The Annals of Statistics
1976, Vol. 4, No. 3, 554-575

WEAK CONVERGENCE OF SEQUENTIAL LINEAR RANK STATISTICS[1]

BY HENRY I. BRAUN

Princeton University

A sequential version of Chernoff-Savage linear rank statistics is introduced as a basis for inference. The principal result is an invariance principle for two-sample rank statistics, i.e., under a fixed alternative the sequence of sequential linear rank statistics converges weakly to a Wiener process. The domain of application of the theorem is quite broad and includes score functions which tend to infinity at the end points much more rapidly than that of the normal scores test. The method of proof involves new results in the theory of multiparameter empirical processes as well as some new probability bounds on the joint behavior of uniform order statistics. Applications of weak convergence are explored; in particular, the extension of the theory of Pitman efficiency to the sequential case.

1. Introduction and summary. A sequential version of two-sample linear rank statistics is introduced and shown to converge weakly to a Wiener process under mild regularity conditions. The method is based on the approach of Pyke and Shorack [14], though no use is made here of the Skorokhod construction of a.s. convergent versions of weakly convergent sequences.

Closely related work is found in Miller and Sen [10], who established the weak convergence to a Wiener process of a sequential version of U-statistics. Another approach to this problem is through a.s. invariance principles. See Sen and Ghosh [17], [18] and Sen [16] for details.

In Section 2, after a brief review of the Pyke–Shorack approach, weak convergence for rank statistics with bounded score functions is demonstrated. In Section 3, the hypotheses are weakened to allow most unbounded score functions. The proof requires estimates on the behavior of a sequence of empirical processes which are of independent interest. Section 4 deals with potential applications to sequential analysis. Related work is alluded to in Hall [8]. A Pitman efficiency result for sequential nonparametric procedures is heuristically derived.

2. Bounded score functions.

2.A. *Preliminaries*. Let $X_1, X_2, \cdots$ be i.i.d. random variables with continuous df F and $Y_1, Y_2, \cdots$ be i.i.d. random variables with continuous df G. Assume that a sample of N observations consists of mX's and nY's. Let $\lambda_N = m/N$. It is assumed throughout that $0 < \bar{\lambda} \leqq \lambda_N \leqq 1 - \bar{\lambda} < 1$ for some fixed $\bar{\lambda} > 0$.

Received March 1974; revised August 1975.

[1] The material of this paper is adapted from the author's Ph. D. dissertation (1974) at Stanford University.

AMS 1970 *subject classifications*. 62E20, 62G10, 62G20, 62L10.

Key words and phrases. Sequential linear rank statistics, weak convergence, uniform order statistics, asymptotic efficiency.

The empirical df based on $X_1, \cdots, X_m$ ($Y_1, \cdots, Y_n$) is denoted by $F_m(G_n)$. The corresponding one-sample empirical processes are defined by

$$U_m(t) = m^{\frac{1}{2}}[F_m F^{-1}(t) - t] \qquad 0 \leqq t \leqq 1$$
$$V_n(t) = n^{\frac{1}{2}}[G_n G^{-1}(t) - t] \qquad 0 \leqq t \leqq 1 .$$

Inverse functions are taken to be left continuous. Let

$$H_N = \lambda_N F_m + (1 - \lambda_N)G_n ; \qquad H_\lambda = \lambda F + (1 - \lambda)G ; \qquad H = H_{\lambda_N} .$$

For a triangular array of constants $(c_{Ni})_{1 \leqq i \leqq N, 1 \leqq N < \infty}$, the corresponding sequence of linear rank statistics is defined by

$$T_N^* = \sum_{i=1}^N c_{Ni} F_m H_N^{-1}(i/N) = m^{-1} \sum_{i=1}^N c_{Ni} R_{Ni} \qquad 1 \leqq N < \infty ,$$

where R_{Ni} = the number of observations among $X_1, X_2, \cdots, X_m$ which do not exceed the ith order statistic (o.s.) of the combined sample. Let ν_N denote the signed measure that places mass c_{Ni} on the point i/N $(1 \leqq i \leqq N)$ and no mass elsewhere. Then define

$$T_N = N^{\frac{1}{2}}(T_N^* - \textstyle\int_0^1 FH^{-1}\, d\nu_N) = \int_0^1 L_N\, d\nu_N \qquad 1 \leqq N \leqq \infty , \tag{2.1}$$

where

$$L_N(t) = N^{\frac{1}{2}}[F_m H_N^{-1}(t) - FH^{-1}(t)] \qquad 0 \leqq t \leqq 1$$

is the two-sample empirical process introduced in [14] in analogy to the one-sample empirical process.

Equation (2.1) gives an integral representation of the normalized linear rank statistic T_N. The following lemma, due to Pyke and Shorack, expresses $L_N(\cdot)$ as a random linear combination of $U_m(\cdot)$ and $V_n(\cdot)$.

LEMMA. (Pyke–Shorack). *With probability one*

$$\begin{aligned} L_N(t) = (1 - \lambda_N)\{&\lambda_N^{-\frac{1}{2}} B_N(t) U_m(FH_N^{-1}(t)) \\ &- (1 - \lambda_N)^{-\frac{1}{2}} A_N(t) V_N(GH_N^{-1}(t))\} + \delta_N(t) \qquad \textit{for all } t \in (0, 1) \end{aligned} \tag{2.2}$$

where

$$\delta_N(t) = A_N(t) N^{\frac{1}{2}}[H_N H_N^{-1}(t) - t]$$
$$A_N(t) = [FH^{-1}(u_t) - FH^{-1}(t)]/[u_t - t] , \qquad u_t = HH_N^{-1}(t)$$

and B_N is defined to satisfy

$$\lambda_N A_N(t) + (1 - \lambda_N) B_N(t) = 1 .$$

In (2.2) $L_N(\cdot)$ is defined by left continuity at otherwise undefined points.

REMARK 2.1. Following Pyke and Shorack we define a process $L_N^{(1)}(\cdot)$ related to $L_N(\cdot)$.

$$\begin{aligned} L_N^{(1)}(t) &= L_N(t) , \quad N^{-1} \leqq t \leqq 1 \\ &= 0 , \qquad 0 \leqq t < N^{-1} . \end{aligned}$$

It is clear that

$$\int_0^1 L_N^{(1)}\, d\nu_N = \int_0^1 L_N\, d\nu_N = T_N$$

so that one may use $L_N^{(1)}$ in the representation for T_N without changing the value of the statistic. For technical reasons it is more convenient to work with $L_N^{(1)}$.

In the usual formulation of the problem, observations, either X's or Y's, arrive sequentially. After each new observation we are allowed to decide whether to stop or continue sampling. The decision is based on the value of the linear rank statistic computed on the totality of the data available. To derive the analytic properties of the procedure, a sequential statistic is defined in a form that is suitable for weak convergence arguments; later on the result is transformed into a form more amenable to statistical applications.

Define a sequence of stochastic processes $\{T_N(\cdot)\}$ by

$$\begin{aligned} T_N(0) &\equiv 0 \\ T_N(s) &= ([Ns]/N)^{\frac{1}{2}}\sigma_0^{-1}T_{[Ns]} \\ &= ([Ns]/N)^{\frac{1}{2}}\sigma_0^{-1}\textstyle\int_0^1 L_{[Ns]}^{(1)}(t)\,d\nu_{[Ns]}(t) \qquad 0 \leqq s \leqq 1\,, \end{aligned}$$

where $[x]$ denotes the greatest integer in x, and σ_0^2 is the variance of the normal distribution to which T_k converges in distribution.

With probability one, $T_N(\cdot) \in D[0, 1]$. In Theorem 2.1 it is shown that $T_N(\cdot)$ converges weakly to a Wiener process on $[0, 1]$ as $N \to \infty$, relative to the Skorokhod topology. This is equivalent to convergence in the uniform topology since the Wiener process is (a.s.) continuous. The key is to express $T_N(\cdot)$ as a continuous function of a single stochastic process—essentially a two-parameter two-sample empirical process constructed from $L_1^{(1)}, L_2^{(1)}, \cdots, L_N^{(1)}$. Having shown that this new process converges weakly, the desired result is obtained by an application of the continuous mapping theorem (see [3], Theorem 5.1).

We now make two key assumptions which will apply throughout this paper.

ASSUMPTION 2.1. The functions FH_λ^{-1} have derivatives a_λ with respect to t for all $t \in (0, 1)$ and for some $\lambda' \in (0, 1)$, $a_{\lambda'}$ is continuous on $(0, 1)$ and has one-sided limits at 0 and 1.

DISCUSSION. This is assumption 4.1 of Pyke and Shorack [14]. They show that it holds under rather wide conditions. (See, for example, their Corollary 4.1.) Now assume that $\lambda_N \to \lambda_0$ (fixed) as $N \to \infty$. Recall that A_N and B_N are the difference quotients of FH^{-1} and GH^{-1} respectively. FH^{-1} and GH^{-1} are absolutely continuous and so have derivatives, a_N and b_N, say, which exist a.e. on $[0, 1]$. The natural limit functions then are the derivatives of $FH_0^{-1}(t)$ and $GH_0^{-1}(t)$, where $H_0 = \lambda_0 F + (1 - \lambda_0)G$. Denote these derivatives by $a_0(t)$ and $b_0(t)$ respectively. Pyke and Shorack show in their Lemmas 4.1 and 4.2 that under Assumption 2.1, a_0 and b_0 are continuous on $[0, 1]$, $\rho(A_N, a_0) \to 0$ (a.s.), and $\rho(B_N, b_0) \to 0$ (a.s.) as $N \to \infty$.

