The Annals of Statistics
1977, Vol. 5, No. 2, 292–315

A GENERAL THEOREM WITH APPLICATIONS ON EXPONENTIALLY BOUNDED STOPPING TIME, WITHOUT MOMENT CONDITIONS[1]

By R. A. Wijsman

University of Illinois at Urbana-Champaign

The stopping time N of a sequential test based on an i.i.d. sequence $X, X_1, X_2, \cdots$ with common distribution P is defined as the first integer $n \geqq 1$ such that $l_1 < L_n < l_2$ is violated, where L_n is a statistic depending only on $X_1, \cdots, X_n$. If for all l_1, l_2 there exist constants $c > 0$, $\rho < 1$ such that $P(N > n) < c\rho^n$, $n = 1, 2, \cdots$, then N is called *exponentially bounded* under P. In the contrary case P is termed *obstructive*.

A general theorem is proved whose conclusion is of the from that N is exponentially bounded under P unless $P\{f(X) = 0\} = 1$, with f a function that depends on the particular testing problem. Among the applications presented there are two new results. The first is for the sequential linear hypotheses F-test. The function f is found, and the distributions for which $P\{f(X) = 0\} = 1$ are shown to be supported on spheres. All these P's, and only these, are obstructive. The second result concerns the sequential two-sample Wilcoxon test for the equality of the distributions of random variables X and Y. Here the result is simple: P is obstructive if and only if $P(X = Y) = 1$.

Two auxiliary results of independent interest are presented. The first is a generalization of the exponential convergence of empirical to theoretical distribution function. The second one shows that if random variables X and Y have joint distribution P and marginal distribution function F, G, respectively, then $P\{F(Y) = G(X)\} = 1$ if and only if $P(X = Y) = 1$.

1. Introduction. Let $X, X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random variables with values in an arbitrary space $\mathscr{X}$ and common distribution P. Suppose one hypothesis (in general composite) is to be tested against another by means of a sequential test that has stopping time N (= random sample size). Consider the following property: there exist constants $c > 0$ and $\rho < 1$ such that

$$P(N > n) < c\rho^n , \qquad n = 1, 2, \cdots . \tag{1.1}$$

Only sequential tests will be considered whose stopping rule is determined by a sequence $L_1, L_2, \cdots$ of statistics, L_n depending only on $(X_1, \cdots, X_n)$, and a pair of stopping bounds $-\infty < l_1 < l_2 < \infty$, such that N is the first integer $n \geqq 1$ for which

$$l_1 < L_n < l_2 \tag{1.2}$$

Received July 1974; revised July 1976.

[1] Research supported by the National Science Foundation under grants GP-28154 and MPS 75-07978.

AMS 1970 *subject classifications*. Primary 62L10; Secondary 62F05, 62G10, 60F10.

Key words and phrases. Stopping time, exponentially bounded, without moment conditions, sequential tests, exponential convergence.

is violated. Invariant sequential probability ratio tests have this form, and all our examples are of that nature, but for the purpose of this paper it is irrelevant how the test arises. If for a particular P and arbitrary choice of l_1, l_2 inequality (1.1) is satisfied (where c, ρ may depend on l_1, l_2), then we shall say (following Berk [1]) that N is *exponentially bounded* under P. In the contrary case P will be termed *obstructive*. It should be kept in mind that P is to be regarded as the true distribution of X and need not belong to the family of distributions (the *model*) that lead to the test in the first place.

In the Wald sequential probability ratio test for testing one simple hypothesis against another the sequence L_n (= log probability ratio of $X_1, \cdots, X_n$) is a random walk, and Stein [11] has shown that N is exponentially bounded under every P, except under those P for which $P(L_1 = 0) = 1$. Thus, in this case a complete characterization of the obstructive distributions has been achieved.

When composite hypotheses are being tested L_n is no longer a random walk and results on the classification of P's are considerably less complete. There is one example in the literature ([12], Section 3) where the family of obstructive P's is completely known. The next best type of result is of the following form: There is a real valued function f on $\mathscr{X}$, and P is obstructive only if

$$P\{f(X) = 0\} = 1 \,. \tag{1.3}$$

The actual family of obstructive P's could possibly be a proper subset of the family satisfying (1.3). The first such result, where N was shown to be exponentially bounded under any P not satisfying (1.3), was Sethuraman's nonpaparametric example [10]. Parametric examples of this sort appeared in [13], Sections 5 and 6 and [15], Section 3. It should be noted that these examples have been treated on a more or less individual basis, rather than as applications of some general theorem. It is true that in [15] the result was obtained as an application of a certain theorem, but the scope of that theorem was very restricted.

