NON-PARAMETRIC STATISTICAL INFERENCE

J. WOLFOWITZ
COLUMBIA UNIVERSITY

1. Introduction

In most statistical problems treated in the literature a datum of the problem is the information that the various distributions of the chance variables involved belong to given families of distribution functions (d.f.'s) completely specified except for one or more parameters. Non-parametric statistical inference is concerned with problems where the d.f.'s are not specified to such an extent, and where their functional form is unknown. This does not preclude some knowledge of the d.f.'s; for example, we may know that they are continuous, uni-modal, bi-modal, and the like.

It is clear that the more information that is available from which to draw inferences the more decisive can our conclusions be, that is, the confidence regions may have smaller average size, the statistical tests will have greater power, and the like. Hence if the functional forms of the d.f.'s are known or if there is good ground for assuming them, it is a loss not to make use of this information. Where this information is not at hand, statistical inference must properly proceed without it. In the latter event the criticism of some statisticians that non-parametric tests are "inefficient" is not valid, because "efficiency" (in the colloquial sense) implies thorough use of available resources, and it cannot be inefficient not to make use of unavailable information. Statistical efficiency must be appraised in the context of available information and, except where uniformly most powerful procedures exist, with respect to specific alternatives.

In the present paper we shall describe briefly a few recent advances in non-parametric theory. Readers who expect a complete, unified theory such as may be found in the analysis of variance will be disappointed; it is impossible to present such a theory because none exists. What has been accomplished thus far is only a series of small advances in various directions. It is as if, faced by a hardy opponent, one lashed out in all directions and succeeded in penetrating the enemy's armor slightly in several places. The analogy is correct to the further extent that many problems in non-parametric inference are of considerable difficulty. It is the hope of the author of the present paper to arouse the interest of the readers of these proceedings, to acquaint them with a few of the developments, and to enlist their aid in solving the multitude of outstanding problems.

2. Estimation

Let $x_1, \cdots, x_n$ be n independent observations on the chance variable X, about whose cumulative d.f., $f(x)$, nothing is known except that it is continuous. We shall discuss the problem of estimating $f(x)$.