The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1661–1674

THE INFORMATION IN A RANK-ORDER AND THE STOPPING TIME OF SOME ASSOCIATED SPRT'S[1]

BY ROBERT H. BERK AND I. RICHARD SAVAGE

University of Michigan and Florida State University

1. Introduction. The following is a contribution to the theory of certain SPRT's based on ranks. Such procedures were suggested by Wilcoxon and further discussed by several writers [1], [2], [4]. We consider the two-sample problems of sequentially testing the null hypothesis $G = F$ against the alternative $\phi: F = \phi_1(K), G = \phi_2(K)$, where ϕ_1 and ϕ_2 are given CDF's on [0, 1] that specify ϕ and K ranges through all CDF's. (In cases of interest, as in the null hypothesis, ϕ is usually of the form $G = \phi(F)$, where ϕ is a CDF on [0, 1].) At stage n, one has n observations from each population, which provide the usual rank-order statistic. It is well-known that under either hypothesis, the distribution of the rank-order depends only on the corresponding ϕ; hence one may compute $\mathbf{L}_n$, the likelihood ratio for the rank-order based on the two hypotheses. One continues sampling until $\mathbf{L}_n$ leaves an interval (B, A), $0 < B < 1 < A < \infty$.

The theory presented by Hall, Wijsman and Ghosh (1965) shows that these procedures are indeed SPRT's: i.e., under either hypothesis, the nth rank-order is sufficient for the first n sampling stages. Theoretical knowledge about these procedures is limited, although some Monte Carlo studies are reported by Bradley, Merchant and Wilcoxon (1966). Sethuraman (1967), improving results obtained jointly with Savage (1966), established (with little restriction on the true distributions) that when ϕ is a Lehmann alternative, the stopping variable defined by the SPRT has a finite moment generating function (in a neighborhood of zero). As a further contribution along these lines, we consider more general ϕ and establish the almost-sure convergence of $n^{-1} \ln \mathbf{L}_n$. The convergence is rapid enough to insure that the stopping variable has a finite moment generating function when the limit (which depends on the true distributions) is not zero. The limit is identified as the difference of two information numbers and in the particular case of a Lehmann alternative, with that obtained by Savage and Sethuraman. A by-product of our development is a strengthening of the Glivenko-Cantelli theorem and a similar theorem for a statistic equivalent to the rank-order (Section 5).

2. The main theorem. Let $(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ denote the two samples at stage n and F^* and G^*, the sampled populations. (F^*, G^*) need not be in either of the hypotheses defining the sequential rank-test. Let $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_{2n})$ denote the rank-order of the combined sample: $\mathbf{Z}_k = 0$ or 1 according as the kth observation in the combined ordered sample is an X or a Y.

Received 26 July 1967.
[1] Work supported by the Office of Naval Research Contract No. Nonr 988(13) and National Science Foundation Grant GP-6008.