EE 602 Term Paper on

Summary of Estimators

Submitted by:

Vishesh lokras(Y8104067)

Avadhesh Kr Yadav(Y8104010)

Introduction

There are two methods to estimate a parameter vector θ

First method is Classical Approach in which the Unknown parameter vector **θ** is a Deterministic Constant**.** Various Estimators in classical Approach are given below:

- Cramer-Rao Lower Bound
- Rao-Blackwell-Lehmann-Scheffe
- Best Linear Unbiased Estimator
- Maximum Likelihood Estimator
- Least Squares Estimator
- Method of Moments

Second method is Bayesian Approach in which the Known parameter vector **θ** is Realization of a Random Vector. Various Estimators in Bayesian Approach are given below:

- Minimum Mean Square Error Estimator
- Maximum A Posteriori Estimator
    - Linear Minimum Mean Square Error Estimator

Cramer Rao Lower Bound(CRLB)

a. Data Model/Assumption

PDF p(**x;θ**) is known.

b. Estimator

If the equality condition for the CRLB

$$\frac{\partial lnp(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = I(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta})$$

is satisfied, then the estimator is

$$\ddot{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x})$$

where **I(θ)** is a p x p matrix dependent only on **θ** and **g(x)** is a p-D function of data x.

c. Optimality/Criterion

$\hat{\theta}$ achieves the CRLB, the lower bound on the variance for any unbiased estimator(and hence is said to be efficient), and is therefore the MVU estimator.

d. Performance

It is **unbiased** and has the **minimum variance**

$$var(\hat{\theta} i) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \quad i=1,2,.....p$$

Where

$$[\mathrm{I}(\theta)]_{ij} = \mathrm{E}\left[\frac{\partial\, lnp(x;\theta)}{\partial \theta_i}\frac{\partial\, lnp(x;\theta)}{\partial \theta_j}\right]$$

e. Comments

An efficient **Estimator may not exist** & this approach may fail

Rao-Blackwell-Lehmann-Scheffe

a. Data Model/Assumption

PDF $p(\mathbf{x};\boldsymbol{\theta})$ is known.

b. Estimator

i. Find a *sufficient statistic* $\mathbf{T}(\mathbf{x})$ by factoring PDF as

$$p(\mathbf{x};\boldsymbol{\theta})=g(\mathbf{T}(\mathbf{x}),\boldsymbol{\theta})h(\mathbf{x})$$

ii. If $E[\mathbf{T}(\mathbf{x})]=\boldsymbol{\theta}$,then $\hat{\boldsymbol{\theta}} = \mathbf{T}(\mathbf{x})$.If not, we must find a p-dimensional function $\mathbf{g}$ so that $E[\mathbf{g}(\mathbf{T})]= \boldsymbol{\theta}$, and then $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{T})$.

c. Optimality/Criterion

$\hat{\boldsymbol{\theta}}$ is the MVU estimator.

d. Performance

Unbiased but **variance depends upon PDF** -no general formula is available

e. Comments

"**Completeness**" of sufficient statistic must be checked. A p-dimensional sufficient statistic may not exist so this method may fail.

Best Linear Unbiased Estimator

a. Data Model/Assumptions

$E[\mathbf{x}]=\mathbf{H\theta}$

where **H** is an N x p (N>p) known matrix and **C**, the covariance matrix of **x**, is known. Equivalently, we have

$$\mathbf{x} = \mathbf{H\theta} + \mathbf{w}$$

where $E(\mathbf{w})=0$ and $\mathbf{C_w} = \mathbf{C}$.

b. Estimator

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

c. Optimality/Error Criterion

$\hat{\boldsymbol{\theta}}_i$ for i = 1,2,….,p has **minimum variance** of all unbiased estimators that are linear in **x.**

d. Performance

$\hat{\boldsymbol{\theta}}_i$ for i = 1,2,…,p is unbiased. The variance is

$$\mathbf{var(\theta_i) = [(H^T C^{-1} H)^{-1}]_{ii}} \qquad i = 1,2,\ldots\ldots,p$$

e. Comments

If **w** is a Gaussian RV ,then is $\hat{\boldsymbol{\theta}}$ also the ***MVU estimator*** (for all linear & non linear functions of **x**).

