The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 1889–1900

ASYMPTOTIC LINEARITY OF A RANK STATISTIC IN REGRESSION PARAMETER

By Jana Jurečková

Charles University, Prague

0. Introduction. Let $(X_1, X_2, \cdots, X_N)$ be an independent random sample from a distribution with finite Fisher's information and let us consider the statistic

$$S_{\Delta N} = \sum_{i=1}^N c_i a_N(R_{Ni}^\Delta)$$

where $R_{N1}^\Delta, R_{N2}^\Delta, \cdots, R_{NN}^\Delta$ is the vector of ranks for random variables $X_1 + \Delta d_1$, $X_2 + \Delta d_2, \cdots, X_N + \Delta d_N$; Δ, c_i and d_i, $1 \leqq i \leqq N$ are real constants. Then $\{S_{\Delta N}; -\infty < \Delta < \infty\}$ forms a random process. We show at first that under some assumptions the realizations of this process are monotone step-functions of Δ and that these realizations are asymptotically linear in Δ in the sense of the formula (3.1) of Theorem 3.1. The asymptotic linearity of $S_{\Delta N}$ may be proved also in the case of K-variate regression, when instead of R_{Ni}^Δ's there will occur the ranks of the values $X_1 + \Delta_1 d_{11} + \Delta_2 d_{21} + \cdots + \Delta_K d_{K1}, \cdots, X_N + \Delta_1 d_{1N} + \cdots + \Delta_K d_{KN}$; the statistic $S_{\Delta N}$ is then an asymptotically linear function of the parameters $\Delta_1, \Delta_2, \cdots, \Delta_K$.

Some possibilities of application are mentioned.

1. Notation and basic assumptions. We shall consider for any positive integer N:

(a) an independent random sample $(X_{N1}, X_{N2}, \cdots, X_{NN})$ from a distribution whose distribution function F has finite Fisher's information, i.e.

$$\int_{-\infty}^{\infty} [f'(x)/f(x)]^2 f(x)\, dx < \infty,$$

where f is the density of the distribution;

(b) a real vector $(c_{N1}, c_{N2}, \cdots, c_{NN})$ (so called regression constants) such that

$$\sum_{i=1}^N (c_{Ni} - \bar{c}_N)^2 > 0. \tag{1.1}$$

$$\lim_{N\to\infty} \max_{1 \leqq i \leqq N} (c_{Ni} - \bar{c}_N)^2 \cdot [\sum_{j=1}^N (c_{Nj} - \bar{c}_N)^2]^{-1} = 0 \tag{1.2}$$

where $\bar{c}_N = (1/N) \sum_{i=1}^N c_{Ni}$.
Condition (1.2) is the so called Noether's condition.

(c) a real vector $(d_{N1}, d_{N2}, \cdots, d_{NN})$ such that

$$\sum_{i=1}^N (d_{Ni} - \bar{d}_N)^2 \leqq M \quad \text{for} \quad N = 1, 2, \cdots \tag{1.3}$$

where $M > 0$ is a constant, $\bar{d}_N = (1/N) \sum_{i=1}^N d_{Ni}$ and

$$\max_{1 \leqq i \leqq N} (d_{Ni} - \bar{d}_N)^2 \to 0 \quad \text{for} \quad N \to \infty. \tag{1.4}$$

(d) a real parameter Δ.

Received 31 May 1968; revised 7 April 1969.

(e) vector of ranks $(R_{N1}^{\Delta}, R_{N2}^{\Delta}, \cdots, R_{NN}^{\Delta})$ corresponding to variables $X_{N1} + \Delta d_{N1}, X_{N2} + \Delta d_{N2}, \cdots, X_{NN} + \Delta d_{NN}$, i.e. R_{Ni}^{Δ} is equal to the number of $(X_{Nj} + \Delta d_{Nj})$'s which are less or equal to $(X_{N1} + \Delta d_{Ni})$ provided all components $X_{N1} + \Delta d_{N1}, \cdots, X_{NN} + \Delta d_{NN}$ are different.

REMARK. For simplicity of notation, we shall omit indices N in X_{Ni}, c_{Ni}, d_{Ni}, and R_{Ni}^{Δ} in the sequel; we hope that this simplification will not cause confusion.

(f) Let us consider the statistic

$$S_{\Delta N} = \sum_{i=1}^{N} c_i a_N(R_i^{\Delta}) \tag{1.5}$$

where the scores $a_N(1), a_N(2), \cdots, a_N(N)$ are generated by a nondecreasing function $\varphi \varepsilon L^2(0, 1)$ such that $\int_0^1 (\varphi(u) - \bar{\varphi})^2 du > 0, \bar{\varphi} = \int_0^1 \varphi(u)\, du$, either by

$$a_N(i) = E\varphi(U^{(i)}), \qquad i = 1, 2, \cdots, N \tag{1.6}$$

or by

$$a_N(i) = \varphi(i/N + 1), \qquad i = 1, 2, \cdots, N \tag{1.7}$$

where $U^{(i)}$ is the ith smallest variable in the sample from the uniform (0, 1) distribution.

2. Monotonicity of $S_{\Delta N}$.

THEOREM 2.1. *Let N be any positive integer. If the assumptions* (a), (d), (e), (f) *are satisfied with real numbers $c_1, c_2, \cdots, c_N$ and $d_1, d_2, \cdots, d_N$ satisfying*

$$(c_i - c_j)(d_i - d_j) \geqq 0, \qquad [(c_i - c_j)(d_i - d_j) \leqq 0] \tag{2.1}$$

for all $i, j = 1, 2, \cdots, N$, then the statistic $S_{\Delta N}$ with the monotone scores $a_N(i)$, $i = 1, 2, \cdots, N$ is a non-decreasing [*(non-increasing)*] *step-function of Δ with probability* 1.