ASSUMPTION 2.2. There exists a signed Lebesgue–Stieltjes measure ν on $(0, 1)$ for which $|\nu|([\varepsilon, 1 - \varepsilon]) < \infty$ for all $\varepsilon > 0$, such that

$$\textstyle\int_0^1 L_N^{(1)}\, d(\nu_N - \nu) \to 0 \quad \text{(a.s.)}.$$

DISCUSSION. Condition (b)(i) of Theorem 4.1 of Pyke and Shorack demands only convergence in probability to zero. Though our Assumption 2.2 is stronger, it is satisfied in many important cases.

For example, suppose $J_N(i/N)$ is the expectation of the ith o.s. in a sample of size N from a population whose cumulative distribution function is the inverse function of J and

$$|J^{(i)}(u)| \leqq K[u(1-u)]^{-i-\frac{1}{2}+\delta} \qquad i = 0, 1, 2 .$$

Pyke and Shorack state ([14], page 769) that Theorem 2 of Chernoff and Savage [5] shows that the above hypotheses are sufficient to imply the conditions of their Corollary 5.1. A close look at the proof of Proposition 5.1(b) ([14], page 768) reveals that it yields the result $\int_0^1 L_N^{(1)}\, d(\nu_n - \nu) = o(1)$ a.s. under these same hypotheses. This shows that Assumption 2.2 is not unduly restrictive.

We are now ready to state the main result of this section.

THEOREM 2.1. *Let Assumptions* 2.1 *and* 2.2 *hold and assume also that* (i) $\lambda_N \to \lambda_0$ (a.s.) *and* (ii) $\int_0^1 d|\nu| \to \infty$. *Then*

$$T_N(\cdot) \to_w \text{ standard Wiener process.}$$

REMARK 2.2. In Theorem 3.2 below we weaken assumption (ii). There it is only required that $\int_0^1 q\, d|\nu| < \infty$ where $q(t)$ is a function that tends to zero sufficiently rapidly as t tends to zero or one. See Section 3 for details.

2.B. *Multiparameter processes.* Extensive use is made of multiparameter processes. The weak convergence of such processes has been investigated; see Bickel and Wichura [2] for details. Our interest focusses on the so-called two-parameter one-sample empirical process, based on a set of observations $X_1, \cdots, X_m$, which is defined by

$$Z_m(s, t) = 0 \qquad 0 \leqq s < 1/m,\ 0 \leqq t \leqq 1$$
$$Z_m(s, t) = ([ms]/m)^{\frac{1}{2}} U_{[ms]}(t) \qquad 1/m \leqq s \leqq 1,\ 0 \leqq t \leqq 1 .$$

In place of $Z_m(s, t)$, the more suggestive notation $s^{\frac{1}{2}} U_m(s, t)$ will often be used. The analogous process for the "Y-sample" is denoted by $Q_n(s, t) = s^{\frac{1}{2}} V_n(s, t)$.

Note that $E[Z_m(s_1, t_1) Z_m(s_2, t_2)] = \min(s_1, s_2) \cdot \min(t_1, t_2)[1 - \max(t_1, t_2)]$. Thus $Z_m(\cdot, \cdot)$ has a covariance structure like a Wiener process in the s-scale and like a tied-down Wiener process in the t-scale. In [2] it is shown that $Z_m(\cdot, \cdot)$ converges weakly to a two-parameter Gaussian process having the same covariance structure and hence continuous sample paths.

REMARK 2.3. For later convenience, we develop notation for some important function spaces. $D^-[0, 1]$ denotes the space of left continuous real-valued functions having no discontinuities of the second kind; $C[0, 1]^2$ denotes the space of continuous real-valued functions defined on $[0, 1]^2$. The definitions of $D[0, 1]^2$ and $D^-[0, 1]^2$ are immediate.

Now define a modified two-parameter two-sample empirical process by

$$\begin{aligned} L_N^{(2)}(s, t) &= 0\,; && s = 0\,; && 0 \leqq t \leqq 1 \\ &= 0\,; && s = 1/N, \cdots, N/N\,; && 0 \leqq t < [Ns]^{-1} \\ &= s^{\frac{1}{2}}[L_{Ns}^{(1)}(t) - \delta_{Ns}(t)]\,; && s = 1/N, \cdots, N/N\,; && [Ns]^{-1} \leqq t \leqq 1 \end{aligned}$$

and, in general,

$$L_N^{(2)}(s, t) = L_N^{(2)}([Ns]/N, t) \qquad 0 \leqq s \leqq 1,\ 0 \leqq t \leqq 1\,,$$

where Ns, written as a subscript, is meant to denote $[Ns]$. Also define two processes which approximate $T_N(\cdot)$.

$$T_N^{(1)}(s) = ([Ns]/N)^{\frac{1}{2}}\sigma_0^{-1} \int_0^1 L_{Ns}^{(1)}(t)\, d\nu(t) \qquad 0 \leqq s \leqq 1\,,$$

and

$$T_N^{(2)} = \sigma_0^{-1} \int_0^1 L_N^{(2)}(s, t)\, d\nu(t) \qquad 0 \leqq s \leqq 1\,.$$

It will be shown that $L_N^{(2)}(\cdot, \cdot)$ converges weakly and hence so does $T_N^{(2)}(\cdot)$. The weak convergence of $T_N(\cdot)$ then follows easily. The first result is obtained in an indirect fashion. First define two-parameter analogs of the processes that appear in the decomposition of $L_N(\cdot)$. For $(s, t) \in [0, 1]^2$ let

$$B_N(s, t) = B_{Ns}(t)\,, \qquad A_N(s, t) = A_{Ns}(t)\,, \qquad \lambda_N(s) = \lambda_{Ns}\,.$$

Also let $\phi_N(\cdot, \cdot)$ and $\psi_N(\cdot, \cdot)$ be random mappings of the unit square into itself given by

$$\phi_N(s, t) = (\lambda_{Ns}s,\ FH_{Ns}^{-1}(t))$$
$$\psi_N(s, t) = ((1 - \lambda_{Ns})s,\ GH_{Ns}^{-1}(t))\,.$$

Thus, on the set where it is not defined to be 0,

$$\begin{aligned} L_N^{(2)}(s, t) = (1 - \lambda_N(s))[&\lambda_N^{-\frac{1}{2}}(s)B_N(s, t)\lambda_N^{-\frac{1}{2}}(s)Z_N(\phi_N(s, t)) \\ &- (1 - \lambda_N(s))^{-\frac{1}{2}}A_N(s, t)(1 - \lambda_N(s))^{-\frac{1}{2}}Q_N(\psi_N(s, t))]\,. \end{aligned}$$

With this representation it is clear what the appropriate limit for $\{L_N^{(2)}(\cdot, \cdot)\}$ must be. For $(s, t) \in [0, 1]^2$, define

$$\phi_0(s, t) = (\lambda_0 s,\ FH_0^{-1}(t))\,, \qquad \psi_0(s, t) = ((1 - \lambda_0)s,\ GH_0^{-1}(t))\,,$$
$$b_0(s, t) = b_0(t)\,, \qquad a_0(s, t) = a_0(t)\,, \qquad \lambda_0(s) = \lambda_0\,.$$

The natural limit is then given by

$$\begin{aligned} L_0(s, t) = (1 - \lambda_0(s))[&\lambda_0^{-\frac{1}{2}}(s)b_0(s, t)\lambda_0^{-\frac{1}{2}}(s)Z_0(\phi_0(s, t)) \\ &- (1 - \lambda_0(s))^{-\frac{1}{2}}a_0(s, t)(1 - \lambda_0(s))^{-\frac{1}{2}}Q_0(\psi_0(s, t))]\,, \end{aligned}$$

where $Z_0(\cdot, \cdot)$ and $Q_0(\cdot, \cdot)$ are the two-parameter Gaussian processes with continuous sample paths to which $Z_N(\cdot, \cdot)$ and $Q_N(\cdot, \cdot)$, respectively, converge. Z_0 and Q_0 are taken independent of each other. Now formally rewrite the above equation as

$$L_0 = h_1(Z_0, Q_0, b_0, a_0, \phi_0, \psi_0, \lambda_0)$$

where the function h_1 is defined implicitly by the above correspondence.

The mechanics of the proof require a third two-parameter process which is intermediate, in some sense, between $L_N^{(2)}(\cdot, \cdot)$ and $L_0(\cdot, \cdot)$. Set $\theta_N = N^{-\frac{1}{2}-\delta}$ where $\delta > 0$. Let

$$L_N^* = h_1(Z_N, Q_N, B_N^*, A_N^*, \phi_N^*, \psi_N^*, \lambda_N^*) \tag{2.3}$$

where

$$\begin{aligned}
\phi_N^*(s, t) &= \phi_0(s, t)\,, && \text{on } [0, \theta_N) \times [0, 1] \\
&= \phi_N(s, t)\,, && \text{on } [\theta_N, 1] \times [0, 1]\,, \\
B_N^*(s, t) &= b_0(s, t)\,, && \text{on } [0, \theta_N) \times [0, 1] \\
&= b_N(s, t)\,, && \text{on } [\theta_N, 1] \times [0, 1]\,, \\
\lambda_N^*(s) &= \lambda_0(s)\,, && \text{on } [0, \theta_N) \times [0, 1] \\
&= \lambda_N(s)\,, && \text{on } [\theta_N, 1] \times [0, 1]\,.
\end{aligned}$$

The definitions of A_N^* and ψ_N^* are similar.

2.C. *Proof of Theorem* 2.1. The proof of the theorem requires three lemmas.

LEMMA 2.1. *Assume* $\lambda_N \to \lambda_0$. *Then* $\rho(FH_N^{-1}, FH_0^{-1}) \to 0$ (a.s.) *and* $\rho(GH_N^{-1}, GH_0^{-1}) \to 0$ (a.s.).

PROOF. Immediate. See [14] for details.