Maximum Likelihood Estimator

a. Data Model/Assumptions

PDF p(**x;θ**) is known

b. Estimator

$\hat{\boldsymbol{\theta}}$ is the value of **θ** maximizing p(**x;θ**), where **x** is replaced by the observed data samples.

c. Optimality/Error Criterion

Not optimal in general. Under certain conditions on the PDF, however, the MLE is efficient for large data records or as N →∞ (asymptotically).Hence, asymptotically it is the MVU estimator.

d. Performance

For finite N depends on PDF-no general formula is available. Asymptotically, under certain conditions

$$\hat{\boldsymbol{\theta}} = \boldsymbol{N}(\boldsymbol{\theta}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$$

e. Comments

If an MVU estimator exists, the MLE procedure will produce it .

Least Squares Estimator

a. Data Model/Assumption

x[n]=s[n;**θ**] + w[n]   n=0,1,……,N-1

where the signal s[n;**θ**] depends explicitly on the unknown parameters. Equivalently, the model is

$$\mathbf{x}=\mathbf{s}(\boldsymbol{\theta}) + \mathbf{w}$$

where **s** is a known N-dimensional function of **θ** and the noise or perturbation **w** has zero mean.

b. Estimator

$\hat{\boldsymbol{\theta}}$ is the value that minimizes

$$J(\theta) = (x - s(\theta))^T(x - s(\theta)) = \sum_{n=0}^{N-1}(x[n] - s[n;\theta])^2$$

c. Optimality/Error Criterion

None In General

d. Performance

Depends on PDF of **w**-no general formula is available

e. Comments

The fact that we are *minimizing a LS error minimizing a LS error criterion does not in general translate into minimizing the estimation error*. Also **w** is a Gaussian RV $N(0,\sigma^2\mathbf{I})$, then the LSE is equivalent to MLE.

Method of Moments

a. Data Model/Assumptions

There are p moments μ_i=E(x^i[n]) for i = 1,2, . . . ,p, which depends upon **θ** in a known way. *The entire PDF need not be known*.

b. Estimator

If **μ = h(θ),** where **h** is an invertible p-D function of **θ** & **μ**=[μ_1 μ_2 . . . μ_p]^T, then

$$\hat{\boldsymbol{\theta}} = \boldsymbol{h}^{-1}(\hat{\boldsymbol{\mu}})$$

c. Optimality/Error Criterion

None in General

d. Performance

For finite N it depends on PDF of **x**. However for large data records (asymptotically), if

$$E(\hat{\theta}_i) = g_i(\boldsymbol{\mu})$$

$$var(\hat{\theta}_i) = \left(\frac{\partial\ g_i}{\partial\hat{\boldsymbol{\mu}}}\right)^T_{\hat{\mu}=\mu} \mathrm{C}_{\hat{\mu}} \left(\frac{\partial\ g_i}{\partial\hat{\boldsymbol{\mu}}}\right)_{\hat{\mu}=\mu}$$

for i = 1,2 … ,p

e. Comments

Usually very ***easy to implement***.

Minimum Mean Square Error Estimator

a. *Data model/Assumptions*

The joint PDF p(**x**,**θ**) is known, where **θ** is considered to be a RV. Usually, p(**x**|**θ**) is specified as the data model and p(**θ**) as the prior PDF for **θ**, so that p(**x,θ**)=p(**x**|**θ**)p(**θ**).

b. Estimator

$\hat{\boldsymbol{\theta}}$ = E(**θ**|**x**) ,where the expectation represents the posterior PDF

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int \quad p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,d\boldsymbol{\theta}}$$

If **x,θ** are jointly Gaussian, $\hat{\boldsymbol{\theta}}$ $=E(\boldsymbol{\theta})+\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x}-E(\mathbf{x}))$

c. Optimality/Error criterion

$\hat{\boldsymbol{\theta}}_i$ minimizes the Bayesian MSE

$Bmse(\hat{\boldsymbol{\theta}}_i)=E[(\theta_i - \hat{\boldsymbol{\theta}}_i)^2]$ i=1 ,2 , . . , p where the Expectation is w.r.t p(**x**, θ_i)

d. Performance

The error $\varepsilon_i = \theta_i - \hat{\theta}_i$ has zero mean & variance

$$\mathrm{var}(\varepsilon_i) = Bmse(\hat{\theta}_i) = \int [\mathbf{C}_{\theta|x}]_{ii}\, p(\mathbf{x})d\mathbf{x}$$

