STATISTICAL SIGNAL PROCESSING

PERFORMANCE LIMITS AND JOINT DETECTOR/ESTIMATORS

P.R. WHITE

INTRODUCTION

SINGLE PARAMETER CRAMÉR-RAO LOWER BOUNDS

MULTI-PARAMETER CRAMÉR-RAO LOWER BOUNDS

COMPOSITE HYPOTHESES

GENERALISED LIKELIHOOD RATIO TESTS

 © ISVR 1999

THE CRAMÉR-RAO LOWER BOUND: SINGLE PARAMETER

The Cramér-Rao Lower Bound (CRLB) is an absolute performance limit on *any* unbiased estimator.

The principle states that if the estimator is unbiased, then the variance of the estimator cannot be less than the CRLB.

Formally:
Any unbiased estimator of a single parameter has a variance that satisfies:

$$\mathrm{var}\left(\hat{\theta}\right) \geq \frac{1}{-E\left[\dfrac{\partial^2 \log\left\{p(\underline{x}\,|\,\theta)\right\}}{\partial\theta^2}\right]} = \frac{1}{E\left[\left(\dfrac{\partial \log\left\{p(\underline{x}\,|\,\theta)\right\}}{\partial\theta}\right)^2\right]}$$

The two expressions on the right hand side of the inequality are equivalent (although to show this it takes some lines of algebra).

Note that both forms of the CRLB involve differentiating the log of the likelihood function, $p(\underline{x}|\theta)$.

If one can find an unbiased estimator whose variance is equal to the CRLB then that is the best unbiased estimator and is referred to as *efficient*.

 © ISVR 1999

EXAMPLE 1

Consider the problem of estimating the parameter A in the following model:

$$x(n)=A+w(n)$$

where we have samples for n=1,2, …, N and $w(n)$ is white and Gaussian, with a variance σ^2.

A natural estimator for A is $\hat{A}=\frac{1}{N}\sum_{n=1}^{N} x(n)$.

Note that $E[\hat{A}]=A$, *i.e.* $\hat{A}$ is unbiased, and that $\mathrm{var}\left[\hat{A}\right]=\frac{\sigma^2}{N}$.

The CRLB for A is derived by first considering the likelihood:

$$p(\underline{x}\,|\,A)=\frac{1}{(2\pi)^{N/2}\sigma^N}e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x(n)-A)^2}$$

Taking logs

$$L=\log\{p(\underline{x}\,|\,A)\}=-\frac{N}{2}\log\{2\pi\sigma^2\}-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x(n)-A)^2$$

Differentiating w.r.t. A:

$$\frac{\partial L}{\partial A}=\frac{1}{\sigma^2}\sum_{n=1}^{N}(x(n)-A)$$

$$\frac{\partial^2 L}{\partial A^2}=\frac{1}{\sigma^2}\sum_{n=1}^{N}-1=-\frac{N}{\sigma^2}$$

 © ISVR 1999

EXAMPLE 1 (CONT'D)

The CRLB is

$$\mathrm{var}\left(\hat{A}\right) \geq \frac{1}{-E\left[\dfrac{\partial^2 L}{\partial A^2}\right]} = \frac{\sigma^2}{N}$$

Thus the estimator $\hat{A} = \frac{1}{N}\sum_{n=1}^{N} x(n)$ has a variance which equals the CRLB. This allows us to be sure that there are no (unbiased) estimators better than $\hat{A}$.

Note that since $\hat{A}$ satisfies

$$\frac{\partial L}{\partial A} = \frac{1}{\sigma^2}\sum_{n=1}^{N} \left(x(n) - A\right) = 0$$

Then $\hat{A}$ is the ML estimate of *A*.

 © ISVR 1999

CRLB: MULTIPLE PARAMETERS

The CRLB presented so far applies when there is only one unknown parameter to be estimated from the data.

