ASYMPTOTIC NORMALITY OF LINEAR RANK STATISTICS UNDER ALTERNATIVES*

By Jaroslav Hájek

CHARLES UNIVERSITY, PRAGUE, AND UNIVERSITY OF CALIFORNIA, BERKELEY

Communicated by Jerzy Neyman, November 10, 1966

Let $X_1,\ldots,X_N$ be independent random variables with continuous distribution functions $F_1,\ldots,F_N$, respectively. Let R_i be the number of observations which are equal to or smaller than X_i, $1 \leqq i \leqq N$. Let $c_1,\ldots,c_N$ and $a(1),\ldots,a(N)$ be arbitrary real numbers. A linear rank statistics is then defined as follows:

$$S = \sum_{i=1}^{N} c_i a(R_i). \tag{1}$$

If $c_i = 1$, $1 \leqq i \leqq m < N$, and $= 0$, $m < i \leqq N$, we have the so-called two-sample problem.

Linear rank statistics are essential in nonparametric statistical theory. The central problem concerning these statistics is to establish conditions under which they are asymptotically normal, either with natural parameters (ES, var S), or with some other parameters (μ, σ^2). Assume that the scores $a(i)$ are generated by a function $\varphi(t)$, $0 < t < 1$, as follows:

$$a(i) = \varphi\left(\frac{i}{N+1}\right),\ 1 \leqq i \leqq N \tag{2}$$

or

$$a(i) = E\varphi(U_N^{(i)}), \tag{3}$$

where $U_N^{(i)}$ is the ith order statistic in a sample from the uniform distribution on (0,1).

Condition A: We have $\varphi(t) = \varphi_1(t) - \varphi_2(t)$, where $\varphi_i(t)$ is nondecreasing, square integrable, and absolutely continuous on every interval $(\epsilon, 1-\epsilon)$, $0 < \epsilon < {}^1/_2$.

Theorem. *Let the scores be given by (2) or (3) with $\varphi(t)$ satisfying Condition A. Then for every $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)$ there exists an N_ϵ such that $N > N_\epsilon$,*

$$\frac{\sum_{i=1}^{N}(c_i - \bar{c})^2}{N \max_{i\leqq i\leqq N} (c_i - \bar{c})^2} > \epsilon_1,\ \bar{c} = \frac{1}{N}\sum_{i=1}^{N} c_i, \tag{4}$$

and

$$\text{var}\, S > \epsilon_2 \sum_{i=1}^{N} (c_i - \bar{c})^2 \tag{5}$$

entail

$$\max_{-\infty<x<\infty} \left\{P(S - ES < x\sqrt{\text{var}\, S}\right\} - (2\pi)^{-1/2}\int_{-\infty}^{x} \exp(-{}^1/_2 y^2)dy| < \epsilon_3. \tag{6}$$

Conclusion (6) also holds with var S replaced by

$$\sigma^2 = \sum_{i=1}^{N} \operatorname{var}\left[\frac{1}{N} \sum_{j=1}^{N} (c_j - c_i) \int_{x_0}^{X_i} \varphi'(N^{-1} \sum_{k=1}^{N} F_k(x)) dF_j(x)\right] \tag{7}$$

Moreover, if

$$\max_{1 \leq i,j < N} \max_{-\infty < x < \infty} |F_i(x) - F_j(x)|$$

is sufficiently small, then var S may be replaced by

$$d^2 = \sum_{i=1}^{N} (c_i - \bar{c})^2 \int_0^1 [\varphi(t) - \bar{\varphi}]^2 dt. \tag{8}$$

This theorem modifies or generalizes the result by Chernoff-Savage[1] and the more recent result by Govindarajulu-LeCam-Raghavachari[2] in three main respects. First, instead of the two-sample case (zero-one c_i's), the general case is considered. Second, the class of scores generating function is larger (considerably compared to Chernoff-Savage and slightly compared to Govindarajulu-LeCam-Raghavachari). Third, the parameters are natural. On the other hand, in the two last-mentioned papers it is proved that ES in (6) may be replaced by

$$\mu = \sum_{i=1}^{N} c_i E\varphi(N^{-1} \sum_{k=1}^{N} F_k(X_i)), \tag{9}$$

which is not asserted in the present theorem.[3] Moreover, Govindarajulu-LeCam-Raghavachari showed that the assertion of the theorem holds uniformly on some subsets of the scores functions.

* This research was partly supported by the National Science Foundation, grant GP-5059.

[1] Chernoff, H., and I. R. Savage, *Ann. Math. Statist.*, **29**, 872 (1958).

[2] Govindarajulu, Z., L. LeCam, and M. Raghavachari, *Proc. Fifth Berkeley Symp. Prob. Statist.*, in press (1966).

[3] Hájek, J., submitted to *Ann. Math. Statist.* (1966).