Tutorial - Performance Bounds for Parameter Estimation

Eyal Nitzan

Department of Signal Processing and Acoustics
Aalto University, Finland

September 29th, 2020

Outline

Outline

Parameter estimation

Fundamental goals:

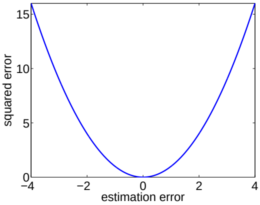
- Estimation methods: computationally manageable estimators under a chosen optimality criterion.
- Performance bounds: tools for performance analysis, system design, and feasibility study.
  - Performance analysis: we compare the performance of an estimator to the bound. Method for establishing optimality of an estimator.
  - System design: we investigate the bound's behavior under different conditions on our system.
  - Feasibility study: we study the optimal performance before implementing a specific estimator.

Parameter estimation

- We have a vector of random observations **y** generated from a probability distribution, which is known up to some unknown parameter.
- Parameter can be continuous, e.g. $\theta \in \mathbb{R}$, $\theta \in (0, 1)$.
  Example: the observations are generated from a Poisson distribution $\mathcal{P}(\theta)$, where the rate parameter $\theta > 0$ is unknown.
- Parameter can be discrete, e.g. $\theta \in \mathbb{Z}$, $\theta \in \{0, 1\}$.
  Example: the observations are generated from a Gaussian distribution $\mathcal{N}(\theta, 1)$, where the expectation parameter $\theta \in \mathbb{Z}$ is unknown.

Parameter estimation

- For simplicity, we will mostly consider a scalar unknown parameter.
- All of the results can be extended to multiple parameters.
- An estimator $\hat{\theta}(\mathbf{y})$ is designed to estimate θ as closely as possible, based on $\mathbf{y}$.
- How do we measure closeness?
- Define the estimation error, $\epsilon \triangleq \hat{\theta} - \theta$.
- Usually we measure closeness using the squared error ϵ^2.



Parameter estimation frameworks

Parameter estimation is divided into two main frameworks:

Non-Bayesian estimation:

- Deterministic unknown parameter.
- Statistical information: the observations' distribution, which is indexed by the parameter.
- If there exists pdf, then we have $f(\mathbf{y};\theta)$.
- Given observation vector $\mathbf{y}$, the function $f(\mathbf{y};\theta)$ is denoted as the likelihood function.
- The likelihood function expresses how likely an observation vector is for different values of θ.

Parameter estimation frameworks

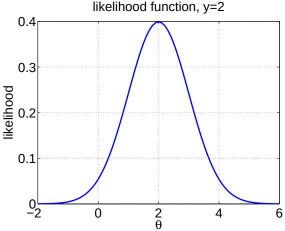
Non-Bayesian estimation (cont'd):

- A popular estimator is the maximum likelihood (ML) estimator, $\hat{\theta}_{\text{ML}}(\mathbf{y}) \triangleq \arg\max_{\theta} f(\mathbf{y};\theta)$.
- In many cases, it is convenient to use the log-likelihood function, $\log f(\mathbf{y};\theta)$.
- The natural logarithm is a strictly increasing function.
- We can maximize the log-likelihood function to obtain ML estimator.

Parameter estimation frameworks

Non-Bayesian estimation (cont'd):

- Example: $y = \theta + w$, $\theta \in \mathbb{R}$ is a deterministic signal that we want to estimate.
- $w \sim \mathcal{N}(0, 1)$ is random noise $\rightarrow y \sim \mathcal{N}(\theta, 1)$



- Maximum is obtained for $\theta = 2$. More generally, $\hat{\theta}_{\mathrm{ML}} = y$.

Parameter estimation frameworks

Non-Bayesian estimation (cont'd):

- How to evaluate the error of the ML estimator (or other non-Bayesian estimators)?
- We usually consider the mean-squared-error (MSE), $\mathrm{E}[(\hat{\theta}-\theta)^2;\theta]$, w.r.t. $f(\mathbf{y};\theta)$.
- In the expectation, integration is only w.r.t. $\mathbf{y}$.
- The non-Bayesian MSE is a function of θ.

Parameter estimation frameworks

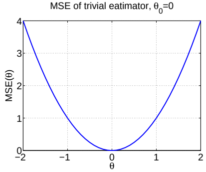
Bayesian estimation:

- Random unknown parameter.
- Statistical information: the observations' distribution given the parameter and the parameter prior distribution.
- If there exists pdfs, then we have $f(\mathbf{y}|\theta)$ and $f(\theta)$.
- We consider the MSE, $\mathrm{E}[(\hat{\theta}-\theta)^2]$, w.r.t. to the joint distribution of $\mathbf{y}$ and θ.
- Example: $y|\theta \sim \mathcal{N}(\theta, 1)$, θ is a random signal that we want to estimate with prior distribution $\mathcal{N}(0, 1)$.

Outline

Non-Bayesian estimation

- Main goal: derivation of uniformly best estimator that attains minimum MSE at any point in the parameter space.
- Find $\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta)^2; \theta] = \arg\min_{\hat{\theta}} \int_{\Omega_{\mathbf{y}}} (\hat{\theta} - \theta)^2 f(\mathbf{y}; \theta)\, \mathrm{d}\mathbf{y}$.
- Problem:
  - Non-Bayesian MSE depends on the parameter θ.
  - Minimization is performed for a fixed value of θ.
  - Unrestricted MSE minimization w.r.t. to the estimator at $\theta = \theta_0$, yields the trivial estimator $\hat{\theta} = \theta_0$.
  - A lower bound on the MSE is 0.



Not very useful

Non-Bayesian estimation

Question: How can we avoid trivial optimal estimator and zero lower bound?

Solution:

Mean-unbiasedness

- Let $b(\theta) \triangleq \mathrm{E}[\hat{\theta} - \theta; \theta]$ be the bias of an estimator $\rightarrow$ Function of θ.
- Let $\text{var}(\theta) \triangleq \mathrm{E}[(\hat{\theta} - \mathrm{E}[\hat{\theta}; \theta])^2; \theta]$ be the variance of an estimator $\rightarrow$ Function of θ.
- $\text{MSE}(\theta) = \text{var}(\theta) + b^2(\theta)$.
- A very common restriction is mean-unbiasedness, $b(\theta) = 0$ $\rightarrow$ expected value of estimator is equal to the parameter.
- The MSE of a mean-unbiased estimator is equal to its variance.
- A more general restriction is allowing specific bias function, $b(\theta)$, which is not necessarily zero. We will discuss it later.

MSE minimization under mean-unbiasedness

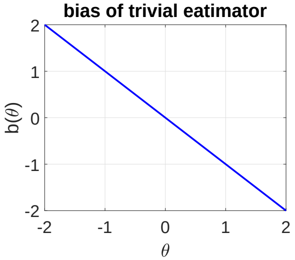
- Non-Bayesian MSE is a function of θ.
- We can try to find $\hat{\theta}$ that minimizes the MSE at a fixed $\theta = \theta_0$.
- In practice, we would like to characterize the optimal performance of estimators that are mean-unbiased for any parameter value.
- Uniform mean-unbiasedness: $b(\theta) = 0, \ \forall \theta \in \Omega_\theta$.
- We consider a constrained minimization problem at $\theta = \theta_0$:

$$\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0], \ \ \text{s.t. } b(\theta) = 0, \forall \theta \in \Omega_\theta$$

- Ω_θ can be a continuous set $\rightarrow$ uncountably infinite number of constraints $\rightarrow$ very difficult problem to solve.
- We need to relax the uniform mean-unbiasedness constraint.