On the other hand, there is a number of examples where the results follow from the application of a theorem that is sufficiently general to cover a wide class of models [1], [12]. However, this greater generality as regards the models was bought at a price: only distributions P could be handled under which certain functions of X have finite moment generating function (m.g.f.) in an interval about 0. In [14], Section 3(a), the obstructive P's among those for which X^2 has finite m.g.f. were exhibited, but nothing was known about P's for which X^2 has infinite m.g.f.

In [4] Lai was the first to demonstrate for the sequential t-test, in the case of hypotheses symmetric about 0, the existence of obstructive P's under which X^2 has infinite m.g.f. It turns out, however, that this is impossible if the hypotheses are not symmetric, and in [16] it was proved that in that case under every unbounded P the stopping time N is exponentially bounded. The method of proof suggested a general theorem that would cover at least all the known examples, and presumably more.

The general theorem mentioned above is presented in Section 2, together with several corollaries for easier application. In Section 3 applications are given to several parametric examples, and in Section 4 to nonparametric examples. The results in subsections 3.3 (sequential linear hypotheses F-test) and 4.1 (sequential two-sample Wilcoxon test) have not appeared in the literature before.

In all examples application of the theory in Section 2 leads eventually to the statement that N is exponentially bounded under P if P does not satisfy (1.3), with f a function that depends on the model. Which, if any, of the distributions P satisfying (1.3) are obstructive has to be investigated on an individual basis. In the examples in subsections 3.1, 3.3, and 4.1 the answer is completely known. The result is particularly simple in the sequential two-sample Wilcoxon test: P is obstructive if and only if $P(X = Y) = 1$, in which case $P(L_n = 0, n = 1, 2, \cdots) = 1$. This result parallels very closely the situation where L_n is a random walk [11].

The method of proof uses the classical idea of Stein [11] as adapted to the case where L_n is not a random walk by Sethuraman [10]. This idea consists in showing that there exists a positive integer r and $p > 0$ such that

$$(1.4) \qquad P\{|L_{n+r} - L_n| > d \,|\, N > n\} > p\,, \qquad n = 1, 2, \cdots,$$

in which $d = l_2 - l_1$. In order to find convenient conditions under which (1.4) holds we consider $L_{n+r} - L_n$ for large n. If this behaves asymptotically as $\sum_{j=1}^{r} f(X_{n+j})$, then L_n behaves asymptotically as a random walk, with steps $f(X_n)$, and exponential boundedness of N can be concluded unless (1.3) holds (the actual proof is of course more delicate). In general, however, the function f will also depend on $(X_1, \cdots, X_n)$. One of the hypotheses of the main theorem is that this dependence is through a statistic taking values in a certain fixed space, and under certain conditions (for instance if that space is compact) the desired conclusion can be drawn anyway. Thus, the theorem can be regarded as a generalization of Theorem 2.2 in [15].

In Section 2 the main theorem (Theorem 2.1) is followed by two corollaries, the second of which being convenient for the application to the nonparametric examples in Section 4. For the applications to the parametric examples in Section 3 the most convenient theorem is Theorem 2.2 (technically a corollary of Theorem 2.1), which assumes more structure than does Theorem 2.1 but also has the hypothesis in easier verifiable form. Theorem 2.3 is an easy and obvious theorem that is often useful to help decide exponential boundedness of N in cases where Theorem 2.1 fails.

Although the statement of these theorems and corollaries is not simple, they are relatively simple to apply. The procedure is to compute $L_{n+r} - L_n$ (this is sometimes the hardest part), replace $X_{n+1}, \cdots, X_{n+r}$ by $y_1, \cdots, y_r$, and let $n \to \infty$. In this process the function f is discovered, and also (if applicable) the statistics U_n and V_n that appear in Theorems 2.1 and 2.2. All that remains is to verify details such as continuity properties of certain functions and exponential boundedness of particular events.

In some problems the expression of L_n in terms of $X_1, \cdots, X_n$ is sufficiently explicit to calculate $L_{n+r} - L_n$ directly. More often, though, this is not possible. However, for the purpose of exponential boundedness of N (and for termination with probability one for that matter) it is permissible to replace L_n by a simpler expression that differs from L_n by a uniformly bounded amount. Thus, in the treatment of the sequential t- and F-tests L_n is replaced in two stages by a function linear in a certain statistic. In the applications in this paper it may not always explicitly be indicated whether the L_n used is the original one or a convenient replacement.

