$$

$$E(\xi^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

and therefore

$$\mathrm{var}(\xi^2) = E(\xi^4) - E^2(\xi^2)$$
$$= 4\mu^2\sigma^2 + 2\sigma^4$$

Hence, $\overline{x}^2$ is an asymptotically efficient estimator of A^2.

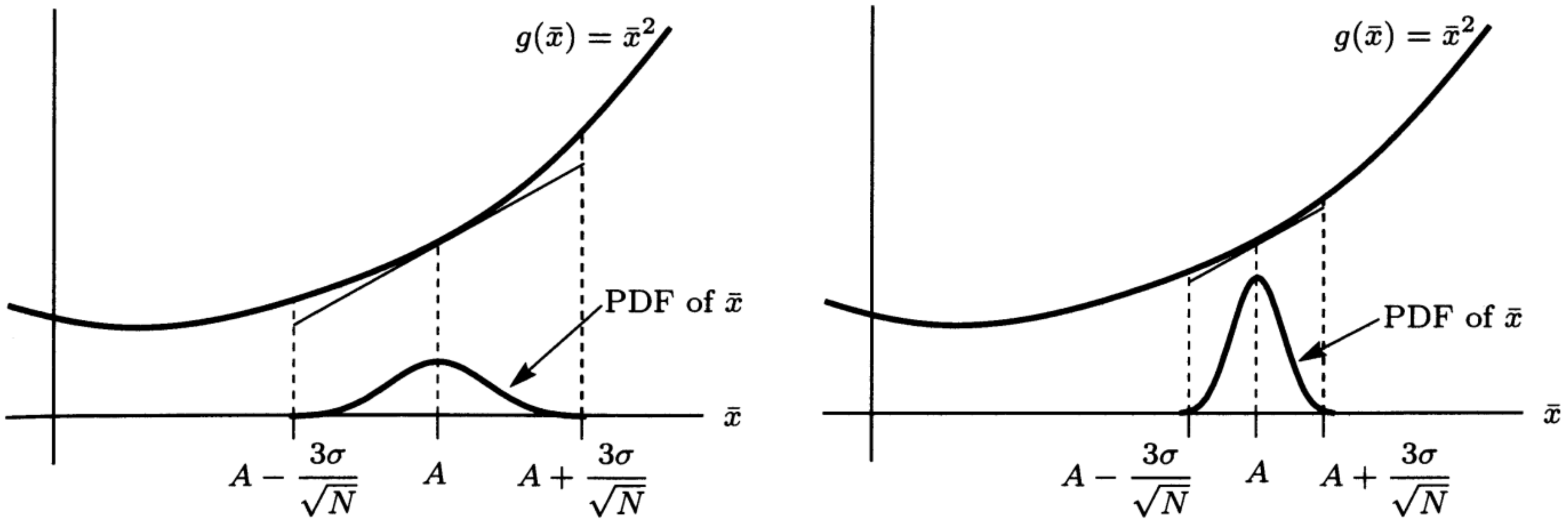
$$\boxed{\mathrm{var}(\hat{A}^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}}$$

$$\mathrm{var}(\overline{x}^2) = E(\overline{x}^4) - E^2(\overline{x}^2)$$
$$= \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}$$

$$\mathrm{var}(\overline{x}^2) \to 4A^2\sigma^2/N, \text{ when } N \to \infty$$

Order	Non-central moment	Central moment
1	μ	0
2	$\mu^2 + \sigma^2$	σ^2
3	$\mu^3 + 3\mu\sigma^2$	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$	$105\sigma^8$

Intuitively, this situation occurs due to the *statistical linearity* of the transformation.



(a) Small N **(b)** Large N

Figure 3.4 Statistical linearity of nonlinear transformations

As N increases, the PDF of $\bar{x}$ becomes more concentrated about the mean A. Therefore, the values of $\bar{x}$ that are observed lie in a small interval about $\bar{x} = A$. Over this small interval the nonlinear transformation is approximately linear.

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots.$$

If we linearize g about A, we have the approximation

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

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

$$\operatorname{var}[g(\bar{x})] = \left[\frac{dg(A)}{dA}\right]^2 \operatorname{var}(\bar{x}) = \frac{(2A)^2\sigma^2}{N} = \frac{4A^2\sigma^2}{N}$$ (the estimator achieves the CRLB)

asymptotically efficient

3.7 Extension to a Vector Parameter

We now extend the results to the case where we wish to estimate a vector parameter $\theta = [\theta_1 \theta_2 \ldots \theta_p]^T$

As derived in Appendix 3B, the CRLB is found as the $[i,i]$ element of the inverse of a matrix or

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

where $\mathbf{I}(\theta)$ is the $p \times p$ Fisher information matrix, defined as

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

for $i = 1, 2, \ldots, p; j = 1, 2, \ldots, p$

Example – DC Level in White Gaussian Noise

Consider the multiple observations

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

where $w[n] \sim N(0,\sigma^2)$, and it is desired to estimate A and σ^2, such that $\theta = [A\sigma^2]^T$.

