Lecture 4

Optimal Estimation Theory and Its Applications

Prof. Yan LIANG

liangyan@nwpu.edu.cn

School of Automation,

Northwestern Polytechnical University

Autumn 2019

Outline

1. Cramér-Rao Bound
2. Expectation Maximization Method

Cramér-Rao Bound

The Fact

Different performance indexes, for example MSE or MLE, derive different optimal estimators.

Question

If there exists the common performance bound of all estimators?

New Concept

Cramér-Rao Bound

Cramér-Rao Bound

(1) Case 1

If $\hat{X}$ is any unbiased estimate of a deterministic variable X based on the measurement Z, then the covariance of the estimation error $\tilde{X} = X - \hat{X}$ is bounded by

$$P_{\tilde{X}} \geqslant J_F^{-1} \tag{1}$$

where the Fisher information matrix is given by

$$J_F = E\left\{ \left[\frac{\partial \ln f_{Z|X}(z|x)}{\partial x} \right] \left[\frac{\partial \ln f_{Z|X}(z|x)}{\partial x} \right]^T \right\} = -E\left[\frac{\partial^2 \ln f_{Z|X}(z|x)}{\partial x^2} \right] \tag{2}$$

Cramér-Rao Bound

Equality holds in Eq.1 if and only if

$$\frac{\partial \ln f_{Z|X}(z|x)}{\partial x} = k(x)(x - \hat{X}) \tag{3}$$

It is assumed that $\partial \ln f_{Z|X}/\partial x$ and $\partial^2 \ln f_{Z|X}/\partial x^2$ exist and are absolutely integrable.
An estimate $\hat{X}$ is efficient if it satisfies the Cramér-Rao bound with equality, that is, if Eq.3 holds.

Cramér-Rao Bound

For linear Gaussian measurement model $Z = HX + V$ with V is a zero-mean Gaussian noise with covariance R, the MLE of X and its covariance are

$$\hat{X}_{ML} = (H^T R^{-1} H)^{-1} H^T R^{-1} z \quad P_{ML} = (H^T R^{-1} H)^{-1}$$

$\hat{X}_{ML}$ is unbiased estimate.

$$J_F = -E\left[\frac{\partial^2 \ln f_{Z|X}(z|x)}{\partial x^2}\right] = H^T R^{-1} H$$

$$P_{\mathsf{ML}} = J_F^{-1}$$

Conclusion: The MLE in this case is efficient.

Cramér-Rao Bound

or

$\hat{X}_{ML} = (H^T R^{-1} H)^{-1} H^T R^{-1} z$ is unbiased estimate.

$$f_{Z|X}(z|x) = \exp\{-(z - Hx)^T R^{-1}(z - Hx)/2\}/\sqrt{2\pi\,|R|}$$

$$\ln f_{Z|X}(z|x) = -(z - Hx)^T R^{-1}(z - Hx)/2 - \ln(2\pi\,|R|)/2$$

$$\frac{\partial \ln f_{Z|X}(z|x)}{\partial x} = H^T R^{-1}(z - Hx) = -H^T R^{-1}(Hx - z)$$

$$= -(H^T R^{-1} H)x + (H^T R^{-1} H)(H^T R^{-1} H)^{-1} H^T R^{-1} z$$

$$= -(H^T R^{-1} H)\left(x - (H^T R^{-1} H)^{-1} H^T R^{-1} z\right)$$

From the expression of the efficient estimate $\dfrac{\partial \ln f_{Z|X}(z|x)}{\partial x} = k(x)(x - \hat{X})$,

we find $\hat{X}_{ML}$ is the efficient estimate.

Cramér-Rao Bound

The ML estimate has several nice properties as the number of independent measurements N goes to infinity

1. $\hat{X}_{ML}$ converges in probability to the correct value of X as $N \to \infty$. (Consistent)
2. $\hat{X}_{ML}$ becomes efficient as $N \to \infty$.
3. $\hat{X}_{ML}$ becomes Gaussian $N(X, PX)$ as $N \to \infty$.

