The Annals of Statistics
1978, Vol. 6, No. 5, 1027–1047

PITMAN EFFICIENCIES OF SEQUENTIAL TESTS AND UNIFORM LIMIT THEOREMS IN NONPARAMETRIC STATISTICS[1]

BY TZE LEUNG LAI

Columbia University and University of Illinois at Urbana-Champaign

In this paper Pitman's method of constructing and comparing tests based on statistics which are asymptotically normal under the null hypothesis and the local alternatives is extended to sequential tests of statistical hypotheses. The asymptotic normality assumption in Pitman's theory is replaced in its sequential analogue by the weak convergence of normalized processes formed from these statistics under the null hypothesis and the local alternatives. Uniform invariance principles are developed for a large class of statistics, and as an immediate corollary of these results, the desired weak convergence assumption is shown to hold. Furthermore uniform large deviation theorems are obtained for the test statistics and these results guarantee that the sequential tests under consideration have finite expected sample sizes under the null hypothesis and the local alternatives. As an illustration of the general method, the two-sample location problem is studied in detail, and the asymptotic relative efficiencies of the sequential Wilcoxon test, the sequential van der Waerden test and the sequential normal scores test relative to the two-sample sequential t-test are easily obtained since one of our key results (Theorem 1) implies that the asymptotic relative efficiencies of these sequential tests coincide with the corresponding Pitman efficiencies of their nonsequential analogues.

1. Introduction. Let T and T^* be two nonsequential tests for the hypothesis $H_0\colon \theta = \theta_0$ against the alternative $\theta > \theta_0$, where θ is a real unknown parameter of a probability distribution P_θ. The relative efficiency of T^* with respect to T is the ratio n/n^*, where n and n^* are the number of observations necessary to give T and T^* the same Type II error probability β at a fixed alternative θ for a given significance level α. The concept of asymptotic relative efficiency is due to Pitman [20]. He considers the limit of n/n^* for a sequence of alternatives θ_N converging to θ_0 in such a way that the Type II error of each of the two tests converges to a limit $\beta < 1$. Pitman has shown that under certain conditions this limit exists and does not depend on α and β. His method is explained and extended by Noether [19] and has been widely used in the study of nonparametric nonsequential tests.

Received June 1976; revised October 1977.

[1] Research supported by National Science Foundation under Grant No. GP-31505-X.

AMS 1970 *subject classifications.* Primary 62L10, 62G10, 62G20, 60F05.

Key words and phrases. Pitman efficiencies, sequential analysis, weak convergence, uniform invariance principles, uniform large deviation probabilities, linear rank statistics, sample quantiles, U-statistics, sequential Wilcoxon test, sequential van der Waerden test, two-sample sequential t-test.

In Section 2, we shall extend Pitman's method to the study of sequential tests. As the sample sizes are now random, we shall replace n and n^* in the definition of relative efficiency by the corresponding expected sample sizes under $H_0: \theta = \theta_0$ or under $H_N: \theta = \theta_N$. Under certain conditions, we shall show that the limiting ratio of the expected sample sizes of two given sequential tests under H_0 is the same as that under H_N, and we shall define the Pitman efficiency of the two sequential tests by this limiting ratio.

A well-known tool in the study of Pitman efficiencies for nonsequential tests is the uniformity in the convergence to normality for the normalized test statistics over a family of distributions P_θ, $\theta \in \Theta$, where Θ contains θ_0 and the sequence of alternatives θ_N. The counterpart of such tools in the sequential case is uniformity in the weak convergence to Brownian motion of the normalized processes formed from the test statistics. In Section 3, certain uniform invariance principles are developed for a large class of test statistics, and as an immediate corollary of these results, we obtain the desired uniformity in the weak convergence of the corresponding stochastic processes.

The uniform invariance principles developed in Section 3 are based on certain representation theorems of the test statistics. As will be shown in Section 4, these representation theorems also yield uniform large deviation probabilities for the test statistics. Such large deviation probabilities guarantee that the sequential tests under consideration have finite expected sample sizes under H_0 and H_N. Special cases of these general results and examples of Pitman efficiencies of sequential tests will be given in Section 5.

The following notations will be used throughout the sequel. Suppose for each $r \in I$ (where $I = \{1, 2, \cdots\}$ or $[1, \infty)$ is an index set), Y_r is a random variable defined on a probability space $(\Omega_r, \mathscr{F}_r, P_r)$ and Y_r converges in distribution to a random variable Y as $r \to \infty$. Then we write $Y_r \Rightarrow_{P_r} Y$ to emphasize that the distribution of Y_r is given by the measure P_r. When there is no confusion as to the measure defining the distribution of Y_r, we shall simply write $Y_r \Rightarrow Y$. If $Y_r(\cdot) = \{Y_r(t), t \geqq 0\}$ are stochastic processes instead of random variables, then we also write $Y_r(\cdot) \Rightarrow Y(\cdot)$ (or $Y_r(\cdot) \Rightarrow_{P_r} Y(\cdot)$) if

(1.1) $\forall h > 0$, the processes $\{Y_r(t), 0 \leqq t \leqq h\}$ and $\{Y(t), 0 \leqq t \leqq h\}$ belong to $D[0, h]$ and $\{Y_r(t), 0 \leqq t \leqq h\}$ converges weakly to $\{Y(t), 0 \leqq t \leqq h\}$.

We shall also let Φ denote the distribution function of the (standard normal) $\mathscr{N}(0, 1)$ distribution and $W(\cdot) = \{W(t), t \geqq 0\}$ denote the standard Wiener process.

2. Sequential analogue of Pitman's method. Let $X_1, X_2, \cdots$ be i.i.d. with distribution P_θ depending on a real unknown parameter θ. We want to test $H_0: \theta = \theta_0$ versus $H_r: \theta = \theta_r$, where $\theta_r = \theta_0 + r^{-\frac{1}{2}}$, $r \geqq r_* (> 0)$, is a family of alternatives. For each sample size n, let $T_n = T_n(X_1, \cdots, X_n)$ be a statistic based on the data so far observed. In analogy with the classical assumptions in the

Pitman theory for the nonsequential case (cf. [4], page 980), suppose that there exists a positive constant d and real-valued functions $\sigma_r(\theta) > 0$ and $\psi_r(\theta)$ such that the following five conditions are satisfied:

(2.1) $$\lim_{r\to\infty} (\psi_r(\theta_r) - \psi_r(\theta_0))/\sigma_r(\theta_0) = d\,;$$

(2.2) $$\lim_{r\to\infty} \sigma_r(\theta_r)/\sigma_r(\theta_0) = 1\,;$$

(2.3) $$\lim_{r\to\infty} \psi_r(\theta_0)/(r\sigma_r(\theta_0)) = 0\,;$$

(2.4) defining $W_{r,\theta}(t) = (T_{[rt]} - t\psi_r(\theta))/\sigma_r(\theta)$, $t \geqq 0$ $(T_0 = 0)$, we have (with the notations of weak convergence as in (1.1)) as $r \to \infty$,

$$W_{r,\theta_0}(\cdot) \Rightarrow_{P_{\theta_0}} W(\cdot) \qquad \text{and} \qquad W_{r,\theta_r}(\cdot) \Rightarrow_{P_{\theta_r}} W(\cdot)\,;$$

(2.5) for every $\varepsilon > 0$, there exist r_ε and a positive function $g_\varepsilon(t)$ such that $\int_1^\infty g_\varepsilon(t)\,dt < \infty$ and for all $r \geqq r_\varepsilon$ and $t \geqq 1$,

$$P_{\theta_0}[W_{r,\theta_0}(t) \geqq \varepsilon t] \leqq g_\varepsilon(t)\,,$$
$$P_{\theta_r}[W_{r,\theta_r}(t) \leqq -\varepsilon t] \leqq g_\varepsilon(t)\,.$$

In the classical Pitman theory, a sequence $\theta_N = \theta_0 + kN^{-\frac{1}{2}}$ of alternatives is considered with the index N being the sample size of the nonsequential test under consideration. For sequential tests, since the sample size is now a random variable, a natural modification of the Pitman formulation is to first parametrize the set of close alternatives as $\theta_r = \theta_0 + r^{-\frac{1}{2}}$ and then see how the sample size distribution varies with r as $r \to \infty$. Thus we have used above a continuous parameter r instead of a discrete index N. Obviously the weak convergence assumption (2.4) implies that

(2.6) $$(T_n - \psi_n(\theta_0))/\sigma_n(\theta_0) \Rightarrow_{P_{\theta_0}} \mathscr{N}(0, 1) \quad \text{and}$$
$$(T_n - \psi_n(\theta_n))/\sigma_n(\theta_n) \Rightarrow_{P_{\theta_n}} \mathscr{N}(0, 1) \quad \text{as} \quad n \to \infty\,.$$

The conditions (2.1), (2.2), and (2.6) are the classical assumptions in the Pitman theory for the nonsequential case. For the corresponding sequential theory, besides strengthening (2.6) into the weak convergence criterion (2.4), we have also added conditions (2.3) and (2.5) to control the tail behavior of the sample size distributions for the sequential tests described below. As in the case of the classical Pitman theory, many commonly used test statistics, when *suitably shifted and scaled*, satisfy conditions (2.1)—(2.5). Some examples will be given in Section 5.

Let $W_\delta(\cdot)$ denote the Wiener process with drift coefficient δ, i.e., $W_\delta(t) = W(t) + \delta t$, $t \geqq 0$. Let $a_r = \frac{1}{2}(\psi_r(\theta_r) + \psi_r(\theta_0))$ and define $\tilde{T}_r(t) = (T_{[rt]} - ta_r)/\sigma_r(\theta_0)$, $t \geqq 0$. The conditions (2.1), (2.2), and (2.4) imply that

(2.7) $$\tilde{T}_r(\cdot) \Rightarrow_{P_{\theta_0}} W_{-d/2}(\cdot) \qquad \text{and} \qquad \tilde{T}_r(\cdot) \Rightarrow_{P_{\theta_r}} W_{d/2}(\cdot) \quad \text{as} \quad r \to \infty\,.$$

As is well known, Wald's sequential probability ratio test (SPRT) in terms of

$W_\delta(\cdot)$ for the null hypothesis $\delta = -d/2$ against the alternative $\delta = d/2$ with prescribed error rates (α, β), $0 < \alpha, \beta < \frac{1}{2}$ (i.e., α and β are the Type I and Type II errors respectively), stops sampling as soon as $W_\delta(t) \geqq d^{-1} \log((1-\beta)/\alpha)$ or $W_\delta(t) \leqq d^{-1} \log(\beta/(1-\alpha))$, and this is the optimum test of the simple null versus the simple alternative. In view of the weak convergence criterion (2.7), if we want to test H_0 versus H_r sequentially using the sequence of statistics T_n, we would use the following sequential test $\mathscr{T}_r$ which stops sampling at stage

$$\tau_r = \inf\{n : T_n \geqq d^{-1}\sigma_r(\theta_0) \log((1-\beta)/\alpha) + na_r/r \text{ or } T_n \leqq d^{-1}\sigma_r(\theta_0) \log(\beta/(1-\alpha)) + na_r/r\}, \tag{2.8}$$

and $\mathscr{T}_r$ accepts H_0 if $T_{\tau_r} \leqq d^{-1}\sigma_r(\theta_0) \log(\beta/(1-\alpha)) + \tau_r a_r/r$ and accepts H_r if otherwise. We shall call the test $\mathscr{T}_r$ an *asymptotic Wald test* of H_0 versus H_r based on the sequence T_n for the given pair (α, β).

