ESTIMATION WITH QUADRATIC LOSS

W. JAMES
FRESNO STATE COLLEGE
AND
CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

It has long been customary to measure the adequacy of an estimator by the smallness of its mean squared error. The least squares estimators were studied by Gauss and by other authors later in the nineteenth century. A proof that the best unbiased estimator of a linear function of the means of a set of observed random variables is the least squares estimator was given by Markov [12], a modified version of whose proof is given by David and Neyman [4]. A slightly more general theorem is given by Aitken [1]. Fisher [5] indicated that for large samples the maximum likelihood estimator approximately minimizes the mean squared error when compared with other reasonable estimators. This paper will be concerned with optimum properties or failure of optimum properties of the natural estimator in certain special problems with the risk usually measured by the mean squared error or, in the case of several parameters, by a quadratic function of the estimators. We shall first mention some recent papers on this subject and then give some results, mostly unpublished, in greater detail.

Pitman [13] in 1939 discussed the estimation of location and scale parameters and obtained the best estimator among those invariant under the affine transformations leaving the problem invariant. He considered various loss functions, in particular, mean squared error. Wald [18], also in 1939, in what may be considered the first paper on statistical decision theory, did the same for location parameters alone, and tried to show in his theorem 5 that the estimator obtained in this way is admissible, that is, that there is no estimator whose risk is no greater at any parameter point, and smaller at some point. However, his proof of theorem 5 is not convincing since he interchanges the order of integration in (30) without comment, and it is not clear that this integral is absolutely convergent. To our knowledge, no counterexample to this theorem is known, but in higher-dimensional cases, where the analogous argument seems at first glance only slightly less plausible, counterexamples are given in Blackwell [2] (which is discussed briefly at the end of section 3 of this paper) and in [14],

This work was supported in part by an ONR contract at Stanford University.