THE TREATMENT OF TIES IN SOME NONPARAMETRIC TESTS[1]

By Joseph Putter

University of California, Berkeley

1. Introduction. Most of nonparametric testing theory is usually presented under the assumption that all the samples involved are drawn from continuous distributions, and that tied observations can therefore be ignored or treated in any convenient way, without affecting the performance characteristic of the test. In practice, however, this assumption is not a realistic one, and the distributions involved are in general to be regarded as discontinuous, either because of intrinsic reasons (integer-valued or otherwise discrete random variables) or because of limitations on the precision of measurements. Therefore, usually, ties will occur with positive probability, and the way they are treated does affect the performance characteristic of the test. The problem of ties has therefore to be considered, in particular with a view to preserving the nonparametric character of the test, and to making sure of setting it up on the desired level of significance.

The usual practice in attacking the problem has been to consider the conditional distributions of the statistics concerned given that the number of observations in each tied group is a fixed constant. This, however, was never explicitly made clear, and these conditional distributions, as well as their variances and other characteristics, are referred to as distributions (or variances, etc.) "when ties are present." In this category belong Kendall's work on ties in rank correlation theory, and Kruskal's theorem concerning a generalized Wilcoxon test (see Section 8).

In this paper, we attack the problem from the standpoint of the ties being random variables. Our main concern is the comparison between the "randomized" and the "nonrandomized" way of treating the ties. In Sections 3 and 4 we consider the one-sided sign test, and show that randomization reduces both the exact power and the asymptotic efficiency of the test. In Sections 5–8 we consider the Wilcoxon test. For small samples the nonrandomized treatment of ties presents practical difficulties, but the asymptotic (large sample) problem can be handled. Again, it is shown that randomization results in reduced efficiency.

2. Notation and theorems used. We shall use the notation $\mathfrak{N}(a, b)$ for normal random variables (with mean a and variance b), and $\mathfrak{B}(n, p)$ for binomials. The symbol $\xrightarrow{P}$ will denote convergence in probability, and $\xrightarrow{L}$ convergence in law (convergence of distributions).

Received March 16, 1954.

[1] Work supported in part by the Office of Naval Research, and submitted in partial satisfaction of the requirements for the degree of Ph.D. at the University of California, Berkeley.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org