The condition (2.5) guarantees that the test $\mathscr{T}_r$ has finite expected sample sizes under H_0 and H_r for all large r. To see this, we note that for $0 < \varepsilon < \frac{1}{2}d$, if $t \geqq t_0$ and $r \geqq r_0$, then

$$\begin{aligned} P_{\theta_0}[\tau_r > rt] &\leqq P_{\theta_0}[T_{[rt]} > d^{-1}\sigma_r(\theta_0) \log(\beta/(1-\alpha)) + a_r[rt]/r] \\ &\leqq P_{\theta_0}[T_{[rt]} - t\psi_r(\theta_0) \geqq \tfrac{1}{2}t(\psi_r(\theta_r) - \psi_r(\theta_0)) - a_r/r \\ &\qquad + d^{-1}\sigma_r(\theta_0) \log(\beta/(1-\alpha))] \\ &\leqq P_{\theta_0}[T_{[rt]} - t\psi_r(\theta_0) \geqq \varepsilon t\sigma_r(\theta_0)] \leqq g_\varepsilon(t), \end{aligned} \tag{2.9}$$

using (2.1), (2.5) and the fact that $a_r/r = o(\sigma_r(\theta_0))$ by (2.1) and (2.3). Likewise for $0 < \varepsilon < \frac{1}{2}d$ and $t \geqq t_0$, $r \geqq r_0$,

$$P_{\theta_r}[\tau_r > rt] \leqq P_{\theta_r}[T_{[rt]} - t\psi_r(\theta_r) \leqq -\varepsilon t\sigma_r(\theta_r)] \leqq g_\varepsilon(t). \tag{2.10}$$

Noting that $a_r/r = o(\sigma_r(\theta_0))$ and that the event $[r^{-1}\tau_r \leqq t]$ depends on T_n for n only up to rt, it is not hard to see from (2.7) that

$$r^{-1}\tau_r \Rightarrow_{P_{\theta_0}} \tau(-d/2) \quad \text{and} \quad r^{-1}\tau_r \Rightarrow_{P_{\theta_r}} \tau(d/2), \tag{2.11}$$

where for any real number δ, we define

$$\tau(\delta) = \inf\{t \geqq 0 : W_\delta(t) \notin [d^{-1} \log(\beta/(1-\alpha)), d^{-1} \log((1-\beta)/\alpha)]\}. \tag{2.12}$$

Furthermore the error probabilities of the test $\mathscr{T}_r$ are given by

$$\lim_{r\to\infty} P_{\theta_0}[\mathscr{T}_r \text{ rejects } H_0] = \alpha, \qquad \lim_{r\to\infty} P_{\theta_r}[\mathscr{T}_r \text{ rejects } H_r] = \beta. \tag{2.13}$$

By (2.9), $r^{-1}\tau_r$ is uniformly integrable under P_{θ_0}. Hence by (2.11),

$$\begin{aligned} E_{\theta_0}\tau_r &\sim rE\tau(-d/2) \\ &= 2rd^{-2}\{(1-\alpha) \log((1-\alpha)/\beta) - \alpha \log((1-\beta)/\alpha)\}. \end{aligned} \tag{2.14}$$

Likewise by (2.10) and (2.11),

$$\begin{aligned} E_{\theta_r}\tau_r &\sim rE\tau(d/2) \\ &= 2rd^{-2}\{(1-\beta) \log((1-\beta)/\alpha) - \beta \log((1-\alpha)/\beta)\}. \end{aligned} \tag{2.15}$$

The results (2.13), (2.14), and (2.15) therefore yield the following extension of Pitman's asymptotic relative efficiency to the sequential case.

THEOREM 1. *Let* $X_1, X_2, \cdots$ *be* i.i.d. *with distribution* P_θ *depending on an unknown real parameter* θ. *Let* $\theta_r = \theta_0 + r^{-\frac{1}{2}}$, $r \geqq r_0\,(> 0)$. *Suppose* $T_n = T_n(X_1, \cdots, X_n)$ *is a sequence of statistics such that there exists a positive constant d and real-valued functions* $\sigma_r(\theta) > 0$ *and* $\psi_r(\theta)$ *satisfying the conditions* (2.1)—(2.5). *Let* $T_n^* = T_n^*(X_1, \cdots, X_n)$ *be another sequence of statistics satisfying* (2.1)—(2.5) *with* d^*, $\sigma_r^*(\theta)$ *and* $\psi_r^*(\theta)$ *in place of* d, $\sigma_r(\theta)$ *and* $\psi_r(\theta)$. *Given* $0 < \alpha$, $\beta < \frac{1}{2}$, *let* $\mathscr{T}_r$ *(respectively* $\mathscr{T}_r^*$*) be the asymptotic Wald test based on the sequence* T_n *(respectively* T_n^**) of* $H_0\colon \theta = \theta_0$ *versus* $H_r\colon \theta = \theta_r$ *corresponding to the pair* (α, β) *of error rates,* i.e., *the stopping rule of* $\mathscr{T}_r$ *is* τ_r *defined by* (2.8) *and the stopping rule* τ_r^* *of* $\mathscr{T}_r^*$ *is similarly defined. Then*

$$\text{(2.16)}\qquad \begin{aligned} &\lim_{r\to\infty} P_{\theta_0}[\mathscr{T}_r \text{ rejects } H_0] = \lim_{r\to\infty} P_{\theta_0}[\mathscr{T}_r^* \text{ rejects } H_0] = \alpha\,,\\ &\lim_{r\to\infty} P_{\theta_r}[\mathscr{T}_r \text{ rejects } H_r] = \lim_{r\to\infty} P_{\theta_r}[\mathscr{T}_r^* \text{ rejects } H_r] = \beta\,. \end{aligned}$$

Moreover, for all $\alpha, \beta \in (0, \frac{1}{2})$,

$$\text{(2.17)}\qquad \begin{aligned} &\lim_{r\to\infty} (E_{\theta_0}\tau_r/E_{\theta_0}\tau_r^*)\\ &\quad = \lim_{r\to\infty} (E_{\theta_r}\tau_r/E_{\theta_r}\tau_r^*) = (d^*/d)^2\\ &\quad = \lim_{r\to\infty} \{(\psi_r^*(\theta_r) - \psi_r^*(\theta_0))/\sigma_r^*(\theta_0)\}^2/\{(\psi_r(\theta_r) - \psi_r(\theta_0))/\sigma_r(\theta_0)\}^2\,. \end{aligned}$$

In view of Theorem 1 and in analogy with Pitman's idea of comparing nonsequential tests, we define the asymptotic relative efficiency of two asymptotic Wald tests $\mathscr{T}_r$ and $\mathscr{T}_r^*$ as follows.

DEFINITION 1. With the same notations and assumptions as in Theorem 1, the asymptotic relative efficiency (A.R.E.), or Pitman efficiency, of $\mathscr{T}_r^*$ relative to $\mathscr{T}_r$ is defined as

$$\text{(2.18)}\qquad \text{A.R.E.} = \lim_{r\to\infty} \frac{\{(\psi_r^*(\theta_r) - \psi_r^*(\theta_0))/\sigma_r^*(\theta_0)\}^2}{\{(\psi_r(\theta_r) - \psi_r(\theta_0))/\sigma_r(\theta_0)\}^2}\,.$$

We note that the right-hand side of (2.18) is exactly the same as the usual expression defining the Pitman efficiency of the corresponding nonsequential tests $\tilde{\mathscr{T}}_m^*$ and $\tilde{\mathscr{T}}_n$ based on the statistics T_m^* and T_n, i.e., $\tilde{\mathscr{T}}_n$ (respectively $\tilde{\mathscr{T}}_m^*$) rejects H_0 if T_n (respectively T_m^*) is large. (See [4], page 980.) In view of (2.2) and (2.4), $(\psi_r(\theta_r) - \psi_r(\theta_0))/\sigma_r(\theta_0)$ can be interpreted as the difference in the asymptotic drift of the process $\{T_{[rt]}, t \geqq 0\}$ (properly scaled by $\sigma_r(\theta_0)$) between the null hypothesis H_0 and the alternative H_r. Therefore the A.R.E. given by (2.18) is simply the limiting squared ratio of the scaled change in the asymptotic drift of $\{T_{[rt]}^*, t \geqq 0\}$ to that of $\{T_{[rt]}, t \geqq 0\}$. Since the scaled asymptotic drifts are the intrinsic quantities determining the asymptotic behavior of the expected sample sizes for the sequential (and nonsequential) tests based on the test statistics under consideration, it is intuitively clear from this point of view why Pitman efficiencies, both sequential and nonsequential, are related to these asymptotic drifts.

While Theorem 1 enables us to compare one asymptotic Wald test with another, it is often desirable to compare an asymptotic Wald test with the SPRT of H_0 versus H_r. Let

$$Z_i^{(r)} = \log\{p_{\theta_r}(X_i)/p_{\theta_0}(X_i)\}\,, \qquad i = 1, 2, \cdots, \tag{2.19}$$

where p_θ is the density of X_1 under P_θ (with respect to some common dominating measure). Suppose the SPRT of H_0 versus H_r has error rates (α, β) with $0 < \alpha$, $\beta < \frac{1}{2}$. Wald's approximation of its expected sample size under H_0 is

$$m_r = \{(1-\alpha)\log((1-\alpha)/\beta) - \alpha\log((1-\beta)/\alpha)\}/|E_{\theta_0}Z_1^{(r)}|\,; \tag{2.20}$$

and Wald's approximation of its expected sample size under H_r is

$$m_r' = \{(1-\beta)\log((1-\beta)/\alpha) - \beta\log((1-\alpha)/\beta)\}/E_{\theta_r}Z_1^{(r)}\,. \tag{2.21}$$

In [24] (pages 156–157), Wald has proved that for any sequential test with error rates (α, β) such that the test terminates with probability 1 under H_0 and H_r, m_r and m_r' as defined by (2.20) and (2.21) are lower bounds of its expected sample sizes under H_0 and H_r respectively. Moreover, the expected sample sizes of the SPRT under H_0 and H_r are asymptotically equal to these lower bounds under weak regularity conditions. If we divide the expression (2.20) (respectively (2.21)) by the asymptotic expected sample size (2.14) (respectively (2.15)) of the asymptotic Wald test $\mathscr{T}_r$, the quantities α and β disappear in the resulting quotient. This leads us to define the asymptotic efficiency of $\mathscr{T}_r$ as follows.

