Connexions module: m11263

Introduction to Estimation Theory*

Don Johnson

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

Abstract

This module introduces estimation theory and its terminology, including bias, consistency, and efficiency.

In searching for methods of extracting information from noisy observations, this chapter describes estimation theory, which has the goal of extracting from noise-corrupted observations the values of disturbance parameters (noise variance, for example), signal parameters (amplitude or propagation direction), or signal waveforms. Estimation theory assumes that the observations contain an information-bearing quantity, thereby tacitly assuming that detection-based preprocessing has been performed (in other words, do I have something in the observations worth estimating?). Conversely, detection theory often requires estimation of unknown parameters: Signal presence is assumed, parameter estimates are incorporated into the detection statistic, and consistency of observations and assumptions tested. Consequently, detection and estimation theory form a symbiotic relationship, each requiring the other to yield high-quality signal processing algorithms.

Despite a wide variety of error criteria and problem frameworks, the optimal detector is characterized by a single result: the likelihood ratio test. Surprisingly, optimal detectors thus derived are usually easy to implement, not often requiring simplification to obtain a feasible realization in hardware or software. In contrast to detection theory, no fundamental result in estimation theory exists to be summoned to attack the problem at hand. The choice of error criterion and its optimization heavily influences the form of the estimation procedure. Because of the variety of criterion-dependent estimators, arguments frequently rage about which of several optimal estimators is "better." Each procedure is optimum for its assumed error criterion; thus, the argument becomes which error criterion best describes some intuitive notion of quality. When more ad hoc, noncriterion-based procedures¹ are used, we cannot assess the quality of the resulting estimator relative to the best achievable. As shown later², bounds on the estimation error do exist, but their tightness and applicability to a given situation are always issues in assessing estimator quality. At best, estimation theory is less structured than detection theory. Detection is science, estimation art. Inventiveness coupled with an understanding of the problem (what types of errors are critically important, for example) are key elements to deciding which estimation procedure "fits" a given problem well.

1 Terminology in Estimation Theory

More so than detection theory, estimation theory relies on jargon to characterize the properties of estimators. Without knowing any estimation technique, let's use parameter estimation as our discussion prototype. The

^{*}Version 1.2: Jul 9, 2003 4:10 pm -0500

 $^{^\}dagger http://creative commons.org/licenses/by/1.0$

¹This governmentese phrase concisely means guessing.

²"Cramer-Rao Bound" http://cnx.org/content/m11266/latest/

Connexions module: m11263

parameter estimation problem is to determine from a set of L observations, represented by the L-dimensional vector r, the values of parameters denoted by the vector θ . We write the **estimate** of this parameter vector as $\theta(r)$, where the "hat" denotes the estimate, and the functional dependence on r explicitly denotes the dependence of the estimate on the observations. This dependence is always present³, but we frequently denote the estimate compactly as θ . Because of the probabilistic nature of the problems considered in this chapter, a parameter estimate is itself a random vector, having its own statistical characteristics. The **estimation error** $\epsilon(r)$ equals the estimate minus the actual parameter value: $\epsilon(r) = \theta(r) - \theta$. It too is a random quantity and is often used in the criterion function. For example, the **mean-squared error** is given by $E\left[\epsilon^T\epsilon\right]$; the minimum mean-squared error estimate would minimize this quantity. The mean-squared error matrix is $E\left[\epsilon\epsilon^T\right]$; on the main diagonal, its entries are the mean-squared estimation errors for each component of the parameter vector, whereas the off-diagonal terms express the correlation between the errors. The **mean-squared estimation error** $E\left[\epsilon^T\epsilon\right]$ equals the trace of the mean-squared error matrix tr $E\left[\epsilon\epsilon^T\right]$.

1.1 Bias

An estimate is said to be **unbiased** if the expected value of the estimate equals the true value of the parameter: $E\left[\theta\mid\theta\right]=\theta$. Otherwise, the estimate is said to be **biased**: $E\left[\theta\mid\theta\right]\neq\theta$. The **bias** $b\left(\theta\right)$ is usually considered to be additive, so that $b\left(\theta\right)=E\left[\theta\mid\theta\right]-\theta$. When we have a biased estimate, the bias usually depends on the number of observations L. An estimate is said to be **asymptotically unbiased** if the bias tends to zero for large L: $\lim_{L\to\infty}b=0$. An estimate's variance equals the mean-squared estimation error **only** if the estimate is unbiased.

An unbiased estimate has a probability distribution where the mean equals the actual value of the parameter. Should the lack of bias be considered a desirable property? 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 are smaller than unbiased ones. In such cases, the biased estimate is usually asymptotically unbiased. Lack of bias is good, but that is just one aspect of how we evaluate estimators.

1.2 Consistency

We term an estimate **consistent** if the mean-squared estimation error tends to zero as the number of observations becomes large: $\lim_{L\to\infty} E\left[\epsilon^T\epsilon\right]=0$. Thus, a consistent estimate must be at least asymptotically unbiased. Unbiased estimates do exist whose errors never diminish as more data are collected: Their variances remain nonzero no matter how much data are available. Inconsistent estimates may provide reasonable estimates when the amount of data is limited, but have the counterintuitive property that the quality of the estimate does not improve as the number of observations increases. Although appropriate in the proper circumstances (smaller mean-squared error than a consistent estimate over a pertinent range of values of L, consistent estimates are usually favored in practice.

1.3 Efficiency

As estimators can be derived in a variety of ways, their error characteristics must always be analyzed and compared. In practice, many problems and the estimators derived for them are sufficiently complicated to render analytic studies of the errors difficult, if not impossible. Instead, numerical simulation and comparison with lower bounds on the estimation error are frequently used instead to assess the estimator performance.

³Estimating the value of a parameter given no data may be an interesting problem in clairvoyance, but not in estimation theory.

Connexions module: m11263 3

An efficient estimate has a mean-squared error that equals a particular lower bound: the Cramér-Rao bound⁴. If an efficient estimate exists (the Cramér-Rao bound is the greatest lower bound), it is optimum in the mean-squared sense: No other estimate has a smaller mean-squared error (see Maximum Likelihood Estimators⁵ for details).

For many problems no efficient estimate exists. In such cases, the Cramér-Rao bound remains a lower bound, but its value is smaller than that achievable by any estimator. How much smaller is usually not known. However, practitioners frequently use the Cramér-Rao bound in comparisons with numerical error calculations. Another issue is the choice of mean-squared error as the estimation criterion; it may not suffice to pointedly assess estimator performance in a particular problem. Nevertheless, every problem is usually subjected to a Cramér-Rao bound computation and the existence of an efficient estimate considered.

^{4&}quot;Cramer-Rao Bound" http://cnx.org/content/m11266/latest/

 $^{{}^5\}text{"Maximum Likelihood Estimators of Parameters"} < \text{http://cnx.org/content/m11269/latest/}>$