2. The main theorems. The following notation and conventions will be employed. If $\mathscr{X}$ is a space, $\mathscr{X}^n$ in the n-fold product of $\mathscr{X}$ with itself, and the points $(x_1, \cdots, x_n) \in \mathscr{X}^n$ will often be denoted x^n. All spaces are measurable, i.e., supplied with a sigma-field. As a rule no notation will be introduced for the various sigma-fields since usually we shall not need them explicitly. A function on a space will be called measurable if it is measurable with respect to the given sigma-field. Product spaces are equipped with the product sigma-field, and in a topological space the sigma-field is generated by the open sets. Euclidean n-space is denoted R^n, and R is the real line.

If A is a subset of a space, A^c denotes the complement of A in that space. If S is a subset of a product space $\mathscr{X} \times \mathscr{Y}$, then for $y \in \mathscr{Y}$ the $\mathscr{X}$-section of S at y is the set $S_y = \{x \in \mathscr{X}: (x, y) \in S\}$. If X^n is a random variable taking values in $\mathscr{X}^n$ and U a measurable function on $\mathscr{X}^n$, then $U(X^n)$ will be called a *statistic*. Let, for $n = 1, 2, \cdots, X^n \in \mathscr{X}^n$ be a random variable on a probability space with probability measure P and A_n an event depending only on X^n, i.e., the inverse image under X^n of a measurable subset of $\mathscr{X}^n$. We shall say that the sequence $A_1, A_2, \cdots$ is *exponentially bounded* if there exists $c > 0$ and $\rho < 1$ such that $PA_n < c\rho^n$, $n = 1, 2, \cdots$. We shall often for short say that A_n is exponentially bounded. A sequence $Y_1, Y_2, \cdots$ of vector valued statistics will be said to *converge exponentially* to 0, in symbols $Y_n \to_{\exp} 0$, if for any $\varepsilon > 0$ the sequence of events $[||Y_n|| > \varepsilon]$, $n = 1, 2, \cdots$, is exponentially bounded, where $||\ ||$ denotes Euclidean norm. If X is a real random variable it is said to have a finite m.g.f. if $E \exp(tX) < \infty$ for t in some interval about 0.

For the purpose of this paper it may be assumed that the stopping bounds in (1.2) satisfy $-l_1 = l_2$, and the common value will be denoted $d/2$. Thus, throughout the remainder of this paper N will be defined by

$$(2.1) \qquad N \quad \text{is the smallest integer} \quad n \geqq 1 \quad \text{such that} \quad |L_n| \geqq d/2 .$$

Throughout this paper A_n will stand for the event

$$(2.2) \qquad A_n = [N > n] , \qquad n = 1, 2, \cdots .$$

Then the definition of the property that N is exponentially bounded given in

Section 1, and expressed by (1.1), can be given equivalently by saying that A_n is exponentially bounded.

Since the statement of Theorem 2.1 is somewhat involved it may be desirable to outline the main ideas in it. If it were possible to approximate $L_{n+r} - L_n$ by a sum $\sum_1^r f(X_{n+j})$, for every positive integer r, then L_n would behave very much like a random walk, and the Stein argument could be used to prove exponential boundedness of N under those P for which $f(X)$ is not degenerate at 0. In some examples such a function f exists, but in general it does not. What seems to be often possible is to approximate $L_{n+r} - L_n$ by a sum of the form $\sum_1^r f(X_{n+j}, U_n)$, where now f is a function of two arguments, and U_n, with values in a fixed space $\mathscr{U}$, is a statistic depending only on $X_1, \cdots, X_n$. Required now of f is not only that $f(X, u)$ be nondegenerate at 0 for every u, but uniformly so in a certain sense. To make that precise one may demand that for some $\delta > 0$ and for any u, $f(x, u) > \delta$ for all x or $< -\delta$ for all x. In general it is not possible to achieve this for all x and u, but only for x in a subset S_u of $\mathscr{X}$, and u in a subset T of $\mathscr{U}$. The argument still goes through if S_u has a positive probability, bounded away from 0 as u runs through T, and if $[U_n \notin T]$ is exponentially bounded. Even this is too strong an assumption for some applications. It suffices that there exist a sequence of events C_n such that $A_n C_n$ implies $[U_n \in T]$ and $A_n C_n^c$ is exponentially bounded.