Definition 2. With the same notations and assumptions as in Theorem 1, the asymptotic efficiency of $\mathscr{T}_r$ is defined as

$$\begin{aligned}&\lim_{r\to\infty}\min\{m_r/E_{\theta_0}\tau_r,\ m_r'/E_{\theta_r}\tau_r\}\\ &\quad= \lim_{r\to\infty}\{(\psi_r(\theta_r)-\psi_r(\theta_0))/\sigma_r(\theta_0)\}^2/\{2r\max(|E_{\theta_0}Z_1^{(r)}|,\ E_{\theta_r}Z_1^{(r)})\}\,.\end{aligned} \tag{2.22}$$

As Example 4 in Section 5 would show, we often have $E_{\theta_r}Z_1^{(r)} \sim |E_{\theta_0}Z_1^{(r)}|$ as $r \to \infty$. If $p_\theta(x) = e^{\theta x - h(\theta)}$ is an exponential family of densities, then letting $Z(\theta) = \log\{p_\theta(X)/p_{\theta_0}(X)\}$, it is easy to see that

$$E_\theta Z(\theta) \sim |E_{\theta_0}Z(\theta)| \sim \tfrac{1}{2}(\theta-\theta_0)^2h''(\theta_0) \qquad \text{as} \quad \theta \to \theta_0\,. \tag{2.23}$$

Since $\theta_r = \theta_0 + r^{-\frac{1}{2}}$, (2.23) implies that

$$E_{\theta_r}Z_1^{(r)} \sim |E_{\theta_0}Z_1^{(r)}| \sim \tfrac{1}{2}r^{-1}h''(\theta_0) \qquad \text{as} \quad r \to \infty\,. \tag{2.24}$$

In Section 5, for the two-sample location problem, we shall show that the two-sample sequential t-test is asymptotically efficient (i.e., its asymptotic efficiency is 1) for the normal model. However, if the normal model is not true, then the sequential t-test can have very low A.R.E. relative to certain sequential rank tests which we shall study in Section 5. On the other hand, for the normal model, these sequential rank tests will be shown to have rather high A.R.E. relative to the sequential t-test.

3. Uniform invariance principles for test statistics. Just as the asymptotic normality condition (2.6) in the classical Pitman theory is satisfied by many commonly used statistics, the stronger weak convergence condition (2.4) also holds for a large class of statistics. There are several approaches to prove such weak convergence. One approach is to check directly tightness and the convergence of the finite-dimensional distributions to the multivariate normal. This can sometimes be done by noting certain probability structures exhibited by $W_{r,\theta}(\cdot)$ and applying known results in the theory of weak convergence. For example, in certain cases, for all large r, $W_{r,\theta}(\cdot)$ after suitable normalization forms a martingale or reversed martingale under P_θ and one can then apply the corresponding weak convergence theorems for martingales and reversed martingales (cf. [9], [18]). Such martingale structures have been utilized by Hall [9] to solve this kind of weak convergence problems for certain nonparametric test statistics.

An alternative approach, due to Hall and Loynes [10], is based on an extension of Le Cam's concept of contiguity. Let $L_n^{(r)} = \prod_{i=1}^n (p_{\theta_r}(X_i)/p_{\theta_0}(X_i))$ denote the likelihood ratio at stage n. To prove (2.4), it suffices to show the weak convergence (as $r \to \infty$) for every $k > 0$ of $(W_{r,\theta_0}(\cdot), \log L_{rk}^{(r)})$ in $D[0, k] \times R$ under $H_0: \theta = \theta_0$ to the Wiener-normal process $(W(\cdot), Z)$ such that the mean of the normal random variable Z is minus one-half of the variance (cf. [10]).

In this section, we shall present another approach to obtain weak convergence results of the type (2.4). To prove (2.4), it suffices to show that for some $\eta > 0$ and for every $\varepsilon > 0$ and $k > 0$, we have (by redefining the random variables on a new probability space if necessary) as $r \to \infty$,

$$(3.1) \qquad P_\theta[\max_{0 \leqq t \leqq k} |W_{r,\theta}(t) - W(t)| > \varepsilon] \to 0 \quad \text{uniformly for} \quad |\theta - \theta_0| \leqq \eta .$$

The uniformity condition (3.1) has itself an interesting implication on the behavior of the sample size distribution and of the power function for the asymptotic Wald test $\mathscr{T}_r$ of $H_0: \theta = \theta_0$ versus $H_r: \theta = \theta_0 + r^{-\frac{1}{2}}$ based on a sequence of statistics T_n satisfying (2.1)—(2.5). While (2.13) shows that the error rates of $\mathscr{T}_r$ converge to the corresponding quantities of Wald's SPRT $\mathscr{T}$ with stopping rule (2.12) for testing $H: \delta = -d/2$ versus $K: \delta = d/2$ for the Wiener process $W_\delta(\cdot)$, it is natural to expect that the power of $\mathscr{T}_r$ at the parameter $\theta = \theta_0 + \xi r^{-\frac{1}{2}}$ would also converge to the power of $\mathscr{T}$ at the point $\delta = (\xi - \frac{1}{2})d$, with the convergence being uniform for ξ in compact sets. We note that condition (3.1) implies not only the weak convergence criterion (2.4), but also the uniformity of such weak convergence, i.e., for all $\varepsilon, h, k > 0$,

$$(3.2) \qquad P_{\theta_0 + \xi r^{-\frac{1}{2}}}[\max_{0 \leqq t \leqq k} |W_{r,\theta_0 + \xi r^{-\frac{1}{2}}}(t) - W(t)| > \varepsilon] \to 0 \quad (\text{as } r \to \infty) \quad \text{uniformly for} \quad |\xi| \leqq h .$$

With the above uniformity of weak convergence, we can establish the desired uniform convergence for the power function of $\mathscr{T}_r$ (see Theorem 2 below). Some applications of Theorem 2 will be given in Section 5.

THEOREM 2. *Let $X_1, X_2, \cdots$ be* i.i.d. *with distribution P_θ depending on an unknown real parameter θ. Suppose $T_n = T_n(X_1, \cdots, X_n)$ is a sequence of statistics such that there exists a positive constant d and real-valued functions $\sigma_r(\theta)$ and $\psi_r(\theta)$ $(r \geqq r_0 > 0)$ satisfying the following assumptions*:

$$(3.3) \qquad \psi_r'(\theta_0) \text{ exists for all } r \geqq r_0 \text{ and } \lim_{r\to\infty} \psi_r'(\theta_0)/(r^{\frac{1}{2}}\sigma_r(\theta_0)) = d\,;$$

$$(3.4) \qquad \lim_{r\to\infty} \psi_r(\theta_0)/(r\sigma_r(\theta_0)) = 0\,;$$

$$(3.5) \qquad \sigma_r(\theta_0 + \xi r^{-\frac{1}{2}})/\sigma_r(\theta_0) \to 1 \quad \text{uniformly for } |\xi| \leqq a \quad (\text{for all } a > 0)\,;$$

$$(3.6) \qquad \{(\xi r^{-\frac{1}{2}})^{-1}(\psi_r(\theta_0 + \xi r^{-\frac{1}{2}}) - \psi_r(\theta_0)) - \psi_r'(\theta_0)\}/(r^{\frac{1}{2}}\sigma_r(\theta_0)) \to 0 \quad (\text{as } r \to \infty)$$
$$\text{uniformly for } 0 < |\xi| \leqq a \quad (\text{for all } a > 0)\,.$$

Moreover, letting $W_{r,\theta}(t) = (T_{[rt]} - t\psi_r(\theta))/\sigma_r(\theta)$, $t \geqq 0$ $(T_0 = 0)$, assume also that (3.2) *holds for all positive ε, k, and h. Let $a_r = \frac{1}{2}\{\psi_r(\theta_0) + \psi_r(\theta_0 + r^{-\frac{1}{2}})\}$. For $0 < \alpha, \beta < \frac{1}{2}$, define the stopping time τ_r by* (2.8). *Let $W_\delta(t) = W(t) + \delta t$, $t \geqq 0$, and define the stopping time $\tau(\delta)$ by* (2.12). *Then setting $d_\xi = (\xi - \frac{1}{2})d$, we have for every compact subset C of the real line, as $r \to \infty$,*

$$(3.7) \qquad P_{\theta_0+\xi r^{-\frac{1}{2}}}[\tau_r \leqq rx] \to P[\tau(d_\xi) \leqq x]$$
$$\text{uniformly for } 0 \leqq x < \infty \text{ and } \xi \in C\,,$$

$$(3.8) \qquad P_{\theta_0+\xi r^{-\frac{1}{2}}}[T_{\tau_r} \leqq d^{-1}\sigma_r(\theta_0)\log(\beta/(1-\alpha)) + \tau_r a_r/r]$$
$$\to P[W_{d_\xi}(\tau(d_\xi)) \leqq d^{-1}\log(\beta/(1-\alpha))] \quad \text{uniformly for } \xi \in C\,.$$

REMARKS. (i) The condition (3.5) is a uniform version of (2.2), while the condition (3.6) is a uniform version of

$$(3.9) \qquad \lim_{r\to\infty} \{r^{\frac{1}{2}}(\psi_r(\theta_0 + \xi r^{-\frac{1}{2}}) - \psi_r(\theta_0)) - \psi_r'(\theta_0)\}/(r^{\frac{1}{2}}\sigma_r(\theta_0)) = 0$$
$$\text{for all } \xi\,.$$

Clearly (2.2), (3.3), and (3.9) imply (2.1). In fact, Pitman and Noether (cf. [19]) originally assumed these three conditions instead of (2.1).

(ii) The uniformity over compact sets in (3.8) can be easily extended to uniformity for ξ in the whole real line if the power function of the test $\mathscr{T}_r$ with stopping rule τ_r is nondecreasing. A similar remark holds for the uniformity in (3.7).

PROOF OF THEOREM 2. From (3.3) and (3.6), it follows that as $r \to \infty$,

$$(3.10) \qquad \begin{aligned} &a_r - \psi_r(\theta_0 + \xi r^{-\frac{1}{2}}) \\ &\quad = \tfrac{1}{2}\{\psi_r(\theta_0 + r^{-\frac{1}{2}}) - \psi_r(\theta_0)\} - \{\psi_r(\theta_0 + \xi r^{-\frac{1}{2}}) - \psi_r(\theta_0)\} \\ &\quad = d\sigma_r(\theta_0)(\tfrac{1}{2} - \xi + o(1))\,, \end{aligned}$$

where the $o(1)$ term is uniform for $\xi \in C$. Using (3.2), (3.4), (3.5), and (3.10), it is easy to see that (3.7) holds. From these relations and (3.7), (3.8) follows easily. □

The method which we use to establish the uniform invariance principle (3.1) for a large class of test statistics is based on the representation of T_n in terms of sums of i.i.d. random variables plus a remainder term. We shall show that for these statistics the remainder term when suitably normalized converges to 0 in an appropriate sense uniformly over a family of distributions satisfying certain regularity conditions. This approach not only establishes (3.1) (and therefore (2.4) as well), but it also has the extra bonus that the representation arguments used can be modified to obtain the uniform large deviation theorems of Section 4 which provide useful tools for establishing (2.5) and studying the ASN function of our tests.