LEMMA 2.2. *Assume* $\lambda_n \to \lambda_0$ *and that Assumption* 2.1 *holds. Then*

$$L_N^*(\cdot, \cdot) \to_w L_0(\cdot, \cdot)\,. \tag{2.4}$$

PROOF. Recall that $L_N^* = h_1(Z_N, Q_N, B_N^*, A_N^*, \phi_N^*, \psi_N^*, \lambda_N^*)$ and $L_0 = h_1(Z_0, Q_0, b_0, a_0, \phi_0, \psi_0, \lambda_0)$. We show first that the vector of arguments for L_N^* converges to the vector of arguments for L_0. The first step is the demonstration of (a) $\rho(\phi_N^*, \phi_0) \to 0$ (a.s.) and (b) $\rho(B_N^*, b_0) \to 0$ (a.s.).

To prove (a) note that

$$\begin{aligned}
&\sup_{(s,t)\in[0,1]^2}|\phi_N^*(s, t) - \phi_0(s, t)| \\
&\qquad \leqq \sup_{[\theta_N,1]} |\lambda_{Ns} s - \lambda_0 s| + \sup_{[\theta_N,1]\times[0,1]} |FH_{Ns}^{-1}(t) - FH_0^{-1}(t)|\,.
\end{aligned}$$

The result now follows since $N\theta_N \to \infty$ as $N \to \infty$ and $\lambda_N \to \lambda_0$ (a.s.), and from Lemma 2.1. To prove (b) note that

$$\sup_{(s,t)\in[0,1]^2} |B_N^*(s, t) - b_0(s, t)| = \sup_{[\theta_N,1]\times[0,1]} |B_{Ns}(t) - b_0(t)|\,.$$

Since $N\theta_N \to \infty$ as $N \to \infty$, (b) follows from the discussion after Assumption 2.1. Of course, similar results hold for A_N^*, ψ_N^*, and λ_N^*.

Recall now that $Z_N \to_w Z_0$, $Q_N \to_w Z_0$, and that Z_N and Q_N are independent for each N. Now apply Theorem 4.4 of Billingsley [3] to assert that

$$(Z_N, Q_N, B_N^*, A_N^*, \phi_N^*, \psi_N^*, \lambda_N^*) \to_w (Z_0, Q_0, b_0, a_0, \phi_0, \psi_0, \lambda_0)$$

relative to the appropriate product topology. Equation (2.4) will then follow immediately from the continuous mapping theorem (cf. [3], Theorem 5.1) if it can be shown that h_1 is indeed continuous with probability 1 (w.p. 1) under the limiting measure. The proof is cumbersome but straightforward. The key is that the limit processes are all continuous w.p. 1, and Z_0 and Q_0 the only random limit processes are independent, so that the operations involved in h_1 (composition, multiplication, and addition) are continuous although we are working in the Skorokhod topology. Details are omitted. □

LEMMA 2.3. *Under the hypotheses of Lemma* 2.2,

$$L_N^{(2)}(\cdot, \cdot) \to_w L_0(\cdot, \cdot) . \tag{2.5}$$

PROOF. First note that $L_N^{(2)}$ and L_N^* are identical except on the set

$$\Lambda_N = \{(s, t): 0 < s < \theta_N, 0 < t < 1\} \cup \{(s, t): \theta_N \leqq s \leqq 1, 0 < t < [Ns]^{-1}\} .$$

As a result the finite dimensional distributions of $L_N^{(2)}$ have the same limit as those of L_N^*. To demonstrate tightness it suffices to show that $L_N^{(2)}$ and L_N^* are both $o_p(1)$ on Λ_N.

Since L_N^* does converge weakly, we can use the fact that $L_N^*(s, 0) \equiv_s 0$ to conclude that $L_N^* = o_p(1)$ on the set $\{(s, t): \theta_N \leqq s \leqq 1, 0 < t < [Ns]^{-1}\}$. On this same set $L_N^{(2)}$ is identically zero. On the set $\{(s, t): 0 < s < \theta_N, 0 < t < 1\}$ the suprema of both processes are $o(1)$. To see this note for example that both $B_N Z_N(\phi_N)$ and $B_N^* Z_N(\phi_N^*)$ are bounded above in absolute value on $[0, \theta_N] \times [0, 1]$ by

$$2[(1 - \tilde{\lambda})\theta_N]^{\frac{1}{2}}[N(1 - \tilde{\lambda})\theta_N]^{\frac{1}{2}}/\tilde{\lambda} \leqq 2(1 - \tilde{\lambda})N^{-\delta}/\tilde{\lambda} = o(1) .$$

We have used the facts that $|B_N(t)| \leqq (1 - \lambda_N)^{-1}, 0 < \tilde{\lambda} \leqq \lambda_N \leqq 1 - \tilde{\lambda} < 1$, and $|U_N(t)| \leqq N^{\frac{1}{2}}$. This completes the proof. □

PROOF OF THEOREM 2.1. Recall that

$$T_N^{(2)}(s) = \sigma_0^{-1} \textstyle\int_0^1 L_N^{(2)}(s, t)\, d\nu(t) \qquad 0 \leqq s \leqq 1 .$$

This may be formally rewritten

$$T_N^{(2)}(\cdot) = h_2(L_N^{(2)}(\cdot, \cdot)) .$$

The operator h_2 is implicitly defined by the correspondence between the two definitions of $T_N^{(2)}(\cdot)$. Since $\int_0^1 d|\nu| < \infty$, it is clear that h_2 is continuous. By Lemma 2.3 $L_N^{(2)} \to_w L_0$ and so an application of the continuous mapping theorem ([3], Theorem 5.1) shows that

$$T_N^{(2)}(\cdot) = h_2(L_N^{(2)}(\cdot, \cdot)) \to_w h_2(L_0(\cdot, \cdot)) . \tag{2.6}$$

It is easy to establish that the latter is indeed a standard Wiener process.

We now show that $T_N^{(1)}$ and $T_N^{(2)}$ have the same weak limits. First,

$$\rho(T_N^{(1)}, T_N^{(2)}) = \max_{1 \leqq k \leqq N} |(k/N)^{\frac{1}{2}}\sigma_0^{-1} \textstyle\int_{k^{-1}}^1 \delta_k(t)\, d\nu(t)| ,$$

since $T_N^{(1)}$ and $T_N^{(2)}$ are step functions with discontinuities at common time points. Recall that $|\delta_k(t)| \leqq \lambda_k^{-1}k^{-\frac{1}{2}}$ and $\int d|\nu| < \infty$ ([14], page 762). Hence

$$\rho(T_N^{(1)}, T_N^{(2)}) \leqq [\textstyle\int d|\nu|] \cdot \max_{1 \leqq k \leqq N} (\lambda_k^{-1}k^{-\frac{1}{2}})k^{\frac{1}{2}}N^{-\frac{1}{2}}\sigma_0^{-1} = o(1) ,$$

and $T_N^{(1)}$ and $T_N^{(2)}$ have the same weak limits.

Finally, employing Assumption 2.2,

$$\rho(T_N, T_N^{(1)}) = \max_{1 \leqq k \leqq N} |(k/N)^{\frac{1}{2}}\sigma_0^{-1} \textstyle\int_0^1 L_k^{(1)}\, d(\nu_k - \nu) \to 0 \quad \text{(a.s.)}.$$

Hence T_N, $T_N^{(1)}$, and $T_N^{(2)}$ have the same weak limits. Applying (2.6), the theorem follows. □

Before proceeding to Section 3, we present an immediate consequence of our theorem. First, it would be of use in statistical applications to have our result available for processes defined on $[0, \infty)$. That is, suppose we defined $T_N(\cdot)$ by

$$T_N(k/N) = (k/N)^{\frac{1}{2}}\sigma_0^{-1} \int_0^1 L_k^{(1)}(t)\, d\nu_k(t) \qquad 1 \leqq k < \infty$$

and in general,

$$T_N(s) = T_N([Ns]/N) \qquad 0 \leqq s < \infty .$$

Then $T_N(\cdot)$ is a process on $[0, \infty)$ and we would like to show that it converges weakly to a Wiener process on $[0, \infty)$, in the extended Skorokhod topology. Now the work of Stone [20] and others has shown that, in this case, weak convergence on $[0, \infty)$ is implied by weak convergence on every compact subinterval of $[0, \infty)$. A little thought shows the latter follows directly from the proof of Theorem 2.1. □

3. Unbounded score functions.

3.A. *Preliminaries*. The aim is to weaken condition (ii) of Theorem 2.1 so that the measure $\nu(\cdot)$ induced by the score function $-J(\cdot)$ satisfies $\int_0^1 |d\nu| < \infty$. This is achieved by demonstrating the weak convergence of $L_N^{(2)}(\cdot, \cdot)$ in the so-called ρ_q metric (see [14] for details), where q is an element of the class of functions specified below. O'Reilly [11] gives necessary and sufficient conditions on a function $q(t): 0 \leqq t \leqq 1$ in order that $Z_N(1, t)/q(t)$ converge weakly to the appropriate Gaussian process. Let Q denote the class of such functions. Then we have

THEOREM 3.1. *Let $q \in Q$. Then*

$$\frac{Z_N(s, t)}{q(t)} \to_w \frac{Z_0(s, t)}{q(t)} . \tag{3.1}$$

PROOF[2]. Bickel and Wichura showed

$$Z_N(s, t) \to_w Z_0(s, t) .$$

Thus, this proof requires only checking that division by $q(t)$ does not cause things to "blow up" near 0 or 1. Formally we must show that

$$p \lim_{\varepsilon \downarrow 0} \limsup_{N\to\infty} \sup_{0 \leqq s \leqq 1; 0 < t \leqq \varepsilon} |Z_N(s, t)|/q(t) = 0 , \tag{3.2}$$

where ε appears in the definition of q. A similar condition for t near 1 is also required, but the proof would be the same as for (3.2). Applying the Banach space version of Skorokhod's inequality [7], it suffices to prove

$$p \lim_{\varepsilon \downarrow 0} \limsup_{N\to\infty} \sup_{0 < t \leqq \varepsilon} |Z_N(1, t)|/q(t) = 0 . \tag{3.3}$$

But (3.3) follows directly from O'Reilly's result. □

[2] The present proof is due to a referee. It is much simpler than the author's own proof which only applied to a subclass of Q.