If there are K unknown parameters, in a vector $\underline{\theta}$, then the CRLB is given by

$$\mathrm{var}\left(\hat{\theta}(k)\right) \geq \left\{ I(\underline{\theta})^{-1} \right\}_{k,k}$$

where the matrix $I(\underline{\theta})$ is referred to as the *Fisher information matrix* whose i^{th}, j^{th} element is given by

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

$\theta(k)$ is the kth element of $\underline{\theta}$.

So that to compute the CRLB in such an example one:

- Computes the Fisher information matrix
- Calculates its inverse
- The CRLB for the kth parameter is the kth term on the diagonal of this inverse.

Note that if K=1 the above result reverts to the original scalar CRLB.

 © ISVR 1999

EXAMPLE 2

Fitting a straight line to data. Consider a model of the form:

$$x(n)=A+Bn+w(n)$$

where $w(n)$ is additive, Gaussian, white noise and A and B are the two parameters to be estimated, $\underline{\theta}=[A,B]^t$.

The likelihood and log likelihood functions are:

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

$$L = \log\left\{\, p(\underline{x}\,|\,\underline{\theta})\right\} = -\frac{N}{2}\log\left\{2\pi\sigma^2\right\} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-A-Bn\right)^2$$

The derivatives we require are:

$$\frac{\partial L}{\partial A} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(x(n)-A-Bn\right) \qquad \frac{\partial L}{\partial B} = \frac{1}{\sigma^2}\sum_{n=1}^{N} n\left(x(n)-A-Bn\right)$$

$$\frac{\partial^2 L}{\partial A^2} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(-1\right) = -\frac{N}{\sigma^2} \qquad \frac{\partial^2 L}{\partial B^2} = \frac{1}{\sigma^2}\sum_{n=1}^{N} n\left(-n\right) = -\frac{1}{\sigma^2}\sum_{n=1}^{N} n^2$$

$$\frac{\partial^2 L}{\partial A \partial B} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(-n\right) = -\frac{1}{\sigma^2}\sum_{n=1}^{N} n$$

 © ISVR 1999

EXAMPLE 2 (CONT'D)

The Fisher information matrix is:

$$I\left(\underline{\theta}\right)=\frac{1}{\sigma^2}\begin{bmatrix} N & \sum_{n=1}^{N} n \\ \sum_{n=1}^{N} n & \sum_{n=1}^{N} n^2 \end{bmatrix}$$

Using the facts that:

$$\sum_{n=1}^{N} n=\frac{N(N+1)}{2} \qquad \text{and} \quad \sum_{n=1}^{N} n^2=\frac{N(N+1)(2N+1)}{6}$$

Then:

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

This is inverted to give:

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

The 2 CRLBs are

$$\text{var}\left(\hat{A}\right)\geq\frac{2\left(2N+1\right)}{N\left(N-1\right)}\sigma^2$$

$$\text{var}\left(\hat{B}\right)\geq\frac{12}{N\left(N^2-1\right)}\sigma^2$$

 © ISVR 1999

DISCUSSION ON EXAMPLE 2

If B is known then this problem reverts to essentially that discussed in example 1, where the CRLB for A was found to be σ^2/N.

The addition of the extra parameter B is seen to increase the uncertainty in estimating A. It is generally true that the CRLB increases as parameters are added to a problem.

Note that the CRLB for B is much smaller than that for A. The CRLB for B behaves like $1/N^3$, whereas the CRLB for A behaves like $1/N$.

Hence we predict that one will be able to estimate the slope of a graph with greater certainty than the intercept.