THEOREM 2.1. *Let $X, X_1, X_2, \cdots$ be* i.i.d. *random variables with values in $\mathscr{X}$ and common distribution P. Let $L_n = L_n(X^n)$ be a sequence of real valued statistics and $U_n = U_n(X^n)$ a sequence of statistics with values in a space $\mathscr{U}$. Let N and A_n be defined by* (2.1) *and* (2.2). *Suppose there exists a measurable subset T of $\mathscr{U}$, a measurable function f: $\mathscr{X} \times T \to R$, a measurable subset S of $\mathscr{X} \times T$, real numbers $p > 0$, $\delta > 0$, $\varepsilon > 0$, and for each integer $r > 0$ a sequence C_n of events, with C_n depending only on X^n, such that the following conditions are satisfied*:

(i) $P(X \in S_u) > p \;\forall\, u \in T$;

(ii) *if $u \in T$, then either $f(x, u) > \delta \;\forall\, x \in S_u$ or $f(x, u) < -\delta \;\forall\, x \in S_u$*;

(iii) *$A_n C_n^c$ is exponentially bounded*;

(iv) *for $n = 1, 2, \cdots$, the event $A_n C_n$ implies the event $[U_n \in T]$ and implies the event*

$$[|L_{n+r}(X^n, y^r) - L_n(X^{n}) - \textstyle\sum_{j=1}^r f(y_j, U_n)| < \varepsilon] \tag{2.3}$$

whenever $y_j \in S_{U_n}$, $j = 1, \cdots, r$.

Then N is exponentially bounded.

PROOF. We have to show that A_n is exponentially bounded. Let d, δ and ε be as given in (2.1) and the hypotheses of the theorem. Take integer $r > 0$ such that

$$r > (d + \varepsilon)/\delta\,. \tag{2.4}$$

Then with this r let C_n be the sequence of events as given in the hypotheses.

Exponential boundedness of A_n follows from the Lemma in [16], by condition (iii) and the fact that A_n is a nonincreasing sequence, if we can show that there exists $p_1 > 0$ such that $P(A_{n+r}C_{n+r} \mid A_n C_n) < 1 - p_1$, $n = 1, 2, \cdots$. It will even be shown that $P(A_{n+r} \mid A_n C_n) < 1 - p_1$, or, equivalently, $P(A^c_{n+r} \mid A_n C_n) > p_1$. Now given A_n, the event A^c_{n+r} is implied by $[|L_{n+r} - L_n| > d]$. Thus, it suffices to prove

$$(2.5) \qquad P(|L_{n+r} - L_n| > d \mid A_n C_n) > p_1 , \qquad n = 1, 2, \cdots .$$

Let D_n denote the event

$$(2.6) \qquad D_n = [X_{n+j} \in S_{U_n}, j = 1, \cdots, r] .$$

If $U_n \in T$, then by (ii) and (2.4) D_n implies

$$(2.7) \qquad |\textstyle\sum_{j=1}^r f(X_{n+j}, U_n)| > d + \varepsilon .$$

Therefore, after replacing in (2.3) the y_j by X_{n+j}, we have that by (iv) and (2.7)

$$(2.8) \qquad A_n C_n D_n \quad \text{implies} \quad [|L_{n+r} - L_n| > d] .$$

Now since the X_{n+j} are independent of $A_n C_n$ and of U_n, and $A_n C_n$ implies $[U_n \in T]$ by (iv), it follows from (i) and (2.6) that

$$(2.9) \qquad P(D_n \mid A_n C_n) > p^r .$$

Then (2.8) and (2.9) together imply (2.5), with $p_1 = p^r$. □

COROLLARY 2.1. *Let $X, X_1, X_2, \cdots$ be* i.i.d. *random variables with values in $\mathscr{X}$ and common distribution P. Let $L_n = L_n(X^n)$ be a sequence of real valued statistics and let N and A_n be defined by* (2.1) *and* (2.2). *Suppose there exists a measurable function $f\colon \mathscr{X} \to R$, a measurable subset S of $\mathscr{X}$ and $\delta > 0$ such that $P(X \in S) > 0$ and $|f(x)| > \delta$ for every $x \in S$. Suppose that for some $\varepsilon > 0$ and every positive integer r there is a sequence of events C_n (depending on r) with*

$$C_n \subset [\sup \{|L_{n+r}(X^n, y^r) - L_n - \textstyle\sum_1^r f(y_j)| : y_1, \cdots, y_r \in S\} < \varepsilon]$$

such that $A_n C_n^c$ is exponentially bounded. Then N is exponentially bounded.