In the demonstration that the introduction of the random time transformation $\phi_N(\cdot, \cdot)$ into $Z_N(\cdot, \cdot)$ does not alter the conclusion of Theorem 3.1, the following result on the behavior of uniform order statistics is crucial.

3.B. *Inequalities for uniform order statistics.*

LEMMA 3.1. *Let* $\{Z_{k,N}\}_{1\leq k\leq N}$ *be the order statistics from a sample of size* N *from the uniform distribution on* $[0, 1]$, *with* $0 \leqq Z_{1,N} \leqq Z_{2,N} \leqq \cdots \leqq Z_{N,N} \leqq 1$. *Let* $Z^*_{k,N} = Z_{k,N} - E(Z_{k,N})$. *Then* $\{Z^*_{k,N}/(N-k+1)\}_{1\leq k\leq N}$ *is a martingale sequence.*

PROOF.

$$\begin{aligned} E[Z^*_{k+1,N} \mid Z^*_{k,N}] &= E[Z_{k+1,N} \mid Z_{k,N}] - E[Z_{k+1,N}] \\ &= Z_{k,N} + \frac{(1 - Z_{k,N})}{N-k+1} - \frac{k+1}{N+1}. \end{aligned}$$

Therefore

$$\begin{aligned} E\left[\frac{Z^*_{k+1,N}}{N-k} \,\middle|\, \frac{Z^*_{k,N}}{N-k+1}\right] &= \frac{Z_{k,N}}{N-k+1} - \frac{k}{(N+1)(N-k+1)} \\ &= \frac{Z^*_{k,N}}{N-k+1}. \end{aligned}$$

Since $\{Z^*_{k,N}\}_{1\leq k\leq N}$ form a Markov process, this suffices to prove the martingale property. □

LEMMA 3.2. *Let* $c > 0$, $0 < \alpha \leqq 1$ *and* $0 < \rho < \frac{7}{8}$. *Given* $\varepsilon > 0$, *there exists* $N(\varepsilon, c, \alpha, \rho)$ *such that*

$$P\{|Z^*_{k,N}| < c(k/N)^\rho,\ 1 \leqq k \leqq [N\alpha] + 1,\ \textit{for all}\ N > N(\varepsilon, c, \alpha, \rho)\} > 1 - \varepsilon .$$

PROOF. Let A_N denote the event

$$\max_{1\leq k\leq [N\alpha]+1} \frac{1}{c}\left(\frac{N}{k}\right)^\rho |Z^*_{k,N}| > 1 .$$

Let

$$\begin{aligned} P_N &= P\{A_N\} \\ &= P\left\{ \max_{1\leq k\leq[N\alpha]+1} \left(\frac{N-k+1}{c}\right)\left(\frac{N}{k}\right)^\rho \left|\frac{Z^*_{k,N}}{N-k+1}\right| > 1\right\} \\ &\leqq \sum_{k=1}^{[N\alpha]+1} (a^8_{k,N} - a^8_{k+1,N}) E\left|\frac{Z^*_{k,N}}{N-k+1}\right|^8, \end{aligned}$$

where

$$\begin{aligned} a_{k,N} &= \left(\frac{N-k+1}{c}\right)\left(\frac{N}{k}\right)^\rho, \qquad & 1 \leqq k \leqq [N\alpha] + 1 \\ &= 0, & [N\alpha] + 2 \leqq k \leqq N. \end{aligned}$$

The inequality follows from applying Theorem 2.1 of Birnbaum and Marshall [4] to the martingale sequence $\{Z^*_{k,N}/(N-k+1)\}_{1\leq k\leq N}$.

It is well known that $Z_{k,N}$ is distributed as Beta $(k, N-k+1)$. A long and tedious calculation shows that

$$E[Z_{k,N}^*]^8 = \frac{7k \sum_{\{(i,j): i+j \leqq 10, i \leqq 7, j \leqq 7\}} d(i,j) N^i k^j}{(N+1)^8 (N+8)(N+7) \cdots (N+2)}$$

where i and j are integers and $d(i,j)$ are constants depending on i and j. Therefore,

$$\begin{aligned} P_N &\leqq \sum_{k=1}^{[N\alpha]+1} a_{k,N}^8 E\left|\frac{Z_{k,N}^*}{N-k+1}\right|^8 \\ &= O\left(\sum_{k=1}^{[N\alpha]+1} \left(\frac{N}{k}\right)^{8\rho} \frac{k^4 N^7}{N^{15}}\right) \\ &= O\left(\sum_{k=1}^{[N\alpha]+1} \left(\frac{N}{k}\right)^{8\rho} \frac{1}{N^{2+\eta}} \frac{k^{6-\eta}}{N^{6-\eta}}\right) \qquad 0 < \eta < 2 \\ &= O\left(\frac{1}{N^{1+\eta}} \int_0^\alpha \frac{dx}{x^{8\rho-6+\eta}}\right) \qquad \text{if } 8\rho - 6 + \eta < 1 \\ &= O\left(\frac{1}{N^{1+\eta}}\right). \end{aligned}$$

Hence $\sum P_N < \infty$ and so by the Borel–Cantelli lemma, $P\{A_N \text{ infinitely often}\} = 0$. It follows that given $\varepsilon > 0$, there exists $N(\varepsilon, c, \alpha, \rho)$ satisfying the conditions of the theorem. Note that the restriction $8\rho - 6 + \eta < 1$ reduces to requiring $\rho < \frac{7}{8}$ since η was arbitrary. □

REMARK 3.1. Wellner [21] has given excellent approximations to the central moments of the Beta distribution. He shows that

$$E|Z_{k,N}^*|^r \leqq C_r (k/N^2)^{r/2}$$

where C_r is a constant depending only on r. These bounds, when combined with the method above, yield the result of Lemma 3.2 for all $\rho < 1$. We will feel free therefore to use this more general result.

Now define two processes that will be needed in the proof of Theorem 3.2. Let

$$\begin{aligned} L_N^{(3)}(s,t) &= L_N^{(2)}(s,t), \qquad 0 \leqq s \leqq 1, \quad [Ns]^{-1} \leqq t \leqq 1 - [Ns]^{-1} \\ &= 0, \qquad \text{elsewhere} \\ W_N(s,t) &= \frac{\lambda_{Ns}^{-\frac{1}{2}} Z_N(\phi_N(s,t))}{q(t)}, \qquad 0 \leqq s \leqq 1, \quad [Ns]^{-1} \leqq t \leqq 1 - [Ns]^{-1} \\ &= 0, \qquad \text{elsewhere.} \end{aligned}$$

Thus on the set where it is not defined to be zero,

$$W_N(s,t) = \lambda_{Ns}^{-\frac{1}{2}} [(\lambda_{Ns} s)^{\frac{1}{2}} U_N(\lambda_{Ns} s, FH_{Ns}^{-1}(t))]/q(t).$$

3.C. *Main result.* It seems appropriate at this point to note that in order to prove that a sequence of measures $\{P_N\}$ on our function space is tight, it is necessary and sufficient to show (see [2]):

For each positive η there exists a K such that

$$P_N\{x : \sup_{v \in [0,1]^2} |x(v)| > K\} < \eta. \tag{3.4}$$

For each positive ε and η there exists a positive μ and an integer N_0 such that

$$P_N\{x: w_x'(\mu) \geqq \varepsilon\} \leqq \eta\,, \qquad N \geqq N_0 \tag{3.5}$$

where

$$w_x'(\mu) = \inf_\Delta \max_{G\in\Delta} \sup_{v,v'\in G} |x(v) - x(v')|$$

and where the infimum extends over all partitions Δ of $[0, 1]^2$ formed by finitely many lines parallel to the coordinate axes, and such that each cell of the grid has diameter at least μ. Such a grid is called a μ-grid. G denotes (the closure of) an arbitrary cell of Δ. Note that the diameter of a rectangle is defined as the length of its shortest side.

Finally define

$$Q^* = \{f \in Q: f(t) \geqq \min(t^\delta, (1-t)^\delta) \quad 0 \leqq t \leqq 1 \quad \text{for some } \delta \in (0, \tfrac{1}{2})\}\,.$$

Though our results apply more generally, the family Q^* is broad enough to cover most cases of interest.

THEOREM 3.2. *Assume that* (i) $\lambda_N \to \lambda_0$ (a.s.), (ii) $\int_0^1 q\, d|\nu| \to \infty$ *for some* $q \in Q^*$ *and that Assumptions* 2.1 *and* 2.2 *hold. Then*

$$T_N(\cdot) \to_w \textit{standard Wiener process.}$$

The proof of the theorem requires three preliminary lemmas.

LEMMA 3.3. *Suppose that in the definition of* $W_N(s, t)$, $q(t) = \min(t^\delta, (1-t)^\delta)$, *for some* $\delta \in (0, \frac{1}{2})$. *Then* (3.4) *holds when* P_N *denotes the measure induced by* $W_N(\cdot, \cdot)$.

