The Cramer-Rao Lower Bound*

Clayton Scott
Robert Nowak

The **Cramer-Rao Lower Bound** (CRLB) sets a lower bound on the variance of **any** unbiased estimator. This can be extremely useful in several ways:

1. If we find an estimator that achieves the CRLB, then we know that we have found an MVUB estimator!
2. The CRLB can provide a benchmark against which we can compare the performance of any unbiased estimator. (We know we're doing very well if our estimator is "close" to the CRLB.)
3. The CRLB enables us to rule-out impossible estimators. That is, we know that it is physically impossible to find an unbiased estimator that beats the CRLB. This is useful in feasibility studies.
4. The theory behind the CRLB can tell us if an estimator exists that achieves the bound.

1 Estimator Accuracy

Consider the likelihood function $p(x \mid \theta)$, where θ is a scalar unknown (parameter). We can plot the likelihood as a function of the unknown, as shown in Figure 1.

Figure 1

The more "peaky" or "spiky" the likelihood function, the easier it is to determind the unknown parameter.

Example 1
Suppose we observe

$$x = A + w$$

where $w \sim$ [U+EF3B] (σ, σ^2) and A is an unknown parameter. The "smaller" the noise w is, the easier it will be to estimate A from the observation x.

Suppose $A = 3$ and $\sigma = 1/3$.

*Version 1.4: May 27, 2004 3:38 pm GMT-5
†http://creativecommons.org/licenses/by/1.0

Image not finished

Figure 2

Given this density function, we can easily rule-out estimates of A greater than 4 or less than 2, since it is very unlikely that such A could give rise to out observation.

On the other hand, suppose $\sigma = 1$.

Image not finished

Figure 3

In this case, it is very difficult to estimate A. Since the noise power is larger, it is very difficult to distinguish A from the noise.

The key thing to notice is that the estimation accuracy of A depends on σ, which in effect determines the peakiness of the likelihood. The more peaky, the better localized the data is about the true parameter.

To quantify the notion, note that the peakiness is effectively measured by the negative of the second derivative of the log-likelihood at its peak, as seen in Figure 4.

Image not finished

Figure 4

Example 2

$$x = A + w$$

$$\log\left(p\left(x \mid A\right)\right) = -\left(\log\left(\sqrt{2\pi\sigma^2}\right)\right) - \frac{1}{2\sigma^2}(x-A)^2 \tag{1}$$

$$\frac{\partial}{\partial A}\left(\log\left(p\left(x \mid A\right)\right)\right) = \frac{1}{\sigma^2}(x-A)$$

$$-\left(\frac{\partial^2}{\partial A^2}\left(\log\left(p\left(x \mid A\right)\right)\right)\right) = \frac{1}{\sigma^2} \tag{2}$$

The curvature increases as σ^2 decreases (curvature=peakiness).

In general, the curavture will depend on the observation data; $-\left(\frac{\partial^2}{\partial\theta^2}\left(\log\left(p\left(x \mid A\right)\right)\right)\right)$ is a function of x. Therefore, an average measure of curvature is more appropriate.

$$-\left(E\left[\frac{\partial^2}{\partial\theta^2}\left(\log\left(p\left(x \mid \theta\right)\right)\right)\right]\right) \tag{3}$$

This average-out randomness due to the data and is a function of θ alone.

We are now ready to state the CRLB theorem.

Theorem 1: Cramer-Rao Lower Bound Theorem

Assume that the pdf $p(x \mid \theta)$ satisfies the "regularity" condition

$$\forall \theta : \left(E\left[\frac{\partial}{\partial \theta}\left(\log\left(p\left(x \mid \theta\right)\right)\right)\right] = 0\right)$$

where the expectation is take with respect to $p(x \mid \theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\sigma\left(\hat{\theta}\right)^2 \geq \frac{1}{-\left(E\left[\frac{\partial^2}{\partial \theta^2}\left(\log\left(p\left(x \mid \theta\right)\right)\right)\right]\right)} \tag{4}$$

where the derivative is evaluated at the true value of θ and the expectation is with respect to $p(x \mid \theta)$. Moreover, an unbiased estimator may be found that attains the bound for all θ if and only if

$$\frac{\partial}{\partial \theta}\left(\log\left(p\left(x \mid \theta\right)\right)\right) = I(\theta)\left(g(\theta) - \theta\right) \tag{5}$$

for some functions g and I.

