EFFICIENT NONPARAMETRIC TESTING AND ESTIMATION

CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

It is customary to treat nonparametric statistical theory as a subject completely different from parametric theory. In this paper, I try to study one of the more obvious connections between the two subjects. Clearly a nonparametric problem is at least as difficult as any of the parametric problems obtained by assuming we have enough knowledge of the unknown state of nature to restrict it to a finite-dimensional set. For a problem in which one wants to estimate a single real-valued function of the unknown state of nature it frequently happens that (in a sense made somewhat more precise in section 2 and, for special cases, in later sections) there is, through each state of nature, a one-dimensional problem which is, for large samples, at least as difficult (to a first approximation) as any other finite-dimensional problem at that point. If a procedure does essentially as well, for large samples, as one could do for each such one-dimensional problem, one is justified in considering the procedure efficient for large samples. If there is no such procedure, one may be forced to adopt a less severe definition of efficiency, as suggested by Wolfowitz [1].

Very few results are obtained here, and, with the exception of the lemma of section 3, they are not rigorous. Also, even for the example of section 4, where a definite procedure is given, the results are not of immediate practical value. The computations required are excessive, and the procedure is not efficient for sample sizes likely to occur in practice.

2. A review of the finite-dimensional case

Let Θ be an open subset of a finite-dimensional Euclidean space. For each $\theta \in \Theta$, let p_θ be a probability density with respect to a σ-finite measure μ on a σ-algebra $\mathcal{B}$ of subsets of a space $\mathcal{X}$. Subject to certain differentiability conditions and other regularity conditions (see for example [2]), the maximum likelihood estimate $\hat{\theta}$ of θ, based on a large sample $X_1, \cdots, X_n$ independently distributed according to p_θ for some unknown $\theta \in \Theta$, has certain desirable properties.

We define Fisher's information matrix $I(\theta)$ at θ by

$$I_{ij}(\theta) = E_\theta \frac{\partial \log p_\theta(X)}{\partial \theta_i} \frac{\partial \log p_\theta(X)}{\partial \theta_j}. \tag{1}$$

If φ is a continuously differentiable real-valued function on Θ, then the asymptotic mean-squared error of $\varphi(\hat{\theta})$ as an estimate of $\varphi(\theta)$ is

$$\frac{1}{n}(\nabla\varphi)' I^{-1} (\nabla\varphi) \tag{2}$$

Research carried out partly under sponsorship of the Office of Naval Research.