The 2×2 Fisher information matrix is

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

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

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

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

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

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

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

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

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

$$E(x[n]-A)^2 = E(x[n]-E(x[n]))^2 = \mathrm{var}(x[n]) = \sigma^2$$

Upon taking the negative expectations, the Fisher information matrix becomes

$$\mathbf{I}(\theta) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \qquad \mathbf{I}^{-1}(\theta) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$$

Although not true in general, for this example the Fisher information matrix is diagonal and hence easily inverted to yield

$$\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{N}$$

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

When Matrix A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Note that the CRLB for $\hat{A}$ is the same as for the case when σ^2 is known due to the diagonal nature of the matrix. Again this is not true in general.

Example – Line Fitting

Consider the problem of line fitting or given the observations

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

where $w[n]$ is WGN, determine the CRLB for the slope B and the intercept A.

The 2×2 Fisher information matrix is

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

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

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

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

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

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

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

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

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

The CRLB is

$$\mathrm{var}(\hat{A}) \ge \frac{2(2N-1)\sigma^2}{N(N+1)}$$

$$\mathrm{var}(\hat{B}) \ge \frac{12\sigma^2}{N(N^2-1)}$$

Since the second-order derivatives do not depend on x, we have immediately that

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

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

Inverting the matrix yields

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

When Matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Note first that the CRLB for A has increased over that obtained when B is known, for in the latter case we have

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

For $N \geq 2$, we have

$$\frac{2(2N-1)\sigma^2}{N(N+1)} - \frac{\sigma^2}{N} = \frac{\sigma^2}{N}\frac{3(N-1)}{(N+1)} > 0$$

This is a quite general result that asserts that *the CRLB always increases as we estimate more parameters*.

A second point is that

$$\frac{\operatorname{CRLB}(\hat{A})}{\operatorname{CRLB}(\hat{B})} = \frac{(2N-1)(N-1)^2}{6} > 1 \text{ for } N \geq 3$$

Hence, B is easier to estimate, its CRLB decreasing as $\frac{1}{N^3}$ as opposed to the $\frac{1}{N}$ dependence for the CRLB of A.

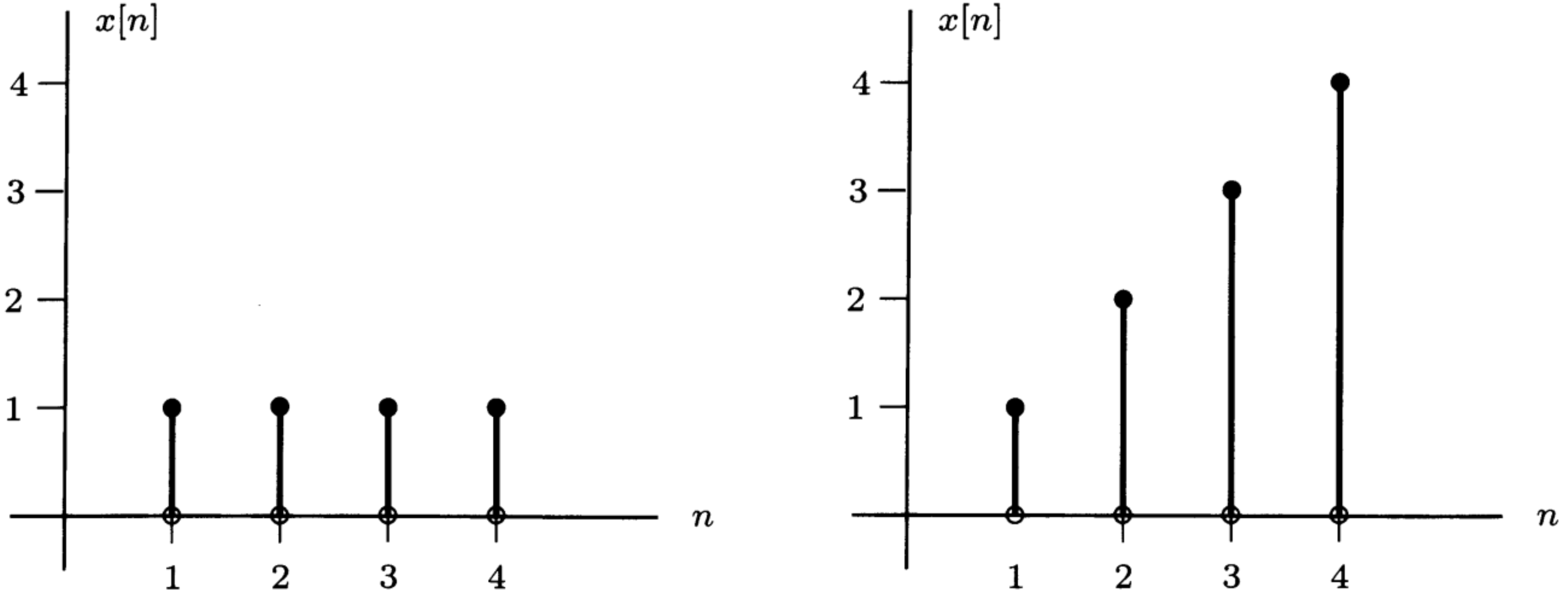
These differing dependences indicate that $x[n]$ is more sensitive to changes in B than to changes in A.

