Fundamentals of Estimation Theory

Muhammad Sarwar Ehsan
Signal Processing and Speech Communication Laboratory
Graz University Technology, Graz Austria

Abstract

Estimation theory is an important mathematical concept used in many communication and signal processing applications. This theory is helpful in estimation of the desired information in the received data and hence is used all range of application from radar to speech processing. In this report we will introduce some of basic definitions and concepts of estimation theory.

1 Introduction

This article introduces some fundamentals of estimation theory. Emphasis here is to introduce some of the basic concepts and definitions without going into mathematical rigor. This problem of parameters estimation is common in many fields where we have to extract parameters form given data. Some common fields in which estimation theory is used are listed below [2]:

- In radar, the delay of the received pulse echo has to be estimated in the presence of noise
- In sonar, the delay of the received signal from each sensor has to estimated in the presence of noise
- In Speech, the parameters of the speech model have to be estimated in the presence of speech/speaker variability and environmental noise
- In image processing, the position and orientation of an object from a camera image has to be estimated in the presence of lighting and background noise
- In biomedicine, the heart rate of a fetus has to be estimated in the presence of sensor and environmental noise Communications where the carrier frequency of a signal has to be estimated for demodulation to the baseband in the presence of degradation noise Control where the position of a powerboat for corrective navigation correction has to be estimated in the presence of sensor and environmental noise
- In seismology, the underground distance of an oil deposit has to be estimated from noisy sound re sections due to di erent densities of oil and rock layers The majority of applications require estimation of an unknown parameter from a collection of observation data x[n] which also includes due to sensor inaccuracies, additive noise, signal distortion (convolutional noise), model imperfections, unaccounted source variability and multiple interfering signals.

All of the above problems share a common thing of estimation of a certain unknown parameter or a group of parameters. Estimation of the parameter can be defined as:

An estimator can be think of as a rule or function that assigns a value to θ for each realization of x. The estimate of θ is the value of θ obtained for a given realization x. Because data is inherently random, therefore, we have to used statistical methods to get these estimates, because in general we do

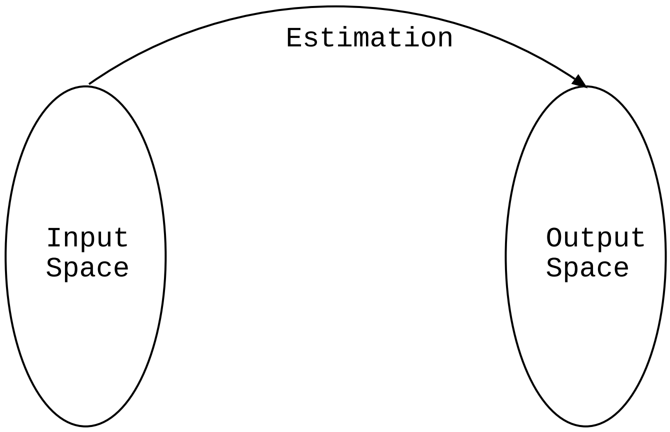

Figure 1: Estimation as a Rule or Function

not know the actual data. We model data by defining probability density function (PDF) as:

$$p(\mathbf{x};\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2\sigma^2}(\mathbf{x}-\theta)^2] \tag{1}$$

In (1), it is important note that the Semicolon denotes the family of PDFs depending on parameter θ. i.e. we will get different PDFs for different values of theta as shown in the figure below.

In General, even the PDf is not known a priori, its selection should be:

- Consistent with problem constraints.
- according to any prior knowledge about the data.
- Computationally feasible to get an estimate.

Due to random nature of data and even estimate, we can only judge the performance of estimator by its statistical properties. Rest of the report is organized as: in section 2, we will introduce a very commonly used estimator known as minimum variance unbiased estimator. In the last section we will discuss cramer rao bound.

2 Minimum Variance Unbiased Estimator

Unbiased estimators are an important class of commonly used estimators. These estimators can be found by applying certain criteria which we will discuss in this section.

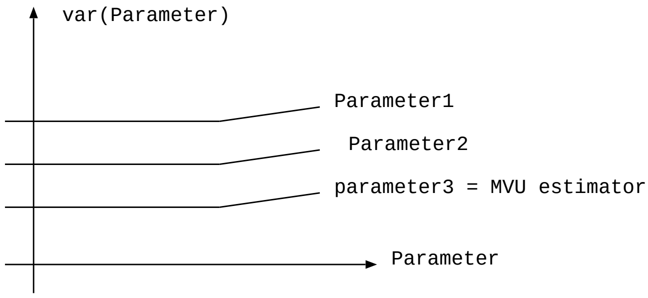

Figure 2: MVU exists

2.1 Biased and Unbiased Estimators

Biased estimator is the estimator which on average do not converge to the true estimate. This type of estimator is used where we Do not not have sufficient amount of data to get reasonably good estimate. Bias is the expected value of deviation of estimated value of parameter from the true value. Mathematically,

