Elements of Parameter Estimation

IV

IV.A Introduction

In Chapters II and III we have considered the design of optimum procedures for deciding between two possible statistical situations on the basis of a random observation $Y$. In many situations arising in practice we are interested not in making a choice between two (or among several) discrete situations, but rather in making a choice among a continuum of possible states of nature. In particular, as in the composite hypothesis-testing problems discussed in Chapter II, we can think of a family of distributions on the observation space, indexed by a parameter or set of parameters. But unlike the case of composite hypothesis testing in which we wish to make a binary decision about the parameter, we wish here to determine as accurately as possible the actual value of the parameter from the observation.

Such problems are known as parameter (or point) estimation problems, and in this chapter we discuss the basic ideas relating to the design of optimum procedures for estimating parameters. As with the hypothesis-testing problem (which incidentally can be thought of as a special case of the parameter estimation problem), a variety of estimation design philosophies can be used, these differing primarily in the amount of prior information known about the parameter and in the performance criteria applied.

In this chapter we discuss two basic approaches to parameter estimation—one, the Bayesian, in which the parameter is assumed to be a random quantity related statistically to the observation, and a second in which the parameter is assumed to be unknown but without being endowed with any probabilistic structure. Of these two approaches, the Bayesian is the most straightforward and so is considered first, in Section IV.B, with nonrandom parameter estimation being considered in the remainder of the chapter.

It should be noted that in this treatment, we consider only the estimation of parameters that are static, i.e., that are constant in time. The estimation of dynamic parameters (i.e., signals) is considered in Chapter V.
IV. Bayesian Parameter Estimation

Throughout this chapter we assume as a model a family of distributions for the random observation $Y$, indexed by a parameter $\theta$ taking values in a parameter set $\Lambda$; i.e., we have the family $\{P_\theta; \theta \in \Lambda\}$, where $P_\theta$ denotes a distribution on the observation space $(\Gamma, \mathcal{G})$. We also assume that the parameter set $\Lambda$ is a subset of $\mathbb{R}^m$ for some $m$. Within this model the goal of the parameter estimation problem is to find a function $\hat{\theta}(y)$ is the “best” guess of the true value of $\theta$ (i.e., the value of $\theta$ for which $Y \sim P_\theta$) based on the observation $Y = y$.

Of course, the solution to this problem depends on the criterion of goodness by which we measure estimation performance; so, as in the hypothesis-testing problem, we begin by assigning costs to our decisions about the parameter. In particular, we suppose that there is a function $C: \Lambda \times \Lambda \rightarrow \mathbb{R}$ such that $C[a, \theta]$ is the cost of estimating a true value of $\theta$ as $a$, for $a$ and $\theta$ in $\Lambda$. Given such a function $C$ we can then associate with an estimator $\hat{\theta}$ a conditional risk or cost averaged over $Y$ for each $\theta \in \Lambda$; i.e., we have

$$R_\theta(\hat{\theta}) = \mathbb{E}_\theta\{C[\hat{\theta}(Y), \theta]\}. \quad (IV.B.1)$$

If we now adopt the interpretation that the actual parameter value $\theta$ is the realization of a random variable $\Theta$, we can define an average or Bayes risk as

$$r(\hat{\theta}) = \mathbb{E}\{R_\Theta(\hat{\theta})\}, \quad (IV.B.2)$$

and the appropriate design goal is to find an estimator minimizing $r(\hat{\theta})$. Such an estimator is known as a Bayes estimate of $\theta$.

Noting that $R_\theta(\hat{\theta}) = \mathbb{E}\{C[\hat{\theta}(Y), \Theta]|\Theta = \theta\}$, we have

$$r(\hat{\theta}) = \mathbb{E}\{C[\hat{\theta}(Y), \Theta]\} = \mathbb{E}\{\mathbb{E}\{C[\hat{\theta}(Y), \Theta]|Y\}\}. \quad (IV.B.3)$$

By inspection of (IV.B.3) we see that the Bayes estimate of $\theta$ can be found (if it exists) by minimizing, for each $y \in \Gamma$, the posterior cost given $Y = y$:

$$\mathbb{E}\{C[\hat{\theta}(y), \Theta]|Y = y\}. \quad (IV.B.4)$$

This is the same procedure as that followed in the Bayesian hypothesis-testing problem (see Section II.E). Note that if we assume that $\Theta$ has a conditional density $w(\theta|y)$ given $Y = y$ for each $y \in \Gamma$, then the Bayes estimate $\hat{\theta}(y)$ corresponding to $y \in \Gamma$ can be sought by minimizing

$$\int_\Lambda C[\hat{\theta}(y), \theta]w(\theta|y)\mu(d\theta). \quad (IV.B.5)$$

The following cases illustrate the application of this criterion.