PROOF. It suffices to consider the case

$$(c_i - c_j)(d_i - d_j) \geqq 0 \quad \text{for} \quad i, j = 1, 2, \cdots, N. \tag{2.2}$$

We may suppose without any loss of generality that the variables are so numerated that

$$d_1 \leqq d_2 \leqq \cdots \leqq d_N. \tag{2.3}$$

Let us fix a vector $(x_1, x_2, \cdots, x_N)$ with different components and let Δ_1 and Δ_2, $\Delta_1 < \Delta_2$ be two values of the parameter Δ such that $S_{\Delta_j N}$ is well-defined for $j = 1, 2$. Let $k < l$ and $R_k^{\Delta_1} < R_l^{\Delta_1}$ then $x_k + \Delta_1 d_k < x_l + \Delta_1 d_l$ and this implies $x_k + \Delta_2 d_k < x_l + \Delta_2 d_l$ in view of (2.3), thus $R_k^{\Delta_2} < R_l^{\Delta_2}$. This means that the permutation $(R_1^{\Delta_2}, \cdots, R_N^{\Delta_2})$ is better ordered than $(R_1^{\Delta_1}, \cdots, R_N^{\Delta_1})$ in the sense of Lehmann's definition in [4]. By a slight generalization of Corollary 2 of Theorem 5 in [4] and with reference to (2.2) and (2.3), it follows that

$$\sum_{i=1}^{N} c_i a_N(R_i^{\Delta_1}) \leqq \sum_{i=1}^{N} c_i a_N(R_i^{\Delta_2})$$

and $S_{\Delta N}$ is a nondecreasing function of Δ.

Let us consider the set

$$C = \{\Delta; x_i + \Delta d_i = x_j + \Delta d_j \text{ for at least one pair } (i, j)\}.$$

The set C is finite; if $C = \{\Delta^{(1)}, \Delta^{(2)}, \cdots, \Delta^{(r)}\}$, then $S_{\Delta N}$ is not defined in the points $\Delta^{(1)}, \Delta^{(2)}, \cdots, \Delta^{(r)}$. Let (Δ_*, Δ^*) be any of the intervals $(-\infty, \Delta^{(1)})$, $(\Delta^{(1)}, \Delta^{(2)}), \cdots, (\Delta^{(r)}, \infty)$ and Δ_0 be any point of this interval; let us consider the points $\Delta \geqq \Delta_0$. Then $S_{\Delta N} = S_{\Delta_0 N}$ holds for all Δ satisfying

$$(2.4) \quad \Delta_0 \leqq \Delta < \Delta_0 + \min_{(i,j)\varepsilon A(\Delta_0)} [x_i + \Delta_0 d_i - (x_j + \Delta_0 d_j)](d_j - d_i)^{-1}$$

where $A(\Delta_0)$ is the set of those pairs (i, j), $(1 \leqq i, j \leqq N)$, for which $x_i + \Delta_0 d_i > x_j + \Delta_0 d_j$ and $d_i < d_j$. If $A(\Delta_0)$ is empty, then $S_{\Delta N}$ is constant for $\Delta \geqq \Delta_0$. If we denote the right-hand side of (2.4) as $\bar{\Delta}$, then by definition of C, $\bar{\Delta} \varepsilon C$ so that $\bar{\Delta} = \Delta^* \cdot S_{\Delta N}$ is then constant on semiclosed interval $[\Delta_0, \Delta^*)$. Since $\Delta_0 > \Delta_*$ is arbitrary, $S_{\Delta N}$ is constant on (Δ_*, Δ^*). The proof is complete.

REMARK. We may complete the definition of $S_{\Delta N}$ at the points of discontinuity as to be continuous either from the left or from the right. We shall suppose that $S_{\Delta N}$ is well-defined for all real Δ in the sequel.

3. Asymptotic linearity of $S_{\Delta N}$ in Δ.

THEOREM 3.1. *Let the assumptions* (a)–(f) *be satisfied. If* $(c_i - c_j)(d_i - d_j) \geqq 0$, $[(c_i - c_j)(d_i - d_j) \leqq 0]$ *for* $i, j = 1, 2, \cdots, N$ *and* $N = 1, 2, \cdots$, *then*

$$(3.1) \quad \lim_{N\to\infty} P\{\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N| \geqq \epsilon(\operatorname{Var} S_{0N})^{\frac{1}{2}}\} = 0$$

for any $\epsilon > 0$ *and* $C > 0$. *Here* b_N *denotes*

$$(3.2) \qquad b_N = \int_0^1 \varphi(u)\varphi(u, f)\, du \cdot [\textstyle\sum_{i=1}^N (c_i - \bar{c})(d_i - \bar{d})] \qquad \textit{and}$$

$$(3.3) \qquad \varphi(u, f) = -f'(F^{-1}(u))/f(F^{-1}(u)).$$

PROOF. We first shall prove the theorem for the scores (1.6). We may suppose without loss of generality that

$$(3.4) \qquad \textstyle\sum_{i=1}^N c_i = 0 \quad \text{and} \quad \sum_{i=1}^N c_i^2 = 1 \quad \text{for} \quad N = 2, 3, \cdots.$$

We shall use the following metric on the space $\mathfrak{M}$ of random variables defined on a probability space $(\Omega, \mathfrak{A}, P)$: $d(X, Y)$, $X, Y \varepsilon \mathfrak{M}$, is defined as

$$(3.5) \qquad d(X, Y) = \inf\{\epsilon > 0; P(|X - Y| \geqq \epsilon) < \epsilon\}.$$

It is known that $d(\cdot, \cdot)$ is a metric on the space of disjoint classes of equivalence of $\mathfrak{M}$ where $X \sim Y$ if and only if $X = Y$ a.e. $[P]$; further,

$$\lim_{n\to\infty} d(X_n, X) = 0$$

if and only if $X_n \to X$ in probability (see e.g. [5]).