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

Changes in B are magnified by n

$$\Delta x[n] \approx \frac{\partial x[n]}{\partial A} \Delta A = \Delta A$$

$$\Delta x[n] \approx \frac{\partial x[n]}{\partial B} \Delta B = n \Delta B$$



(a) $A = 0,\ B = 0$ to $A = 1,\ B = 0$ **(b)** $A = 0,\ B = 0$ to $A = 0,\ B = 1$

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

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

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

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}}-\mathbf{I}^{-1}(\boldsymbol{\theta})\geq \mathbf{0}$$

where $\geq \mathbf{0}$ *is interpreted as meaning that the matrix is* *positive semidefinite*. *The Fisher information matrix* $\mathbf{I}(\boldsymbol{\theta})$ *is given as*

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

where the derivatives are evaluated at the true value of $\boldsymbol{\theta}$ *and the expectation is taken with respect to* $p(\mathbf{x};\boldsymbol{\theta})$. *Furthermore, an unbiased estimator may be found that attains the bound in that* $\mathbf{C}_{\hat{\boldsymbol{\theta}}}=\mathbf{I}^{-1}(\boldsymbol{\theta})$ *if and only if*

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

for some p-dimensional function $\mathbf{g}$ *and some* $p\times p$ *matrix* $\mathbf{I}$. *That estimator, which is the MVU estimator, is* $\hat{\boldsymbol{\theta}}=\mathbf{g}(\mathbf{x})$, *and its covariance matrix is* $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

If the entries in the column vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

are random variables, each with finite variance, then the covariance matrix Σ is the matrix whose (*i*, *j*) entry is the covariance

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{E}[X_i X_j] - \mu_i \mu_j,$$

where the operator E denotes the expected (mean) value of its argument, and

$$\mu_i = \mathrm{E}(X_i)$$

is the expected value of the *i*-th entry in the vector **X**. In other words,

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

Positive-definite matrix

From Wikipedia, the free encyclopedia

Not to be confused with Positive matrix and Totally positive matrix.

In linear algebra, a symmetric $n \times n$ real matrix M is said to be **positive definite** if the scalar $z^{\mathrm{T}} M z$ is strictly positive for every non-zero column vector z of n real numbers. Here z^{T} denotes the transpose of z.[1]

Example – Line Fitting

Consider the problem of line fitting or given the observations

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

where $w[n]$ is WGN, determine the CRLB for the slope B and the intercept A.

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

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

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

Inverse of the covariance matrix

where

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

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

Hence, the conditions for equality are satisfied and $[\hat{A}\hat{B}]^T$ is an efficient and therefore MVU estimator.

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

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

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

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

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

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

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

$$(1) * \frac{2(N-1)}{2} - (2) \quad get\hat{A}$$

$$(1) * \frac{(N-1)}{2} - (2) \quad get\hat{B}$$

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

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

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

quiz： Consider a generalization of line fitting problem as described in Example 3.7, termed polynomial or curve fitting. The data model is:

$$x[n]=\sum_{k=0}^{p-1} A_k n^k + w[n]$$

for n=0,1....N-1. As before,w[n] is WGN with variance σ^2.It is desired to estimate $\{A_0, A_1, \ldots, A_{p-1}\}$. Find the Fisher information matrix for this problem.

3.8 Vector Parameter CRLB for Transformations

Assume that it is desired to estimate $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ for $\mathbf{g}$, an r-dimensional function. Then, as shown in Appendix 3B

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

where, as before, $\geq \mathbf{0}$ is to be interpreted as positive semidefinite. $\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the $r \times p$ **Jacobian matrix** defined as

$$\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}$$