$$b(\theta) = E(\hat{\theta} - \theta) \tag{2}$$

An example of biased estimator is a modified sample mean defined as:

$$\breve{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n] \tag{3}$$

which on average gives:

$$E(\breve{A}) = \frac{1}{2} A \tag{4}$$

An unbiased estimator is the Estimator which on average yields the true value of parameters. Mathematically,

$$E(\hat{\theta}) = \theta \tag{5}$$

A sample mean estimator is an unbiased estimator define as:

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \tag{6}$$

which yield on average

$$E(\hat{A}) = A \tag{7}$$

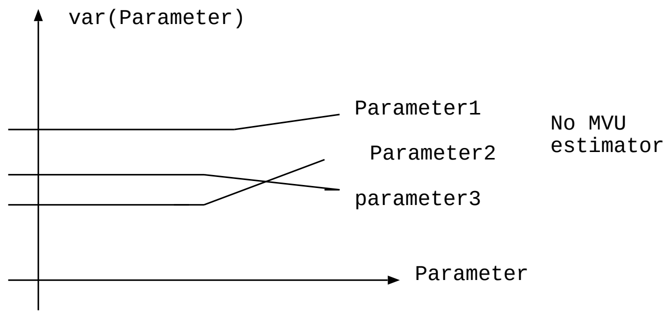

Figure 3: MVU does not exist

Unbiased estimator does not necessarily mean a good estimator but on average it will converge to true value. Some times biased estimators can also perform well, specially when do not have large data to estimate.

2.2 Minimum Variance Unbiased Estimator (MVUE)

Mean Square Error (MSE) estimator is the natural choice as an optimality criterion for the search of an Unbiased estimator, therefore, we define MSE as:

$$mse(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{8}$$

but, unfortunately it leads to biased estimator such that:

$$mse(\hat{\theta}) = var(\hat{\theta}) + b^2(\theta) \tag{9}$$

To obtain minimum variance unbiased estimator, we constrain bias term to zero and minimize the variance.

2.3 Existence of the Minimum Variance Unbiased Estimator

Fig. 2 and 3 show a situation were MVUE exits or not respectively. It is important to note that, MVUE exists only if its variance is minimum for all values of the parameter as shown in Fig. 2. In Fig. 3 estimator 2 has lower variance for the small values of parameter but estimator 3 has lower variance for the large values of parameter, therefore, none of the estimator is MVUE. This condition can be put in the following way also.

MVUE exits only if class of PDF does not change with Parameters.

2.4 Finding the Minimum Variance Unbiased Estimator

Even if MVUE exits for a given problem, it is difficult to find it. We can follow one of the following paths to find MVUE.

- Determine the Cramer-Rao Lower Bound (CRLB) and see if some of the estimator satisfies it.
- Apply Rao Blackwell Lehmann Scheffe (RBLS) theorem if no MUVE satisfies CRLB.
- By applying linear constraint we can further restrict the class of estimators.

First two approaches may produce the MVUE, while third approach will yield it only if the MUV estimator is linear in data.

3 Cramer Rao Lower Bound

It was stated earliar that, Cramer Rao lower bound is used to determine the MVUE. If we are able to find CRLB for some estimator then it is not only MVUE but also efficient. On the other hand, there exist some MVUE which do not satisfy CRLB, therefore, they are inefficient estimators. We depict have these situations in Fig. 4 and 5.

3.1 Some Definitions

Likelihood Function: When PDF is viewed as function of unknown parameter with fixed data then it is known as likelihood function. Logarithm of this function is known as loglikelihood function. Mathematically:

$$p(x[0];A) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2\sigma^2}(x[0]-A)^2] \tag{10}$$

Sharpness of likelihood function is used to determine how accurately we can find the unknown parameter. to quantify this notion, second derivative of the logarithm of the likelihood function is used.

Fisher Information: Second derivative of Loglikelihood Function is used to determine the sharpness. Expected value of the second derivative of Loglikelihood Function is known as Fisher Information.