The proof of the theorem is divided into several steps.

(i) For $k = 1, 2, \cdots$, let us consider the functions $\varphi^{(k)}(u)$ defined as

$$(3.6) \qquad \varphi^{(k)}(u) = \varphi(i/k + 1) \qquad (i - 1)/k \leqq u < i/k$$

for $i = 1, 2, \cdots, k$.

Each function $\varphi^{(k)}$ is non-decreasing and bounded on the interval (0, 1). By Lemma V.1.6.a of [1] we have

$$\lim_{k\to\infty} \int_0^1 [\varphi^{(k)}(u) - \varphi(u)]^2 \, du = 0. \tag{3.7}$$

Let us consider the statistics

$$S_{\Delta N}^{(k)} = \sum_{i=1}^N c_i a_N^{(k)}(R_i^\Delta) \tag{3.8}$$

where the scores $a_N^{(k)}(i)$ are generated by $\varphi^{(k)}$ about (1.6).

LEMMA 3.1. *Corresponding to every positive number ϵ, there is a positive integer k_0 such that for any positive integer $k > k_0$ an index $N_0(k)$ may be found such that*

$$d(S_{0N}, S_{0N}^{(k)}) < \epsilon \quad \text{for any} \quad N > N_0(k). \tag{3.9}$$

PROOF. The inequality (V.1.6.6) of [1] implies

$$\begin{aligned} E[S_{0N} - S_{0N}^{(k)}]^2 &\leqq (1/N - 1)[\textstyle\sum_{i=1}^N c_i^2] \sum_{j=1}^N [a_N(j) - a_N^{(k)}(j)]^2 \\ &= (N/N - 1) \textstyle\int_0^1 [\varphi_N(u) - \varphi_N^{(k)}(u)]^2 \, du \end{aligned} \tag{3.10}$$

where $\varphi_N(u) = a_N(i)$, $(i - 1)/N \leqq u < i/N$, and $\varphi_N^{(k)}(u) = a_N^{(k)}(i)$, $(i - 1)/N \leqq u < i/N$, $i = 1, 2, \cdots, N$.

Let us fix $\epsilon > 0$; by (3.7) there exists k_0 such that for all $k > k_0$

$$\int_0^1 [\varphi^{(k)}(u) - \varphi(u)]^2 \, du < \epsilon^3/96. \tag{3.11}$$

On the other hand, Theorem V.1.4.b of [1] implies the existence of $N_0(k)$ to every k such that for any $N > N_0(k)$

$$\begin{aligned} &\int_0^1 [\varphi_N(u) - \varphi(u)]^2 \, du < \epsilon^3/192 \\ &\int_0^1 [\varphi_N^{(k)}(u) - \varphi^{(k)}(u)]^2 \, du < \epsilon^3/192. \end{aligned} \tag{3.12}$$

(3.10), (3.11) and (3.12) imply $E[S_{0N} - S_{0N}^{(k)}]^2 < (\frac{1}{2}\epsilon)^3$. (3.9) then follows from the inequality

$$P\{|S_{0N} - S_{0N}^{(k)}| \geqq \tfrac{1}{2}\epsilon\} \leqq E[S_{0N} - S_{0N}^{(k)}]^2 \cdot (\tfrac{1}{2}\epsilon)^{-2} < \tfrac{1}{2}\epsilon$$

which holds for $k > k_0$ and $N > N_0(k)$. □

(ii) Let us consider the statistics

$$T_{0N}^{(k)} = \sum_{i=1}^N c_i \varphi^{(k)}(F(X_i)) \quad \text{for} \quad k = 1, 2, \cdots. \tag{3.13}$$

It is a consequence of the proof of Theorem V.1.5.a of [1] that

$$\lim_{N\to\infty} d(S_{0N}^{(k)}, T_{0N}^{(k)}) = 0 \quad \text{for any} \quad k = 1, 2, \cdots. \tag{3.14}$$

(iii) LEMMA 3.2. *Let $T_{\Delta N}^{(k)}$ denote the statistic*

$$T_{\Delta N}^{(k)} = \sum_{i=1}^N c_i \varphi^{(k)}[F(X_i + \Delta(d_i - \bar{d}))]. \tag{3.15}$$

Then

$$\lim_{N\to\infty} d(T_{0N}^{(k)}, T_{\Delta N}^{(k)} - ET_{\Delta N}^{(k)}) = 0 \tag{3.16}$$

for any $k = 1, 2, \cdots$ *and any real* Δ.