PROOF. Since $|f| > \delta$ on S and $P(X \in S) > 0$, there is a measurable subset S_1 of S and $p > 0$ such that $P(X \in S_1) > p$ and either $f > \delta$ or $f < -\delta$ on S_1. Then in the hypotheses of Theorem 2.1 the space $\mathscr{U}$ can be taken arbitrarily and f in that theorem can be chosen to be the function f in the hypotheses of the corollary, so that f does not depend on the second argument u. The set S in the hypotheses of Theorem 2.1 can be taken to be $S_1 \times \mathscr{U}$. The hypotheses of the corollary then imply those of Theorem 2.1. □

COROLLARY 2.2. *Let $X, X_1, X_2, \cdots$ be* i.i.d. *random variables with values in $\mathscr{X}$ and common distribution P. Let $L_n = L_n(X^n)$ be a sequence of real valued statistics and let N be defined by* (2.1). *Suppose there exists a measurable function $f\colon \mathscr{X} \to R$, a measurable subset S of $\mathscr{X}$ with $P(X \in S) > 0$ and $f(x) \neq 0$ whenever $x \in S$, and a* *set* $S_0 \supset S$, *such that for every positive integer r the function*

$$K_n = \sup\{|L_{n+r}(X^n, y^r) - L_n - \textstyle\sum_1^r f(y_j)| : y_1, \cdots, y_r \in S_0\} \tag{2.10}$$

is measurable. If $K_n \to_{\exp} 0$, *then N is exponentially bounded.*

Proof. Since $|f| > 0$ on S, there is $S_1 \subset S$ with $P(X \in S_1) > 0$ and $\delta > 0$ such that $|f| > \delta$ on S_1. This set S_1 can then be substituted for the set S in the hypotheses of Corollary 2.1. In those hypotheses $\varepsilon > 0$ can be taken arbitrarily and $C_n = [|K_n| < \varepsilon]$. Since by assumption $K_n \to_{\exp} 0$, we have that C_n^c is exponentially bounded, so a fortiori $A_n C_n^c$ is exponentially bounded. Thus, the hypotheses of Corollary 2.2 imply those of Corollary 2.1. □

The following theorem is technically a corollary of Theorem 2.1 since it reaches the same conclusion while making stronger assumptions. These assumptions are usually easier to verify in parametric applications than are those of Theorem 2.1. An explanation is in order of the statistic V_n that appears in the statement of the theorem. In parametric problems it is often the case that $L_{n+r}(X^n, y^r) - L_n$ depends on X^n through (U_n, V_n), with values in the fixed space $\mathscr{U} \times \mathscr{V}$, where (at least on A_n) V_n has an exponential limit v_0 and U_n lies eventually in the compact set T (in the exponential sense). Then for suitably chosen neighborhood G of v_0, C_n in Theorem 2.1 is taken as $[U_n \in T] \cap [V_n \in G]$.

Theorem 2.2. *Let* $X, X_1, X_2, \cdots$ *be* i.i.d. *random variables with values in a sigma-compact topological space* $\mathscr{X}$, *possessing common distribution P. For* $n = 1, 2, \cdots$ *let* $L_n = L_n(X^n)$, $U_n = U_n(X^n)$, *and* $V_n = V_n(X^n)$ *be statistics with values in* R, $\mathscr{U}$, *and* $\mathscr{V}$, *respectively, with* $\mathscr{U}$ *and* $\mathscr{V}$ *topological spaces. Let N and* A_n *be defined by* (2.1) *and* (2.2). *Suppose there exists a compact subset T of* $\mathscr{U}$, *a real valued continuous function f on* $\mathscr{X} \times T$, *a point* $v_0 \in \mathscr{V}$, *and for* $r = 1, 2, \cdots$ *a real valued measurable function* l_r *on* $\mathscr{X}^r \times \mathscr{U} \times \mathscr{V}$ *which is continuous at* (y^r, u, v_0) *for every* $y^r \in \mathscr{X}^r$, $u \in T$, *such that the following conditions are satisfied*:

(i) $A_n \cap [U_n \notin T]$ *is exponentially bounded*:
(ii) $A_n \cap [V_n \notin G]$ *is exponentially bounded* $\forall$ *neighborhood G of* v_0;
(iii) $L_{n+r} - L_n = l_r(X_{n+1}, \cdots, X_{n+r}, U_n, V_n)$ *on* $A_n \cap [U_n \in T]$;
(iv) $l_r(y^r, u, v_0) = \sum_1^r f(y_j, u)$, $y^r \in \mathscr{X}^r$, $u \in T$;
(v) $P\{f(X, u) = 0\} < 1$ $\forall$ $u \in T$.