Pointwise mean-unbiasedness

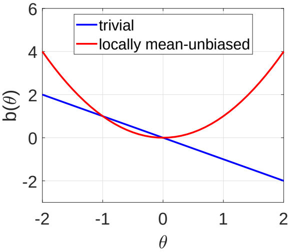
- We can consider only estimators that satisfy $b(\theta_0) = 0$.
- This is pointwise mean-unbiasedness. Can we avoid the trivial estimator?
- Consider the trivial estimator $\hat{\theta} = \theta_0$ for $\theta_0 = 0$:



- Pointwise mean-unbiasedness is satisfied by the trivial estimator $\hat{\theta} = \theta_0$.
- We need a more restrictive constraint than pointwise mean-unbiasedness.

Local mean-unbiasedness

- Local mean-unbiasedness in the vicinity of a fixed $\theta = \theta_0$: $b(\theta_0) = 0$, $b'(\theta_0) = 0$.
- Consider the trivial estimator $\hat{\theta} = \theta_0$ for $\theta_0 = 0$:



- The trivial estimator $\hat{\theta} = \theta_0$ does not satisfy this restriction!

Local mean-unbiasedness

- We can try to solve

$$\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0], \quad \text{s.t. } b(\theta_0) = 0, \; b'(\theta_0) = 0$$

- Note: a lower bound for locally mean-unbiased estimators will also be a lower bound for uniformly mean-unbiased estimators.
- For local mean-unbiasedness restriction, the solution is simple and a very useful performance bound...

Local mean-unbiasedness

- We can try to solve

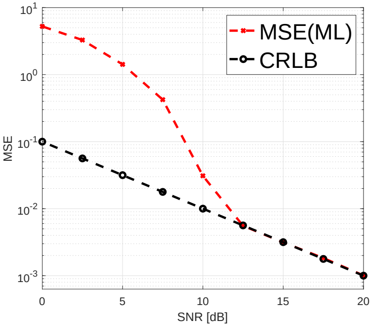
$$\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0], \quad \text{s.t. } b(\theta_0) = 0, \; b'(\theta_0) = 0$$

- Note: a lower bound for locally mean-unbiased estimators will also be a lower bound for uniformly mean-unbiased estimators.
- For local mean-unbiasedness restriction, the solution is simple and a very useful performance bound...

Cramer-Rao Lower Bound (CRLB)

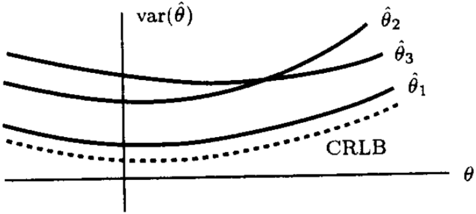
CRLB background

- Historically the first and currently the most popular lower bound on the MSE (variance) of mean-unbiased estimators.
- Characterizes the asymptotic performance of the maximum likelihood estimator, i.e. for high SNR or a large number of observations.



CRLB background

- Similar to the MSE, the CRLB is a function of θ. It is derived separately for each value of θ.



- How is it derived? We will consider two approaches:
  1. Using Cauchy-Schwartz inequality with a specific choice of auxiliary function.
  2. Directly solving the constrained minimization problem.

CRLB derivation at $\theta = \theta_0$

Theorem: Assume that

- $\log f(\mathbf{y};\theta)$ is twice differentiable at $\theta = \theta_0$.
- The order of integration w.r.t. $\mathbf{y}$ and differentiation w.r.t. θ can be interchanged at $\theta = \theta_0$.
- The Fisher information $J(\theta) \triangleq \mathrm{E}[l_{\mathbf{y}}^2(\theta);\theta]$ is nonzero at $\theta = \theta_0$, $l_{\mathbf{y}}(\theta) \triangleq \frac{\partial}{\partial\theta}\log f(\mathbf{y};\theta)$ is the score function.

Consider an estimator $\hat{\theta}$, which is locally mean-unbiased in the vicinity of $\theta = \theta_0$. Then,

$$\mathrm{E}[(\hat{\theta} - \theta_0)^2;\theta_0] \geq \mathrm{CRLB}(\theta_0) \triangleq \frac{1}{J(\theta_0)}$$

with equality *iff*

$$l_{\mathbf{y}}(\theta_0) = J(\theta_0)(\hat{\theta} - \theta_0).$$

CRLB Interpretation

Interpretation of CRLB:

- Derivative of a function $\rightarrow$ measure of function sensitivity to small change in its argument.
- Score function $\rightarrow$ derivative of the log-likelihood function.
- In the context of parameter estimation:
  high sensitivity of the log-likelihood function to the parameter
  $\rightarrow$ more information that we can use for estimation.
- The Fisher information can be viewed as squared norm of the score function.
- The CRLB is the reciprocal of the Fisher information.
- High sensitivity $\rightarrow$ high Fisher information $\rightarrow$ low CRLB $\rightarrow$
  We can better estimate the parameter.

CRLB derivation

Cauchy-Schwartz approach:

- An inner product between two random variables X and Y can be defined as $\mathrm{E}[XY]$.
- We can apply Cauchy-Schwartz inequality $\mathrm{E}^2[XY] \leq \mathrm{E}[X^2]\mathrm{E}[Y^2]$ with equality *iff* $X = cY, c \in \mathbb{R}$.
- In order to obtain an MSE lower bound, we can apply Cauchy-Schwartz inequality with $X = \hat{\theta} - \theta_0$ and a carefully chosen auxiliary function Y.
- We will choose, currently based on intuition, $Y = l_{\mathbf{y}}(\theta_0)$.

CRLB derivation

Cauchy-Schwartz approach (cont'd):

- Applying Cauchy-Schwartz inequality, we obtain

$$\mathrm{E}[(\hat{\theta}-\theta_0)^2;\theta_0] \geq \frac{\mathrm{E}[(\hat{\theta}-\theta_0)l_{\mathbf{y}}(\theta_0);\theta_0]}{J(\theta_0)}.$$

- Let's consider the numerator:

$$\begin{aligned}\mathrm{E}[(\hat{\theta}-\theta_0)l_{\mathbf{y}}(\theta_0);\theta_0] &= \mathrm{E}[(\hat{\theta}-\theta)l_{\mathbf{y}}(\theta);\theta]|_{\theta=\theta_0} \\ &= \int_{\Omega_{\mathbf{y}}}(\hat{\theta}-\theta)\frac{\frac{\partial}{\partial\theta}f(\mathbf{y};\theta)}{f(\mathbf{y};\theta)}f(\mathbf{y};\theta)\,\mathrm{d}\mathbf{y}|_{\theta=\theta_0} \\ &= \int_{\Omega_{\mathbf{y}}}(\hat{\theta}-\theta)\frac{\partial}{\partial\theta}f(\mathbf{y};\theta)\,\mathrm{d}\mathbf{y}|_{\theta=\theta_0}\end{aligned}$$

CRLB derivation

Cauchy-Schwartz approach (cont'd):