PROOF. We have

$$\text{(3.17)}\quad \operatorname{Var}(T_{0N}^{(k)} - T_{\Delta N}^{(k)}) \leqq \sum_{i=1}^{N} c_i^2 \int \{\varphi^{(k)}(F(x)) - \varphi^{(k)}[F(x + \Delta(d_i - \bar{d}))]\}^2 \, dF(x).$$

The convergence

$$\text{(3.18)}\qquad F(x + \Delta(d_i - \bar{d})) \to F(x) \quad \text{for} \quad N \to \infty$$

is uniform for $x \varepsilon (-\infty, \infty)$ and $i = 1, 2, \cdots, N$ because of the continuity of F of (1.4). The set A of the points of discontinuity of $\varphi^{(k)}$ is at most countable, for $\varphi^{(k)}$ is non–decreasing, square-integrable and bounded on (0, 1). Hence the convergence

$$\text{(3.19)}\qquad \lim_{N\to\infty} \varphi^{(k)}[F(x + \Delta(d_i - \bar{d}))] = \varphi^{(k)}(F(x))$$

holds uniformly for $i = 1, 2, \cdots, N$ almost everywhere with respect to F, as the exceptional set $B = F^{-1}(A)$ satisfies $\int_B dF(x) = \int_A du = 0$. Thus Lebesgue's theorem may be applied to the integrals in (3.17) which will tend to zero uniformly for $i = 1, 2, \cdots, N$. The desired result then follows from Chebyshev's inequality and from $ET_{0N}^{(k)} = 0$.

(iv) LEMMA 3.3. *It holds that*

$$\text{(3.20)}\quad \lim_{N\to\infty} d(S_{\Delta N}^{(k)}, T_{\Delta N}^{(k)}) = 0 \quad \textit{for} \quad k = 1, 2, \cdots \quad \textit{and for any real} \quad \Delta.$$

PROOF. Let P_N be the probability distribution with the density

$$p_N = \prod_{i=1}^{N} f(x_i)$$

and $Q_{\Delta N}$ the probability distribution with the density

$$q_{\Delta N} = \prod_{i=1}^{N} f(x_i - \Delta(d_i - \bar{d})).$$

The densities $q_{\Delta N}$ are contiguous to the densities p_N, as follows from Theorem VI.2.1 and from the remarks about (VI.2.4.15) of [1] (for the definition of the contiguity see also [1]).

We denote $S_{0N}^{(k)} = S_0^{(k)}(X_1, \cdots, X_N)$ and $T_{0N}^{(k)} = T_0^{(k)}(X_1, \cdots, X_N)$ for purposes of this proof. By (3.14),

$$P_N\{|S_0^{(k)}(X_1, \cdots, X_N) - T_0^{(k)}(X_1, \cdots, X_N)| \geqq \eta\} \to 0 \quad \text{for} \quad N \to \infty$$

holds for any $\eta > 0$. The contiguity implies

$$Q_{\Delta N}\{|S_0^{(k)}(X_1, \cdots, X_N) - T_0^{(k)}(X_1, \cdots, X_N)| \geqq \eta\} \to 0 \text{ for } N \to \infty,$$

which may be written as

$$P_N\{|S_0^{(k)}(X_1 + \Delta d_1, \cdots, X_N + \Delta d_N) - T_0^{(k)}(X_1 + \Delta(d_1 - \bar{d}), \cdots, X_N + \Delta(d_N - \bar{d}))| \geqq \eta\} \to 0;$$

for the statistic $S_{0N}^{(k)}$, depending only on the ranks, is invariant to the translation of the whole sample. The last relation may be written as

$$P\{|S_{\Delta N}^{(k)} - T_{\Delta N}^{(k)}| \geqq \eta\} \to 0 \quad \text{for} \quad N \to \infty \quad \text{and for any} \quad \eta > 0. \qquad \square$$

(v) LEMMA 3.4. *There exists a positive integer k_1 such that for any $k > k_1$ and for any real Δ*

$$\lim_{N\to\infty} d(T_{0N}^{(k)}, T_{\Delta N}^{(k)} - \Delta b_N^{(k)}) = 0 \tag{3.21}$$

where

$$b_N^{(k)} = \int_0^1 \varphi^{(k)}(u)\varphi(u, f)\, du[\textstyle\sum_{i=1}^N c_i(d_i - \bar{d})]. \tag{3.22}$$

PROOF. (3.7) implies

$$\lim_{k\to\infty} \int_0^1 (\varphi^{(k)}(u) - \bar{\varphi}^{(k)})^2\, du = \int_0^1 (\varphi(u) - \bar{\varphi})^2\, du > 0,$$

the positiveness being a consequence of assumption (f) of Section 1. It implies the existence of k_1 such that $\int_0^1 (\varphi^{(k)}(u) - \bar{\varphi}^{(k)})^2\, du > 0$ for all $k > k_1$. According to Theorem VI.2.4 and the remarks about (VI.2.4.15) of [1], the statistic $T_{\Delta N}^{(k)}$ is for any $k > k_1$ and for any real Δ asymptotically normal $(\Delta b_N^{(k)}, (\sigma^{(k)})^2)$, where

$$(\sigma^{(k)})^2 = \textstyle\sum_{i=1}^N c_i^2 \cdot \int_0^1 (\varphi^{(k)}(u) - \bar{\varphi}^{(k)})^2\, du = \int_0^1 (\varphi^{(k)}(u) - \bar{\varphi}^{(k)})^2\, du.$$

On the other hand, the statistic $T_{0N}^{(k)}$ is asymptotically normal $(0, (\sigma^{(k)})^2)$ for all $k > k_1$, as follows from Theorem V.1.5.a of [1]. Returning to Lemma 3.2, we see that

$$\lim_{N\to\infty} [ET_{\Delta N}^{(k)} - \Delta b_N^{(k)}]^2 \cdot (\sigma^{(k)})^{-2} = 0 \quad \text{for} \quad k > k_1;$$

and the lemma follows from Lemma 3.2 again.