Then N is exponentially bounded.

Proof. We shall show that the hypotheses of Theorem 2.2 imply those of Theorem 2.1. From (v) it follows that for every $u \in T$ there exists $\delta(u) > 0$, $p(u) > 0$, and $e(u)$ which equals either 1 or -1, such that

$$P\{e(u)f(X, u) \geq 2\delta(u)\} > 2p(u) . \tag{2.11}$$

Since $\mathscr{X}$ is sigma-compact, there exists a compact $\mathscr{X}_0 \subset \mathscr{X}$ with $P(X \in \mathscr{X}_0^c) < p(u)$. Define

$$S(u) = \mathscr{X}_0 \cap \{x \in \mathscr{X} : e(u)f(x, u) \geq 2\delta(u)\} , \tag{2.12}$$

then since f is continuous, $S(u)$ is compact. Furthermore,

$$P\{X \in S(u)\} > p(u) , \tag{2.13}$$

$$e(u)f(x, u) \geqq 2\delta(u) \quad \text{if} \quad x \in S(u) . \tag{2.14}$$

Let u_0 be an arbitrary point of T. Since f is continuous and $S(u_0)$ compact there exists a neighborhood $F(u_0)$ of u_0 such that

$$|f(x, u) - f(x, u_0)| < \delta(u_0) , \qquad x \in S(u_0),\ u \in F(u_0) . \tag{2.15}$$

Write (2.14) for $u = u_0$ and combine with (2.15) to obtain

$$e(u_0)f(x, u) > \delta(u_0) \quad \text{if} \quad x \in S(u_0) , \quad u \in F(u_0) . \tag{2.16}$$

Since T is compact it can be covered by a finite subcollection of the neighborhoods $F(u_0)$, $u_0 \in T$, say by $F(u_1), \cdots, F(u_m)$. Put

$$\delta = \min \{\delta(u_i) : i = 1, \cdots, m\} \tag{2.17}$$

$$p = \min \{p(u_i) : i = 1, \cdots, m\} . \tag{2.18}$$

Write (2.16) for $u_0 = u_1, \cdots, u_m$ and replace on the right-hand side the $\delta(u_i)$ by their minimum:

$$e(u_i)f(x, u) > \delta \quad \text{if} \quad x \in S(u_i) , \quad u \in F(u_i) , \quad i = 1, \cdots, m . \tag{2.19}$$

We may replace the $F(u_i)$ by disjoint, measurable sets $F'(u_i) \subset F(u_i)$ whose union is T (some of the $F'(u_i)$ may be empty). Define

$$S = \bigcup \{S(u_i) \times F'(u_i) : i = 1, \cdots, m\} . \tag{2.20}$$

If $(x, u) \in S$, then $x \in S(u_i)$, $u \in F'(u_i)$, for a unique i, $1 \leqq i \leqq m$, and for this u and i, $S_u = S(u_i)$. It follows then from (2.19) that (ii) of Theorem 2.1 is fulfilled. Furthermore, for the u and i of the preceding sentence, after writing (2.13) for $u = u_i$ and replacing $p(u_i)$ by the left-hand side of (2.18), we have $P(X \in S_u) = P\{X \in S(u_i)\} > p$, verifying (i) of the hypotheses of Theorem 2.1. Now put

$$S_0 = \bigcup \{S(u_i) : i = 1, \cdots, m\} , \tag{2.21}$$

then S_0 is a compact subset of $\mathscr{X}$. Since for every $u \in T$, S_u is one of the $S(u_i)$, it follows that

$$S_u \subset S_0 , \quad u \in T . \tag{2.22}$$

Thus, for every $u \in T$, $(x, u) \in S$ implies $x \in S_0$. Now $S_0 \times T$ is compact, and so is $S_0{}^r \times T$ as a subset of $\mathscr{X}^r \times \mathscr{U}$. By hypothesis l_r as a function on $S_0{}^r \times T \times \mathscr{V}$ is continuous at (y^r, u, v_0) for every $y^r \in S_0{}^r$, $u \in T$. Therefore, given $\varepsilon > 0$, there exists a neighborhood G of v_0 such that

$$|l_r(y^r, u, v) - l_r(y^r, u, v_0)| < \varepsilon , \qquad y^r \in S_0{}^r,\ u \in T,\ v \in G , \tag{2.23}$$

or, equivalently, by (iv)

$$|l_r(y^r, u, v) - \textstyle\sum_1^r f(y_j, u)| < \varepsilon , \qquad y_1, \cdots, y_r \in S_0,\ u \in T,\ v \in G . \tag{2.24}$$