The corresponding estimator is MVUB and is given by $\hat{\theta} = g(x)$, and the minimum variance is $\frac{1}{I(\theta)}$.

Example

$$x = A + w$$

where

$$w \sim \text{[U+EF3B]}\left(0, \sigma^2\right)$$

$$\theta = A$$

$$\forall A : \left(E\left[\frac{\partial}{\partial \theta}\left(\log(p)\right)\right] = E\left[\frac{1}{\sigma^2}(x - A)\right] = 0\right)$$

$$\text{CRLB} = \frac{1}{-\left(E\left[\frac{\partial^2}{\partial \theta^2}\left(\log(p)\right)\right]\right)} = \frac{1}{\frac{1}{\sigma^2}} = \sigma^2$$

Therefore, any unbiased estimator $\hat{A}$ has $\sigma\left(\hat{A}\right)^2 \geq \sigma^2$. But we know that $\hat{A} = x$ has $\sigma\left(\hat{A}\right)^2 = \sigma^2$. Therefore, $\hat{A} = x$ is the MVUB estimator.

NOTE:

$$\theta = A$$

$$I(\theta) = \frac{1}{\sigma^2}$$

$$g(x) = x$$

Proof:

First consider the reguarity condition:

$$E\left[\frac{\partial}{\partial \theta}\left(\log\left(p\left(x \mid \theta\right)\right)\right)\right] = 0$$

NOTE:

$$E\left[\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right]=\int\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)p\left(x\mid\theta\right)d\theta=\int\frac{\partial}{\partial\theta}\left(p\left(x\mid\theta\right)\right)d\theta$$

Now assuming that we can interchange order of differentiation and integration

$$E\left[\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right]=\frac{\partial}{\partial\theta}\left(\int p\left(x\mid\theta\right)d\theta\right)=\frac{\partial}{\partial\theta}\left(1\right)=0$$

So the regularity condition is satisfied whenever this interchange is possible[1]; *i.e.*, when derivative is well-defined, fails for uniform density.

Now lets derive the CRLB for a scalar parameter $\alpha=g\left(\theta\right)$, where the pdf is $p\left(x\mid\theta\right)$. Consider any unbiased estimator of α:

$$\hat{\alpha}\in E\left[\hat{\alpha}\right]=\alpha=g\left(\theta\right)$$

Note that this is equivalent to

$$\int\hat{\alpha}p\left(x\mid\theta\right)dx=g\left(\theta\right)$$

where $\hat{\alpha}$ is unbiased. Now differentiate both side

$$\int\hat{\alpha}\frac{\partial}{\partial\theta}\left(p\left(x\mid\theta\right)\right)dx=\frac{\partial}{\partial\theta}g\left(\theta\right)$$

or

$$\int\hat{\alpha}\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)p\left(x\mid\theta\right)dx=\frac{\partial}{\partial\theta}g\left(\theta\right)$$

Now, exmploiting the regularity condition,

$$\int\left(\hat{\alpha}-\alpha\right)\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)p\left(x\mid\theta\right)dx=\frac{\partial}{\partial\theta}g\left(\theta\right)\qquad(6)$$

since

$$\int\alpha\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)p\left(x\mid\theta\right)dx=\alpha E\left[\log\left(p\left(x\mid\theta\right)\right)\right]=0$$

Now apply the **Cauchy-Schwarz inequality** to the integral above (6):

$$\left(\frac{\partial}{\partial\theta}g\left(\theta\right)\right)^{2}=\left(\int\left(\hat{\alpha}-\alpha\right)\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)p\left(x\mid\theta\right)dx\right)^{2}$$

$$\left(\frac{\partial}{\partial\theta}g\left(\theta\right)\right)^{2}\leq\int\left(\hat{\alpha}-\alpha\right)^{2}p\left(x\mid\theta\right)dx\int\left(\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right)^{p\left(x\mid\theta\right)}d\theta$$