The following theorem gives a uniform invariance principle for sums of i.i.d. random variables.

THEOREM 3. *Let $\mathscr{F}$ be a family of distribution functions with mean 0 and unit variance such that the following uniform square integrability condition holds*:

$$\sup_{F\in\mathscr{F}} \int_{|x|\geqq a} x^2\, dF(x) \to 0 \qquad \text{as} \quad a \to \infty .$$

Let $X_1, X_2, \cdots$ be i.i.d. *random variables with a common distribution function $F \in \mathscr{F}$. For $r \geqq 1$ and $t \geqq 0$, define $\zeta_r(t) = r^{-\frac{1}{2}} \sum_{i=1}^{[rt]} X_i$. Then for every $\varepsilon > 0$ and $k > 0$, we have (by redefining the random variables on a new probability space if necessary) as $r \to \infty$,*

$$P_F[\max_{0\leqq t\leqq k} |W(t) - \zeta_r(t)| \geqq \varepsilon] \to 0 \qquad \text{uniformly for} \quad F \in \mathscr{F} .$$

REMARK. The proof is similar to that of Theorem 2 of [16] and makes use of truncation and the Skorohod embedding scheme. Obviously the uniform square integrability condition of Theorem 3 is satisfied if

$$\sup_{F\in\mathscr{F}} E_F|X_1|^{2+\delta} < \infty \qquad \text{for some} \quad \delta > 0 .$$

We now apply Theorem 3 and standard representation theorems to obtain uniform invariance principles of the type (3.1) for linear rank statistics, sample quantiles and U-statistics.

THEOREM 4. *Suppose $X_1, X_2, \cdots$ are* i.i.d. *random variables with a common continuous distribution function F and are independent of $Y_1, Y_2, \cdots$ which are* i.i.d. *with a common continuous distribution function G. Let $F_n(x) = n^{-1} \sum_1^n I_{[X_i\leqq x]}$ and $G_m(x) = m^{-1} \sum_1^m I_{[Y_i\leqq x]}$ be the empirical distribution functions. Let $J\colon [0, 1] \times [0, 1] \to R$ be twice continuously differentiable except possibly at the points $(0, 0)$ and $(1, 1)$ such that for some $\delta > 0$ and $C > 0$,*

$$\begin{aligned} &|\partial^2 J/\partial x^2| + |\partial^2 J/\partial y^2| + |\partial^2 J/\partial x\, \partial y| \\ (3.11) \qquad &\leqq C(\{\max(x, y)\}^{-\frac{5}{2}+\delta} + \{\max(1-x, 1-y)\}^{-\frac{5}{2}+\delta}) , \\ &\qquad\qquad\qquad\qquad\qquad\qquad 0 < x, y < 1 . \end{aligned}$$

Let (m_n) be a nondecreasing sequence of positive integers such that $\lim_{n\to\infty} n/m_n = \lambda > 0$, and let $J_n\colon \{0, 1/n, \cdots, 1\} \times \{0, 1/m_n, \cdots, 1\} \to R$ be a sequence of functions satisfying

$$(3.12) \qquad n^{-1} \sum_{i=1}^n \sup_{y\in\{0,1/m_n,\cdots,1\}} |J_n(i/n, y) - J(i/n, y)| = o(n^{-\frac{1}{2}}) .$$

Define the generalized Chernoff–Savage statistic

$$\Gamma_n = \int_{-\infty}^{\infty} J_n(F_n(x), G_{m_n}(x))\, dF_n(x) \tag{3.13}$$

and represent it as

$$\begin{aligned}\Gamma_n = \int_{-\infty}^{\infty} J(F(x), G(x))\, dF(x) + n^{-1} \textstyle\sum_1^n (h_{F,G}(X_i) - E_F h_{F,G}(X_i)) \\ + m_n^{-1} \textstyle\sum_1^{m_n} (h^*_{F,G}(Y_i) - E_G h^*_{F,G}(Y_i)) + R_n\,,\end{aligned} \tag{3.14}$$

where $h_{F,G}(u) = J(F(u), G(u)) - \int_{u_0}^u (\partial J/\partial x)(F(t), G(t))\, dF(t)$ *and* $h^*_{F,G}(u) = -\int_{u_0}^u (\partial J/\partial y)(F(t), G(t))\, dF(t)$. *Then for every* $\varepsilon > 0$, *as* $m \to \infty$,

$$P_{F,G}[\sup_{j \geq m} j^{\frac{1}{2}}|R_j| > \varepsilon] \to 0 \qquad \text{uniformly for} \quad F, G \in \mathscr{C}\,, \tag{3.15}$$

where $\mathscr{C}$ *denotes the class of all continuous distribution functions. Let* $V_n = n(\Gamma_n - \int_{-\infty}^{\infty} J(F(x), G(x))\, dF(x))/\sigma(F, G)$, *where* $\sigma^2(F, G) = \mathrm{Var}_F\, h_{F,G}(X_1) + \lambda\, \mathrm{Var}_G\, h^*_{F,G}(Y_1)$. *Let* $\mathscr{C}(a) = \{(F, G) : F, G \in \mathscr{C}$ *and* $\sigma^2(F, G) \geq a\}$. *Then for any* $\varepsilon > 0$, $a > 0$ *and* $k > 0$, *by redefining the random variables on a new probability space if necessary, we have as* $r \to \infty$,

$$P_{F,G}[\max_{0 \leq t \leq k} |r^{-\frac{1}{2}} V_{[rt]} - W(t)| \geq \varepsilon] \to 0 \qquad \text{uniformly for} \quad (F, G) \in \mathscr{C}(a)\,. \tag{3.16}$$

REMARK. As shown in [14] (pages 834–842), statistics of the type (3.13) in fact cover a large class of two-sample linear rank statistics. Chernoff and Savage (cf. page 986 and Corollary 1 of [4]) have shown that $P_{F,G}[n^{\frac{1}{2}}|R_n| \geq \varepsilon] \to 0$ uniformly for $F, G \in \mathscr{C}$. Our conclusion (3.15) strengthens this uniform convergence in probability into uniform almost sure convergence. While Chernoff and Savage [4] use the uniform convergence in probability to obtain uniform convergence to normality (which then implies a result of the type (2.6)) for certain linear rank statistics, our uniform almost sure convergence result (3.15) leads to the uniform invariance principle (3.16) and therefore gives a result of the type (2.4) for such linear rank statistics. Braun [2] has raised the problem of obtaining a weak convergence analogue of the Chernoff–Savage theorem on uniform convergence to normality for certain linear rank statistics. He points out (cf. page 54 of [2]) that the methods which he developed in [2] and [3] to prove weak convergence of linear rank statistics do not seem to be able to produce the stronger uniformity result. The method we use below to prove the uniform invariance principle (3.16) is to apply Theorem 3 and to show (3.15) involving the remainder R_n. As shown by Chernoff and Savage ([4], page 977), if $\delta' > 0$ satisfies $(2 + \delta')(\delta - \frac{1}{2}) > -1$, then

$$\sup_{F,G \in \mathscr{C}} \{E_F|h_{F,G}(X_1)|^{2+\delta'} + E_G|h^*_{F,G}(Y_1)|^{2+\delta'}\} < \infty\,. \tag{3.17}$$

Hence Theorem 3 is applicable to the sample sums $\sum_1^n h_{F,G}(X_i)$ and $\sum_1^m h^*_{F,G}(Y_j)$.

PROOF OF THEOREM 4. To prove the uniform almost sure convergence result (3.15), we let $L = \sup\{n \geq 1 : n^{\frac{1}{2}}|R_n| > \varepsilon\}$ ($\sup \emptyset = 0$). We shall show that

$$\sup_{F,G \in \mathscr{C}} E_{F,G} L^\gamma < \infty \qquad \text{for all} \quad \gamma < \delta\,, \tag{3.18}$$

where δ is as defined by (3.11). Since

$$P_{F,G}[\sup_{j\geqq m} j^{\frac{1}{2}}|R_j| > \varepsilon] = P_{F,G}[L \geqq m] \leqq m^{-\gamma} E_{F,G} L^{\gamma} ,$$

it is clear that (3.18) implies (3.15).

To prove (3.18), we note that the estimates in [14] where we show the finiteness of $E_{F,G}L^{\gamma}$ are all uniform for $F, G \in \mathscr{C}$ (see the proof of Theorem 2(ii) in [14] (pages 836–840), and set there $\mu = \frac{1}{2}$). In this connection, it should be pointed out that Lemma 3 of [14] which links the probability estimates with $E_{F,G}L^{\gamma}$ can be strengthened as follows: Let $Z_1, Z_2, \cdots$ be any sequence of random variables and set $\tau(\zeta, \varepsilon) = \sup\{n \geqq 1 : |Z_n| \geqq \varepsilon n^{\zeta}\}$ where $\varepsilon > 0$ and ζ is real. Then for $p > 0$ and $\alpha > 0$, there exist universal constants A_p depending only on p and $A_{p,\alpha}$ depending only on p and α such that

$$E\tau^p(\zeta, \varepsilon) \leqq A_p \textstyle\sum_1^\infty n^p P[|Z_n| \geqq \varepsilon n^{\zeta}] ;$$
$$E\tau^p(\alpha, \varepsilon) \leqq A_{p,\alpha} \textstyle\sum_1^\infty n^{p-1} P[\max_{j\leqq n} |Z_j| \geqq \frac{1}{4}\varepsilon n^{\alpha}] ;$$
$$E\tau^p(\zeta, \varepsilon) \leqq A_{p,\alpha} \textstyle\sum_1^\infty n^{p-1} P[\max_{j\leqq n} j^{\alpha-\zeta}|Z_j| \geqq \frac{1}{4}\varepsilon n^{\alpha}]$$

(cf. Lemma 2 of [5]). Hence (3.18) holds.