- Using the product rule for derivatives:

$$\mathrm{E}[(\hat{\theta}-\theta_0)l_{\mathbf{y}}(\theta_0);\theta_0] = \int_{\Omega_{\mathbf{y}}} \frac{\partial}{\partial\theta}\left((\hat{\theta}-\theta)f(\mathbf{y};\theta)\right) \mathrm{d}\mathbf{y}|_{\theta=\theta_0} - \int_{\Omega_{\mathbf{y}}} \left(\frac{\partial}{\partial\theta}(\hat{\theta}-\theta)\right) f(\mathbf{y};\theta)\,\mathrm{d}\mathbf{y}|_{\theta=\theta_0}$$

- Under the regularity assumptions:

$$\mathrm{E}[(\hat{\theta}-\theta_0)l_{\mathbf{y}}(\theta_0);\theta_0] = b'(\theta_0)+1$$

- For locally mean-unbiased estimators $b'(\theta_0)=0$:

$$\mathrm{E}[(\hat{\theta}-\theta_0)l_{\mathbf{y}}(\theta_0);\theta_0] = 1$$

CRLB derivation

Cauchy-Schwartz approach (cont'd):

- Finally, we obtain

$$\mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0] \geq \frac{1}{J(\theta_0)} \qquad \text{CRLB is derived!}$$

- What about the equality condition?

$$\hat{\theta} - \theta_0 = c(\theta_0) l_{\mathbf{y}}(\theta_0)$$

  $c(\theta_0)$ is a constant w.r.t. **y**

- Let's find $c(\theta_0)$:

$$\hat{\theta} - \theta = \theta_0 - \theta + c(\theta_0) l_{\mathbf{y}}(\theta_0)$$

CRLB derivation

Cauchy-Schwartz approach (cont'd):

- Taking expectation at θ:

$$\mathrm{E}[\hat{\theta} - \theta; \theta] = \theta_0 - \theta + c(\theta_0)\mathrm{E}[l_{\mathbf{y}}(\theta_0); \theta]$$

- Applying derivative at $\theta = \theta_0$:

$$\begin{aligned} 0 = b'(\theta_0) &= -1 + c(\theta_0)\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}[l_{\mathbf{y}}(\theta_0); \theta]|_{\theta=\theta_0} \\ &= -1 + c(\theta_0)J(\theta_0) \end{aligned}$$

- We get

$$c(\theta_0) = \frac{1}{J(\theta_0)}$$

Attaining CRLB

Efficient estimator:

- The estimator $\hat{\theta} = \theta + \frac{1}{J(\theta)} l_{\mathbf{y}}(\theta)$ attains the CRLB $\forall\theta$.
- This estimator is a function of $\theta \rightarrow$ not practical...
- In case $\hat{\theta} \neq \text{func}(\theta)$, then it is an efficient estimator with MSE equal to CRLB.
- The efficient estimator coincides with the maximum likelihood estimator.
- Example: Gaussian variance estimation
  $\mathbf{y} \sim \mathcal{N}(\mathbf{0}_N, \theta \mathbf{I}_N),\ \theta > 0$.
- Let's derive the CRLB...

CRLB for Gaussian variance estimation

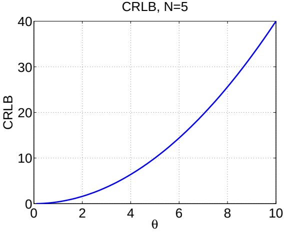
- The likelihood function is $f(\mathbf{y};\theta) = \frac{1}{(2\pi\theta)^{\frac{N}{2}}} e^{-\frac{\sum_{n=1}^{N} y_n^2}{2\theta}}$.
- Taking the natural logarithm:
  $\log f(\mathbf{y};\theta) = -\frac{N}{2}\log\theta - \frac{\sum_{n=1}^{N} y_n^2}{2\theta} + \text{const.}$
- Derivative w.r.t. θ:
  $l_{\mathbf{y}}(\theta) = -\frac{N}{2\theta} + \frac{\sum_{n=1}^{N} y_n^2}{2\theta^2} \rightarrow l_{\mathbf{y}}(\theta) = \frac{N}{2\theta^2}\left(\frac{1}{N}\sum_{n=1}^{N} y_n^2 - \theta\right)$.
- The equality condition of Cauchy-Schwartz is satisfied.
- The Fisher information is $J(\theta) = \frac{N}{2\theta^2}$
- We get an efficient estimator $\hat{\theta}_{\text{eff}} = \frac{1}{N}\sum_{n=1}^{N} y_n^2$ that attains $\text{CRLB}(\theta) = \frac{2\theta^2}{N}$.

CRLB for Gaussian variance estimation

- Let's make sure that we got the CRLB right.
- Taking the square of the score function:
  $l^2_{\mathbf{y}}(\theta) = \frac{N^2}{4\theta^4}\left(\frac{1}{N^2}\sum_{n=1}^{N}\sum_{k=1}^{N} y_n^2 y_k^2 - 2\theta\frac{1}{N}\sum_{n=1}^{N} y_n^2 + \theta^2\right)$
- Notice: $\mathrm{E}[y_n^2] = \theta$, $\mathrm{E}[y_n^2 y_k^2] = \begin{cases}\theta^2, & n \neq k \\ 3\theta^2 & n = k\end{cases}$.
- Taking the expected value of $l^2_{\mathbf{y}}(\theta)$:
  $J(\theta) = \frac{N^2}{4\theta^4}\left(\frac{1}{N^2}(3N\theta^2 + N(N-1)\theta^2) - 2\theta^2 + \theta^2\right) \rightarrow$
  $J(\theta) = \frac{N^2}{4\theta^4}\left(\frac{1}{N^2}(2N\theta^2 + N^2\theta^2) - \theta^2\right) = \frac{N^2}{4\theta^4}\frac{2\theta^2}{N} = \frac{N}{2\theta^2} \rightarrow$
  $\mathrm{CRLB}(\theta) = \frac{2\theta^2}{N}$, We got it right.

CRLB analysis for Gaussian variance estimation

- We obtained $\text{CRLB}(\theta) = \frac{2\theta^2}{N}$.
- As $N \to \infty$ the bound approaches zero.
- Estimation performance is better as the number of observations increases.
- What can we say about the bound dependence on θ:



- As θ increases, the bound increases as well.
- Variance of observations is high $\to$ observations are less “reliable”.
- Our estimation performance is worse.

CRLB alternative derivation

Constrained minimization approach:

- Going back to the choice of auxiliary function $Y = l_{\mathbf{y}}(\theta_0)$.
- How can we justify this choice?
- We consider a constrained minimization problem at $\theta = \theta_0$:

$$\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0]$$

$$\text{s.t. } b(\theta_0) = 0, b'(\theta_0) = 0$$

- Formulating a Lagrangian:

$$\int_{\Omega_{\mathbf{y}}} (\hat{\theta} - \theta_0)^2 f(\mathbf{y}; \theta_0)\,\mathrm{d}\mathbf{y} - 2\lambda_0 \int_{\Omega_{\mathbf{y}}} (\hat{\theta} - \theta_0) f(\mathbf{y}; \theta_0)\,\mathrm{d}\mathbf{y} - 2\lambda_1 \int_{\Omega_{\mathbf{y}}} ((\hat{\theta} - \theta_0) l_{\mathbf{y}}(\theta_0) - 1) f(\mathbf{y}; \theta_0)\,\mathrm{d}\mathbf{y}$$

CRLB alternative derivation

Constrained minimization approach (cont'd):

- completing the square of the Lagrangian:

$$\int_{\Omega_{\mathbf{y}}} \left(\hat{\theta} - \theta_0 - (\lambda_0 + \lambda_1 l_{\mathbf{y}}(\theta_0))\right)^2 f(\mathbf{y};\theta_0)\,\mathrm{d}\mathbf{y} + \text{extra terms}$$

- We obtain the minimizer $\hat{\theta}_{\text{opt}} = \theta_0 + \lambda_0 + \lambda_1 l_{\mathbf{y}}(\theta_0)$.
- From the constraint $b(\theta_0) = 0$ we get $\lambda_0 = 0$.
- From the constraint $b'(\theta_0) = 0$ we get $\lambda_1 = \frac{1}{J(\theta_0)}$.
- The optimal estimator is $\hat{\theta}_{\text{opt}} = \theta_0 + \frac{1}{J(\theta_0)} l_{\mathbf{y}}(\theta_0)$.
- This is a justification to the auxiliary function choice $Y = l_{\mathbf{y}}(\theta_0)$.
- Moreover, this is an alternative (less known) derivation of CRLB.