If **x, θ** are jointly Gaussian, then the error is Gaussian with zero mean & variance

$$\mathrm{var}(\varepsilon_i)=[\mathbf{C}_{\boldsymbol{\theta\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}]_{ii}$$

Maximum A Posteriori Estimator

a. *Data Model/Assumptions*

Same as for the MMSE estimator.

b. *Estimator*

$\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ that maximizes p($\boldsymbol{\theta}|\mathbf{x}$) or, equivalently the value that maximizes p($\mathbf{x}|\boldsymbol{\theta}$)p($\boldsymbol{\theta}$). If $\mathbf{x},\boldsymbol{\theta}$ are jointly Gaussian,

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta})+\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x}-E(\mathbf{x}))$$

c. *Optimality/Error Criterion*

Minimizes the ***"hit or miss"*** cost function.

d. *Performance*

Depends on PDF-no general formula available. If $\mathbf{x},\boldsymbol{\theta}$ are jointly Gaussian, then the performance is identical to that of MMSE estimator.

e. *Comments*

For PDF's whose ***mean & mode are the same***, the MMSE & MAP estimators will be identical, i.e., the Gaussian PDF, for example.

Linear Minimum Mean Square Error Estimator

a. *Data Model/Assumptions*

The first two moments of the joint PDF p(x,) are known or the mean and covariance

$$\begin{bmatrix} E(\boldsymbol{\theta}) \\ E(\boldsymbol{x}) \end{bmatrix} = \begin{pmatrix} C_{\theta\theta} & C_{\theta x} \\ C_{x\theta} & C_{xx} \end{pmatrix}$$

b. *Estimator*

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x}))$$

c. Optimality/Error Criterion

$\hat{\boldsymbol{\theta}}_i$ has the minimum Bayesian MSE of all estimators that are *linear* function of x

d. Performance

The error $\varepsilon_i = \theta_i - \hat{\boldsymbol{\theta}}_i$ has zero mean & variance

$$\text{var}(\varepsilon_i) = \text{Bmse}(\hat{\boldsymbol{\theta}}_i) = [\mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}]_{ii}$$