It is easy to see that for every fixed l, $\max_{1\leqq j\leqq l} |\Gamma_j| \leqq \max\{|J_n(x, y)| : n = 1, \cdots, l; x = 1/n, \cdots, 1; y = 0, 1/m_n, \cdots, 1\} < \infty$. Hence (3.15) implies that for every $\varepsilon > 0$ and $k > 0$, as $r \to \infty$,

$$P_{F,G}[\max_{0\leqq t\leqq k} r^{\frac{1}{2}}|R_{[rt]}| \geqq \varepsilon] \to 0 \qquad \text{uniformly for} \quad F, G \in \mathscr{C} . \tag{3.19}$$

We note that $\inf_{(F,G)\in\mathscr{C}(a)} \sigma(F, G) > 0$. Moreover by (3.17), $\sigma(F, G)$ is uniformly bounded for $(F, G) \in \mathscr{C}(a)$. Hence the desired conclusion (3.16) follows easily from (3.14), (3.19), and Theorem 2. □

THEOREM 5. *Let $0 < p < 1$ and let $\mathscr{F}$ be a family of distribution functions F on the real line such that the equation $F(\xi) = p$ has a unique solution $\xi = \xi_F$ and there exists a positive constant b for which*

$$\sup_{F\in\mathscr{F}} \sup_{|x|\leqq b} |F''(\xi_F + x)| < \infty , \tag{3.20a}$$

$$\inf_{F\in\mathscr{F}} F'(\xi_F) > 0 , \qquad \sup_{F\in\mathscr{F}} F'(\xi_F) < \infty . \tag{3.20b}$$

Let (k_n) be a sequence of positive integers such that $1 \leqq k_n \leqq n$ and

$$k_n = np + O(n^{\frac{1}{4}}) . \tag{3.21}$$

Let $X_1, X_2, \cdots$ be i.i.d. *with a common distribution function $F \in \mathscr{F}$ and let Z_n be the k_nth order statistic among $X_1, \cdots, X_n$.*

(i) *Letting $F_n(x) = n^{-1} \sum_1^n I_{[X_i\leqq x]}$, we represent Z_n as*

$$Z_n = \xi_F + \{(p - F_n(\xi_F))/F'(\xi_F)\} + R_n . \tag{3.22}$$

Then given any $0 < \delta < \frac{3}{4}$, there exist $c_\delta > 0$ and n_δ such that for all $n \geqq n_\delta$,

$$P_F[|R_n| \geqq n^{-\frac{3}{4}+\delta}] \leqq \exp(-c_\delta n^{4\delta/3}) \qquad \text{for all} \quad F \in \mathscr{F} . \tag{3.23}$$

(ii) *Assume further that the family $\mathscr{F}$ also satisfies the following condition*:

$$F(\xi_F - x) \to 0 \quad \textit{and} \quad F(\xi_F + x) \to 1 \quad \textit{uniformly for} \quad F \in \mathscr{F} \quad \textit{as} \quad x \to \infty\,. \tag{3.24}$$

Let $\sigma_F^2 = p(1-p)/(F'(\xi_F))^2$ and let $V_n = n(Z_n - \xi_F)/\sigma_F$. Then for any $\varepsilon > 0$ and $k > 0$, we have (by redefining the random variables on a common probability space if necessary) as $r \to \infty$,

$$P_F[\max_{0 \le t \le k} |r^{-\frac{1}{2}} V_{[rt]} - W(t)| \geqq \varepsilon] \to 0 \quad \textit{uniformly for} \quad F \in \mathscr{F}\,. \tag{3.25}$$

PROOF. Let $I_n(F) = [\xi_F - n^{-\frac{1}{2}+\frac{2}{3}\delta}, \xi_F + n^{-\frac{1}{2}+\frac{2}{3}\delta}]$ and set $H_n = \sup_{x \in I_n(F)} |(F_n(x) - F_n(\xi_F)) - (F(x) - F(\xi_F))|$. From (3.20a) and (3.20b), it follows that $\sup_{F \in \mathscr{F}} \sup_{|\xi_F - x| \le b} F'(x) < \infty$. Making use of this fact and an argument similar to that of Bahadur ([1], pages 578–579), it can be shown that there exist n_δ and c_δ such that for all $n \geqq n_\delta$ and $F \in \mathscr{F}$,

$$P_F[H_n \geqq \tfrac{1}{2} n^{-\frac{3}{4}+\delta}] \leqq \tfrac{1}{2} \exp(-c_\delta n^{4\delta/3})\,, \tag{3.26}$$

$$P_F[Z_n \notin I_n(F)] \leqq \tfrac{1}{2} \exp(-c_\delta n^{4\delta/3})\,. \tag{3.27}$$

Let $\varepsilon_n = F(Z_n) - \{F(\xi_F) + (Z_n - \xi_F)F'(\xi_F)\}$. If $|Z_n - \xi_F| \leqq \min\{b, n^{-\frac{1}{2}+\frac{2}{3}\delta}\}$, then $|\varepsilon_n| \leqq (\sup_{F \in \mathscr{F}} \sup_{|x| \le b} |F''(\xi_F + x)|) n^{-1+\frac{4}{3}\delta} = o(n^{-\frac{3}{4}+\delta})$ since $\delta < \frac{3}{4}$, and so

$$\begin{aligned} p + O(n^{-\frac{3}{4}}) = k_n/n &= F_n(\xi_F) + F(Z_n) - F(\xi_F) + \theta_n H_n \quad (|\theta_n| \leqq 1) \\ &= F_n(\xi_F) + (Z_n - \xi_F)F'(\xi_F) + \theta_n H_n + \varepsilon_n\,. \end{aligned} \tag{3.28}$$

From (3.26), (3.27), and (3.28), we obtain (3.23).

The condition (3.24) implies that for every fixed l,

$$\max_{1 \le j \le l} \sup_{F \in \mathscr{F}} P_F[|Z_j - \xi_F| \geqq x] \to 0 \quad \text{as} \quad x \to \infty\,. \tag{3.29}$$

Since $\inf_{F \in \mathscr{F}} F'(\xi_F) > 0$, (3.25) follows from (3.22), (3.23), (3.29), and Theorem 2. □

THEOREM 6. *Let $h: R^m \to R$ be a symmetric kernel*, i.e., $h(x_1, \cdots, x_m) = h(x_{\nu(1)}, \cdots, x_{\nu(m)})$ *for all permutations ν of $\{1, \cdots, m\}$. Suppose $X_1, X_2, \cdots$ are* i.i.d. *with a common distribution function $F \in \mathscr{G}$ such that*

$$\sup_{F \in \mathscr{G}} E_F h^2(X_1, \cdots, X_m) < \infty\,. \tag{3.30}$$

Let $U_n = \binom{n}{m}^{-1} \sum' h(X_{i_1}, \cdots, X_{i_m})$ where $\sum'$ stands for summation over all choices of $\{i_1, \cdots, i_m\}$ such that $1 \leqq i_j \leqq n$ and $i_j \neq i_k$ if $j \neq k$. Define

$$\begin{aligned} &\psi(F) = E_F h(X_1, \cdots, X_m)\,, \qquad h_F(x) = E_F h(x, X_2, \cdots, X_m) - \psi(F)\,, \\ &\hat{U}_n = \psi(F) + (m/n) \textstyle\sum_{i=1}^n h_F(X_i)\,, \qquad R_n = U_n - \hat{U}_n\,. \end{aligned} \tag{3.31}$$

Then for every $\varepsilon > 0$,

$$\sup_{F \in \mathscr{G}} P_F[\sup_{j \ge n} j|R_j|/(\log j) \geqq \varepsilon] \to 0 \quad \textit{as} \quad n \to \infty\,. \tag{3.32}$$

Consequently for every $\varepsilon > 0$,

$$\sup_{F \in \mathscr{G}} P_F[\max_{m \le j \le n} j|R_j| \geqq \varepsilon \log n] \to 0 \quad \textit{as} \quad n \to \infty\,. \tag{3.33}$$

REMARK. The result (3.32) gives the uniform almost sure convergence to 0 of $nR_n/(\log n)$. When $\mathscr{G}$ is a singleton (i.e., in the case of a single fixed distribution F), this result was obtained by Sen [21]. The result (3.33), which follows easily from (3.32), leads to a uniform invariance principle for U-statistics which we shall present in Theorem 8. To prove (3.32), we make use of the Chow-Hájek-Rényi inequality as in ([21], page 391) to obtain that for $n \geqq 3$

$$(3.34) \qquad P_F[\sup_{j \geqq n} j|R_j|/(\log j) \geqq \varepsilon] \leqq \varepsilon^{-2} \sum_{j=n}^{\infty} (j/\log j)^2 (E_F R_j^2 - E_F R_{j+1}^2)\,.$$

Since R_n is a U-statistic with mean 0 corresponding to the kernel $h_F^* = h - h_F - \psi(F)$ (cf. [7]), we can apply Hoeffding's formula (cf. (5.13) of [12]) to evaluate $E_F R_j^2$ and thereby obtain that $E_F R_j^2 - E_F R_{j+1}^2 \leqq C(F) j^{-3}$, where $\sup_{F \in \mathscr{G}} C(F) < \infty$ in view of (3.30). (See also (3.11) of [21].) Hence (3.32) follows from (3.34).

4. Uniform large deviation theorems for normalized test statistics. In this section, we shall obtain certain uniform large deviation probabilities of the following type: For some $\eta > 0$ and all $\varepsilon > 0$,

$$(4.1) \qquad P_\theta[|W_{r,\theta}(t)| \geqq \varepsilon t] \leqq g_\varepsilon(t) \qquad \text{for all } |\theta - \theta_0| \leqq \eta\,, \quad t \geqq 1 \text{ and all large } r \text{ (say } r \geqq r_\varepsilon)\,,$$

where $W_{r,\theta}(t)$ is the normalized statistic defined in (2.4) and g_ε is a positive function such that $\int_1^\infty g_\varepsilon(t)\,dt < \infty$. Obviously (4.1) implies not only (2.5), but also the following stronger result:

(4.2) For every $\varepsilon > 0$ and $h > 0$, there exist $r_{\varepsilon,h} > 0$ and a positive function g_ε such that $\int_1^\infty g_\varepsilon(t)\,dt < \infty$ and $P_{\theta_0+\xi r^{-\frac{1}{2}}}[|W_{r,\theta_0+\xi r^{-\frac{1}{2}}}(t)| \geqq \varepsilon t] \leqq g_\varepsilon(t)$ for all $t \geqq 1$, $|\xi| \leqq h$ and $r \geqq r_{\varepsilon,h}$.

This kind of uniform large deviation probabilities leads to the following analogue of Theorem 2 concerning the uniform convergence of the normalized ASN function for asymptotic Wald tests.

THEOREM 7. *With the same notations and assumptions as in Theorem* 2, *assume further that* (4.2) *holds. Then setting $d_\xi = (\xi - \frac{1}{2})d$ as in Theorem* 2, *we have for every compact subset C of $R - \{\frac{1}{2}\}$, as $r \to \infty$,*

$$(4.3) \qquad E_{\theta_0+\xi r^{-\frac{1}{2}}}(\tau_r/r) \to E\tau(d_\xi) \qquad \textit{uniformly for} \quad \xi \in C\,.$$

PROOF. Take $\varepsilon > 0$, $a > \frac{1}{2} + 3\varepsilon$ and $b < \frac{1}{2} - 3\varepsilon$. Then for $b \leqq \xi \leqq \frac{1}{2} - 3\varepsilon$ we obtain using (3.4), (3.5), (3.10), (4.2) and an argument similar to that used in (2.9) that for all large t and r (say $t \geqq t_0$, $r \geqq r_0$),

$$(4.4) \qquad \begin{aligned} P_{\theta_0+\xi r^{-\frac{1}{2}}}[\tau_r > rt] &\leqq P_{\theta_0+\xi r^{-\frac{1}{2}}}[T_{[rt]} - t\psi_r(\theta_0 + \xi r^{-\frac{1}{2}}) \geqq \tfrac{1}{2}t(\tfrac{1}{2} - \xi)d\sigma_r(\theta_0)] \\ &\leqq P_{\theta_0+\xi r^{-\frac{1}{2}}}[W_{r,\theta_0+\xi r^{-\frac{1}{2}}}(t) \geqq d\varepsilon t] \leqq g_{d\varepsilon}(t)\,. \end{aligned}$$

Likewise if $\frac{1}{2} + 3\varepsilon \leqq \xi \leqq a$, then for all large t and r,

$$P_{\theta_0+\xi r^{-\frac{1}{2}}}[\tau_r > rt] \leqq P_{\theta_0+\xi r^{-\frac{1}{2}}}[W_{r,\theta_0+\xi r^{-\frac{1}{2}}}(t) \leqq -d\varepsilon t] \leqq g_{d\varepsilon}(t) . \tag{4.5}$$

From (3.7), (4.4), and (4.5), the desired conclusion follows easily. □

We now establish uniform large deviation probabilities of the type (4.1) for U-statistics, sample quantiles and Chernoff–Savage statistics. The following lemma, which is a consequence of a theorem of von Bahr ([23], page 811), establishes the corresponding results for sample sums.