$\sigma(\hat{\alpha})^{2}$ is $\int\left(\hat{\alpha}-\alpha\right)^{2}p\left(x\mid\theta\right)dx$, so

$$\sigma(\hat{\alpha})^{2}\geq\frac{\left(\frac{\partial}{\partial\theta}g\left(\theta\right)\right)^{2}}{E\left[\left(\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right)^{2}\right]}\qquad(7)$$

Now we note that

$$E\left[\left(\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right)^{2}\right]=-\left(E\left[\frac{\partial^{2}}{\partial\theta^{2}}\left(\log\left(p\left(x\mid\theta\right)\right)\right)\right]\right)$$

[1] This is simply the **Fundamental Theorem of Calculus** applied to $p\left(x\mid\theta\right)$. So long as $p\left(x\mid\theta\right)$ is absolutely continuous with respect to the **Lebesgue measure**, this is possible.

Why? Regularity condition.

$$E\left[\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\right]=\int\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)p\left(x\ |\theta\right)dx=0$$

Thus,

$$\frac{\partial}{\partial\theta}\left(\int\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)p\left(x\ |\theta\right)dx\right)=0$$

or

$$\int\frac{\partial^2}{\partial\theta^2}\left(\log\left(p\left(x\ |\theta\right)\right)\right)p\left(x\ |\theta\right)+\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\frac{\partial}{\partial\theta}\left(p\left(x\ |\theta\right)\right)dx=0$$

Therefore,

$$-\left(E\left[\frac{\partial^2}{\partial\theta^2}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\right]\right)=\int\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)p\left(x\ |\theta\right)dx=E\left[\left(\frac{\partial}{\partial\theta}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\right)^2\right]$$

Thus, (7) becomes

$$\sigma(\hat{\alpha})^2\geq\frac{\left(\frac{\partial}{\partial\theta}g\left(\theta\right)\right)^2}{-\left(E\left[\frac{\partial^2}{\partial\theta^2}\left(\log\left(p\left(x\ |\theta\right)\right)\right)\right]\right)}$$

NOTE: If $g\left(\theta\right)=\theta$, then numerator is 1.

Example: DC Level in White Guassian Noise

$$\forall n,n\in\{1,\dots,N\}:\left(x_n=A+w_n\right)$$

where

$$w_n\sim\text{[U+EF3B]}\left(0,\sigma^2\right)$$

$$p\left(x\ |A\right)=\frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}}e^{-\left(\frac{1}{\sigma^2}\sum_{n=1}^{N}\left(\left(x_n-A\right)^2\right)\right)}$$

$$\frac{\partial}{\partial A}\left(\log\left(p\left(x\ |A\right)\right)\right)=\frac{\partial}{\partial A}\left(-\left(\log\left(\left(2\pi\sigma^2\right)^{\frac{N}{2}}\right)\right)-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(\left(x_n-A\right)^2\right)\right)=\frac{1}{\sigma^2}\sum_{n=1}^{N}\left(x_n-A\right)$$

$$E\left[\frac{\partial}{\partial A}\left(\log\left(p\left(x\ |A\right)\right)\right)\right]=0$$

$$\frac{\partial^2}{\partial A^2}\left(\log\left(p\left(x\ |A\right)\right)\right)=-\left(\frac{N}{\sigma^2}\right)$$

Therefore, the variance of any unbiased estimator satisfies:

$$\sigma\left(\hat{A}\right)^2\geq\frac{\sigma^2}{N}$$

The sample-mean estimator $\hat{A}=\frac{1}{N}\sum_{n=1}^{N}x_n$ attains this bound and therefore is MVUB.

Corollary 1:
When the CRLB is attained

$$\sigma\left(\hat{\theta}\right)^2=\frac{1}{I\left(\theta\right)}$$

where

$$I(\theta) = -\left(E\left[\frac{\partial^2}{\partial\theta^2}(\log(p(x\mid\theta)))\right]\right)$$

The quantity $I(\theta)$ is called **Fisher Information** that x contains about θ.