 © ISVR 1999

EXAMPLE 3: ESTIMATION OF SINE WAVE

The signal model is

$$x(n) = A\cos(2\pi f_0 n + \phi) + w(n), \quad n=0,1,\ldots, N\text{-}1$$

where $w(n)$ is Gaussian white noise. This model has 3 parameters: amplitude A, frequency f_0 and phase ϕ, so that $\underline{\theta}=[A, f_0, \phi]^t$

The log likelihood function is given by:

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

One can show that the Fisher information matrix satisfies:

$$I\left(\underline{\theta}\right) \approx \frac{1}{\sigma^2}\begin{bmatrix} \frac{N}{2} & 0 & 0 \\ 0 & 2A^2\pi^2\sum_{n=0}^{N-1} n^2 & \pi A^2\sum_{n=0}^{N-1} n \\ 0 & \pi A^2\sum_{n=0}^{N-1} n & \frac{NA^2}{2} \end{bmatrix}$$

Upon inversion one can show that the CRLBs are:

$$\text{var}\left\{\hat{A}\right\} \geq \frac{2\sigma^2}{N}, \ \text{var}\left\{\hat{f}_0\right\} \geq \frac{12}{SNR\, 4\pi^2 N(N^2-1)}, \ \text{var}\left\{\hat{\phi}\right\} \geq \frac{2(2N-1)}{SNR\, N(N+1)}$$

where $SNR=A/(2\sigma^2)$.

 © ISVR 1999

DISCUSSION OF EXAMPLE 3

The CRLB for the amplitude is similar in form to that found for the offset parameter in example 1. There is a factor of 2 difference; this arises because the mean squared value of a sine wave is ½.

The Fisher information has zeros along the first row and column, except in the diagonal position. This implies that the estimation of amplitude is unaffected by knowledge of the phase or frequency.

If the phase and frequency were known, then the CRLB for amplitude remains as $2\sigma^2/N$. This leads to the surprising result the estimation of the amplitude of a sine wave is not helped by knowing the sine wave's frequency.

The CRLB for frequency is inversely proportional to the SNR.

This implies that for sufficiently high SNRs one can gain arbitrarily good estimates of frequency.

This implies that the ability to locate the frequency of a sinusoid need not be controlled by the duration of the data segment, as conventional Fourier analysis predicts.

 © ISVR 1999

JOINT DETECTION AND ESTIMATION PROBLEMS (COMPOSITE TESTS)

Recap: To detect a signal the Likelihood Ratio (LR) test can be used. It is defined as:

$$\frac{p(\underline{x}\,|\,H_1)}{p(\underline{x}\,|\,H_0)} > \frac{\{C_{10}\text{-}C_{00}\}P(H_0)}{\{C_{01}\text{-}C_{11}\}P(H_1)}$$

where $p(\underline{x}|H_0)$ is the likelihood function in the absence of a signal and $p(\underline{x}|H_1)$ is the likelihood function assuming a signal is present.

We have assumed that under H_1 the signal is precisely known.

Commonly one seeks to detect a signal in which some of the parameters are unknown. This lack of knowledge may extend to the noise, *i.e.* to H_0, as well.

e.g. one may seek to construct a detector of a sinusoid in noise but not have knowledge of the frequency of the sinusoid *a priori*.

Under such circumstances one is faced with, potentially, having to estimate the unknown model parameters and then to perform detection.

These two steps can be merged to form a *Generalised Likelihood Ratio* (GLR) test.

 © ISVR 1999

GENERALISED LIKELIHOOD RATIO TESTS

The principle of a GLR test is:

- Estimate the unknown model parameters using maximum likelihood. This may have to be conducted under both H_0 and H_1.
- Apply a LR test as if the estimated parameters were the actual parameters.

GLR detectors do not in general achieve optimal performance but do tend to result in reasonable performance.

A GLR test can be written as:

$$\frac{p\left(\underline{x} \mid \underline{\theta}_1\right)}{p\left(\underline{x} \mid \underline{\theta}_0\right)} > \lambda$$

where $\underline{\theta}_0$ are the parameters associated with the null hypothesis estimated, via maximum likelihood, and $\underline{\theta}_1$ are the parameters associated with the alternative hypothesis, again estimated via ML.