LEMMA 1. *Let* $q \geqq 2$. *Let* $X, X_1, X_2, \cdots$ *be* i.i.d. *random variables with a common distribution function* $F \in \mathscr{F}$, *where* $\mathscr{F}$ *is a family of distribution functions such that*

$$E_F X = 0 \qquad \textit{for all} \quad F \in \mathscr{F} \quad \textit{and} \quad \sup_{F\in\mathscr{F}} E_F|X|^q < \infty . \tag{4.6}$$

Then there exists a positive constant B *such that*

$$E_F|\textstyle\sum_{i=1}^n X_i|^q \leqq Bn^{q/2} \qquad \textit{for all} \quad n \geqq 1 \quad \textit{and} \quad F \in \mathscr{F} .$$

THEOREM 8 (*U-statistics*). *With the same notations as in Theorem* 6, *let* $\mathscr{F}_\delta$ *be a family of distribution functions such that*

$$\sup_{F\in\mathscr{F}_\delta} E_F|h(X_1, \cdots, X_m)|^{2+\delta} < \infty \qquad \textit{and} \qquad \inf_{F\in\mathscr{F}_\delta} \mathrm{Var}_F\, h_F(X_1) > 0 , \tag{4.7}$$

where $\delta > 0$. *Let* $\sigma_F{}^2 = m^2 \mathrm{Var}_F\, h_F(X_1)$ *and define* $V_n = n(U_n - \psi(F))/\sigma_F$ *for* $n \geqq m$ *and* $V_n = 0$ *if* $n < m$. *Then given any* $\varepsilon > 0$ *and* $k > 0$, *we have* (*by redefining the random variables on a new probability space if necessary*) *as* $r \to \infty$,

$$P_F[\max_{0\leqq t\leqq k} |r^{-\frac{1}{2}} V_{[rt]} - W(t)| \geqq \varepsilon] \to 0 \qquad \textit{uniformly for} \quad F \in \mathscr{F}_\delta . \tag{4.8}$$

Furthermore there exists a positive constant C *such that*

$$P_F[r^{-\frac{1}{2}}|V_{[rt]}| \geqq t] \leqq Ct^{-(2+\delta)/2} \qquad \textit{for all} \quad t \geqq 1 , \quad r \geqq m \quad \textit{and} \quad F \in \mathscr{F}_\delta . \tag{4.9}$$

PROOF. In view of (4.7), the uniform invariance principle (4.8) follows easily from Theorems 3 and 6. We now prove (4.9). Let $2l$ be the largest positive even integer such that $2 + \delta \geqq 2l$. Then $\sup_{F\in\mathscr{F}_\delta} E_F|h(X_1, \cdots, X_m)|^{2l} < \infty$. Making use of this and the fact that R_n is a U-statistic with kernel $h_F{}^* = h - h_F - \psi(F)$, it can be proved by an argument due to Grams and Serfling (cf. Theorem 1 of [7]) that there exists a positive constant A such that $E_F R_n{}^{2l} \leqq An^{-2l}$ for all $n \geqq m$ and $F \in \mathscr{F}_\delta$. Therefore by the Chebyshev inequality, for $t \geqq 1$ and $r \geqq m$,

$$P_F[r^{-\frac{1}{2}}|[rt]R_{[rt]}|/\sigma_F \geqq \tfrac{1}{2}t] \leqq (4/\sigma_F)^{2l} A t^{-2l} . \tag{4.10}$$

By (4.7), Lemma 1 is applicable with $q = 2 + \delta$. From Lemma 1 and the Chebyshev inequality, there exists a positive constant B such that

$$P_F[mr^{-\frac{1}{2}}|\textstyle\sum_{i=1}^{[rt]} h_F(X_i)|/\sigma_F \geqq \frac{1}{2}t] \leqq (2m/\sigma_F)^{2+\delta} B t^{-(2+\delta)/2} \tag{4.11}$$

for all $t \geqq 1$, $r \geqq m$, and $F \in \mathscr{F}_\delta$. Since $2l > (2 + \delta)/2$ and $\inf_{F\in\mathscr{F}_\delta} \sigma_F > 0$, (4.9) follows from (4.10) and (4.11). □

THEOREM 9 (*Sample quantiles*). *Let $0 < p < 1$. With the same notations as in Theorem 5, let $\mathscr{F}^*$ be a family of distribution functions F on the real line such that the equation $F(\xi) = p$ has a unique solution $\xi = \xi_F$ and there exists a positive constant b for which*

$$(4.12) \qquad \inf_{F\in\mathscr{F}^*}\inf_{|x|\leqq b} F'(\xi_F + x) > 0 \qquad \text{and} \qquad \sup_{F\in\mathscr{F}^*} F'(\xi_F) < \infty\,.$$

Then given $\varepsilon > 0$, there exist $\eta > 0$ and $r_0 \geqq 1$ such that

$$(4.13) \qquad P_F[r^{-\frac{1}{2}}|V_{[rt]}| \geqq \varepsilon t] \leqq e^{-\eta t} \qquad \text{for all} \quad t \geqq 1\,, \quad r \geqq r_0 \quad \text{and} \quad F \in \mathscr{F}^*\,.$$

PROOF. Let $B(n, \rho)$ denote the binomial random variable with mean $n\rho$ ($n =$ number of trials). We note that for $t, r \geqq 1$,

$$P_F[Z_{[rt]} \leqq \xi_F - \tfrac{1}{2}\varepsilon\sigma_F r^{-\frac{1}{2}}] = P[B([rt], \pi_r) \geqq k_{[rt]}]\,,$$

where $\pi_r = F(\xi_F - \frac{1}{2}\varepsilon\sigma_F r^{-\frac{1}{2}}) = p - \frac{1}{2}\varepsilon\sigma_F r^{-\frac{1}{2}} g(r, F)$ and $\inf_{r\geqq r_0}\inf_{F\in\mathscr{F}^*} g(r, F) > 0$ in view of (4.12). A similar result holds for $P_F[Z_{[rt]} \geqq \xi_F + \frac{1}{2}\varepsilon\sigma_F r^{-\frac{1}{2}}]$. Noting that $\inf_{F\in\mathscr{F}^*}\sigma_F > 0$ by (4.12), the desired conclusion then follows by an easy application of Bernstein's inequality (cf. [22], pages 204–205). □

THEOREM 10 (*Generalized Chernoff–Savage statistics*). *With the same notations and assumptions as in Theorem 4, for every $a > 0$ and $\varepsilon > 0$, there exist $\zeta > 0$ and $r_0 \geqq 1$ such that*

$$(4.14) \qquad P_{F,G}[r^{-\frac{1}{2}}|V_{[rt]}| \geqq \varepsilon t] \leqq e^{-\zeta t}$$
$$\text{for all} \quad t \geqq 1\,, \quad r \geqq r_0 \quad \text{and} \quad (F, G) \in \mathscr{C}(a)\,.$$

While our estimates in [14] of the remainder term R_n in the representation (3.14) of generalized Chernoff–Savage statistics have readily given us the uniform invariance principle in Theorem 4, they are not sharp enough to prove Theorem 10. Making use of these estimates and Lemma 1, we are only able to obtain a uniform algebraic rate of convergence of the form

$$P_{F,G}[r^{-\frac{1}{2}}|V_{[rt]}| \geqq t] \leqq t^{-\theta(\delta)} \qquad \text{for all} \quad t \geqq 1\,, \quad r \geqq r_0 \quad \text{and} \quad (F, G) \in \mathscr{C}(a)\,,$$

where $\theta(\delta)$ is a positive constant depending on the δ given by (3.11). To prove Theorem 10, we need a different representation of Chernoff–Savage statistics together with some new estimates on the tail of the empirical distribution function. The details of the proof are given in [17], where it is also shown that these new estimates also provide another proof of the uniform invariance principle (3.16) for generalized Chernoff–Savage statistics under weaker conditions than Theorem 4.

5. Applications and examples. In this section, we shall illustrate how the general ideas of Section 2 and the general theorems of Sections 3 and 4 can be applied to specific testing problems. While the same kind of argument would work for most other problems, here we shall only consider the two-sample location problem to indicate the argument used. Our basic tools in the subsequent analysis are therefore Theorems 4 and 10 on generalized Chernoff–Savage statistics. Obviously by applying our corresponding results for U-statistics and sample quantiles, we can similarly construct and analyze asymptotic Wald tests based on these statistics.

Suppose $X_1, X_2, \cdots$ are i.i.d. with a common continuous distribution function F and are independent of $Y_1, Y_2, \cdots$ which are i.i.d. with a common continuous distribution function G. We wish to test sequentially $H: G = F$ versus $K: G(x) = F(x+\theta)$ for all x and some positive θ. Some commonly used nonsequential tests of H versus K include the t-test, the Wilcoxon test, the normal scores test and the van der Waerden test. We shall now describe their sequential analogues in the following examples and show that the test statistics used in these sequential tests satisfy the conditions of Section 2. For simplicity, we shall assume the vector-at-a-time sampling scheme, i.e., at each stage prior to stopping, a pair of observations (X_i, Y_i) is taken. We shall let $R_1^{(n)}, \cdots, R_n^{(n)}$ denote the ordered ranks of the X's in the combined sample of $2n$ observations $X_1, \cdots, X_n, Y_1, \cdots, Y_n$. Also let

$$F_n(x) = n^{-1} \textstyle\sum_1^n I_{[X_i \leqq x]}\,, \qquad G_n(x) = n^{-1} \sum_1^n I_{[Y_i \leqq x]}$$

denote the empirical distribution functions.

EXAMPLE 1 (*Sequential Wilcoxon test*). Consider the Wilcoxon statistic

$$T_n = n^{-1} \textstyle\sum_{i=1}^n R_i^{(n)} = n \int_{-\infty}^{\infty} J(F_n(x), G_n(x))\, dF_n(x)\,,$$

where $J(x, y) = x + y$. Obviously J satisfies (3.11) with $\delta = \frac{5}{2}$. Hence the assumptions of Theorems 4 and 10 are satisfied. To test H versus K, take positive numbers b, b', and c and stop sampling at stage

$$\tau = \inf\{n \geqq 1 : T_n \geqq cn + b \quad\text{or}\quad T_n \leqq cn - b'\}\,. \tag{5.1}$$

We reject H iff $T_\tau \geqq c\tau + b$.