Computation of CRLB

- Sometimes, the parametric model is complicated.
- It is not simple to compute the Fisher information (or equivalently the CRLB).
- For the commonly assumed Gaussian observations, a very useful formula has been derived.
- Assume an observation vector $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}(\theta), \mathbf{C}(\theta))$.
- Slepian-Bangs formula:

$$J(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\mu}^T(\theta)\mathbf{C}^{-1}(\theta)\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\mu}(\theta) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{-1}(\theta)\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{C}(\theta)\mathbf{C}^{-1}(\theta)\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{C}(\theta)\right)$$

Computation of CRLB

- Example: Gaussian model with known variational coefficient.
- Commonly used in statistics, analytical chemistry, economics, etc.
- $\mathbf{y} \sim \mathcal{N}(\theta \mathbf{1}_N, \theta^2 \mathbf{I}_N)$.
- We compute $\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\mu}(\theta) = \mathbf{1}_N$, $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{C}(\theta) = 2\theta \mathbf{I}_N$, $\mathbf{C}^{-1}(\theta) = \frac{1}{\theta^2}\mathbf{I}_N$.
- We obtain the Fisher information and the CRLB:

$$J(\theta) = \frac{1}{\theta^2}\mathbf{1}_N^T \mathbf{1}_N + \frac{2}{\theta^2}\mathrm{Tr}(\mathbf{I}_N) = \frac{3N}{\theta^2} \rightarrow \mathrm{CRLB}(\theta) = \frac{\theta^2}{3N}$$

- Estimation performance is better as N increases.
- Estimation performance is worse as $|\theta|$ increases.

Fisher information for statistically independent observations

- A more common form of the Fisher information is $J(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{y};\theta);\theta\right]$.
- Assume that we have K statistically independent observation vectors $\mathbf{y}_k,\ k = 1,\ldots,K$.
- Denote by $J_k(\theta)$ the Fisher information resulting from observation vector $\mathbf{y}_k$.
- The overall Fisher information is $J(\theta) = \sum_{k=1}^{K} J_k(\theta)$.
- Explanation: let $\mathbf{y} \triangleq [\mathbf{y}_1^T,\ldots,\mathbf{y}_K^T]^T$.
- Due to statistical independency $f(\mathbf{y};\theta) = \prod_{k=1}^{K} f(\mathbf{y}_k;\theta)$.

Fisher information for statistically independent observations

- The Fisher information

$$J(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{y};\theta);\theta\right] = \sum_{k=1}^{K}\left(-\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{y}_k;\theta);\theta\right]\right)$$
$$= \sum_{k=1}^{K} J_k(\theta)$$

- For i.i.d. observation vectors, $J_k(\theta) = J_1(\theta),\ k = 1, \ldots, K$.
  - The Fisher information $J(\theta) = KJ_1(\theta)$, i.e. $J(\theta) = O(K)$.
  - $\mathrm{CRLB}(\theta) = \frac{1}{KJ_1(\theta)} \underset{K\to\infty}{\longrightarrow} 0$.

CRLB derivation for $g(\theta)$

- We are interested to estimate a differentiable function $g(\theta)$ using a mean-unbiased estimator $\hat{g}$.
- The MSE of $\hat{g}$ is lower bounded by the following CRLB:

$$CRLB_g(\theta) = \frac{(g'(\theta))^2}{J(\theta)}$$

- Equality is obtained *iff*

$$\hat{g} - g(\theta) = \frac{1}{J(\theta)} g'(\theta) l_{\mathbf{y}}(\theta)$$

- For $g(\theta) = \theta$ we return to the conventional CRLB and efficient estimator.

CRLB derivation for $g(\theta)$

- Example: $y = e^\theta + w$, $\theta \in \mathbb{R}$.
- $g(\theta) = e^\theta$ is a deterministic signal that we want to estimate.
- We are not interested in the actual value of θ.
- $w \sim \mathcal{N}(0, \sigma^2)$ is random noise, σ^2 is known.
- The Fisher information is $J(\theta) = \frac{e^{2\theta}}{\sigma^2}$, $g'(\theta) = e^\theta$.
- In this case $\hat{g} = y$ is an efficient estimator that attains $\text{CRLB}_g(\theta) = \sigma^2$.
- There is no efficient estimator of θ.

Biased CRLB

- We allow the estimator to have specific bias function $b(\theta)$ with derivative $b'(\theta)$.
- The variance of such estimator is lower bounded

$$\text{var}(\theta) \geq \frac{(1 + b'(\theta))^2}{J(\theta)}$$

- The bound is attained *iff* $l_{\mathbf{y}}(\theta) = \frac{J(\theta)}{1+b'(\theta)}(\hat{\theta} - \theta - b(\theta))$.
- MSE is a direct measure of estimation error.
- Usually, we are more interested in $\text{MSE}(\theta) = \text{var}(\theta) + b^2(\theta)$.
- We obtain

$$\text{MSE}(\theta) \geq \frac{(1 + b'(\theta))^2}{J(\theta)} + b^2(\theta)$$

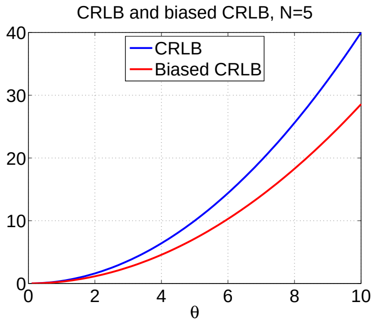
Biased CRLB

- What can we gain by allowing biased estimators?
- In many estimation problems, there is a bias-variance tradeoff.
- low variance $\rightarrow$ high bias, low bias $\rightarrow$ high variance.
- A biased estimator may have a uniformly lower MSE than CRLB.
- Example: Gaussian variance estimation

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}_N, \theta \mathbf{I}_N), \;\; \text{CRLB}(\theta) = \frac{2\theta^2}{N}$$

Biased CRLB

- In this example, we can find a biased-efficient estimator $\hat{\theta}_{\text{b-eff}} = \frac{1}{N+2}\sum_{n=1}^{N} y_n^2$ with bias function $b(\theta) = -\frac{2}{N+2}\theta$.



- The MSE of this estimator is $\frac{2\theta^2}{N+2} < \text{CRLB}(\theta),\ \forall\theta > 0$.

An alternative to CRLB

- In some cases, the regularity assumptions of the CRLB are not satisfied.
- For example, the likelihood function may not be differentiable.
- Hammersley, Chapman, and Robbins (1950) proposed a less restrictive bound, HCRLB, for mean-unbiased estimators:

$$\text{HCRLB}(\theta) = \sup_h \frac{h^2}{\mathrm{E}\left[\left(\frac{f(\mathbf{y};\theta+h)}{f(\mathbf{y};\theta)} - 1\right)^2 ; \theta\right]}$$

- $\theta + h$ is called a test-point.