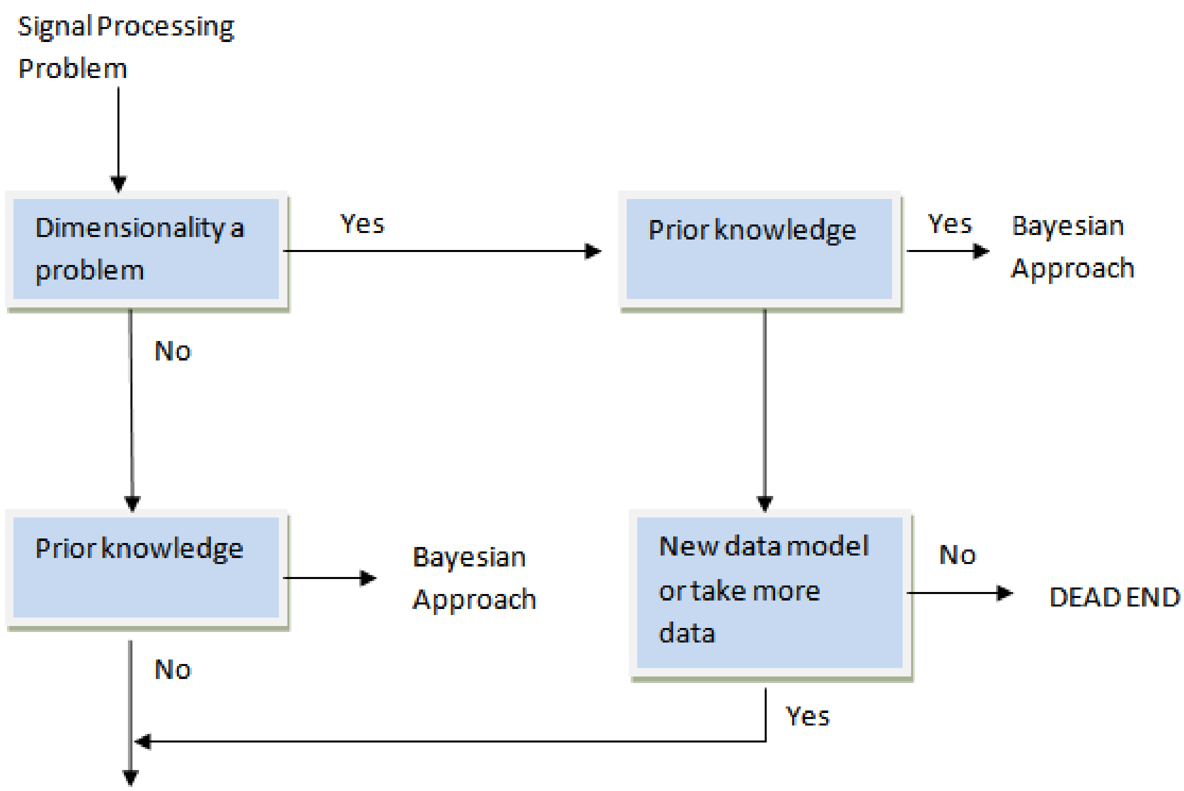
e. Comments

If **x,θ** are jointly Gaussian, this is identical to the MMSE & MAP estimators.

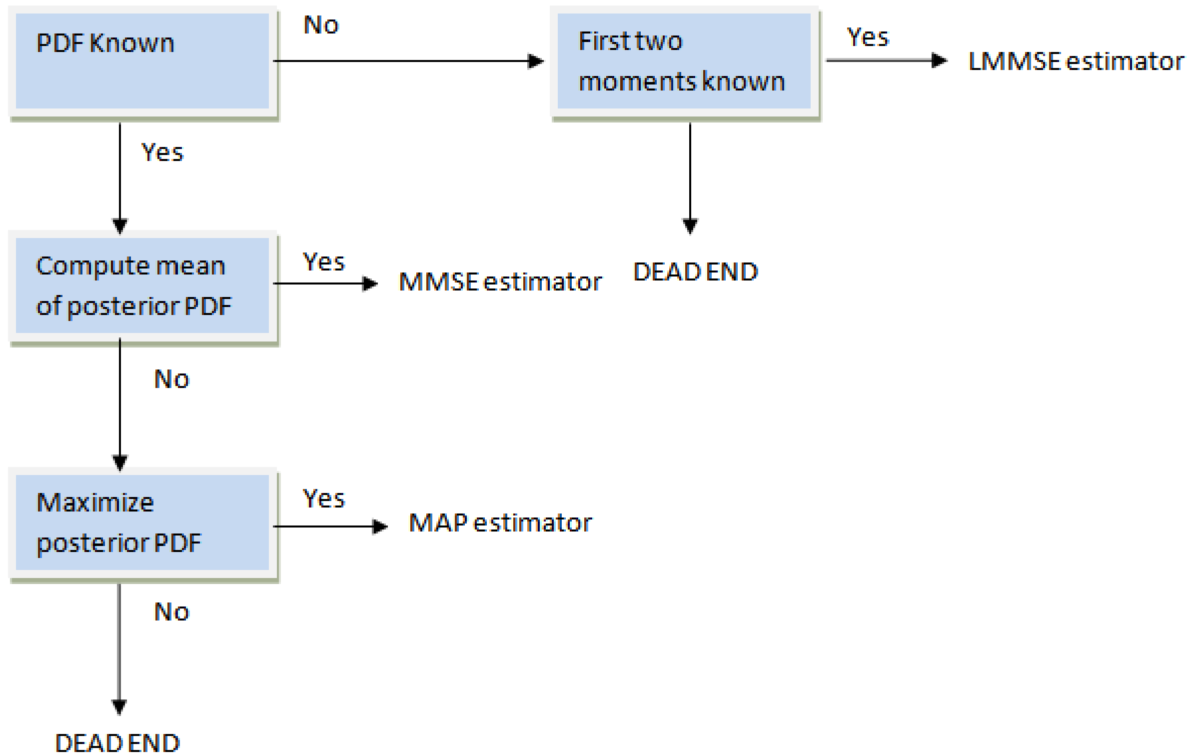
Choosing an Estimator

- The Choice of an estimator should begin with the search for an optimal estimator that is computationally feasible.
- If the search proves futile, then suboptimal estimators should be investigated.
- Even if the estimator is optimal for the given data model, its performance may not be adequate. In such a case, the data model may need to be modified.