PROOF. Proving (3.4) requires showing that given $\eta > 0$, there exists K such that

$$P\{|W_N(s, t)| < K, (s, t) \in [0, 1]^2\} > 1 - \eta \qquad \text{for } N \geqq 1\,. \tag{3.6}$$

Let $\theta_N = N^{-\frac{3}{4}}$ and $0 < \alpha < \tilde{\lambda}$. Define regions $R_1 = [0, \theta_N] \times [0, 1]$, $R_2 = [\theta_N, 1] \times [\alpha, 1-\alpha]$ and $R_3 = [\theta_N, 1] \times \{[0, \alpha] \cup [1-\alpha, 1]\}$. Then $[0, 1]^2 = R_1 \cup R_2 \cup R_3$ and (3.6) will follow from showing $P\{|W_N(s, t)| < K, (s, t) \in R_i\} > 1 - \eta$ for $N \geqq 1$ $(i = 1, 2, 3)$. For $(s, t) \in R_1$,

$$\begin{aligned}\sup_{[Ns]^{-1} \leqq t \leqq 1-[Ns]^{-1}} |\lambda_{Ns}^{-\frac{1}{2}}[(\lambda_{Ns}s)^{\frac{1}{2}} U_N(\lambda_{Ns}s, FH_{Ns}^{-1}(t))]/q(t)| &\leqq s^{\frac{1}{2}}(Ns\lambda_{Ns})^{\frac{1}{2}}(Ns)^\delta \\ &\leqq \theta_N^{1+\delta} N^{\frac{1}{2}+\delta} \leqq N^{(\delta-1)/4}\,.\end{aligned}$$

Thus $|W_N(s, t)| = o(1)$ on R_1. Next, the weak convergence of $\lambda_{Ns}^{-\frac{1}{2}} Z_N(\phi_N(s, t))$ and the fact that $q(t)$ is bounded away from 0 on R_2 together imply that $|W_N(s, t)|$ is bounded in probability on R_2. R_3 will require more work.

Noting the symmetry about $t = \frac{1}{2}$ and that it suffices to work with N sufficiently large, the problem can be reduced to showing

$$P\{|W_N(s, t)| < K, (s, t) \in [\theta_N, 1] \times [0, \alpha]\} > 1 - \eta\,, \qquad N \geqq N_0\,. \tag{3.7}$$

Let $\bar{q}(t) = \min(t^{\bar{\delta}}, (1-t)^{\bar{\delta}})$ where $\delta < \bar{\delta} < \frac{1}{2}$. Then (3.7) is implied by the following two inequalities to be established:

$$(3.8)\qquad P\left\{\left|\frac{s^{\frac{1}{2}}U_N(\lambda_{Ns}s,\, FH_{Ns}^{-1}(t))}{\bar{q}(FH_{Ns}^{-1}(t))}\right| < K,\, \theta_N \leqq s \leqq 1,\, [Ns]^{-1} \leqq t \leqq \alpha\right\} > 1 - \eta/2\,, \qquad N \geqq N_0$$

and

$$(3.9)\qquad P\{\bar{q}(FH_{Ns}^{-1}(t)) < q(t),\, \theta_N \leqq s \leqq 1,\, [Ns]^{-1} \leqq t \leqq \alpha\} > 1 - \eta/2\,, \qquad N \geqq N_0\,.$$

We first prove (3.8).

In Theorem 3.1 it was shown that

$$s^{\frac{1}{2}}U_N(s, t)/\bar{q}(t) \to_w s^{\frac{1}{2}}U_0(s, t)/\bar{q}(t) \qquad \text{on} \quad [0, 1]^2\,.$$

It follows that

$$(3.10)\qquad s^{\frac{1}{2}}U_N(\lambda_{Ns}s, t)/\bar{q}(t) \to_w s^{\frac{1}{2}}U_0(\lambda_0 s, t)/\bar{q}(t) \qquad \text{on} \quad [0, 1]^2\,,$$

since the transformation $(s, t) \to (\lambda_{Ns}s, t)$ does not affect tightness. But (3.10) implies that

$$(3.11)\qquad P\{|s^{\frac{1}{2}}U_N(\lambda_{Ns}s, t)/\bar{q}(t)| < K;\, (s, t) \in [0, 1]^2\} > 1 - \eta/2\,, \qquad N \geqq N_0\,,$$

when N_0 is chosen sufficiently large. (3.8) is an immediate consequence of (3.11). We now prove (3.9).

The statement: $\bar{q}(FH_{Ns}^{-1}(t)) < q(t)$ for $[Ns]^{-1} \leqq t \leqq \alpha$, is implied by the statement: $FH_{Ns}^{-1}(t) - t < ct^{\delta/\bar{\delta}}$ for $[Ns]^{-1} \leqq t \leqq \alpha$, if c is a constant chosen to be suitably (depending on α, δ, $\bar{\delta}$) smaller than one. Since $H_{Ns}^{-1}(t)$ is left continuous and constant between points of jump, in order to prove (3.9) it suffices to show

$$(3.12)\qquad \begin{aligned} P\{&FH_{Ns}^{-1}(1/Ns) - 1/Ns < c(1/Ns)^{\delta/\bar{\delta}},\\ &FH_{Ns}^{-1}(k/Ns) - ((k-1)/Ns) < c((k-1)/Ns)^{\delta/\bar{\delta}};\\ &2 \leqq k \leqq [Ns\alpha] + 1,\, s \in D_N\} > 1 - \eta/2 \qquad \text{for} \quad N \geqq N_0\,, \end{aligned}$$

where $D_N = \{k/N\colon k \text{ an integer};\, \theta_N \leqq k/N \leqq 1\}$. Note that in (3.12) the cases $k \geqq 2$ are treated differently from the case $k = 1$ because of the left-continuity of the inverse function which implies that

$$\lim_{t \downarrow (k-1)/Ns} (FH_{Ns}^{-1}(t) - t) = FH_{Ns}^{-1}(k/Ns) - (k-1)/Ns \qquad k \geqq 2\,.$$

The strategy is to modify (3.12) so that Lemma 3.2 can be brought to bear. Note that $H_{Ns}^{-1}(k/Ns) \leqq F_{m(Ns)}^{-1}(k/m(Ns))$, $1 \leqq k \leqq m(Ns)$, and $FF_{m(Ns)}^{-1}(k/m(Ns)) = Z_{k,m(Ns)} = k$th o.s. in a sample of size $m(Ns)$ from a uniform $[0, 1]$ distribution and that

$$\bar{\lambda}/m(Ns) < 1/Ns < (1 - \bar{\lambda})/m(Ns)\,.$$

Thus to prove (3.12) it suffices to show

$$(3.13)\qquad \begin{aligned} P\{&Z_{1,m(Ns)} - 1/m(Ns) < c'(1/m(Ns))^{\delta/\bar{\delta}},\\ &Z_{k,m(Ns)} - (k-1)/m(Ns) < c'((k-1)/m(Ns))^{\delta/\bar{\delta}};\\ &2 \leqq k \leqq m(Ns),\, s \in D_N\} > 1 - \eta/2 \qquad \text{for} \quad N \geqq N_0\,, \end{aligned}$$

where c' is a constant chosen suitably smaller than $c\tilde{\lambda}^{\delta/\bar{\delta}}$. But (3.13), in turn, is implied by

$$(3.14)\qquad P\{Z_{k,m(Ns)} - k/(m(Ns)+1) < c''(k/m(Ns))^{\delta/\bar{\delta}}; \\ 1 \leqq k \leqq m(Ns),\ s \in D_N\} > 1 - \eta/2 \qquad \text{for} \quad N \geqq N_0\,,$$

where c'' is a constant chosen suitably smaller than c'. Finally, noting that $m(N\theta_N) \to \infty$ as $N \to \infty$, it is clear that (3.14) is certainly implied by

$$P\{Z_{k,N} - k/(N+1) < c''(k/N)^{\delta/\bar{\delta}};\ 1 \leqq k \leqq N \ \text{ for all } \ N > N(\eta, c'', \delta/\bar{\delta})\} \\ > 1 - \eta/2\,.$$

This last statement, though, follows from Lemma 3.2 with $\delta/\bar{\delta}$ here replaced by ρ there. Note that $\delta/\bar{\delta} < 1$, so that $\rho < 1$ is required. See Remark 3.1. We have thus proven (3.9) and so the lemma. □

REMARK 3.2. This method of handling the introduction of $\phi_N(\cdot, \cdot)$ applies to a more general class of random time transformations.

REMARK 3.3. In the next lemma, we will have occasion to use the following result: Given $\eta, \varepsilon > 0$ there exists $\alpha > 0$ such that

$$(3.15)\qquad P\{|W_N(s,t)| < \varepsilon;\ (s,t) \in [0,1] \times ([0,\alpha] \cup [1-\alpha, 1])\} > 1 - \eta \\ \text{for} \quad N \geqq N(\varepsilon, \eta, \alpha)\,.$$

The proof is essentially the same as that of (3.6) except that the proof of the analog of (3.8) does not hold unless α can be chosen suitably small. Then the weak convergence of $Z_N(s,t)/\bar{q}(t)$ together with a continuity argument can be used to prove the analog of (3.11) with K replaced by a fixed $\varepsilon > 0$. The proof of (3.9) carries over unchanged.

LEMMA 3.4. *Suppose that in the definition of* $W_N(s,t)$, $q(t) = \min(t^\delta, (1-t)^\delta)$, *for some* $\delta \in (0, \frac{1}{2})$. *Then* (3.5) *holds where* P_N *denotes the measure induced by* $W_N(\cdot, \cdot)$.