An alternative to CRLB

- In a similar manner to CRLB, this bound is obtained by using Cauchy-Schwartz inequality.
- It uses an approximation of the score function $\frac{f(\mathbf{y};\theta+h)-f(\mathbf{y};\theta)}{hf(\mathbf{y};\theta)}$.
- Differentiability of the likelihood function is not required.
- $\lim\limits_{h\to 0} \frac{f(\mathbf{y};\theta+h)-f(\mathbf{y};\theta)}{hf(\mathbf{y};\theta)} = l_{\mathbf{y}}(\theta)$.
- Consequently, for $h \to 0$ HCRLB$(\theta) \to$ CRLB(θ).
- HCRLB is tighter than or equal to CRLB.

HCRLB example

- Example: $y = \theta + w$, $\theta \in \mathbb{Z}$ is a deterministic integer signal that we want to estimate, $w \sim \mathcal{N}(0, \sigma^2)$ is random noise.
- Parameter is discrete, likelihood function is not differentiable.
- CRLB cannot be used. HCRLB can be used instead:

$$\text{HCRLB} = \frac{1}{e^{1/\sigma^2} - 1}$$

- The HCRLB is lower than the CRLB (for continuous parameter).
- The discrete nature of the parameter is side information $\rightarrow$ MSE is reduced.

MSE minimization under uniform mean-unbiasedness

- Let's return to the uniform mean-unbiasedness restriction.
- We were interested to solve:
  $\hat{\theta}_{\text{opt}} = \arg\min_{\hat{\theta}} \mathrm{E}[(\hat{\theta} - \theta_0)^2; \theta_0], \quad \text{s.t. } b(\theta) = 0, \forall \theta \in \Omega_\theta$
- Barankin (1946) solved this problem.
- The solution is based on sampling the parameter space at M test-points $\theta_1, \ldots, \theta_M$.
- Mean-unbiasedness is required only at these test-points $\rightarrow$ we get M constraints
- M can be arbitrarily large, so eventually we cover the entire parameter space.
- Barankin discovered that the optimal solution is based on linear combinations of the likelihood ratio function $\frac{f(\mathbf{y};\theta)}{f(\mathbf{y};\theta_0)}$ sampled at the test-points.

Barankin lower bound

- The solution is named the Barankin lower bound (BLB):

$$\text{MSE}(\theta) \geq \sup_{a_1,\ldots,a_M,\theta_1,\ldots,\theta_M} \frac{\left(\sum_{m=1}^{M} a_m(\theta_m - \theta)\right)^2}{\mathrm{E}\left[\left(\sum_{m=1}^{M} a_m \frac{f(\mathbf{y};\theta_m)}{f(\mathbf{y};\theta)}\right)^2 ; \theta\right]}$$

- The bound is valid for any choice of M, real coefficients $a_1, \ldots, a_M$, and test-points $\theta_1, \ldots, \theta_M$.
- This is the tightest lower bound on the MSE of uniformly mean-unbiased estimators.
- Problem: usually, we are not able to determine the optimal choice of coefficients and test-points $\rightarrow$ BLB cannot be computed.
- What can we learn from the BLB?

Barankin lower bound

- For obtaining intuition about the BLB, it is useful to consider the special case $M = 2$, $a_1 = -1$, $a_2 = 1$ $\theta_1 = \theta$, $\theta_2 = \theta + h$.
- We obtain

$$\text{MSE}(\theta) \geq \sup_h \frac{h^2}{\mathrm{E}\left[\left(\frac{f(\mathbf{y};\theta+h)}{f(\mathbf{y};\theta)} - 1\right)^2 ; \theta\right]} = \text{HCRLB}(\theta)$$

- HCRLB is a special case of BLB.
- Consequently, CRLB is also a special case of BLB.
- To obtain a tight bound, we choose h that maximizes the HCRLB.

Barankin lower bound

- High value of h increases the numerator but usually the denominator is increased as well.
- For $h \to 0$ the bound tends to the CRLB.
- In high SNR, usually the choice $h \to 0$ is optimal.
- Equivalently, in high SNR, the CRLB is the tightest bound on the MSE of uniformly mean-unbiased estimators.

Barankin lower bound

- In low SNR, we can sometimes find $h \nrightarrow 0$ for which $f(\mathbf{y};\theta+h)$ and $f(\mathbf{y};\theta)$ are very similar (their ratio is close to 1).
- It is difficult to distinguish between θ and $\theta+h$.
- Estimation performance is worse.
- This phenomenon is ignored by CRLB.
- In this case, the HCRLB is tighter than CRLB and better characterizes the optimal performance of uniformly mean-unbiased estimators.

CRLB for other risks

- In some problems, MSE is inappropriate.
- For example, when the likelihood function is periodic.
- The parameter can be an angle or phase of a signal.
- We need to consider the periodicity and use 2π-periodic cost function.
- The cyclic-error, $2 - 2\cos(\hat{\theta} - \theta)$, measures the square euclidean distance on a circle.
- The mean-cyclic-error (MCE), $\mathrm{E}[2 - 2\cos(\hat{\theta} - \theta); \theta]$, is an appropriate risk.
- Is mean-unbiasedness appropriate in this periodic case?

CRLB for other risks

- It was shown by Todros, Winik, and Tabrikian (2015) that the Barankin bound is infinite for periodic likelihood function.
- There are no mean-unbiased estimators.
- We need alternative unbiasedness conditions and corresponding CRLB on the MCE.
- Lehmann (1951) proposed a generalization of mean-unbiasedness to arbitrary cost functions.
- An estimator $\hat{\theta}$ is said to be a uniformly Lehmann-unbiased estimator of θ w.r.t. the cost function $L(\cdot, \cdot)$ if

$$\mathrm{E}[L(\hat{\theta}, \eta); \theta] \geq \mathrm{E}[L(\hat{\theta}, \theta); \theta], \ \forall \theta, \eta$$

CRLB for other risks

- Under the squared-error cost function, $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, the Lehmann-unbiasedness is reduced to the conventional mean-unbiasedness:

$$\mathrm{E}[\hat{\theta} - \theta] = 0, \ \forall \theta$$

- Under the cyclic-error cost function, $L(\hat{\theta}, \theta) = 2 - 2\cos(\hat{\theta} - \theta)$, the Lehmann-unbiasedness conditions are:

$$\mathrm{E}[\sin(\hat{\theta} - \theta)] = 0, \ \mathrm{E}[\cos(\hat{\theta} - \theta)] \geq 0, \ \forall \theta$$

- These conditions were developed by Routtenberg and Tabrikian (2014) and are named cyclic-unbiasedness conditions.