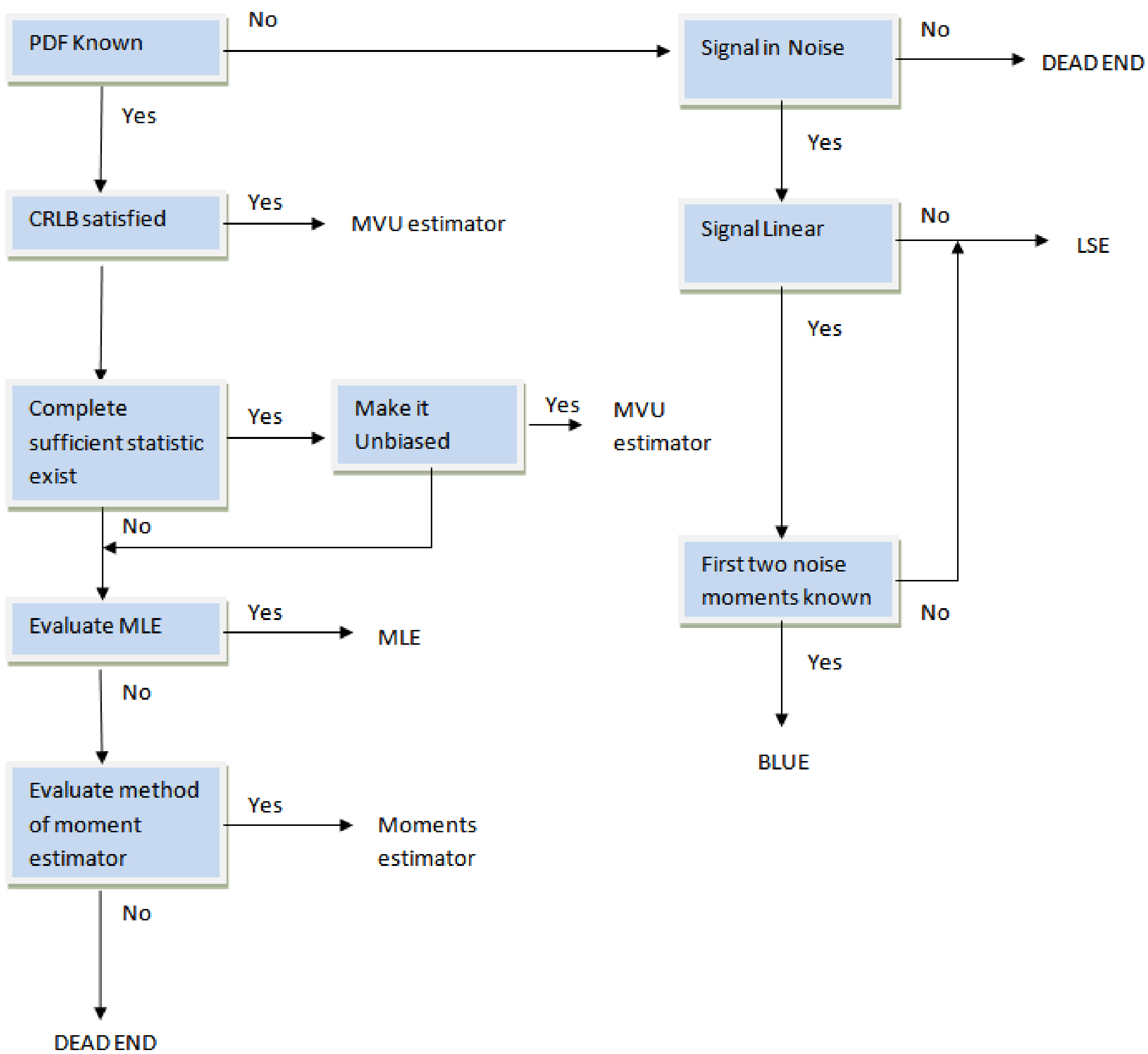
Classic v/s Bayesian



Bayesian Approach



Classical Approach



Final Comments

- Selection of Good Data Model
  - Complex enough to describe principle features
  - Simple enough to allow an estimator that is optimal & easily feasible.
- Determination of optimal estimators not possible
  - Search of an MVU estimator in classical approach
- Implementation of optimal estimator not possible
  - MMSE estimator in Bayesian estimation.