Proof:

By CRLB Theorem,

$$\sigma\left(\hat{\theta}\right)^2 = \frac{1}{-\left(E\left[\frac{\partial^2}{\partial\theta^2}(\log(p(x\mid\theta)))\right]\right)}$$

and

$$\frac{\partial}{\partial\theta}(\log(p(x\mid\theta))) = I(\theta)\left(\hat{\theta}-\theta\right)$$

This yields

$$\frac{\partial^2}{\partial\theta^2}(\log(p(x\mid\theta))) = \frac{\partial}{\partial\theta}I(\theta)\left(\hat{\theta}-\theta\right) - I(\theta)$$

which in turn yields

$$-\left(E\left[\frac{\partial^2}{\partial\theta^2}(\log(p(x\mid\theta)))\right]\right) = I(\theta)$$

So,

$$\sigma\left(\hat{\theta}\right)^2 = \frac{1}{I(\theta)}$$

The CRLB is not always attained.

Example 3: Phase Estimation

$$\forall n, n\in\{1,\dots,N\} : (x_n = A\cos(2\pi f_0 n+\phi)+w_n)$$

The amplitude and frequency are assumed known

$$w_n \sim \texttt{[U+EF3B]}\left(0,\sigma^2\right)$$

idd.

$$p(x\mid\phi) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}}e^{-\left(\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - A\cos(2\pi f_0 n+\phi))\right)}$$

$$\frac{\partial}{\partial\phi}(\log(p(x\mid\phi))) = \left(-\left(\frac{A}{\sigma^2}\right)\right)\sum_{n=1}^{N}\left(x_n\sin(2\pi f_0 n+\phi) - \frac{A}{2}\sin(4\pi f_0 n+\phi)\right)$$

$$\frac{\partial^2}{\partial\phi^2}(\log(p(x\mid\phi))) = \left(-\left(\frac{A}{\sigma^2}\right)\right)\sum_{n=1}^{N}(x_n\cos(2\pi f_0 n+\phi) - A\cos(2\pi f_0 n+2\phi))$$

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

Since $I(\phi) = -\left(E\left[\frac{\partial^2}{\partial\phi^2}(\log(p(x\mid\phi)))\right]\right)$,

$$I(\phi) \approx \frac{NA^2}{2\sigma^2}$$

because $\forall f_0, 0<f_0<k : \left(\frac{1}{N}\sum(\cos(4\pi f_0 n)) \approx 0\right)$ Therefore,

$$\sigma\left(\hat{\phi}\right)^2 \geq \frac{2\sigma^2}{NA^2}$$

In this case, it can be shown that there does not exist a g such that

$$\frac{\partial}{\partial\phi}\left(\log\left(p\left(x\mid\phi\right)\right)\right)\neq I\left(\phi\right)\left(g\left(x\right)-\phi\right)$$

Therefore, an unbiased phase estimator that attains the CRLB does not exist.

However, a MVUB estimator may still exist–only its variance will be larger than the CRLB.

2 Efficiency

An estimator which is unbiased and attains the CRLB is said to be **efficient**.

Example 4

Sample-mean estimator is efficient.

Example 5

Supposed three unbiased estimators exist for a param θ.

Figure 5

Figure 6

Example 6: Sinusoidal Frequency Estimation

$$\forall f_0, 0<f_0<1/2:\left(s_n\left(f_0\right)=A\cos\left(2\pi f_0 n+\phi\right)\right)$$

$$\forall n, n\in\{1,\dots,N\}:\left(x_n=s_n\left(f_0\right)+w_n\right)$$

A and ϕ are known, while f_0 is unknown.

$$\sigma\left(\hat{f_0}\right)^2\geq\frac{\sigma^2}{A^2\sum_{n=1}^{N}\left(\left(2\pi n\sin\left(2\pi f_0 n+\phi\right)\right)^2\right)}$$

Suppose $\frac{A^2}{\sigma^2}=1$ (SNR), where $N=10$ and $\phi=0$.