CRLB on MCE

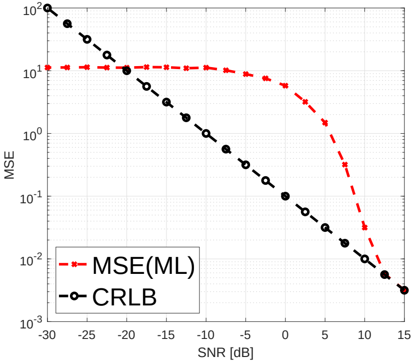
- Routtenberg and Tabrikian (2014) derived the cyclic CRLB on the MCE of cyclic-unbiased estimators.
- The MCE of a cyclic-unbiased estimator $\hat{\theta}$ is lower bounded by

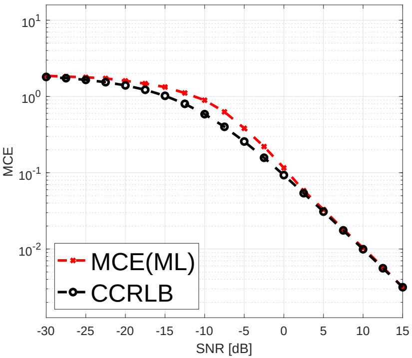
$$\text{MCE}(\theta) \geq \text{CCRLB}(\theta) \stackrel{\triangle}{=} 2 - 2(1 + \text{CRLB}(\theta))^{-\frac{1}{2}}$$

- Phase estimation: $y_n = Ae^{j\theta} + w_n, \ n = 1, \ldots, N$
- $\theta \in [-\pi, \pi)$, unknown deterministic phase.
- $w_n \sim \mathcal{CN}(0, \sigma^2)$ is circular complex Gaussian random noise, σ^2 is known.

CRLB on MCE

- Maximum likelihood estimator is cyclic-unbiased and is not mean-unbiased in this example.





CRLB for vector parameter

- Multivariate CRLB:

$$\text{MSE}(\boldsymbol{\theta}) \succeq \text{CRLB}(\boldsymbol{\theta}) \triangleq \mathbf{J}^{-1}(\boldsymbol{\theta})$$

- $\mathbf{J}(\boldsymbol{\theta}) \triangleq \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{y};\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}}^T \log f(\mathbf{y};\boldsymbol{\theta}); \boldsymbol{\theta}\right]$ is the Fisher information matrix (FIM).
- The inequality is in the sense of positive semidefinite matrices.

CRLB for vector parameter

Subvector estimation:

- We have an unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T$.
- We are interested in $\boldsymbol{\theta}_1$.
- The FIM can be expressed as $\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{1,1}(\boldsymbol{\theta}) & \mathbf{J}_{1,2}(\boldsymbol{\theta}) \\ \mathbf{J}_{2,1}(\boldsymbol{\theta}) & \mathbf{J}_{2,2}(\boldsymbol{\theta}) \end{bmatrix}$
- $\mathbf{J}_{m,k}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}_m} \log f(\mathbf{y}; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}_k}^T \log f(\mathbf{y}; \boldsymbol{\theta}); \boldsymbol{\theta}\right], \ m, k = 1, 2.$

CRLB for vector parameter

Subvector estimation (cont'd):

- In case θ_2 is known:

$$\mathrm{CRLB}_1 = \mathbf{J}_{1,1}^{-1}$$

- In case θ_2 is unknown:

$$\mathrm{CRLB}_1 = \mathbf{J}_{1,1}^{-1} + \underbrace{\mathbf{J}_{1,1}^{-1}\mathbf{J}_{1,2}(\mathbf{J}_{2,2} - \mathbf{J}_{2,1}\mathbf{J}_{1,1}^{-1}\mathbf{J}_{1,2})^{-1}\mathbf{J}_{2,1}\mathbf{J}_{1,1}^{-1}}_{\text{positive semidefinite matrix}}$$

- We have an additional term.
- The CRLB is usually higher when we have additional unknown parameters.

CRLB for vector parameter

- If $\mathbf{J}_{1,2} = \mathbf{J}_{2,1} = \mathbf{0}$ the parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are decoupled in terms of CRLB.
- In this case, the CRLB for estimation of $\boldsymbol{\theta}_1$ is unchanged, no matter if $\boldsymbol{\theta}_2$ is known or not.
- Example: $y_n = \mu + w_n, \;\; n = 1, \ldots, N$, $\mu \in \mathbb{R}$ is unknown mean, $w_n \sim \mathcal{N}(0, \sigma^2)$ is random noise, the variance σ^2 is unknown.
- $\boldsymbol{\theta} \triangleq [\mu, \sigma^2]^T$, $\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$, $\text{CRLB}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$
- In the Gaussian case, there is no coupling between mean and variance in terms of the CRLB.

Singular FIM

- CRLB is the inverse of the FIM. But, what if the FIM is singular?
- In this case, we cannot compute a bound for $\boldsymbol{\theta}$.
- But, we can obtain a meaningful bound for the projection of $\boldsymbol{\theta}$ on a subspace $\mathbf{V}$, i.e. $\mathbf{V}^T\boldsymbol{\theta}$.
- Eigenvalue decomposition (EVD): $\mathbf{J}(\boldsymbol{\theta}) = \mathbf{U}\Lambda\mathbf{U}^T$.
- $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$, matrix of eigenvectors.
- $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_M)$, diagonal matrix of eigenvalues.

Singular FIM

- Assume that $\lambda_m > 0,\ m = 1, \dots, M-K$, $\lambda_m = 0,\ m = M-K+1, \dots, M$.
- $\mathbf{V} \triangleq [\mathbf{u}_1, \dots, \mathbf{u}_{M-K}]$ contains eigenvectors with nonzero eigenvalues.
- CRLB for estimation of $\mathbf{V}^T\boldsymbol{\theta}$:

$$\mathrm{CRLB}_{\mathbf{V}}(\boldsymbol{\theta}) = \mathrm{diag}\left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_{M-K}}\right)$$

Singular FIM

- We only estimate the components of $\boldsymbol{\theta}$ in the subspace of eigenvectors with nonzero eigenvalues.
- Nonzero eigenvalues imply that the observations provide information for estimating the components of $\boldsymbol{\theta}$ in the corresponding directions (eigenvectors).
- Example: $y = \theta_1 + \theta_2 + w$, $w \sim \mathcal{N}(0, 1)$, $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$
- $\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, Singular matrix.

Singular FIM

- EVD: $\mathbf{J} = \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right)\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right)$
- $\mathbf{V} = \frac{1}{\sqrt{2}}[1, 1]^T$, $\mathbf{V}^T\boldsymbol{\theta} = \frac{1}{\sqrt{2}}(\theta_1 + \theta_2)$, $\text{CRLB}_{\mathbf{V}} = \frac{1}{2}$.
- The observation only provides information on the sum of the elements of $\boldsymbol{\theta}$.
- We cannot estimate each element of $\boldsymbol{\theta}$, but we can estimate the sum of its elements.

CRLB under parametric constraints

- Consider estimation of parameter vector $\boldsymbol{\theta} \in \mathbb{R}^M$.
- It is known that $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{0}_K, \; K < M$.
- Define $\mathbf{F}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}^T}\mathbf{f}(\boldsymbol{\theta})$.
- Can we better estimate $\boldsymbol{\theta}$ when we have this side information about the constraints?
- Gorman and Hero (1990) developed a constrained version of the CRLB:

$$\mathrm{MSE}(\boldsymbol{\theta}) \succeq \mathrm{COCRLB}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \mathbf{J}^{-1} - \underbrace{\mathbf{J}^{-1}\mathbf{F}^T(\mathbf{F}\mathbf{J}^{-1}\mathbf{F}^T)^{-1}\mathbf{F}\mathbf{J}^{-1}}_{\text{positive semidefinite matrix}}$$

- The COCRLB is lower than the unconstrained CRLB.

