The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 1–23

A DISTRIBUTION FREE VERSION OF THE SMIRNOV TWO SAMPLE TEST IN THE p-VARIATE CASE[1]

BY P. J. BICKEL

University of California, Berkeley

1. Introduction. One of the classic problems of the theory of nonparametric inference is testing whether two samples come from the same or different populations. If the observations are univariate and we suppose only that the parent populations are governed by a continuous distribution function, genuinely distribution free and asymptotically consistent tests have been proposed by Smirnov, Von Mises, and Lehmann among others. For a review of the enormous literature that has grown up about these and related tests we refer to [1] and [3]. The multivariate case seems to have been studied far less fully. The obvious generalization of the univariate methods leads to procedures which are not distribution free (see [17]). Lehmann [12] has proposed a genuinely distribution free test consistent against all alternatives in this instance also, but his method involves post experimental randomization as an intrinsic factor and is therefore not fully satisfactory. Rosenblatt in [14] has given an ingenious solution to the goodness of fit problem in the multivariate case but his device does not seem to carry over to the two sample problem. Recently, David and Fix [2] and Sen and Chatterjee [16] have remarked that the classical permutation principle of Fisher, as used for instance by Wald and Wolfowitz in [19], when applied to rank tests in multivariate problems leads to procedures which are distribution free and which can, in principle at least, be tabled.

Our approach consists of combining this remark with the generalization of the classical Smirnov test to obtain a procedure which we show is distribution free and consistent against all alternatives. To achieve this aim we require a theorem on the "conditional" convergence of the empirical distribution function which is analogous to known results on permutation distributions (cf. [19], [8]). This theorem is then applied to yield known results of Kiefer and Wolfowitz [11] and Dudley [6].

The paper is organized as follows: Section 2 contains a formal statement of the model and our proposed procedure. In Section 3 we introduce the stochastic process considerations we need, state the main theorems on convergence of stochastic processes ((3.1) and (3.2)) and derive the consistency of our procedure from these as well as the previously mentioned results of Dudley and Kiefer and Wolfowitz. Section 4 deals with some auxiliary results on sampling from a finite population necessary for the proof of Theorem 3.1. Two of these results are due to Hájek ([8] and [9]), and one to Wald, Wolfowitz, Noether and Hoeffding

Received 11 September 1967.

[1] This paper was prepared with the partial support of the Office of Naval Research Contract N00014-67-A-0114-0004.