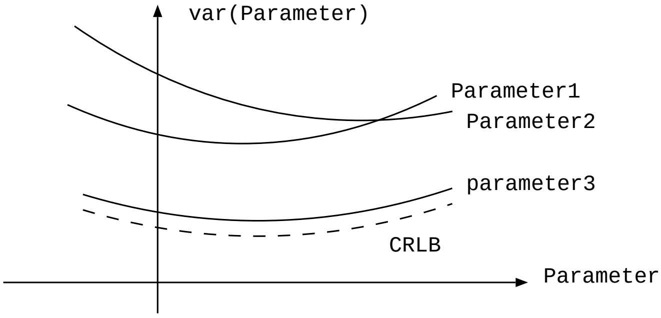

Figure 4: An efficient MVUE

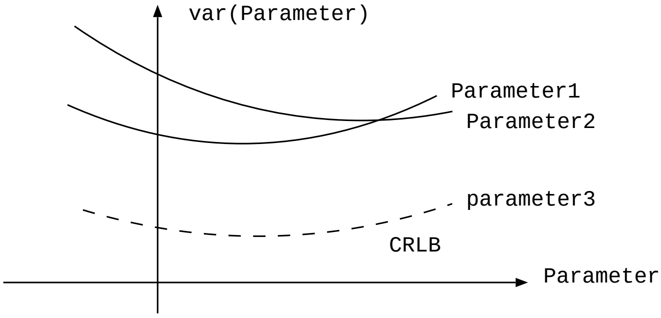

Figure 5: An inefficient MVUE

$$I(\theta) = -E[\frac{\partial^2 \ln p(x[0];\theta)}{\partial \theta^2}] \tag{11}$$

Fisher Information has following properties:

- Always nonnegative.
- Additive for independent observations.
- For completely dependent data Fisher Information does not change.

3.2 Cramer Rao Lower Bound (CRLB)

If a process has regular PDF then variance of any unbiased estimator must satisfy CRLB, mathematically:

$$var(\hat{\theta}) \geq \frac{1}{-E[\frac{\partial^2 \ln p(x[0];\theta)}{\partial \theta^2}]} \tag{12}$$

CRLB is attained when variance is equal to the reciprocal of Fisher Information.

Asymptotic CRLB: For the cases where it is difficult to evaluate the CRLB, then asymptotic CRLB is calculated by using asymptotic Fisher Information which is defined as:

$$[\mathbf{I}(\theta)]_{ij} = \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial P_{xx}(f;\theta)}{\partial \theta_i}\frac{\partial P_{xx}(f;\theta)}{\partial \theta_j} df \tag{13}$$

where $P_{xx}(f;\theta)$ is the PSD of the process with explicit dependence on θ. It is also assumed that process $x[n]$ is zero mean.

3.3 An Application of CRLB

Here we apply CRLB to estimate the parameters of autoregressive (AR) process. It is a commonly used model for speech processing in which data is modelled by a causal all pole filter excited by white gaussian noise. In this model the excitation noise is inherent in the model to make the process wide sense stationary. This model is also known as linear predictive coding model [1]. Here we want to apply the above concepts to get model parameters from the received data. For this purpose, we first define an AR process as:

$$A(z) = 1 + \sum_{m=1}^{p} a[m]z^{-m} \tag{14}$$

and estimated power spectral density is given as:

$$\hat{P}_{xx}(f) = \frac{\hat{\sigma}_u^2}{|1 + \sum_{m=1}^{p} \hat{a}[m]exp(-j2)|^2} \tag{15}$$

The PSD implied by the AR model is

$$P_{xx}(f;\theta) = \frac{\sigma_u^2}{|A(f)|^2} \tag{16}$$

and estimated Fisher Information matrix elements are:

$$[\mathbf{I}(\theta)]_{kl} = \frac{N}{2}\sum_{\frac{-1}{2}}^{\frac{1}{2}} \frac{1}{\sigma_u^4} df = \frac{N}{2\sigma_u^4} \tag{17}$$

and Fisher Information Matrix is:

$$\mathbf{I}(\theta) = \begin{bmatrix} \frac{N}{\sigma_u^2}\mathbf{R}_{xx} & \mathbf{0} \\ \mathbf{0}^T & \frac{N}{2\sigma_u^4} \end{bmatrix} \tag{18}$$

where $[\mathbf{R}_{xx}]_{ij} = r_{xx}[i-j]$ is a $p \times p$ toeplitz autocorrelation matrix and $\mathbf{0}$ is a $p \times 1$ vectors of zeros. Upon inverting the Fisher Information matrix we have

$$var(\hat{a}[k]) \geq \frac{\sigma_u^2}{N}[\mathbf{R}_{xx}^{-1}]_{kk} k = 1, 2, ..., p \tag{19}$$

$$var(\hat{\sigma}_u^2) \geq \frac{2\sigma_u^4}{N} \tag{20}$$

References

[1] J. R. Deller, J. G. Proakis, and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals.* Macmillan Publishing Company, New York, 1993.

[2] Steven M. Kay. *Fundamentals of Statistical Signal Processing Vol. 1 (Estimation Theory).* Prentice Hall, Englewood Cliffs NJ, 1993.