PROOF. Equation (3.15) implies that given $\varepsilon, \eta > 0$ there exists $\alpha(\varepsilon, \eta)$ and $N(\varepsilon, \eta, \alpha)$ such that

$$(3.16)\qquad P\{|W_N(s,t)| < \varepsilon/2;\ (s,t) \in [0,1] \times ([0,\alpha] \cup [1-\alpha, 1])\} > 1 - \eta \\ \text{for} \quad N \geqq N(\varepsilon, \eta, \alpha)\,.$$

For such an α, consider the parameter set $[0,1] \times [\alpha, 1-\alpha]$. The proof of Lemma 2.3 can be applied to show that $Z_N(\phi_N)$ converges weakly and hence the corresponding sequence of distributions must be tight. Thus the distribution of W_N restricted to the domain $[0,1] \times [\alpha, 1-\alpha]$ must also be tight. But this tightness implies the analog of (3.5) on the domain $[0,1] \times [\alpha, 1-\alpha]$ with x replaced by W_N. Choose $\mu < \alpha$ and let Δ be any μ-grid on $[0,1] \times [\alpha, 1-\alpha]$. Then Δ can be extended in the obvious way to be a μ-grid on $[0,1] \times [0,1]$, by simply adjoining the lines $t = 0$ and $t = 1$. Call the extended grid Δ^*. Let

x denote the process W_N. Then

$$\max_{G\in\Delta}\sup_{v,v'\in G}|x(v)-x(v')|<\varepsilon \tag{3.17}$$

implies

$$\max_{G\in\Delta^*}\sup_{v,v'\in G}|x(v)-x(v')|<\varepsilon$$

if x lies in the set displayed in (3.16) which has probability greater than $1-\eta$. Since the probability of the set of x's satisfying (3.17) is also greater than $1-\eta$ for N sufficiently large, therefore given ε, $\eta>0$ there exists a $\mu>0$ such that

$$P_N\{x\colon w_x'^*(\mu)>\varepsilon\}<2\eta \qquad \text{for } N \text{ sufficiently large,}$$

where $w_x'^*(\mu)=\inf_{\Delta^*}\max_{G\in\Delta^*}\sup_{v,v'\in G}|x(v)-x(v')|$. This proves (3.5) since $w_x'^*(\mu)\geqq w_x'(\mu)$ where the latter term denotes the infimum taken over all μ-grids on $[0,1]\times[0,1]$. □

LEMMA 3.5. *Under the hypotheses of Theorem* 3.2,

$$L_N^{(3)}(s,t)/q(t)\to_w L_0(s,t)/q(t)\,.$$

PROOF. The proof of Lemma 2.3 implies that the finite dimensional distributions of $W_N(\cdot,\cdot)$ converge to those of $s^{\frac12}U_0(\lambda_0 s, FH_0^{-1}(t))/q(t)$. Lemmas 3.3 and 3.4 together imply that the sequence $\{W_N\}$ is tight. Hence

$$W_N(s,t)\to_w s^{\frac12}U_0(\lambda_0 s, FH_0^{-1}(t))/q(t)\,.$$

It is now relatively easy to show (using methods developed in Section 2)

$$(1-\lambda_{Ns})\lambda_{Ns}^{-\frac12}B_N(s,t)W_N(s,t)\to_w (1-\lambda_0)\lambda_0^{-\frac12}b_0(s,t)s^{\frac12}U_0(\lambda_0 s, FH_0^{-1}(t))/q(t)\,.$$

A similar result naturally holds for the analogous term involving A_N and V_N. Although addition is in general not a continuous operation in the Skorokhod topology, the two summands in L_0 are both a.s. continuous and Skorokhod convergence to continuous limits is equivalent to convergence in the supremum metric. This fact together with the convergence results implies that

$$\frac{L_N^{(3)}(\cdot,\cdot)}{q(\cdot)}\to_w\frac{L_0(\cdot,\cdot)}{q(\cdot)}\,. \qquad □$$

PROOF OF THEOREM 3.2. The proof follows the lines established in Theorem 2.1. Define a new process by

$$T_N^{(3)}(s)=\sigma_0^{-1}\int_0^1[L_N^{(3)}(s,t)/q(t)]q(t)\,d\nu(t) \qquad 0\leqq s\leqq 1\,,$$

which may be formally rewritten as

$$T_N^{(3)}(\cdot)=h_3(L_N^{(3)}(\cdot,\cdot)/q(\cdot))$$

where the argument of $q(\cdot)$ corresponds to the second argument of $L_N^{(3)}(\cdot,\cdot)$. The operator h_3 is implicitly defined. Since the measure "$q\,d\nu$" has finite absolute variation on $[0,1]$, it is clear that h_3 is continuous. Using the result of Lemma 3.5 and the continuous mapping theorem, we have

$$T_N^{(3)}(\cdot)=h_3(L_N^{(3)}(\cdot,\cdot)/q(\cdot))\to_w h_3(L_0(\cdot,\cdot)/q(\cdot))\,. \tag{3.18}$$

The latter is, of course, a standard Wiener process. We now show that $T_N^{(1)}(\cdot)$ and $T_N^{(3)}(\cdot)$ have the same weak limits.

$$\rho(T_N^{(1)}(\cdot), T_N^{(3)}(\cdot)) \leqq \max_{1\leqq k\leqq N} \left|\frac{(k/N)^{\frac{1}{2}}}{\sigma_0} \int_{k^{-1}}^{1-k^{-1}} \frac{\delta_k(t)}{q(t)} q(t)\, d\nu(t)\right| + \max_{1\leqq k\leqq N} \left|\frac{(k/N)^{\frac{1}{2}}}{\sigma_0} \int_{1-k^{-1}}^{1} \frac{L_k(t)}{q(t)} q(t)\, d\nu(t)\right|.$$

Using the assumption that $\int q\, d|\nu| < \infty$, and the results that

$$\sup_{k^{-1}\leqq t\leqq 1-k^{-1}} |\delta_k(t)/q(t)| = o(1) \qquad \text{and} \qquad \sup_{1-k^{-1}\leqq t\leqq 1} |L_k(t)/q(t)| = o(1)$$

(see [14], page 763), apply the result that $c_k = o(1)$ implies that $\sup_{1\leqq k\leqq N} [(k/N)^{\frac{1}{2}} c_k] = o(1)$, to conclude that

$$\rho(T_N^{(1)}(\cdot), T_N^{(3)}(\cdot)) \to 0 \quad \text{(a.s.)}.$$

Finally, near the end of the proof of Theorem 2.1 we established under Assumption 2.2, that $\rho(T_N(\cdot), T_N^{(1)}(\cdot)) \to 0$ (a.s.). Hence $T_N(\cdot)$ and $T_N^{(3)}(\cdot)$ have the same weak limits. Together with (3.18) this proves the theorem. □

REMARK 3.4. We note that in both Theorems 2.1 and 3.2, the crucial step was to prove that a certain two-parameter process, $L_N^{(2)}(\cdot, \cdot)$ in the first case, $L_N^{(3)}(\cdot, \cdot)/q(\cdot)$ in the second case, converged weakly to the appropriate limit process. The methods used, however, were quite different. In the latter case the proof was more difficult because it required rather sharp bounds on the behavior of the order statistics from a uniform [0, 1] sample. Our particular result in this area, Lemma 3.2, should be of some interest in its own light.

4. Sequential applications.

4.A. *Introduction.* We now examine some applications of Theorem 3.2. In particular, we investigate the selection and evaluation of procedures to be used in conjunction with a sequential linear rank statistic and present a heuristic approach to determining the asymptotic relative efficiency of two competing statistics.

Let T_k denote a linear rank statistic calculated on k observations. As data accumulate, plot T_k against k. It follows from Theorem 3.2 that, when properly rescaled, the resulting process converges weakly to a Wiener process. In fact we propose to act as if the observed graph were actually a realization of such a process, and thus select sequential procedures analogous to those appropriate to the related Wiener process problem. Theorem 3.2 can be used to find approximations to the operating characteristics of this or other procedures based on T_k. One caution: we shall treat the observed graph as a process with known variance, although this is not actually the case; that is, we shall replace $\sigma^2(\theta)$ by $\sigma^2(0)$ when testing $H_1: \theta \geqq 0$ against $H_2: \theta < 0$. For large sample sizes, the relevant range of the parameter is small and the range of the variance is correspondingly so. Nevertheless, we may be losing valuable information on the unknown parameter.

4.B. *Sequential analysis for the Wiener process.* Here we consider the related Wiener process problem in detail. Suppose θ is an unknown parameter and we wish to discriminate between the hypothesis $H_1: \theta \geqq 0$ and $H_2: \theta < 0$. The usual loss structure involves two components: a cost c per unit time of sampling and a loss $l|\theta|$, incurred by making a wrong decision when θ is the true parameter. Since we are interested primarily in values of θ near 0, choosing the loss to be proportional to the magnitude of θ seems reasonable. For a given procedure δ, the risk function is given by

$$R(\theta, \delta) = cE_\theta(T_\delta) + l|\theta|\varepsilon_\delta(\theta) , \tag{4.1}$$

where T_δ is the sampling time for δ and $\varepsilon_\delta(\theta)$ is the probability of accepting the wrong hypothesis when θ is the true parameter value and δ is used.

Suppose the data consists of a Wiener process ξ with mean drift $\mu(\theta)$, variance $\sigma^2(\theta)$ and that $\mu(\theta)$ is approximately proportional to θ for θ in a neighborhood of 0. We can then consider the *restricted* version of the problem for which the procedures δ are confined to the use of ξ. Here the aim would be to test H_1: $\mu(\theta) \geqq 0$ against H_2: $\mu(\theta) < 0$ with reparameterized risk function,

$$\tilde{R}(\mu, \delta) = cE_\mu(T_\delta) + \tilde{l}|\mu|\varepsilon_\delta(\mu) . \tag{4.2}$$

By choosing $\tilde{l} = l/\mu'(0)$, $\tilde{R}(\mu, \delta)$ is nearly equal to $R(\theta, \delta)$ and the approximation is good in the relevant range of θ. The restricted setup is closely related to the *transformed* problem of testing

$$\tilde{H}_1: \mu \geqq 0 \quad \text{against} \quad \tilde{H}_2: \mu < 0 , \tag{4.3}$$

using the risk function $\tilde{R}(\mu, \delta)$ with the Wiener process ξ, mean μ and variance $\sigma^2(0)$, as data.

For simplicity, the choice of a stopping rule for the transformed problem is made from the class of Wald strategies. These involve a pair of straight line barriers, parallel to the time axis and at heights $\pm a$, for some constant a. The process is observed until it crosses one of the barriers, and a decision is made in favor of $\tilde{H}_1$ or $\tilde{H}_2$ according as the upper or lower barrier is hit. Denote the above scheme by δ_a.