(vi) The following lemma establishes a property of contiguous sequences which we shall use in the sequel.

LEMMA 3.5. *Let $\{P_N\}$ and $\{Q_N\}$ be two sequences of probability measures on measure spaces $\{X_N, \mathfrak{A}_N, \mu_N\}$ with the densities p_N and q_N corresponding to measures μ_N. If the densities q_N are contiguous to p_N, then, corresponding to every $\epsilon > 0$, there is a positive number $\delta > 0$ such that $Q_N(A_N) < \epsilon$ is satisfied for almost all N for every sequence of sets $\{A_N\}$, $A_N \in \mathfrak{A}_N$ $(N = 1, 2, \cdots)$ satisfying $P_N(A_N) < \delta$ for almost all N.*

REMARK. We say that a proposition holds for almost all N if it holds for all but a finite number of values of N.

PROOF. By definition of contiguity, we have $Q_N(A_N) \to 0$ whenever $P_N(A_N) \to 0$ for $N \to \infty$. Suppose that it is possible, for some $\epsilon_0 > 0$ and for any positive integer k, to find a sequence $\{A_{N,k}\}_{N=1}^{\infty} = \mathfrak{A}_k$, $A_{N,k} \in \mathfrak{A}_N$, $N = 1, 2, \cdots$ of sets such that

$$P_N(A_{N,k}) < (\tfrac{1}{2})^k \quad \text{for} \quad N > N_0(k) \tag{3.23}$$

but

$$Q_N(A_{N,k}) \geqq \epsilon_0 \tag{3.24}$$

for infinitely many values of N. (3.24) implies that there exists $N_k^* > N_0(k)$ for $k = 1, 2, \cdots$ such that $Q_{N_k^*}(A_{N_k^*,k}) \geqq \epsilon_0$ and that the numbers N_k^*, $k = 1, 2, \cdots$ may be chosen so that $N_k^* < N_{k+1}^*$, $k = 1, 2, \cdots$. Let $\{B_N\}$ be the sequence of sets:

$$\begin{aligned} B_N &= A_{N_k^*,k} \qquad N = N_k^*, \qquad k = 1, 2 \cdots \\ &= \varnothing \qquad \text{for other } N. \end{aligned} \tag{3.25}$$

We have $P_N(B_N) \to 0$ for $N \to \infty$ but $Q_N(B_N) \nrightarrow 0$. Since this contradicts the assumption of contiguity of q_N to p_N, the proof of the lemma is complete.

LEMMA 3.6. *There exists a positive integer k_2 to any fixed Δ and to any $\epsilon > 0$ such that for every integer $k > k_2$, there is a positive integer $N_2(k)$ and for any $N > N_2(k)$*

$$d(S_{\Delta N}, S_{\Delta N}^{(k)}) < \epsilon. \tag{3.26}$$

PROOF. Let P_N, $Q_{\Delta N}$ be the probability distributions with the densities p_N, $q_{\Delta N}$ which were defined in the proof of Lemma 3.3. By Lemma 3.5, in view of the contiguity of $q_{\Delta N}$ to p_N, there corresponds $\delta > 0$ to any $\epsilon > 0$ such that $Q_{\Delta N}(A_N) < \frac{1}{2}\epsilon$ for almost all N for any sequence of events for which $P_N(A_N) < \delta$ for almost all N. By Lemma 3.1, corresponding to the number $\eta = \min(\frac{1}{2}\epsilon, \delta)$, there is a positive integer k_2 such that for every $k > k_2$ there exists $N_1(k)$ and for all $N > N_1(k)$ it holds that (using the notation of the proof of Lemma 3.3)

$$P_N\{|S_0(X_1, \cdots, X_N) - S_0^{(k)}(X_1, \cdots, X_N)| \geqq \tfrac{1}{2}\epsilon\} < \delta.$$

Lemma 3.5 then establishes the existence of $N_2(k)$ for every $k > k_2$ such that

$$Q_{\Delta N}\{|S_0(X_1, \cdots, X_N) - S_0^{(k)}(X_1, \cdots, X_N)| \geqq \tfrac{1}{2}\epsilon\} < \tfrac{1}{2}\epsilon$$

for all $N > N_2(k)$. This may be rewritten as

$$P_N\{|S_0(X_1 + \Delta d_1, \cdots, X_N + \Delta d_N) - S_0^{(k)}(X_1 + \Delta d_1, \cdots, X_N + \Delta d_N)| \geqq \tfrac{1}{2}\epsilon\} < \tfrac{1}{2}\epsilon$$

which means that (3.26) holds for $k > k_2$ and $N > N_2(k)$. □

(vii) LEMMA 3.7. *There exists a positive integer k^*, corresponding to any fixed Δ and any $\epsilon > 0$, such that an index $N^*(k)$ may be found to any $k > k^*$ and for all $N > N^*(k)$ it holds that*

$$d(S_{0N}, S_{\Delta N} - \Delta b_N^{(k)}) < \epsilon. \tag{3.27}$$

PROOF. We may write

$$\begin{aligned} & d(S_{0N}, S_{\Delta N} - \Delta b_N^{(k)}) \\ & \leqq d(S_{0N}, S_{0N}^{(k)}) + d(S_{0N}^{(k)}, T_{0N}^{(k)}) + (T_{0N}^{(k)}, T_{\Delta N}^{(k)} - \Delta b_N^{(k)}) \\ & \quad + d(T_{\Delta N}^{(k)} - \Delta b_N^{(k)}, S_{\Delta N}^{(k)} - \Delta b_N^{(k)}) + d(S_{\Delta N}^{(k)} - \Delta b_N^{(k)}, S_{\Delta N} - \Delta b_N^{(k)}) \end{aligned} \tag{3.28}$$

for any k, $N = 1, 2, \cdots$. The desired result then follows from Lemmas 3.1, 3.3, 3.4, 3.6 and from (3.14).