Stochastic convergence: the sequence $Y_1, Y_2, \cdots$ of RVs (random variables) converges in probability to $RV\ \ Y$ if for all $\varepsilon > 0$

$$P(||YN - -Y|| > \varepsilon) \to 0\ for\ N \to \infty$$

Cramér-Rao Bound

(1)Case 2

If $\hat{X}$ is any estimate of a stochastic variable X based on measurement Z, then the covariance of the estimation error $\tilde{X} = X - \hat{X}$ is bounded by

$$P_{\tilde{X}} \geqslant L^{-1} \tag{4}$$

where the Fisher information matrix is given by

$$L = E\left\{\left[\frac{\partial \ln f_{XZ}(x,z)}{\partial x}\right]\left[\frac{\partial \ln f_{XZ}(x,z)}{\partial x}\right]^T\right\} = -E\left[\frac{\partial^2 \ln f_{XZ}(x,z)}{\partial x^2}\right] \tag{5}$$

Cramér-Rao Bound

Equality holds in Eq.4 if and only if

$$\frac{\partial \ln f_{XZ}(x,z)}{\partial x} = k(x - \hat{X}) \tag{6}$$

an estimate $\hat{X}$ is efficient if Eq.3 holds.
It is assumed that $\partial \ln f_{Z|X}/\partial x$ and $\partial^2 \ln f_{Z|X}/\partial x^2$ exist and are absolutely integrable with respect to both variables, and .

$$\lim_{x\to+\infty} B(x) f_X(x) = 0, \quad \lim_{x\to-\infty} B(x) f_X(x) = 0 \tag{7}$$

with $B(x) \underline{\underline{\Delta}} \int_{-\infty}^{+\infty} (x - \hat{X}) f_{Z|X}(z|x) \mathrm{d}z$

Cramér-Rao Bound

To a deterministic variable, its Cramér-Rao Bound

- depends on the likelihood function $f_{Z|X}$
- requires the unbiasedness
- has the efficiency condition about $k(x)$

$$\frac{\partial \ln f_{Z|X}(z|x)}{\partial x} = k(x)(x - \hat{X})$$

To a random variable, its Cramér-Rao Bound

- depends on the joint PDF f_{XZ}
- requires $\lim_{x\to+\infty} B(x)f_X(x) = 0, \quad \lim_{x\to-\infty} B(x)f_X(x) = 0$
- has the efficiency condition about k

$$\frac{\partial \ln f_{XZ}(x,z)}{\partial x} = k(x - \hat{X})$$

Cramér-Rao Bound

To a random variable, we have

$$\frac{\partial \ln f_{XZ}(x,z)}{\partial x} = \frac{\partial \ln f_{X|Z}(x|z)}{\partial x} + \frac{\partial \ln f_Z(z)}{\partial x} = \frac{\partial \ln f_{X|Z}(x|z)}{\partial x}$$

From $\dfrac{\partial \ln f_{XZ}(x,z)}{\partial x} = k(x - \hat{X}),$ we have

$$f_{X|Z}(x|z) = e^{-(k/2)x^T x + k\hat{X}^T x + c}$$

It is concluded that its efficient estimate necessarily have the Gaussian posterior conditional PDF $f_{X|Z}$

Expectation-maximization algorithm

MLE in the case of data missing or partly-unobservable data

Consider the observation data Z_{obs}, and partly-unobservable data Z_{mis}.

They constitutes the data set $Z = Z_{obs} \cup Z_{mis}$

Here $p\{Z_{mis}|Z_{obs}, X\}$ and $p\{Z_{obs}, Z_{mis}|X\}$are known.

Our aim is to obtain the ML estimate, i.e., arg $\left\{\max_X p\{Z_{obs}|X\}\right\}$

based on the available data Z_{obs} and apriori information

$p\{Z_{mis}|Z_{obs}, X\}$ and $p\{Z_{obs}, Z_{mis}|X\}$.