Define

$$C_n = [U_n \in T] \cap [V_n \in G]\,. \tag{2.25}$$

Then $A_n C_n^c$ is exponentially bounded by (i) and (ii). Thus, (iii) of the hypotheses of Theorem 2.1 is verified.

It remains to verify (iv) of Theorem 2.1. It is immediate that $A_n C_n \Rightarrow C_n \Rightarrow [U_n \in T]$, by (2.25). It remains to be shown, using (iii), that

$$|l_r(y^r, U_n, V_n) - \textstyle\sum_1^r f(y_j, U_n)| < \varepsilon \qquad \text{if} \quad y_1, \cdots, y_r \in S_{U_n} \tag{2.26}$$

is implied by $A_n C_n$. Now $A_n C_n$ implies $[U_n \in T]$ and $[V_n \in G]$, by (2.25), and then (2.26) follows from (2.24) and (2.22). □

The following theorem is much easier to prove than Theorem 2.1 but is also much less powerful. However, it often is successful in some cases where Theorem 2.1 or its corollaries fail. The theorem has been used before in the literature, e.g. in [8].

THEOREM 2.3. *Suppose that N is defined by* (2.1) *and* $L_n/n \to_{\text{exp}} a \neq 0$; *then N is exponentially bounded.*

PROOF. Take $0 < \varepsilon < |a|/2$. Define $C_n = [|(L_n/n) - a| < \varepsilon]$ and A_n by (2.2). Then $A_n \subset A_n C_n \cup C_n^c$. In this union C_n^c is exponentially bounded by hypothesis. The event C_n implies $|L_n/n| > |a|/2$ so that if $n > d/|a|$, then $|L_n| > d/2$, implying A_n^c by (2.1) and (2.2). Hence $A_n C_n$ is empty for $n > d/|a|$. □

3. Parametric applications. In this section several examples are given of the use of Theorems 2.2 and 2.3, especially the former. Subsection 3.3 treats the sequential F-test and is a new result. The results in subsections 3.1 and 3.2 have appeared in [12] and [13]. Subsection 3.1 has been included because it is one of the simplest illustrations of Theorem 2.2, and subsection 3.2 because the proof of Proposition 3.2.1 using Theorem 2.2 is so much simpler than the proof in [12]. Moreover, this proposition is basic to other problems (see also Remark 2 following the proposition). Two more parametric examples of exponentially bounded N are known: the sequential t-test [16], and a generalization of two testing situations treated in [13], Sections 5 and 6, and in [15], Section 1. It will be very briefly indicated here how Theorem 2.2 is used to deal with those problems.

In the problem treated in [13] and [15], mentioned above, $X, X_1, X_2, \cdots$ are i.i.d. random vectors in R^k and $L_n = ||\sum_1^n X_i|| + a \sum_1^n ||X_i|| + bn + c \log(1 + ||\sum_1^n X_i||)$. After computing $L_{n+r} - L_n$, and denoting $S_n = \sum_1^n X_i$, one sees that the statistic U_n in Theorem 2.2 should be taken as $U_n = S_n/||S_n||$. Furthermore, $V_n = ||S_n||$, $v_0 = \infty$, and $f(y, u) = u'y + a||y|| + b$ (prime denotes transpose), in which $u, y \in R^k$, and $||u|| = 1$. (This f was also found in [15], equation (3.2).) Thus, N is exponentially bounded unless $P(u'X + a||X|| + b = 0) = 1$ for some $u \in R^k$ with $||u|| = 1$. Of the family of P's satisfying this condition certain subfamilies have been classified as obstructive or not in [14].