CRLB under parametric constraints

- Example: $\mathbf{y} = \boldsymbol{\theta} + \mathbf{w},\ \mathbf{w} \sim \mathcal{N}(\mathbf{0}_M, \sigma^2 \mathbf{I}_M)$.
- $\hat{\boldsymbol{\theta}} = \mathbf{y}$ is an efficient estimator attaining $\text{CRLB}(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}_M$.
- It is known that $\boldsymbol{\theta}$ satisfies $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} = \mathbf{0}_K,\ K < M$.
- Let $\mathbf{N}_\mathbf{A} = \mathbf{I}_M - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$ denote the orthogonal projection matrix onto the null space of $\mathbf{A}$.
- $\hat{\boldsymbol{\theta}} = \mathbf{N}_\mathbf{A}\mathbf{y}$ is a constrained efficient estimator attaining

$$\text{COCRLB}(\boldsymbol{\theta}) = \sigma^2 \mathbf{N}_\mathbf{A} \preceq \sigma^2 \mathbf{I}_M = \text{CRLB}(\boldsymbol{\theta})$$

- Projection of the efficient estimator on the null space of $\mathbf{A} \rightarrow$ constrained efficient estimator with lower MSE matrix.

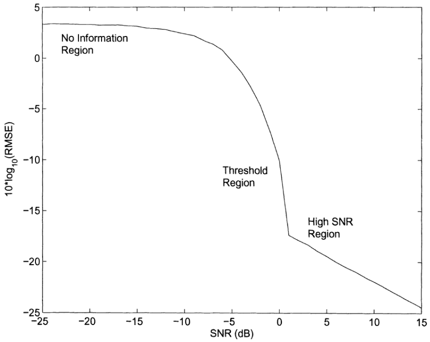
Outline

Bayesian MSE estimation

- We are given observation vector $\mathbf{y}$ from conditional distribution with pdf $f(\mathbf{y}|\theta)$.
- We are interested to estimate the unknown parameter θ.
- θ is a random variable with known prior pdf $f(\theta)$.
- We are usually interested in MSE $\triangleq \mathrm{E}[(\hat{\theta}-\theta)^2]$.
- The expectation is w.r.t. the joint pdf $f(\mathbf{y},\theta)$.
- MMSE estimator: $\hat{\theta}_{\text{MMSE}} = \mathrm{E}[\theta|\mathbf{y}]$, MMSE: $\mathrm{E}\left[\left(\mathrm{E}[\theta|\mathbf{y}]-\theta\right)^2\right]$
- Another popular estimator is the maximum a-posteriori (MAP) estimator, $\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} f(\theta|\mathbf{y})$.
- When $f(\theta|\mathbf{y})$ is symmetric and unimodal, MAP and MMSE estimators coincide.

Common behavior of MAP/MMSE estimators

- In nonlinear estimation problems, the MSE (or root MSE) of the MMSE/MAP estimator is usually described:



- Given prior distribution on θ, MMSE is the optimal performance that one can attain.

Bayesian MSE lower bounds

- Problem: in many cases, the computation of MMSE is intractable.
- We would like to characterize the optimal performance as closely as possible.
- Bayesian lower bounds on the MSE of any estimator can be used as benchmarks.
- Two main classes:
  - Weiss-Weinstein class - based on Cauchy-Schwartz inequality: including Bayesian Cramér-Rao bound (BCRB) and Weiss-Weinstein bound (WWB).
  - Ziv-Zakai class - based on the relation between MSE and probability of error: including Ziv-Zakai bound (ZZB).

BCRB

- Van Trees (1968) developed a Bayesian analogue to the CRLB.

$$\mathrm{E}[(\hat{\theta} - \theta)^2] \geq \text{BCRB} \triangleq \frac{1}{J}$$

- $J \triangleq \mathrm{E}[l^2(\mathbf{y}, \theta)]$ is the Bayesian Fisher information, $l(\mathbf{y}, \theta) \triangleq \frac{\partial}{\partial \theta} \log f(\mathbf{y}, \theta)$.
- BCRB is the simplest and most popular Bayesian lower bound but it has some drawbacks:
  - It may not be attained by the MAP or MMSE estimators, even asymptotically.
  - Requires restrictive regularity assumptions, e.g. differentiability of $f(\theta)$.
  - Fails to predict the threshold region of MMSE/MAP estimator.

Overcoming BCRB drawbacks

- The asymptotic performance of the MAP estimator is characterized by the expected value of the CRLB

$$\text{ECRB} \triangleq \mathrm{E}[\text{CRLB}(\theta)]$$

- The ECRB is not a lower bound so it can be higher than the MSE of the MAP/MMSE estimator.
- ECRB can only be used as an asymptotic benchmark.

Overcoming BCRB drawbacks

- It is necessary to find a lower bound with mild regularity assumptions that is able to predict the threshold region of the MMSE/MAP estimator.
- Two lower bounds satisfy these requirements:
  - WWB: derived by Weiss and Weinstein (1985)
  - ZZB: derived by Ziv and Zakai (1969)
- For example, unlike the BCRB, these bounds can be used for estimation of a discrete random parameter.
- It is not clear which of these bounds is tighter in general.

WWB

- The WWB is obtained by applying Cauchy-Schwartz inequality $\mathrm{E}^2[XY] \leq \mathrm{E}[X^2]\mathrm{E}[Y^2]$ with $X = \hat{\theta} - \theta$ and $Y = L^s(\mathbf{y}, \theta + h, \theta) - L^{1-s}(\mathbf{y}, \theta - h, \theta)$.
- $L(\mathbf{y}, \theta_1, \theta_2) \triangleq \frac{f(\mathbf{y},\theta_1)}{f(\mathbf{y},\theta_2)}$
- The choice of Y is based on Chernoff distance between pdfs: $cd(s) \triangleq \int_{\Omega_\mathbf{y}} f_0^{1-s}(\mathbf{y}) f_1^s(\mathbf{y})\, \mathrm{d}\mathbf{y} = \mathrm{E}_{f_0}\left[\left(\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})}\right)^s\right]$.

WWB

- The WWB is given by

$$\mathrm{E}[(\hat{\theta}-\theta)^2] \geq \text{WWB} \triangleq \frac{h^2 \mathrm{E}^2[L^s(\mathbf{y},\theta+h,\theta)]}{\mathrm{E}[(L^s(\mathbf{y},\theta+h,\theta)-L^{1-s}(\mathbf{y},\theta-h,\theta))^2]}$$

- The bound should be maximized w.r.t. $s \in (0,1)$ and $h \in \mathbb{R}$. In many cases, the choice $s=\frac{1}{2}$ is optimal.
- For $h \to 0$, WWB coincides with BCRB $\to$ WWB is tighter than BCRB.