LEMMA 3.8. *If Δ is a real and ϵ a positive number, then*

$$\lim_{N\to\infty} P\{|S_{\Delta N} - S_{0N} - \Delta b_N| \geqq \epsilon\} = 0$$

where $b_N = \sum_{i=1}^N c_i(d_i - \bar{d}) \cdot \int_0^1 \varphi(u)\varphi(u, f)\, du$.

PROOF. Schwarz's inequality and (1.3) imply

$$\begin{aligned} |b_N^{(k)} - b_N| &= |\textstyle\sum_{i=1}^N c_i(d_i - \bar{d}) \cdot \int_0^1 [\varphi^{(k)}(u) - \varphi(u)]\varphi(u, f)\, du| \\ &\leqq M^{\frac{1}{2}} |\textstyle\int_0^1 (\varphi^{(k)}(u) - \varphi(u))\varphi(u, f)\, du|. \end{aligned}$$

The right-hand side of the inequality tends to zero for k tending to infinity from (3.7) and (a) and thus $\lim_{k\to\infty} b_N^{(k)} = b_N$ uniformly in N. The proof of the lemma then follows from Lemma 3.7.

(viii) We consider now the scores (1.7). We use the notation

$$a_N(i) = E(\varphi(U^{(i)})) \qquad a_N^*(i) = \varphi(i/N+1)$$
$$S_{\Delta N} = \sum_{i=1}^N c_i a_N(R_i^\Delta) \qquad S_{\Delta N}^* = \sum_{i=1}^N c_i a_N^*(R_i^\Delta).$$

By Lemma V.1.6.a and Theorem V.1.6.a of [1] it holds that

$$\lim_{N\to\infty} P\{|S_{0N} - S_{0N}^*| \geqq \epsilon\} = 0 \quad \text{for any} \quad \epsilon > 0. \tag{3.29}$$

It follows in view of the contiguity of $q_{\Delta N}$ to p_N that

$$\lim_{N\to\infty} P\{|S_{\Delta N} - S_{\Delta N}^*| \geqq \epsilon\} = 0 \quad \text{for any} \quad \epsilon > 0. \tag{3.30}$$

(3.29), (3.30) and Lemma 3.8 then imply

$$\lim_{N\to\infty} P\{|S_{\Delta N}^* - S_{0N}^* - \Delta b_N| \geqq \epsilon\} = 0 \quad \text{for any} \quad \epsilon > 0$$

which means that Lemma 3.8 is right also for the scores (1.7).

(ix) We now complete the proof of Theorem 2.1.

Let C, ϵ, η be any positive numbers. Consider a partition of the interval $-C, C]$, $-C = \Delta_0 < \Delta_1 < \cdots < \Delta_r = C$, such that

$$|(\Delta_i - \Delta_{i-1}) \textstyle\int_0^1 \varphi(u)\varphi(u,f)\,du| < \frac{1}{2}\epsilon\cdot M^{-\frac{1}{2}}, \tag{3.31}$$

for $i = 1, 2, \cdots, r$, where $M > 0$ is the constant satisfying $\sum_{i=1}^N (d_i - \bar{d})^2 \leqq M$ $(N = 1, 2, \cdots)$ by (1.3). Lemma 3.8 guarantees existence of N_0 such that for any $N > N_0$

$$P\{|S_{\Delta_i N} - S_{0N} - \Delta_i b_N| \geqq \tfrac{1}{4}\epsilon\} < \eta/r + 1 \tag{3.32}$$

for $i = 1, 2, \cdots, r$.

Let Δ be a point of interval $[-C, C]$. Then there exists i_0, $1 \leqq i_0 \leqq r$, such that $\Delta_{i_0-1} \leqq \Delta \leqq \Delta_{i_0}$ and the following inequality is satisfied:

$$|S_{\Delta N} - S_{0N} - \Delta b_N| \leqq |S_{\Delta i_0 N} - S_{0N} - \Delta_{i_0} b_N| + |b_N(\Delta_{i_0} - \Delta_{i_0-1}) + |S_{\Delta i_0-1,N} - S_{0N} - \Delta_{i_0-1} b_N|. \tag{3.33}$$

We now prove (3.33). Suppose that the constants c_i and d_i $(i = 1, 2, \cdots, N)$ satisfy $(c_i - c_j)(d_i - d_j) \geqq 0$ $(i, j = 1, 2, \cdots, N)$. Then the statistic $S_{\Delta N}$ is a non-decreasing function of Δ by Theorem 2.1. If $S_{\Delta N} - S_{0N} - \Delta b_N \geqq 0$, then

$$|S_{\Delta N} - S_{0N} - \Delta b_N| \leqq S_{\Delta_{i_0},N} - S_{0N} - \Delta b_N \leqq |S_{\Delta i_0 N} - S_{0N} - \Delta_{i_0} b_N| + |b_N(\Delta_{i_0} - \Delta_{i_0-1})| + |S_{\Delta_{i_0-1},N} - S_{0N} - \Delta_{i_0-1} b_N|.$$

The case $S_{\Delta N} - S_{0N} - \Delta b_N < 0$ is handled quite analogously. The same conclusions would be reached in the case where the constants c_i and d_i $(i = 1, 2, \cdots, N)$ satisfied $(c_i - c_j)(d_i - d_j) \leqq 0$ $(i, j = 1, 2, \cdots, N)$, when the statistic $S_{\Delta N}$ is a non-increasing function of Δ.