Let

$$g(u) = [e^u + 1]^{-1} \qquad u > 0$$

and

$$h(u) = e^u - 1 \qquad u > 0 .$$

Then (cf. DeGroot [6])

$$\tilde{R}(\mu, \delta_a) = g(2\mu a/\sigma^2)[ca\mu^{-1}h(2\mu a/\sigma^2) + \tilde{l}|\mu|] . \tag{4.4}$$

Making the transformations

$$a = (\tilde{l}c^{-1}\sigma^4)^{\frac{1}{3}}x \qquad \text{and} \qquad \mu = (\tilde{l}^{-1}c\sigma^2)^{\frac{1}{3}}y ,$$

we obtain

$$\begin{aligned}\tilde{R}(\mu, \delta_a) &= (c\tilde{l}^2\sigma^2)^{\frac{1}{3}}g(2xy)[xy^{-1}h(2xy) + y] , \qquad \mu > 0 \\ &=_\Delta (c\tilde{l}^2\sigma^2)^{\frac{1}{3}}\bar{R}(y, x) ,\end{aligned}$$

where $\tilde{R}(\cdot, \cdot)$ is the risk function in the normalized problem for which the parameters c, $\bar{l}$ and σ^2 are equal to 1. Thus $(c\bar{l}^2\sigma^2)^{\frac{1}{3}}$ is a key parameter in the study of $\tilde{R}(\mu, \delta_a)$.

It is of some interest to investigate the minimax symmetric Wald procedure. DeGroot [6] has shown it exists and is in fact minimax among all decision procedures. Let a^* denote the height of the minimax barrier and μ^* the least favorable value of μ. Then $a^* = \rho(\bar{l}\sigma^4/c)^{\frac{1}{3}}$ and $\mu^* = \pm\rho'(c\sigma^2/\bar{l})^{\frac{1}{3}}$, where ρ and ρ' are constants with approximate values 0.426 and 1.38 respectively. A simple calculation shows that

$$\tilde{R}(\mu^*, \delta_{a^*}) \cong 0.488(c\bar{l}^2\sigma^2)^{\frac{1}{3}} . \tag{4.5}$$

Further, $\tilde{R}(0, \delta_a) = \lim_{\mu\to 0} \tilde{R}(\mu, \delta_a) = ca^2/\sigma^2$, so that

$$\tilde{R}(0, \delta_{a^*}) \cong 0.181(c\bar{l}^2\sigma^2)^{\frac{1}{3}} .$$

Note that if $c \to 0$, then $\mu^* = O(c^{\frac{1}{3}})$ and $a^* = O(c^{-\frac{1}{3}})$.

A second approach is to obtain the Wald procedure which is Bayes against a suitable prior. Specifically, assume a symmetric normal prior for the parameter μ; that is, $\mu \sim N(0, \sigma_0^2)$. This implies that $y \sim N(0, \sigma_1^2)$ where $\sigma_1 = \sigma_0(\bar{l}c^{-1}\sigma^{-2})^{\frac{1}{3}}$. Set $v = 2xy$ and $\alpha = (8\sigma_1^2x^2)^{-1}$. The Bayes risk is given by

$$r(a, \sigma_0) = (c\bar{l}^2\sigma^2)^{\frac{1}{3}}r(x, \sigma_1) ,$$

where

$$\begin{aligned} r(x, \sigma_1) &= 2(2\pi\sigma_1^2)^{-\frac{1}{2}}\{\textstyle\int_0^\infty xy^{-1}g(2xy)h(2xy)\exp(-y^2/2\sigma_1^2)\,dy \\ &\qquad + \textstyle\int_0^\infty yg(2xy)\exp(-y^2/2\sigma_1^2)\,dy\} \\ &= 2(2\pi\sigma_1^2)^{-\frac{1}{2}}\{x\textstyle\int_0^\infty v^{-1}g(v)h(v)e^{-\alpha v^2}\,dv + \frac{1}{4}x^{-2}\int_0^\infty vg(v)e^{-\alpha v^2}\,dv\} \\ &= 2(2\pi\sigma_1^2)^{-\frac{1}{2}}\{xH_1 + \tfrac{1}{4}x^{-2}H_2\} . \end{aligned} \tag{4.6}$$

The quantities H_1 and H_2 are implicitly defined by the final equality. Since a closed form solution seems difficult to obtain for finite σ_1, we find an approximation to the Bayes procedure as $\sigma_1 \to \infty$ ($c \to 0$).

It is clear that $\lim_{\alpha\to 0} H_2 = \int_0^\infty vg(v)\,dv$. After dividing the region of integration and expanding the integrand we find that

$$\lim_{\alpha\to 0} H_1 = \int_0^1 v^{-1}g(v)h(v)\,dv + \lim_{\alpha\to 0}\int_1^\infty v^{-1}e^{-\alpha v^2}\,dv - 2\int_1^\infty v^{-1}g(v)\,dv .$$

The second integral requires separate consideration. Setting $w = \alpha v^2$, we have

$$\int_1^\infty v^{-1}e^{-\alpha v^2}\,dv = \tfrac{1}{2}\{\textstyle\int_\alpha^1 w^{-1}e^{-w}\,dw + \int_1^\infty w^{-1}e^{-w}\,dw\} .$$

The final integral is finite, whilst

$$\int_\alpha^1 w^{-1}e^{-w}\,dw = \int_\alpha^1 w^{-1}h(-w)\,dw + \int_\alpha^1 w^{-1}\,dw .$$

The first integral on the RHS converges as $\alpha \to 0$ and the second integral equals $-\log\alpha$. We have thus shown that

$$xH_1 + \tfrac{1}{4}x^{-2}H_2 \simeq x[K_1 - \tfrac{1}{2}\log\alpha] + \tfrac{1}{4}x^{-2}K_2 \tag{4.7}$$

where

$$K_1 = \int_0^1 v^{-1}g(v)h(v)\,dv - 2\int_1^\infty v^{-1}g(v)\,dv + \tfrac{1}{2}[\int_0^1 w^{-1}h(-w)\,dw + \int_1^\infty w^{-1}e^{-w}\,dw]$$

and

$$K_2 = \int_0^\infty vg(v)\,dv\,.$$

The approximation is valid for $\alpha = (8\sigma_1^2x^2)^{-1}$ close to 0.

Substitute (4.7) into (4.6), differentiate the resulting approximation to $r(x, \sigma_1)$ with respect to x, and set equal to 0 to obtain

$$K_1 + \tfrac{3}{2}\log 2 + \log\sigma_1 + \log x + 1 - \tfrac{1}{2}K_2x^{-3} = 0\,.$$

An approximate solution is given by

$$X_B = (K_2/2\log\sigma_1)^{\frac{1}{3}} \qquad \text{(valid as } \sigma_1 \to \infty)\,.$$

The form of the solution justifies neglecting the $\log x$ term. The corresponding Bayes risk $r(a_B, \sigma_0)$ can be shown to be approximately proportional to

$$\frac{(c\bar{l}^2\sigma^2)^{\frac{1}{3}}}{\sigma_0(\bar{l}c^{-1}\sigma^{-2})^{\frac{1}{3}}}[\log\sigma_0(\bar{l}c^{-1}\sigma^{-2})^{\frac{1}{3}}]^{\frac{2}{3}} \tag{4.8}$$

and hence is $O(-c\log c)^{\frac{2}{3}}$.

Whilst one might ordinarily prefer to use procedures which do not employ Wald barriers, they do not appear amenable to the present analysis. The remainder of the paper is thus confined to examining some features of the Bayes and minimax Wald procedures as they apply to the sequence of nonparametric statistics.

4.C. *Connection with sequential rank tests.* Consider the two-sample location shift problem. Let $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$ be independent observations drawn sequentially from two populations with continuous distributions F and G, respectively. Suppose that $G(x) = F(x + \theta)$. In order to test $H_1: \theta \geqq 0$ against $H_2: \theta < 0$, employ a sequential version of a linear rank statistic T. Let J be the limiting score function of T and assume $\lambda_N \to \lambda_0$. Set

$$m(\theta) = \int_0^1 J[\lambda_0 F(x) + (1 - \lambda_0)F(x + \theta)]F(dx)\,.$$

Writing $T_k(\theta) = k^{\frac{1}{2}}[T_k^* - m(\theta)]$, Theorem 3.2 implies that, when θ is the true value of the parameter, the sequential statistic based on $T_k(\theta)$ converges weakly to a standard Wiener process. Since $T_k(\theta)$ can not be observed because θ is unknown, it is customary to use the statistic $T_k = T(0) = k^{\frac{1}{2}}[T_k^* - m(0)]$. Thus when $\theta \neq 0$, the observed graph is approximated by a Wiener process with mean drift $\mu(\theta) = m(\theta) - m(0)$ and variance $\sigma^2(\theta) \simeq \sigma^2(0)$. The sense of the approximation can be made rigorous by considering a sequence of problems, indexed by N, in which $\theta = O(N^{-\frac{1}{2}})$ and time is scaled by a factor of N^{-1}. It seems reasonable therefore, to act as if the sequential statistic actually were a Wiener process with mean $\mu(\theta)$ and variance $\sigma^2(0)$, and to employ sequential procedures appropriate to Wiener process problems.

In order to make use of the discussion around equations (4.1) to (4.3), we must first show that $\mu(\theta)$ is indeed proportional to θ near 0. Assume that $J(\cdot)$ can be expanded by Taylor's formula up to the second derivative, that $J'(x) \geqq 0$ for $x \in (0, 1)$, that F has a first derivative and that there exists a function $D(x)$ satisfying

(i) $|(F(x+\theta) - F(x))/\theta| \leqq D(x)$ for θ sufficiently small, and all x, and
(ii) $\int_0^1 D(x)J'(F(x))F(dx) < \infty$.