 © ISVR 1999

EXAMPLE 4

Consider the following (familiar) model

$$H_0: \quad x(n)=w(n)$$

$$H_1: \quad x(n)=w(n)+A$$

for n=1,2,…,N, we shall assume that A is unknown.

Step 1) Find the ML estimate of A.

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

The log likelihood function, L, is:

$$L=-N\log\left\{2\pi\sigma^2\right\}/2-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-A\right)^2$$

The ML estimate is given by:

$$\frac{\partial L}{\partial A}=\frac{1}{\sigma^2}\sum_{n=1}^{N}\left(x(n)-A\right)=0 \quad \Rightarrow \quad \hat{A}=\frac{1}{N}\sum_{n=1}^{N}x(n)$$

Step 2) The LR test is (see example 2 in detection notes)

$$\frac{1}{N}\sum_{n=1}^{N}x(n)>\frac{\sigma^2}{AN}\log\left\{\lambda\right\}+\frac{A}{2}$$

Substituting $A=\hat{A}$ and rearranging gives:

$$\left(\frac{1}{N}\sum_{n=1}^{N}x(n)\right)^2>\frac{2\sigma^2}{N}\log\left\{\lambda\right\}$$

 © ISVR 1999

EXAMPLE 5

Consider a model:

$$H_0: \quad x(n)=w(n)$$

$$H_1: \quad x(n) = A\cos(2\pi f_0 n + \phi) + w(n)$$

where the phase, ϕ, is unknown and n=1,2,…,N.

Step 1) To find the ML estimator of ϕ.

$$p(\underline{x} \mid H_1) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\sum_{n=1}^{N} \left(x(n)-A\cos(2\pi f_0 n+\phi)\right)^2/(2\sigma^2)}$$

The log likelihood is:

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

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

This implies that:

$$\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n + \phi) = A\sum_{n=1}^{N} \cos(2\pi f_0 n + \phi)\sin(2\pi f_0 n + \phi)$$

The RHS of this equation is approximately zero, if f_0 is not too close to 0 or ½. Expanding the LHS using standard trig. identities gives:

$$\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n)\cos(\phi) = -\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n)\sin(\phi)$$

Hence

$$\phi = -\tan^{-1}\left\{\frac{\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n)}{\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n)}\right\}$$

 © ISVR 1999

EXAMPLE 5 (CONT'D)

Step 2) the LR test is simply a special case of the matched filter:

$$A\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n+\phi) > \sigma^2 \log\{\lambda\} + \frac{1}{2}\sum_{n=1}^{N} s(n)^2$$

The second term on the RHS can be approximated by $NA^2/4$.

Using the trig. identities again we can write:

$$\cos(\phi)\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n) - \sin(\phi)\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n) > \frac{\sigma^2}{A}\log\{\lambda\} + \frac{AN}{4}$$

Thus the GLR test simply consists of performing the above test with ϕ computed from the previous page, but

$$\tan(\phi) = \frac{\sin(\phi)}{\cos(\phi)} = \frac{-\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n)}{\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n)}$$

So that

$$\sin(\phi) = -\frac{1}{q}\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n)$$

and

$$\cos(\phi) = \frac{1}{q}\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n)$$

where q is a constant, such that

$$q^2 = \left(\sum_{n=1}^{N} x(n)\sin(2\pi f_0 n)\right)^2 + \left(\sum_{n=1}^{N} x(n)\cos(2\pi f_0 n)\right)^2$$

 © ISVR 1999

EXAMPLE 5 (CONT'D)

The GLR detector becomes

$$q\cos(\phi)^2 + q\sin(\phi)^2 = q > \frac{\sigma^2}{A}\log\{\lambda\} + \frac{AN}{4}$$

Recalling the definition of q we see that the final detector is phase insensitive, or *incoherent*.

It can be realised by:

- Applying a matched filter, whose phase is selected arbitrarily.
- The output of the matched filter is then squared and averaged. This process being referred to as envelope detection.

The performance of the incoherent detector differs little from the ideal matched filter.

 © ISVR 1999