By Theorem 4, we can write

$$\begin{aligned} T_n = n\{\tfrac{1}{2} + \textstyle\int_{-\infty}^{\infty} G(x)\, dF(x)\} + \sum_1^n \{G(X_i) - \int_{-\infty}^{\infty} G(x)\, dF(x)\} \\ - \textstyle\sum_1^n \{F(Y_i) - \int_{-\infty}^{\infty} F(x)\, dG(x)\} + nR_n\,, \end{aligned} \tag{5.2}$$

where R_n satisfies (3.15). Since $G(x) = F(x+\theta)$, we set

$$\mu_F(\theta) = \tfrac{1}{2} + \textstyle\int_{-\infty}^{\infty} G(x)\, dF(x) = \frac{1}{2} + \int_{-\infty}^{\infty} F(x+\theta)\, dF(x)\,, \tag{5.3}$$

$$v_F(\theta) = \operatorname{Var}_F G(X_1) + \operatorname{Var}_G F(Y_1)\,. \tag{5.4}$$

The continuity of F implies that $\mu_F(\theta)$ and $v_F(\theta)$ are continuous functions of θ and $v_F(\theta) > 0$ for all θ. We note that for all F,

$$\mu_F(0) = 1\,, \qquad v_F(0) = \tfrac{1}{6}\,. \tag{5.5}$$

Let $\theta_0 = 0$ and define for $r > 0$

$$\psi_r(\theta) = r\mu_F(\theta)\,, \qquad \sigma_r(\theta) = \{rv_F(\theta)\}^{\frac{1}{2}}\,; \tag{5.6a}$$

$$W_{r,\theta}(t) = (T_{[rt]} - t\psi_r(\theta))/\sigma_r(\theta)\,, \qquad t \geqq 0\,. \tag{5.6b}$$

Writing the probability measure $P_{F,G}$ as P_θ for fixed F, we obtain from Theorem

4 that (3.1) holds for all η, ε, $k > 0$. Moreover, Theorem 10 implies that given $\varepsilon, \eta > 0$, (4.1) holds with $g_\varepsilon(t) = e^{-\zeta t}$. Obviously condition (3.4) is satisfied, and by the continuity of v_F, condition (3.5) also holds.

Assume that F satisfies the following condition:

$$\mu_F'(0) = \lim_{\theta\to 0} \theta^{-1} \textstyle\int_{-\infty}^{\infty} (F(x+\theta) - F(x))\, dF(x) \quad \text{exists and is positive.} \tag{5.7}$$

Then (3.3) and (3.6) obviously also hold with

$$d(= \lim_{r\to\infty} \psi_r'(0)/\{r^{\frac{1}{2}}\sigma_r(0)\} = \lim_{r\to\infty} \{\psi_r(r^{-\frac{1}{2}}) - \psi_r(0)\}/\sigma_r(0)) = 6^{\frac{1}{2}}\mu_F'(0)\,. \tag{5.8}$$

Hence conditions (2.1)—(2.5) are satisfied and by our results in Section 2, given any continuous distribution function F satisfying (5.7) and $0 < \alpha, \beta < \frac{1}{2}$, the sequential Wilcoxon test with stopping rule (5.1) is an asymptotic Wald test with asymptotic error rates (α, β) of $H_0: \theta = 0$ versus $H_r: \theta = r^{-\frac{1}{2}}$ if in (5.1) we set $\tau = \tau_r$ and

$$\begin{aligned} b &= b_r = (6\mu_F'(0))^{-1} r^{\frac{1}{2}} \log((1-\beta)/\alpha)\,,\\ b' &= b_r' = (6\mu_F'(0))^{-1} r^{\frac{1}{2}} \log((1-\alpha)/\beta)\,,\\ c &= c_r = \tfrac{1}{2}\{\mu_F(r^{-\frac{1}{2}}) + \mu_F(0)\} = 1 + \tfrac{1}{2}r^{-\frac{1}{2}}\mu_F'(0) + o(1)\,. \end{aligned} \tag{5.9}$$

Since we have shown that the conditions of Theorems 2 and 7 are satisfied, the uniform convergence properties (3.7), (3.8), and (4.3) for the normalized ASN and power functions of the sequential Wilcoxon test hold.

While (4.3) gives the uniform convergence of $E_{\xi r^{-\frac{1}{2}}}(\tau_r/r)$ for ξ in compact subsets of $R - \{\frac{1}{2}\}$, it turns out that for the sequential Wilcoxon test, we also have uniform convergence of $E_{\xi r^{-\frac{1}{2}}}(\tau_r/r)$ for ξ in compact neighborhoods of $\frac{1}{2}$. To see this, take $0 < \rho < \frac{1}{2}$ and define

$$L_\rho = \sup\{n \geqq 1 : |nR_n| \geqq n^\rho\} \quad (\sup \varnothing = 0)\,, \tag{5.10}$$

where R_n is as defined in (5.2). Then using the uniform estimates in [14, page 836] (which are applicable since $J(x, y) = x + y$ is a C^∞ function on the whole of the unit square $[0, 1] \times [0, 1]$), it can be shown that given any $\gamma > 0$, $\sup_{F,G\in\mathscr{G}} EL_\rho^\gamma < \infty$. Therefore by a simple modification of the argument used in the proof of Theorem 1 of [15], we obtain the desired uniform convergence of $E_{\xi r^{-\frac{1}{2}}}(\tau_r/r)$ for ξ in compact subsets of R.

Some of the above asymptotic results for the sequential Wilcoxon test have been obtained recently by other methods by Holm [13].

EXAMPLE 2 (*Sequential van der Waerden test*). Define the functions J and J_n $(n = 1, 2, \cdots)$ on $[0, 1] \times [0, 1]$ by

$$\begin{aligned} &J(1,1) = J(0,0) = J_n(0,0) = 0\,, &&J_n(1,1) = \Phi^{-1}(2n/(2n+1))\,,\\ &J(x,y) = \Phi^{-1}(\tfrac{1}{2}(x+y))\,, &&J_n(x,y) = \Phi^{-1}(n(x+y)/(2n+1)) \end{aligned}$$

for $(x, y) \notin \{(0, 0), (1, 1)\}$. Let

$$T_n = \textstyle\sum_{i=1}^n \Phi^{-1}(R_i^{(n)}/(2n+1)) = n \int_{-\infty}^{\infty} J_n(F_n(x), G_n(x))\, dF_n(x)$$

denote the van der Waerden statistic. Clearly J satisfies (3.11) with $\delta = \frac{1}{2}$, and by Lemma 5(ii) of [14], J_n satisfies (3.12). Hence the assumptions of Theorems 4 and 10 are again satisfied. With the sequence $\{T_n\}$ of van der Waerden statistics, define the stopping rule τ by (5.1). The sequential van der Waerden statistic stops sampling at stage τ and rejects H iff $T_\tau \geqq c\tau + b$.

By Theorem 4, we have the representation (3.14) for $\Gamma_n = T_n/n$ with the functions $h_{F,G}$ and $h^*_{F,G}$ as in Theorem 4. Since $G(x) = F(x + \theta)$, we set

$$(5.11) \qquad \mu_F(\theta) = \int_{-\infty}^{\infty} J(F(x), G(x))\, dF(x) = \int_{F(x)+F(x+\theta)<2} \Phi^{-1}(\tfrac{1}{2}(F(x) + F(x+\theta)))\, dF(x)\,,$$

$$(5.12) \qquad v_F(\theta) = \mathrm{Var}_F\, h_{F,G}(X_1) + \mathrm{Var}_G\, h^*_{F,G}(Y_1)\,.$$

With this choice of μ_F and v_F, define ψ_r and σ_r as in (5.6a). We note that

$$(5.13) \qquad \mu_F(0) = 0\,, \qquad v_F(0) = \tfrac{1}{2}\,.$$

Let $\theta_0 = 0$ and assume that F satisfies the following condition:

$$(5.14) \qquad \mu_F'(0) = \lim_{\theta\to 0} \theta^{-1} \int_{F(x)+F(x+\theta)<2} \Phi^{-1}(\tfrac{1}{2}(F(x) + F(x+\theta)))\, dF(x) \quad \text{exists and is positive.}$$

Then assumptions (2.1)—(2.5), and in fact also the stronger assumptions in Theorems 2 and 7, are again satisfied with

$$(5.15) \qquad d(= \lim_{r\to\infty} \psi_r'(0)/\{r^{\frac{1}{2}}\sigma_r(0)\} = \lim_{r\to\infty} \{\psi_r(r^{-\frac{1}{2}}) - \psi_r(0)\}/\sigma_r(0)) = 2^{\frac{1}{2}}\mu_F'(0)\,.$$

Hence given any continuous distribution function F satisfying (5.14) and $0 < \alpha$, $\beta < \frac{1}{2}$, the sequential van der Waerden test with stopping rule (5.1) is an asymptotic Wald test with asymptotic error rates (α, β) of $H_0: \theta = 0$ versus $H_r: \theta = r^{-\frac{1}{2}}$ if in (5.1) we set

$$(5.16) \qquad b = (2\mu_F'(0))^{-1} r^{\frac{1}{2}} \log((1-\beta)/\alpha)\,,$$
$$b' = (2\mu_F'(0))^{-1} r^{\frac{1}{2}} \log((1-\alpha)/\beta)\,, \qquad c = \tfrac{1}{2}\mu_F(r^{-\frac{1}{2}}) \sim \tfrac{1}{2} r^{-\frac{1}{2}}\mu_F'(0)\,.$$

EXAMPLE 3 (*Sequential normal scores test*). Let J be defined as in Example 2. For $j = 1, \cdots, 2n$, let $u_n(j) = E\Phi^{-1}(U_{jn})$, where U_{jn} is the jth order statistic in a sample of size $2n$ from the uniform distribution on $[0, 1]$. Define for $r, s \in \{0, n^{-1}, 2n^{-1}, \cdots, 1\}$,

$$J_n(r, s) = u_n(n(r + s)) \quad \text{if} \quad (r, s) \neq (0, 0)\,, \qquad J_n(0, 0) = 0\,.$$

By Lemma 5(iii) of [14], condition (3.12) still holds for the present choice of J_n and J. The sequential normal scores test is similar to the sequential van der Waerden test, the only change being the replacement of the van der Waerden statistic by the normal scores statistic $\sum_{i=1}^{n} u_n(R_i^{(n)})$. Since Theorems 4 and 10 are again applicable, the results of Example 2 also hold for the sequential normal scores test.