Then

$$\begin{aligned} m(\theta) &= \textstyle\int_0^1 J[F(x) + (1-\lambda_0)[F(x+\theta) - F(x)]]F(dx) \\ &= \textstyle\int_0^1 \{J[F(x)] + (1-\lambda_0)[F(x+\theta) - F(x)]J'[F(x)] \\ &\quad + \tfrac{1}{2}(1-\lambda_0)^2[F(x+\theta) - F(x)]^2 J''[F(x) + \phi(1-\lambda_0) \\ &\quad \times [F(x+\theta) - F(x)]]\}F(dx) \qquad \text{for some } \phi \in (0, 1)\,. \end{aligned}$$

Therefore under the present assumptions, as θ tends 0,

$$\mu(\theta) = m(\theta) - m(0) \simeq \theta \textstyle\int_0^1 (1-\lambda_0)F'(x)J'[F(x)]F(dx)\,.$$

Since $\mu(\theta)$ has the same sign as θ, tests for the sign of $\mu(\theta)$ are also tests for the sign of θ.

The principal aim of this section is to compare the performance of two sequential statistics based on different linear rank statistics. It seems reasonable to mimic the approach taken in computing the Pitman efficiency of two nonparametric tests in the fixed-sample situation. It has already been shown that the location shift problem, for a distribution F, with risk structure given by (4.1), can be transformed into an equivalent one involving a test for the sign of the drift of a sequential statistic. Furthermore, for small values of θ, the risk structure of the transformed problem, with parameter $\mu(\theta)$, is approximately equivalent to that of the original problem involving θ (cf. (4.2)).

We now indulge in some heuristics. In the transformed problem, regarding the observed sequential statistic as the realization of a Wiener process, compute the maximum risk incurred by employing the minimax procedure. To compare two different statistics, carry out the above calculation for each of them, and then compute the limiting ratio of the maximum risks as the cost of sampling tends to 0. Suppressing functional dependence on the cost c and employing (4.2) and (4.5), yields

$$\begin{aligned} \lim_{c\to 0} \frac{R^*_{(1)}(\mu^*_{(1)}, \delta_{a^*_{(1)}})}{R^*_{(2)}(\mu^*_{(2)}, \delta_{a^*_{(2)}})} &= \lim_{c\to 0} \frac{\tilde{R}_{(1)}(\mu^*_{(1)}, \delta_{a^*_{(1)}})}{\tilde{R}_{(2)}(\mu^*_{(2)}, \delta_{a^*_{(2)}})} \\ &= \lim_{c\to 0} [c\tilde{l}^2_{(1)}\sigma^2_{(1)}(0)/c\tilde{l}^2_{(2)}\sigma^2_{(2)}(0)]^{\frac{1}{3}} \\ &= \left[\frac{l^2\sigma^2_{(1)}(0)}{[\mu'_{(1)}(0)]^2}\Big/\frac{l^2\sigma^2_{(2)}(0)}{[\mu'_{(2)}(0)]^2}\right]^{\frac{1}{3}} \\ &= [E_2/E_1]^{\frac{1}{3}}\,, \end{aligned} \tag{4.9}$$

where $E_i = [\mu'_{(i)}(0)/\sigma_{(i)}(0)]^2$ is the efficacy of the linear rank statistic underlying the ith sequential statistic ($i = 1, 2$). E_2/E_1 is the Pitman asymptotic relative efficiency (a.r.e.) of the two statistics.

Now, it is hoped that (4.9) gives a reasonable approximation to the ratio of "maximum risks" incurred when $\delta_{a^*_{(i)}}$ is used, the data are the *actual* sequential statistics ($i = 1, 2$) and the cost of sampling is small. A similar, though simpler, problem arises in computing the Pitman a.r.e. in the fixed sample case. There, in order to show that the ratio of efficacies corresponds to the limiting ratio of sample sizes, one must establish a uniformity of convergence to normality. Here, even more is required.

Briefly, in the present setup the cost of sampling c indexes the sequence of problems much as N did in Sections 2 and 3. The result required is that the sequential statistics converge weakly to the Wiener process as $c \to 0$ (and so $\theta \to 0$ at an appropriate rate). This would follow from showing that the weak convergence in Theorem 3.2 was uniform in a set of pairs of distributions $\{(F(x), F(x+\theta)) : \theta \in [0, \Delta]\}$ (cf. [15]). Convergence of the relevant boundary crossing probabilities and convergence in distribution of the sampling time would be immediate consequences. However, the risk function also involves the expected sampling time and in order to show that (4.9) is meaningful for the original problem, the convergence of the expected sampling times is required. Unfortunately, neither this latter result nor the uniformity of weak convergence have been obtained as yet in general.

A calculation similar to that in (4.9) may be made when considering Bayes procedures. The heuristic approximation does not yield nearly as neat an answer. Assume a zero-mean, normal prior on the unknown parameter θ. This in turn induces prior distributions on the parameter $\mu_{(1)}$ and $\mu_{(2)}$, the mean drifts of the Wiener processes approximating the sequential statistics being compared. In the relevant range of the parameters, these priors should be nearly normal with zero means and variances (say) $\sigma^2_{0,(1)}$ and $\sigma^2_{0,(2)}$, respectively. After applying the results obtained in (4.8) for the related Wiener process problems (but neglecting the logarithmic factor entirely), the limiting ratio of Bayes risks of two competing statistics can be computed to be

$$\begin{aligned}
\lim_{c\to 0} \frac{r(a_B{}^{(1)}, \sigma_{0,(1)})}{r(a_B{}^{(2)}, \sigma_{0,(2)})} &\simeq [c^2\tilde{l}_{(1)}\sigma^4_{(1)}(0)/c^2\tilde{l}_{(2)}\sigma^4_{(2)}(0)]^{\frac{1}{3}}\sigma_{0,(2)}/\sigma_{0,(1)} \\
&= \left[\frac{l\sigma_{(1)}(0)}{\mu'_{(1)}(0)}\Big/\frac{l\sigma_{(2)}(0)}{\mu'_{(2)}(0)}\right]^{\frac{1}{3}} [\sigma_{(1)}(0)\sigma_{0,(2)}/\sigma_{(2)}(0)\sigma_{0,(1)}] \qquad (4.10)\\
&= [E_2/E_1]^{\frac{1}{6}}[\sigma_{(1)}(0)\sigma_{0,(2)}/\sigma_{(2)}(0)\sigma_{0,(1)}] \,.
\end{aligned}$$

Acknowledgments. The author would like to thank Professor Herman Chernoff for his advice during the course of the research and preparation of this paper. He would also like to thank Dr. Louis Gordon and Sid Resnick for many helpful conversations and the referees and associate editor for their valuable comments.

Note added in proof. Since the completion of this paper, an excellent related article by T. L. Lai (*Ann. Statist.* **3** 825–845) has appeared. Professor Lai, among other things, refines the Chernoff–Savage analysis to obtain an invariance principle and a law of the iterated logarithm for linear rank statistics. The chief advantage of this approach is that the result is uniform in F and G. A minor drawback is that the score function J must have a second derivative satisfying certain regularity conditions.

On the other hand, the weak convergence approach does not in principle require J to have even a first derivative. In practice, this means that score functions for which the first derivative fails to exist at a finite number of points can be dealt with. Such functions often arise in testing for scale alternatives. Examples are the tests of Ansari–Freund and Siegel–Tukey.

Both Lai's paper and this one involve careful study of the tails of the empirical process. The results obtained are, however, not equivalent and will undoubtedly find different uses in future work. The applications to sequential analysis also have quite a different flavor.

REFERENCES

[1] BICKEL, P. J. (1967). Contributions to the theory ot order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 575–591. Univ. of California Press.

[2] BICKEL, P. and WICHURA, M. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] BIRNBAUM, Z. W. and MARSHALL, A. W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* **32** 687–703.

[5] CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[6] DEGROOT, M. H. (1960). Minimax sequential tests of some composite hypotheses. *Ann. Math. Statist.* **31** 1193–1200.

[7] FERNANDEZ, P. J. (1970). A weak convergence theorem for random sums of independent random variables. *Ann. Math. Statist.* **41** 710–712.

[8] HALL, W. J. (1969). A sequential Wilcoxon test. (Abstract). *Ann. Math. Statist.* **40** 1879.

[9] KIEFER, J. (1972). Skorokhod embedding of multivariate rv's and the sample df. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **24** 1–35.

[10] MILLER, R. G. and SEN, P. K. (1972). Weak Convergence of U-statistics and von Mises' differentiable statistical function. *Ann. Math. Statist.* **43** 31–41.

[11] O'REILLY, N. E. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability* **2** 642–651.

[12] PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–224.

[13] PRUITT, W. E. and OREY, S. (1973). Sample functions of the N-parameter Wiener process. *Ann. Probability* **1** 138–163.

[14] PYKE, R. and SHORACK, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

[15] PYKE, R. and SHORACK, G. R. (1969). A note on Chernoff–Savage theorems. *Ann. Math. Statist.* **40** 1116–1119.

[16] SEN, P. K. (1974). Weak convergence of generalized U-statistics. *Ann. Probability* **2** 90–102.

[17] SEN, P. K. and GHOSH, M. (1972). On strong convergence of regression rank statistics. *Sankhyā Ser. A* **34** 335–348.

[18] SEN, P. K. and GHOSH, M. (1973). A Chernoff-Savage representation of rank order statistics for stationary ϕ-mixing processes. *Sankhyā Ser. A* **35** 153–171.

[19] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

[20] STONE, C. (1963). Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proc. Amer. Math. Soc.* **14** 694–696.

[21] WELLNER, J. (1974). Convergence of the sequential uniform empirical process with bounds for centered beta rv's and a log-log law. Technical report no. 31, Univ. of Washington.

DEPARTMENT OF STATISTICS
FINE HALL
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08540