Expectation-maximization algorithm

Denote $L(X) = \ln p\{Z_{obs}|X\}$. To obtain the iterative estimate, we define the estimate after the k - th iteration as $\hat{X}^{(k)}$, and have

$$L(X) - L(\hat{X}^{(k)}) = \ln \frac{p\{Z_{obs}|X\}}{p\{Z_{obs}|\hat{X}^{(k)}\}} = \ln \sum_{Z_{mis}} \frac{p\{Z_{obs}, Z_{mis}|X\}}{p\{Z_{obs}|\hat{X}^{(k)}\}}$$

$$= \ln \sum_{Z_{mis}} \frac{p\{Z_{obs}|Z_{mis}, X\}p\{Z_{mis}|X\}}{p\{Z_{obs}|\hat{X}^{(k)}\}}$$

If the operator ln and $\sum$ can exchange the order, then the computation will be easily implemented.

Expectation-maximization algorithm

$$L(X) - L(\hat{X}^{(k)}) = \ln \sum_{Z_{mis}} \frac{p\{Z_{obs}|Z_{mis}, X\}p\{Z_{mis}|X\}p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\}}{p\{Z_{obs}|\hat{X}^{(k)}\}p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\}}$$

From the Jensen Inequality $\sum_j \lambda_j = 1 \Rightarrow \ln(\sum_j \lambda_j y_j) \geqslant \sum_j \lambda_j \ln y_j$ and the fact that $\sum_{Z_{mis}} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\} = 1$, we have

$$L(X) - L(\hat{X}^{(k)}) \geqslant$$
$$\sum_{Z_{mis}} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\} \ln \left[\frac{p\{Z_{mis}|X\}p\{Z_{obs}|Z_{mis}, X\}}{p\{Z_{obs}|\hat{X}^{(k)}\}p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\}} \right]$$

In other words, the operators exchange their order via the Jensen Inequality.

Expectation-maximization algorithm

$$L(X) \geqslant$$

$$L(\hat{X}^{(k)}) + \sum_{Z_{mis}} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\} \ln \left[\frac{p\{Z_{mis}|X\} p\{Z_{obs}|Z_{mis}, X\}}{p\{Z_{obs}|\hat{X}^{(k)}\} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\}} \right]$$

Through conditional probability integral (called conditional mean), the unobservable data is smoothed.

Given the initial estimate value $\hat{X}^{(0)}$, iteratively find the maximum point of $\hat{X}^{(k+1)} = \arg\max_{X} l(X, \hat{X}^{(k)})$ until $\hat{X}^{(k+1)} = \hat{X}^{(k)}$

Expectation-maximization algorithm

We have

$$\hat{X}^{(k+1)} = \arg\max_X$$

$$\left[L(\hat{X}^{(k)}) + \sum_{Z_{mis}} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\} \ln \left[\frac{p\{Z_{mis}|X\} p\{Z_{obs}|Z_{mis}, X\}}{p\{Z_{obs}|\hat{X}^{(k)}\} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\}} \right] \right]$$

$$= \arg\max_X \left[\sum_{Z_{mis}} p\{Z_{mis}|Z_{obs}, \hat{X}^{(k)}\} \ln \left[p\{Z_{mis}|X\} p\{Z_{obs}|Z_{mis}, X\} \right] \right]$$

$$= \arg\max_X \left[E_{Z_{mis}|Z_{obs}, \hat{X}^{(k)}} \ln \left[p\{Z_{obs}, Z_{mis}|X\} \right] \right]$$

It is proved that such iterative computation always converge.

Remark: $p\{Z_{mis}|Z_{obs}, X\} = p\{Z_{mis}, Z_{obs}|X\} / p\{Z_{obs}|X\}$ is known

Expectation-maximization algorithm

The implementation of EM algorithm includes two steps

1. Expectation (E-Step)

$$Q(X|\hat{X}^{(k+1)}) = E_{Z_{mis}|Z_{obs},\hat{X}^{(k)}} \ln \left[p\{Z_{obs}, Z_{mis}|X\}\right]$$

2. Maximization(M-Step)

$$\hat{X}^{(k+1)} = \arg\max_{X} \left[Q(X|\hat{X}^{(k+1)})\right]$$

McLachlan G J, Krishnan T. **The EM Algorithm and Extension**. New York: Wiley,1997