The inequality (3.33) implies

(3.34) $\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N| \leqq 2 \max_{0 \leqq i \leqq r} |S_{\Delta_i N} - S_{0N} - \Delta_i b_N| + \frac{1}{2}\epsilon.$.

It may be shown that $\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N|$ is a random variable so that the probability $P\{\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N| \geqq \epsilon\}$ is well-defined. (3.31) and (3.32) imply

(3.35) $$P\{\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N| \geqq \epsilon\} \leqq \sum_{i=0}^{r} P\{|S_{\Delta_i N} - S_{0N} - \Delta_i b_N| \geqq \tfrac{1}{4}\epsilon\} < \eta$$

for $N > N_0$.

By Theorem V.1.6.a of [1] the statistic S_{0N} is asymptotically normal $N(0, \sigma^2)$ with

$$\sigma^2 = \sum_{i=1}^{N} c_i^2 \int_0^1 (\varphi(u) - \bar{\varphi})^2 \, du = \int_0^1 (\varphi(u) - \bar{\varphi})^2 \, du$$

or $\sigma^2 = \operatorname{Var} S_{0N}$; thus $\operatorname{Var} S_{0N} \to \int_0^1 (\varphi(u) - \bar{\varphi})^2 \, du$ for $N \to \infty$; this, together with (3.35), implies that

$$\lim_{N \to \infty} P\{\max_{|\Delta| \leqq C} |S_{\Delta N} - S_{0N} - \Delta b_N| \geqq \epsilon (\operatorname{Var} S_{0N})^{\frac{1}{2}}\} = 0.$$

The proof of Theorem 3.1 is complete.

REMARK. The theorem remains valid also in the case when the constants c_i and d_i satisfy the following weaker assumptions:

$$c_i = c_i' + c_i'', \qquad d_i = d_i' + d_i'', \qquad i = 1, 2, \cdots, N, \; N = 1, 2, \cdots$$

where either $\sum_{i=1}^{N} (c_i' - \bar{c}')^2 = 0$ for all but a finite number of N or $\sum_{i=1}^{N} (c_i' - \bar{c}')^2 > 0$ for all but a finite number of N; the cases are analogous for c_i'', d_i', d_i''.

Moreover, we suppose that the (c_i')'s and the (c_i'')'s satisfy the Noether's condition (1.2) and that the (d_i')'s and the (d_i'')'s satisfy (1.3) and (1.4) and that for all pairs (i, j), $1 \leqq i, j \leqq N$, it holds that

$$(c_i' - c_j')(d_i' - d_j') \geqq 0 \qquad (c_i'' - c_j'')(d_i' - d_j') \leqq 0$$
$$(c_i' - c_j')(d_i'' - d_j'') \leqq 0 \qquad (c_i'' - c_j'')(d_i'' - d_j'') \geqq 0.$$

On the other hand, the theorem also remains valid when the function φ, generating the scores, is a difference of two monotone square integrable functions φ_1, φ_2.

4. Numerical illustration.

Consider the statistic

(4.1) $$S_{\Delta N} = \sum_{i=1}^{N} c_i a_N(R_{Ni}^{\Delta})$$

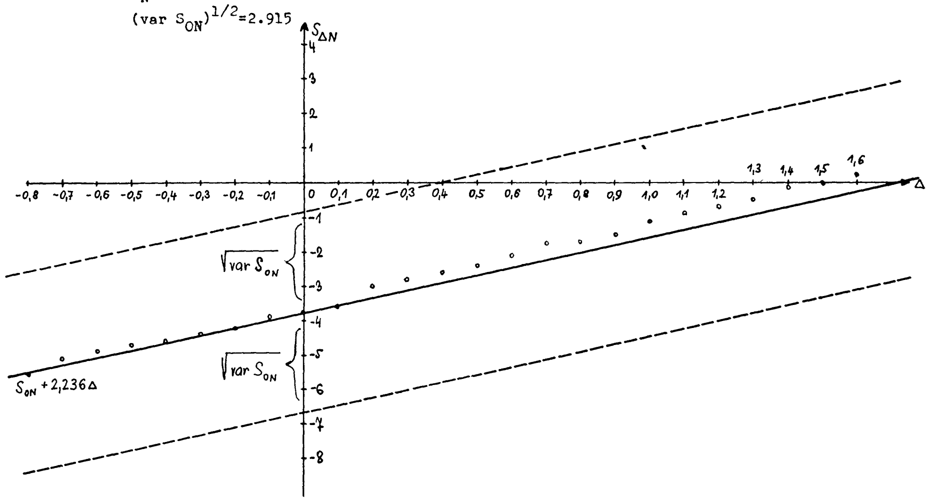
with $N = 2n$ and

(4.2)
$$\begin{aligned} c_i &= 0 & i &= 1, 2, \cdots, n, \\ &= 1 & i &= n+1, n+2, \cdots, N; \\ d_i &= 0 & i &= 1, 2, \cdots, n, \\ &= 1/n^{\frac{1}{2}} & i &= n+1, n+2, \cdots, N; \end{aligned}$$