EXAMPLE 4 (*Two-sample sequential t-test*). Let k_n be a sequence of positive constants such that $k_n \sim 2n$. Let $T_0 = T_1 = 0$ and for $n \geqq 2$ define

$$\bar{X}_n = n^{-1} \textstyle\sum_1^n X_i\,, \qquad \bar{Y}_n = n^{-1} \sum_1^n Y_i\,, \tag{5.17}$$
$$s_n^2 = k_n^{-1}\{\textstyle\sum_1^n (X_i - \bar{X}_n)^2 + \sum_1^n (Y_i - \bar{Y}_n)^2\}\,, \qquad T_n = n(\bar{X}_n - \bar{Y}_n)/s_n\,.$$

With T_n thus defined, define the stopping rule τ as in (5.1). The two-sample sequential t-test stops sampling at stage τ and rejects H iff $T_\tau \geqq c\tau + b$.

Assume that F satisfies the following condition:

$$\textstyle\int_{-\infty}^{\infty} x^4\, dF(x) < \infty\,. \tag{5.18}$$

Let $\theta_0 = 0$, $\sigma^2(F) = \mathrm{Var}_F\, X_1\,(> 0$ since F is continuous) and define for $r > 0$,

$$\psi_r(\theta) = r\theta/\sigma(F)\,, \qquad \sigma_r(\theta) = (2r)^{\frac{1}{2}}\,. \tag{5.19}$$

Clearly ψ_r and σ_r thus defined satisfies (3.3)—(3.6) with

$$d(= \lim_{r\to\infty} \{\psi_r(r^{-\frac{1}{2}}) - \psi_r(0)\}/\sigma_r(0)) = (2\sigma^2(F))^{-\frac{1}{2}}\,. \tag{5.20}$$

Since $G(x) = F(x + \theta)$, we can write $Y_i = X_i^* - \theta$, where X_i^* has distribution function F. We note that for $n \geqq 2$ and $rt \geqq 2$,

$$s_n^2 = k_n^{-1}\{\textstyle\sum_1^n (X_i - \bar{X}_n)^2 + \sum_1^n (X_i^* - \bar{X}_n^*)^2\} \to \sigma^2(F) \quad \text{a.s.,} \qquad \text{and} \tag{5.21}$$

$$\begin{aligned} W_{r,\theta}(t)(&= \{T_{[rt]} - t\psi_r(\theta)\}/\sigma_r(\theta)) \\ &= -r^{\frac{1}{2}}t\theta/(2^{\frac{1}{2}}\sigma(F)) + (2rs^2_{[rt]})^{-\frac{1}{2}}\{\textstyle\sum_{i=1}^{[rt]} (X_i - X_i^*) + [rt]\theta\}\,. \end{aligned} \tag{5.22}$$

From (5.21) and (5.22), it follows that for all ε, h, $k > 0$, we have (by redefining the random variables on a new probability space if necessary) as $r \to \infty$,

$$P[\max_{0\leqq t\leqq k} |W_{r,\xi r^{-\frac{1}{2}}}(t) - W(t)| > \varepsilon] \to 0 \qquad \text{uniformly for} \quad |\xi| \leqq h\,. \tag{5.23}$$

Hence (3.2) holds.

Using (5.18) and Theorem 5 of [5], we obtain that for all $\delta > 0$,

$$\textstyle\int_2^\infty f_\delta(t)\, dt < \infty \qquad \text{where} \quad f_\delta(t) = P[\sup_{n\geqq t} |s_n^2 - \sigma^2(F)| > \delta]\,. \tag{5.24}$$

In view of (5.22), given ε, $h > 0$, we can choose $\delta > 0$ sufficiently small and B, r_1 sufficiently large such that for all $r \geqq r_1$, $t \geqq 2$ and $|\xi| \leqq h$,

$$\begin{aligned} P[|W_{r,\xi r^{-\frac{1}{2}}}(t)| \geqq \varepsilon t] &\leqq f_\delta(t) + P[|\textstyle\sum_{i=1}^{[rt]} (X_i - X_i^*)| \geqq \varepsilon t r^{\frac{1}{2}}\sigma(F)] \\ &\leqq f_\delta(t) + Bt^{-2}\,. \end{aligned} \tag{5.25}$$

The last inequality above follows from the finiteness of EX_1^4 (see Lemma 1). Hence (4.2) holds.

We have therefore shown that conditions (2.1)—(2.5), and in fact also the stronger assumptions of Theorems 2 and 7, are again satisfied. An alternative proof of the weak convergence criterion (2.4) for the two-sample t-statistic T_n has been given earlier by Hall [8] using the Hall–Loynes sequential extension of the contiguity concept.

By our results in Section 2, given any continuous distribution function F

satisfying (5.18) and $0 < \alpha, \beta < \frac{1}{2}$, the two-sample sequential t-test is an asymptotic Wald test with asymptotic error rates (α, β) of $H_0: \theta = 0$ versus $H_r: \theta = r^{-\frac{1}{2}}$ if in (5.1) we set $\tau = \tau_r$ and

$$b = b_r = 2r^{\frac{1}{2}}\sigma(F)\log((1-\beta)/\alpha), \tag{5.26}$$
$$b' = b_r' = 2r^{\frac{1}{2}}\sigma(F)\log((1-\alpha)/\beta), \qquad c = c_r = (2\sigma(F))^{-1}r^{-\frac{1}{2}}.$$

Letting $\theta_r = r^{-\frac{1}{2}} (r > 0)$, we now show that the above two-sample sequential t-test is an asymptotically efficient test of $H_0: \theta = 0$ versus $H_r: \theta = \theta_r$ for the case where the underlying distributions are normal. Let F be the distribution function of the $\mathcal{N}(\mu, \sigma^2)$ distribution and so G is the distribution function of the $\mathcal{N}(\mu - \theta, \sigma^2)$ distribution. We assume that σ^2 is known but μ and θ are both unknown. Let $W_n = X_n - Y_n$, $n = 1, 2, \cdots$, so that W_n has the $\mathcal{N}(\theta, 2\sigma^2)$ distribution. Let $p_0(w) = (4\pi\sigma^2)^{-\frac{1}{2}}\exp(-w^2/4\sigma^2)$ and $p_\theta(w) = p_0(w - \theta)$. Then $(W_1, \cdots, W_n)$ is Fraser-sufficient for θ for every sample size n and the optimum test of $H_0: \theta = 0$ versus $H_r: \theta = \theta_r$ in this normal model is Wald's SPRT involving the log likelihood ratio

$$Z_i^{(r)} = \log\{p_{\theta_r}(W_i)/p_0(W_i)\} = (2r^{-\frac{1}{2}}W_i - r^{-1})/(4\sigma^2)$$

(cf. [6], pages 246–248 and 252–253). It is easy to see that

$$E_{\theta_r} Z_1^{(r)} = r^{-1}/4\sigma^2 = -E_0 Z_1^{(r)}. \tag{5.27}$$

Putting (5.20) and (5.27) in (2.22), we obtain that the asymptotic efficiency of the two-sample sequential t-test is equal to 1.

In the preceding examples, we have shown that for the sequence of Wilcoxon statistics, or van der Waerden statistics, or normal scores statistics, or two-sample t-statistics, conditions (2.1)—(2.5) are all satisfied. Let us now compare the sequential rank tests in Examples 1, 2, 3 with the two-sample sequential t-test which we have seen is asymptotically efficient for the normal model. Let ε denote the asymptotic relative efficiency of the sequential rank test under consideration with respect to the two-sample sequential t-test. Our results in Section 2 say that ε is given by (2.18) and is therefore equal to the Pitman efficiency for the corresponding fixed sample size tests based on these statistics. Hence in the particular case of the sequential Wilcoxon test, ε is $3/\pi$ $(= .95)$ when F (and therefore G also) is a normal distribution. When F is not necessarily normal but is continuous and satisfies (5.7) and (5.18), ε is always $\geqq .864$ (cf. [11]). In particular, if F has a density f, then (5.8) and (5.20) imply that $\varepsilon = 12\sigma^2(F)\{\int_{-\infty}^{\infty} f^2(x)\,dx\}^2$ which can be made as large as we please by choosing $\sigma(F)$ large while $\int_{-\infty}^{\infty} f^2(x)\,dx$ remains bounded (cf. [11]). The situation is even more favorable for the sequential van der Waerden test. Here ε is 1 when F is normal and is always $\geqq 1$ when F has a density f and satisfies (5.14) and (5.18) (cf. [4]). In fact, (5.15) and (5.20) imply that $\varepsilon = \sigma^2(F)\{\int_{-\infty}^{\infty} J_0'(F(x))f^2(x)\,dx\}^2$, where $J_0 = \Phi^{-1}$, and so we can again choose F to make ε as large as we please. The same results obviously also hold for the sequential normal scores test which is asymptotically equivalent to the sequential van der Waerden test.

REFERENCES

[1] BAHADUR, R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[2] BRAUN, H. I. (1974). Applications of weak convergence to sequential nonparametric statistics. Stanford Univ. Tech. Report.

[3] BRAUN, H. I. (1976). Weak convergence of sequential linear rank statistics. *Ann. Statist.* **4** 554–575.

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

[5] CHOW, Y. S. and LAI, T. L. (1975). Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Amer. Math. Soc.* **208** 51–72.

[6] GHOSH, B. K. (1970). *Sequential Tests of Statistical Hypotheses.* Addison-Wesley, Reading.

[7] GRAMS, W. F. and SERFLING, R. J. (1973). Convergence rates for U-statistics and related statistics. *Ann. Statist.* **1** 153–160.

[8] HALL, W. J. (1973). Two asymptotically efficient t-tests. *Proc. Prague Symp. Asymptotic Statist.* **2** (Jaroslav Hâjek, ed.), 89–108. Charles Univ., Prague.

[9] HALL, W. J. (1975). On construction of asymptotic sequential tests. Univ. of Rochester Tech. Report.

[10] HALL, W. J. and LOYNES, R. M. (1977). Weak convergence of processes related to likelihood ratios. *Ann. Statist.* **5** 330–341.

[11] HODGES, J. L. and LEHMANN, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[12] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.

[13] HOLM, S. (1975). Sequential inversion sum tests. *Scand. J. Statist.* **2** 1–10.

[14] LAI, T. L. (1975). On Chernoff-Savage statistics and sequential rank tests. *Ann. Statist.* **3** 825–845.

[15] LAI, T. L. (1975). A note on first exit times with applications to sequential analysis. *Ann. Statist.* **3** 999–1006.

[16] LAI, T. L. (1977). First exit times from moving boundaries for sums of independent random variables. *Ann. Probability* **5** 210–221.

[17] LAI, T. L. (1977). Uniform weak convergence and tail probabilities of normalized linear rank statistics with applications to sequential rank tests. Columbia Univ. Tech. Report.

[18] LOYNES, R. M. (1975). On weak convergence of martingales. Sheffield Univ. Tech. Report.

[19] NOETHER, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68.

[20] PITMAN, E. J. G. (1949). Lecture notes on nonparametric inference. Columbia Univ.

[21] SEN, P. K. (1974). Almost sure behavior of U-statistics and von Mises' differentiable statistical functions. *Ann. Statist.* **2** 387–395.

[22] USPENSKY, J. V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.

[23] VON BAHR, B. (1965). On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36** 808–818.

[24] WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16** 117–186.

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
MATHEMATICS BUILDING
NEW YORK, N.Y. 10027