ZZB

- consider the Bayesian detection problem:

$$
\begin{aligned}
H_0 &: \mathbf{y} \sim f(\mathbf{y}|\theta),\ Pr(H_0) = \frac{f(\theta)}{f(\theta) + f(\theta + h)} \\
H_1 &: \mathbf{y} \sim f(\mathbf{y}|\theta + h),\ Pr(H_1) = 1 - Pr(H_0)
\end{aligned}
$$

- Let $P_{\min}(\theta, \theta + h)$ denote the minimum probability of error obtained from the optimum likelihood ratio test.
- The ZZB is given by:

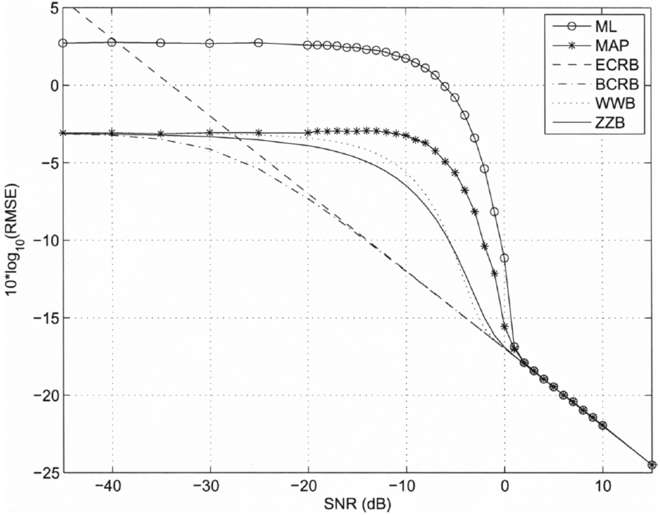
$$
\begin{aligned}
\mathrm{E}[(\hat{\theta} - \theta)^2] &\geq \mathrm{ZZB} \\
&\triangleq \frac{1}{2} \int_0^{\infty} \left(\int_{-\infty}^{\infty} (f(\theta) + f(\theta + h)) P_{\min}(\theta, \theta + h)\, \mathrm{d}\theta \right) h\, \mathrm{d}h
\end{aligned}
$$

- The minimum probability of error can be difficult to compute.

Example - Bayesian frequency estimation

- $y_n = Ae^{jn\theta+\phi} + w_n, \; n = 1, \ldots, N$
- $\theta \in [-\pi, \pi)$, unknown frequency with generalized symmetric beta prior pdf.
- The amplitude A and the phase ϕ are known.
- $w_n \sim \mathcal{CN}(0, \sigma^2)$ is circular complex Gaussian random noise, σ^2 is known.
- We define SNR $\triangleq \frac{A^2}{\sigma^2}$.

Example - Bayesian frequency estimation



BCRB for stochastic filtering

- Stochastic filtering: Bayesian estimation problem.
- We want to estimate a current system state (random variable) based on current and previous random observations.
- In general, the MMSE estimator and its performance are difficult to compute both analytically and numerically.
- BCRB is a commonly used tool for performance analysis of stochastic filters.

state space model

$$\begin{cases} \theta_n = a_n(\theta_{n-1}, w_n) \\ \mathbf{y}_n = \mathbf{h}_n(\theta_n, \boldsymbol{\nu}_n) \end{cases}, \ \forall n \in \mathbb{N},$$

- $\theta_n \in \mathbb{R}$ - random state
- $\theta_0 \in \mathbb{R}$ - initial random state
- $\boldsymbol{\theta}^{(n)} \triangleq [\theta_0, \ldots, \theta_n]^T$ - vector of augmented states
  Number of unknown parameters increases with time
- $\mathbf{y}_n \in \mathbb{R}^K$ - observation vector
- $\mathbf{y}^{(n)} \triangleq [\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T]^T$ - vector of augmented observations
- $w_n \in \mathbb{R}$ and $\boldsymbol{\nu}_n \in \mathbb{R}^N$ - system and observation noise, respectively
- $a_n : \mathbb{R}^2 \to \mathbb{R}$ - state transition function
- $\mathbf{h}_n : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^K$ - observation function

BCRB for stochastic filtering

- Multivariate BCRB: BCRB $\triangleq \mathbf{J}^{-1}$.
- $\mathbf{J} \triangleq \mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{y}, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}}^T \log f(\mathbf{y}, \boldsymbol{\theta})]$ is the Bayesian FIM (BFIM).
- At time step n, there are $n+1$ unknown parameters, $\theta_0, \ldots, \theta_n$.
- We are only interested in estimation of the current state θ_n.
- $f_n \triangleq f(\mathbf{y}^{(n)}, \boldsymbol{\theta}^{(n)})$ denotes the joint pdf at time step n.
- $\mathbf{J}_n \triangleq \mathrm{E}\left[\frac{\partial \log f_n}{\partial \boldsymbol{\theta}^{(n)}} \frac{\partial^T \log f_n}{\partial \boldsymbol{\theta}^{(n)}}\right] \in \mathbb{R}^{(n+1)\times(n+1)}$, denotes the nth step BFIM.
- The corresponding BCRB is $\left[\mathbf{J}_n^{-1}\right]_{n+1,n+1}$.

Computation of BCRB for stochastic filtering

- Problem: at each time step n, the BCRB requires the inversion of the BFIM $\mathbf{J}_n$.
- This task can be very difficult for large n since the size of $\mathbf{J}_n$ grows linearly with n.
- Tichavsky, Muravchik, and Nehorai (1998) proposed a recursive computation of the BCRB at each time step n that does not require inversion of $\mathbf{J}_n$.
- Due to the Markovian nature of the problem:

$$f_{n+1} = \underbrace{f_n}_{\text{Previous step}} \underbrace{f(\theta_{n+1}|\theta_n)}_{\textit{Dynamics}} \underbrace{f(\mathbf{y}_{n+1}|\theta_{n+1})}_{\text{New observation}}, \ \forall n \geq 0,$$

- nth step BCRB is computed recursively based on this relation.

Recursive computation of BCRB

- Define the nth step Fisher information $\xi_n \triangleq \frac{1}{[\mathbf{J}_n^{-1}]_{n+1,n+1}}$.
- The sequence $\{\xi_n\}$ obeys the following recursion

$$\xi_{n+1} = D_{n,2,2} - \frac{D_{n,1,2}^2}{\xi_n + D_{n,1,1}}, \ \forall n = 0, 1, 2, \ldots$$

$$\xi_0 = -\mathrm{E}\left[\frac{\mathrm{d}^2 \log f(\theta_0)}{\mathrm{d}\theta_0^2}\right], \ D_{n,1,1} \triangleq -\mathrm{E}\left[\frac{\partial^2 \log f(\theta_{n+1}|\theta_n)}{\partial\theta_n^2}\right]$$

$$D_{n,1,2} \triangleq -\mathrm{E}\left[\frac{\partial^2 \log f(\theta_{n+1}|\theta_n)}{\partial\theta_n \partial\theta_{n+1}}\right]$$

$$D_{n,2,2} \triangleq -\mathrm{E}\left[\frac{\partial^2 \log f(\theta_{n+1}|\theta_n)}{\partial\theta_{n+1}^2}\right] - \mathrm{E}\left[\frac{\partial^2 \log f\left(\mathbf{x}_{n+1}|\theta_{n+1}\right)}{\partial\theta_{n+1}^2}\right]$$

- At each time step n, we can compute ξ_n based on ξ_{n-1} and substitute it in the nth step BCRB.

Outline

Conclusion

- Performance bounds for parameter estimation are useful tools for performance analysis.
- These bounds improve our understanding of the problem at hand, before resorting to specific estimation techniques.
- Performance bounds have been around for decades and are still an active area of research.

Conclusion

Examples:

- Misspecified performance bounds: maybe our parametric model is not accurate. How well can we estimate? Example: estimation of Gaussian mean with "known" variance. We assume $y \sim \mathcal{N}(\theta, \sigma_1^2)$ but in fact $y \sim \mathcal{N}(\theta, \sigma_2^2)$.
- Semiparametric performance bounds: combination of parametric and nonparametric approaches. How well can we estimate the parameters under nonparametric uncertainty? Example: estimation of the mean value of a pdf in the set of elliptically symmetric pdfs.

Questions

?