Figure 7: Some frequencies are easier to estimator (lower CRLB, but not necessarily just lower bound) than others.

3 CRLB for Vector Parameter

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}$$

$\hat{\theta}$ is unbiased, *i.e.*,

$$\forall i, i \in \{1, \dots, p\} : \left(E\left[\hat{\theta}_i\right] = \theta_i \right)$$

4 CRLB

$$\sigma\left(\hat{\theta}_i\right)^2 \geq \left[(I(\theta))^{-1} \right]_{ii}$$

where

$$\forall i \wedge j : \left([I(\theta)]_{ij} = -\left(E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} (\log(p(\mathbf{x} \mid \theta))) \right] \right) \right)$$

$I(\theta)$ is the **Fisher Information Matrix**.

Theorem 2: Cramer-Rao Lower Bound - Vector Parameter
Assume the pdf $p(\mathbf{x} \mid \phi)$ satisfies the "regularity" condition

$$\forall \theta : \left(E\left[\frac{\partial}{\partial\theta} (\log(p(\mathbf{x} \mid \theta))) \right] = 0 \right)$$

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

$$C_{\hat{\theta}} - (I(\theta))^{-1} \geq 0$$

(meaning $C_{\hat{\theta}} - (I(\theta))^{-1}$ is p.s.d.) The Fisher Information matrix is

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

Furthermore, $\hat{\theta}$ attains the CRLB ($C_{\hat{\theta}} = (I(\theta))^{-1}$) iff

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

and

$$\hat{\theta} = \mathbf{g}(\mathbf{x})$$

Example: DC Level in White Guassian Noise

$$\forall n, n \in \{1, \dots, N\} : (x_n = A + w_n)$$

A is unknown and $w_n \sim$ [U+EF3B] $\left(0, \sigma^2\right)$, where σ^2 is unknown.

$$\theta = \begin{pmatrix} A \\ \sigma^2 \end{pmatrix}$$

$$\log\left(p\left(\mathbf{x} \mid \theta\right)\right) = -\left(\frac{N}{2}\log\left(2\pi\right)\right) - \frac{N}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=2}^{N}\left(\left(x_n - A\right)^2\right)$$

$$\frac{\partial}{\partial A}\left(\log\left(p\left(\mathbf{x} \mid \theta\right)\right)\right) = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(x_n - A\right)$$

$$\frac{\partial}{\partial \sigma^2}\left(\log\left(p\left(\mathbf{x} \mid \theta\right)\right)\right) = -\left(\frac{N}{\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{n=1}^{N}\left(\left(x_n - A\right)^2\right)$$

$$\frac{\partial^2}{\partial A^2}\left(\log\left(p\left(\mathbf{x} \mid \theta\right)\right)\right) = -\left(\frac{N}{\sigma^2}\right) \rightarrow -\left(\frac{N}{\sigma^2}\right)$$

$$\frac{\partial^2}{\partial A \partial \sigma^2}\left(\log\left(p\left(\mathbf{x} \mid \theta\right)\right)\right) = -\left(\frac{1}{\sigma^4}\sum_{n=1}^{N}\left(x_n - A\right)\right) \rightarrow 0$$

$$\frac{\partial^2}{\partial {\sigma^2}^2}\left(\log\left(p\left(\mathbf{x} \mid \theta\right)\right)\right) = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{n=1}^{N}\left(\left(x_n - A\right)^2\right) \rightarrow -\left(\frac{N}{2\sigma^4}\right)$$

Which leads to

$$I\left(\theta\right) = \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}$$

$$\sigma\left(\hat{A}\right)^2 \geq \frac{\sigma^2}{N}$$

$$\sigma\left(\widehat{\sigma^2}\right)^2 \geq \frac{2\sigma^4}{N}$$

Note that the CRLB for $\hat{A}$ is the same whether or not σ^2 is known. This happens in this case due to the diagonal nature of the Fisher Information Matrix.

In general the Fisher Information Matrix is not diagonal and consequently the CRLBs will depend on other unknown parameters.

Glossary

Definition 1: idd
independent and identically distributed