ECSE 6520: Estimation and Detection Theory

Estimation Theory

Class Notes - 6

February 9, 2014

Contents

1 Estimation: Main Ingredients

We will use the following notion as in the previous lectures:

- $X \in \mathcal{X}$ is random measurements or observation
- $\mathcal{X} \subseteq \mathbb{R}^N$ is the sample space of measurement realizations x.
- $\theta \in \Theta$ is the unknown parameter vector of interest.
- $\Theta \subseteq \mathbb{R}^p$ is the parameter space.
- $f_\theta(x)$ or $f(x;\theta)$ is the pdf of X for given θ (a known function).
- With these definitions, the objective of parameter estimation is to design a function

  $$\hat{\theta} : \mathcal{X} \to \Theta$$

  called *estimator* that maps measurements into parameter space.

  - The function $\hat{\theta}$ is the estimator and
  - The point $\hat{\theta}(x)$ for a particular measurement x is called the *estimate*. Note that estimate $\hat{\theta}(x)$ is a random variable.

2 Estimation Categories

There are several ways we can classify estimators and estimation strategies. Here are some of the main ones:

- **Optimality Criterion -** The primary element in an optimality criterion is whether the unknown parameter is viewed as random or *deterministic*.

  θ is *deterministic*: Classical (frequentist) estimation

  - Minimum variance unbiased estimation

- Maximum likelihood estimation
- Method of moments (We will not cover this topic.)
- Least squares

θ is *random* : Bayesian estimation

- Minimum mean square error estimation
- Minimum absolute deviation estimation
- Maximum a posteriori estimation

- **Form of the estimator -** The primary distinction regarding the form of the estimator is whether it is *linear* or *nonlinear*. A linear estimator has the form

$$\hat{\theta}(x) = c^T x.$$

Such estimates arise frequently in conjunction with multivariate Gaussian density function. Because of their simplicity, sometimes the estimator design criterion is constrained with the linear case.

- **Off-line vs On-line estimators -** On-line estimators are those that can be updated with the incoming data, i.e, they are recursive. Off-line estimators require all data to be available upfront to produce an estimate.

3 Classical Estimation

In classical estimation, we assume θ is non-random, that is, unknown but fixed.

Definition 1 ***Bias of an estimator -*** *The bias of $\hat{\theta}$ is*

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

Definition 2 ***Unbiased estimator -*** *We say $\hat{\theta}$ is unbiased if*

$$Bias(\hat{\theta}) = 0 \quad for\ all\ \theta \in \Theta.$$

Otherwise, we say $\hat{\theta}$ is biased.

- An unbiased estimate has a probability distribution where the mean equals the actual value of the parameter.
- If many unbiased estimates are computed from statistically independent sets of observations having the same parameter value, the average of these estimates will be close to this value.
- This property does not mean that the estimate has less error than a biased one; there exist biased estimates whose mean-squared errors (see below) are smaller than unbiased ones.
- Lack of bias is good, but that is just one aspect of how we evaluate estimators.

Definition 3 ***Mean Squared Error -*** *Let $\hat{\theta}$ be an estimate of θ. The mean square error (MSE) of $\hat{\theta}$ is*

$$\begin{aligned} MSE(\hat{\theta}) &:= E[\|\hat{\theta} - \theta\|^2] \\ &= E_X[\|\hat{\theta}(x) - \theta\|^2]. \end{aligned}$$

The estimation error ε is equal to the estimate minus the actual parameter value: $\varepsilon(x) = \hat{\theta} - \theta$. The MSE equals the trace of the mean-squared error matrix $E_X[\varepsilon(x)\varepsilon(x)^T]$, i.e., $MSE(\hat{\theta}) = tr\{E_X[\varepsilon(x)\varepsilon(x)^T]\}$.

Definition 4 ***Covariance of an estimator -*** *The covariance of $\hat{\theta}$ is*

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

Note that in the textbook page 94, the "variance" of a random vector $\hat{\theta}$ is defined as

$$E[\|\hat{\theta} - E[(\hat{\theta})\|^2].$$

This is an unusual convention. To be exact

$$E[\|\hat{\theta} - E[(\hat{\theta})\|^2] = tr\{Cov(\hat{\theta})\} = \sum_{i=1}^{p} Var(\hat{\theta}_i).$$

Definition 5 ***Variance of an scalar estimator -*** *The variance of $\hat{\theta}$ is*

$$Var(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2].$$

Definition 6 ***Asymptotically unbiased and consistent estimator -*** *Let $\{\hat{\theta}_N\}_{N=1}^{\infty}$ be a family of estimators. We say that $\hat{\theta}_N$ is asymtotically unbiased if*

$$Bias(\hat{\theta}_N) \to 0 \quad as \quad N \to \infty \quad for\ all\ \theta \in \Theta$$

and consistent (in the mean squared sense) if

$$MSE(\hat{\theta}_N) \to 0 \quad as \quad N \to \infty \quad for\ all\ \theta \in \Theta.$$

Consistent estimator means MSE tends to zero as the number of observations/measurements becomes large.

Definition 7 ***Efficient estimator -*** *An estimator is said to be efficient if it is unbiased and its covariance achieves a particular lower bound: the Cramer-Rao lower bound (CRLB).*

This is the definition adapted by the book. In other books, you may find the definition of "efficiency" in terms of MSE which is given below.

Definition 8 ***Efficient estimator -*** *An efficient estimate has a MSE that equals a particular lower bound: the Cramer-Rao lower bound (CRLB).*

Comments below apply to the definition of efficiency with respect to MSE.

- If an efficient estimate exists (the Cramer-Rao bound is the greatest lower bound), it is optimum in the mean-squared sense: No other estimate has a smaller mean-squared error.
- For many problems no efficient estimate exists. In such cases, the Cramer-Rao bound remains a lower bound, but its value is smaller than that achievable by any estimator. How much smaller is usually not known.
- Practitioners frequently use the CRLB in comparisons with numerical error calculations.
- Another issue is with the choice of the mean-squared error as the estimation criterion; this criterion may not be sufficient or suitable to assess estimator performance in a particular problem. Nevertheless, every problem is usually subjected to a CRLB computation and the existence of an efficient estimate considered.