ECSE 6520: Estimation and Detection Theory

Cramer-Rao Lower Bound

Class Notes - 8

February 17th, 2014

Contents

1 Importance of Cramer-Rao Lower Bound

- CRLB is the minimum achievable variance for an unbiased estimator. Hence, if an unbiased estimator achieves CRLB, it is MVUE. (But, the opposite is not necessarily true.)
- CRLB provides a benchmark against which we can check the performance of any estimator.
- The CRLB can be defined for both random and non-random parameters. However the CRLB is more useful for non-random parameters as it can be used to establish optimality or near optimality of an unbiased candidate estimator.
- The theory behind CRLB tells us exactly when CRLB is achievable.

2 Fisher Information Matrix and Cramer-Rao Lower Bound

For the rest of this topic, we will make the following assumptions:

- Θ is an open subset in $\mathbb{R}^p$.
- $f_\theta(x)$ is smooth and differentiable in θ.

Definition 1 ***Score Function*** **-** *Let $f_\theta(x)$ be the pdf of X and let $f_\theta(x)$ be smooth and differentiable in θ. We call*

$$s(\theta, x) = \frac{\partial}{\partial\theta} \ln f_\theta(x)$$

the score function.

Note that $\ln f_\theta(x)$ is referred to as the *log-likelihood function*.

Definition 2 ***Fisher Information Matrix*** **-** *Let $f_\theta(x)$ be the pdf of X and let $f_\theta(x)$ be smooth and differentiable in θ. We call*

$$\mathbf{F}(\theta) := E[(\frac{\partial}{\partial\theta} \ln f_\theta(x))(\frac{\partial}{\partial\theta} \ln f_\theta(x))^T]$$

the Fisher information matrix.

Theorem 1 ***Cramer-Rao Lower Bound*** **-** *Let $X \in \mathcal{X} \subseteq \mathbb{R}^N$, $\theta \in \Theta$ and let Θ be an open subset of $\mathbb{R}^p$. Let the pdf of X, f_θ, be a smooth differentiable function of θ and $\hat{\theta}$ be an unbiased estimate of θ. Assume that the covariance of $\hat{\theta}$ and the Fisher information matrix $\mathbf{F}(\theta)$ are non-singular. Then*

$$Cov(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \geq \mathbf{F}^{-1}(\theta).$$

Furthermore, an unbiased estimate is efficient, that is,

$$E[\hat{\theta}] = \theta$$

$$Cov(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] = \mathbf{F}^{-1}(\theta)$$

if and only if

$$\mathbf{F}(\theta)(\hat{\theta} - \theta) = s(\theta, x).$$

The CRLB for the scalar parameter can be stated as follows:

Theorem 2 ***CRLB for the scalar case*** **-** *Let the pdf of X, $f_\theta(x)$ be a smooth differentiable function of θ. If $\hat{\theta}$ is an unbiased estimate of θ, then*

$$Var(\hat{\theta}) \geq \frac{1}{\mathbf{F}(\theta)}$$

where $\mathbf{F}(\theta) = E[(\frac{\partial}{\partial\theta} \ln f_\theta(x))^2]$ is the Fisher information.

Some Facts relevant to the proof of CRLB theorem

- Note that $X \in \mathcal{X}$ can be discrete or continuous.

- $Cov(\hat{\theta}) \geq \mathtt{F}^{-1}(\theta)$ means $Cov(\hat{\theta}) - \mathtt{F}^{-1}(\theta)$ is non-negative definite.

- The score function $s(\theta, x)$ has zero mean, i.e.,

$$E[\frac{\partial}{\partial \theta} \ln f_\theta(x)] = 0.$$

  (This is true only if a certain integration and differentiation operators can be interchanged. For most pdf's this is true.)

- The covariance of the score function

$$\mathtt{F}(\theta) := E[s(\theta, x)s(\theta, x)^T] = E[(\frac{\partial}{\partial \theta} \ln f_\theta(x))(\frac{\partial}{\partial \theta} \ln f_\theta(x))^T]$$

  is equal to the Fisher information matrix.

- The Fisher information matrix $\mathtt{F}(\theta)$ can be alternatively expressed as

$$\mathtt{F}(\theta) = -E[\frac{\partial}{\partial \theta}(\frac{\partial}{\partial \theta} \ln f_\theta(x))^T].$$

- The ijth element of the Fisher information matrix is given by

$$\begin{aligned} \mathtt{F}_{ij} &= E[(\frac{\partial}{\partial \theta_i} \ln f_\theta(x))(\frac{\partial}{\partial \theta_j} \ln f_\theta(x))] \\ &= E[-\frac{\partial^2}{\partial \theta_i \theta_j} \ln f_\theta(x)]. \end{aligned}$$

- The cross-covariance between the score function and the error between the unbiased estimate and the parameter is identity, that is

$$E[s(\theta, x)(\hat{\theta}(x) - \theta)^T] = I.$$

- An efficient estimate does not always exist.

- If an efficient estimate exists, then it is MVUE.

3 Properties of CRLB

It is insightful to study the properties of the CRLB in the scalar case.

- **Property 1**: The Fisher information is a measure of the average curvature profile of the log likelihood function $\ln f_\theta(x)$ near θ.

- **Property 2**: Let $\mathtt{F}_N(\theta)$ be the Fisher information matrix for a sample of N i.i.d. random variables $X_1, X_2, ..., X_N$. Then,

$$F_N(\theta) = NF_1(\theta).$$

  As a result, $Var[\theta_i] = \mathcal{O}(1/N)$ is expected for "good" estimators.

- **Property 3**: *Asymptotic Efficiency* - If an estimator is asymptotically unbiased and its covariance decays with optimal rate

$$\lim_{N\to\infty} NCov(\hat{\theta}) = [\mathtt{F}_1(\theta)]^{-1},$$

  then, $\hat{\theta}$ is said to be asymptotically efficient.

  In particular, if $f_\theta(x) = a(\theta)b(x)\exp\{c^T(\theta)\tau(x)\}$ and $E[\tau(x)] = \theta$, then

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} \tau(x_i)$$

  is an unbiased and efficient estimator of θ.

- **Property 4**: Under the assumptions of the CRLB,

$$\sum_{ij} Cov(\hat{\theta}) \geq \sum_{ij} [\mathtt{F}^{-1}(\theta)]_{ij}$$

  and

$$Var(\hat{\theta}_i) \geq \mathtt{F}^{-1}(\theta)_{ii}.$$

  The vector CRLB implies scalar lower bounds on each component of $\hat{\theta}$.

- **Property 5**: If an estimator $\hat{\theta}$ satisfies

$$\nabla_\theta \ln f_\theta(x) = K_\theta(\hat{\theta} - \theta),$$

then we can immediately conclude the followings:

1. $\hat{\theta}$ is unbiased.
2. $\hat{\theta}$ is efficient and thus its components are UMVE.
3. The covariance of $\hat{\theta}$ is given by the inverse of the Fisher information matrix $\mathtt{F}(\theta)$.
4. $K_\theta = \mathtt{F}(\theta)$ and $Cov(\hat{\theta}) = K_\theta^{-1}$.