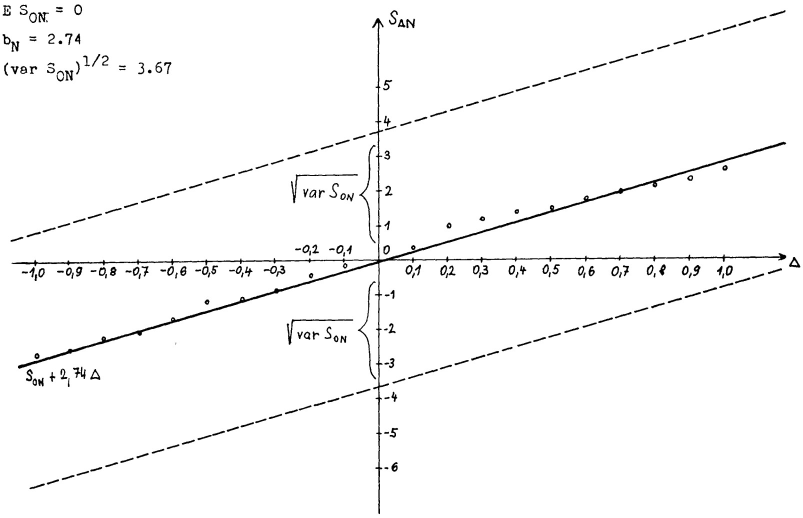
the scores are of the form $a_N(i) = \Phi^{-1}(i/N + 1)$ where Φ is the distribution function of the normal distribution $N(0, 1)$. The statistic S_{0N} is the test statistic of van den Waerden's test.

The following graphs show values of the statistic $S_{\Delta N}$ determined from samples of sizes $N = 40, 60, 80$ from a $N(0, 1)$ distribution.

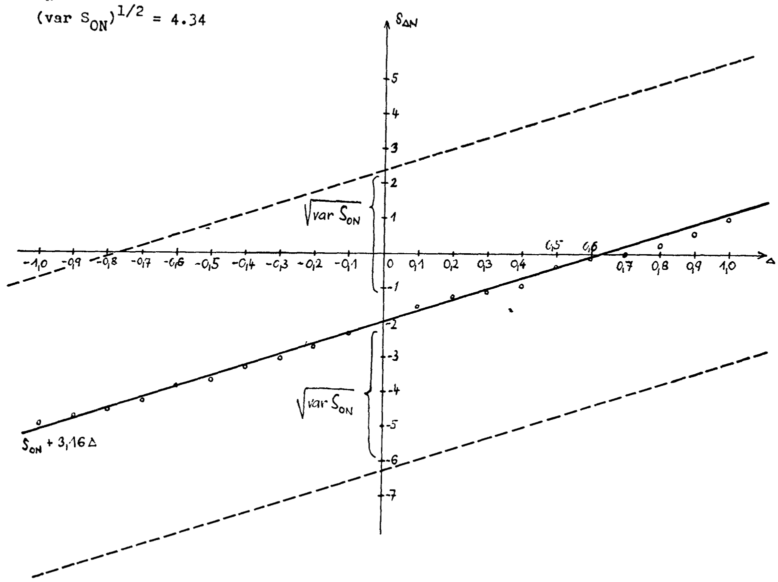
Graph 1. $N = 2n = 40$
$E\,S_{ON} = 0$
$b_N = 2{,}236$
$(\text{var } S_{ON})^{1/2} = 2.915$



Graph 2. $N = 2n = 60$
$E\,S_{ON} = 0$
$b_N = 2.74$
$(\text{var } S_{ON})^{1/2} = 3.67$



Graph 3. $N = 2n = 80$
$E\, S_{ON} = 0$
$b_N = 3.16$
$(\text{var } S_{ON})^{1/2} = 4.34$



5. Some possibilities for application. The first possible application of the asymptotic linearity of the statistic of type $S_{\Delta N}$ is in the construction of a simultaneous nonparametric estimate of regression coefficients $\Delta_1, \Delta_2, \cdots, \Delta_K$ in the regression equation $Y_i = Y_i^0 + \alpha + \Delta_1 x_{1i} + \cdots + \Delta_K x_{Ki}$, $i = 1, 2, \cdots, N$, where $(Y_1^0, \cdots, Y_N^0)$ is an independent random sample from a continuous distribution. The idea of the estimate is similar to that of Hodges and Lehmann [2] who propose a nonparametric estimate of the location. An estimate based on the statistics of type $S_{\Delta N}$ is proposed in the author's thesis [3] which, up to this time, remains unpublished.

The second application may be in the problem of testing the hypothesis that two populations differ only in location against the alternative that they may differ also in scale. Sukhatme [7] proposes to use in this case some test-statistics for a two-sample test of scale with variables centered by estimates of location parameters. If it is supposed that the basic distribution is symmetric then applicability of many usual two-sample tests of scale may be justified just by the asymptotic linearity, for the coefficient b_N is then equal to zero.

6. Acknowledgment. The author wishes to thank Professor J. Hájek for his helpful suggestions during the preparation of her dissertation.

REFERENCES

[1] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academia, Prague.
[2] Hodges, J. L. and Lehmann, E. L. (1963). Estimation of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[3] JUREČKOVÁ, J. (1967). Nonparametric estimate of regression coefficients. Unpublished dissertation. Czechoslovak Academy of Sciences, Prague.
[4] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.
[5] LOÈVE, M. (1962). *Probability Theory* (Russian translation). Moskva.
[6] NATANSON, I. P. (1957). *Theory of Functions of a Real Variable* (in Russian). Moskva.
[7] SUKHATME, B. (1958). Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29** 60–78.