Exponential boundedness of N in the sequential t-test has been considered in various stages of generality by Berk [1], Section 4; Wijsman [12], Section 4, [14], Section 3(a); Lai [4], Section 3; and Wijsman [16]. Here the X's are real valued and $\mu/\sigma = \gamma_1$ is to be tested against $\mu/\sigma = \gamma_2$, with $\gamma_1 \neq \gamma_2$ given. The test is expressible in terms of the sequence $T_n = n^{-1} \sum_1^n X_i/(n^{-1} \sum_1^n X_i^2)^{\frac{1}{2}}$, and the log probability ratio can be replaced (following Lai [4]) by $L_n = n(T_n - \lambda)$, in which λ is a constant ($|\lambda| \leqq 1$) depending on γ_1 and γ_2, and $\lambda = 0$ if and only if $\gamma_1 = -\gamma_2$ (symmetric hypotheses). The distribution $P(X = 0) = 1$ is to be excluded, otherwise T_n would be undefined. After computing $L_{n+r} - L_n$ ([16], equations (20) and (23)), and denoting $S_n = \sum_1^n X_i^2$, it is found for the application of Theorem 2.2 that $U_n = S_n/n$, $V_n = (n, T_n, S_n)$, and $v_0 = (\infty, \lambda, \infty)$. Furthermore, $\mathscr{U} = [0, \infty]$, and for T one can take $[a, \infty]$ with any a such that $0 < a < EX^2$ (observe that $0 < EX^2 \leqq \infty$). Checking conditions (i) and (ii) proceeds as in subsection 3.3. Letting $n \to \infty$ one finds

$$(3.1) \qquad \begin{aligned} f(y, u) &= -(\lambda/2) - (\lambda/2)u^{-1}y^2 + u^{-\frac{1}{2}}y && \text{if} \quad u < \infty \\ &= -\lambda/2 && \text{if} \quad u = \infty\,. \end{aligned}$$

If $|\gamma_1| \neq |\gamma_2|$ (asymmetric hypotheses), then $\lambda \neq 0$ so that $P\{f(X, u) = 0\} = 1$ can happen only with $u < \infty$, and reads then $P(\lambda X^2 - 2u^{\frac{1}{2}}X + \lambda u = 0) = 1$. This is a two-point distribution so that X^2 has a finite m.g.f. An application of Theorem 2.3 then narrows the possibly obstructive P's down to the family of two-point distributions found in [12], equation (4.19). These distributions were shown in [14], Section 3(a), to be obstructive. On the other hand, if $\gamma_1 = -\gamma_2$ (symmetric hypotheses), then $\lambda = 0$ so that $f(y, u)$ could be $= 0$ simply by having $u = \infty$. That this can indeed happen was shown by Lai [4], Section 3, who demonstrated in the symmetric hypotheses case the existence of unbounded obstructive distributions. At present no complete description of all such obstructive distributions has been given.

3.1. *Sequential test about the variance in a normal population with unknown mean.* Under the model let $X, X_1, X_2, \cdots$ be i.i.d. $N(\mu, \sigma^2)$, μ, σ unknown, and the problem is to test sequentially one value of σ against another. Put $\bar{X}_n = (1/n) \sum_1^n X_i$, then the log probability ratio (apart from a nonzero multiplicative constant) is

$$(3.1.1) \qquad L_n = \textstyle\sum_1^n (X_i - \bar{X}_n)^2 - (n-1)a^2$$

(see [12], Section 3), in which $a > 0$ is a constant depending on the hypotheses to be tested. From (3.1.1) compute

$$(3.1.2) \qquad \begin{aligned} L_{n+r} - L_n = {} & \textstyle\sum_{j=1}^r X_{n+j}^2 - 2n(n+r)^{-1}\bar{X}_n \sum_{j=1}^r X_{n+j} \\ & - (n+r)^{-1}(\textstyle\sum_{j=1}^r X_{n+j})^2 + n(n+r)^{-1}r\bar{X}_n^2 - ra^2\,. \end{aligned}$$

From this expression one sees that for large n, $L_{n+r} - L_n$ ought to behave as $\sum_{j=1}^r (X_{n+j}^2 - 2\bar{X}_n X_{n+j} + \bar{X}_n^2 - a^2)$ and this suggests that one should try to apply

Theorem 2.2 with $U_n = \bar{X}_n$, $V_n = n$, and

$$f(y, u) = y^2 - 2uy + u^2 - a^2 \qquad \text{if } |u| < \infty\,, \tag{3.1.3}$$
$$= \infty \qquad \text{if } |u| = \infty\,.$$

Although the possible values of U_n are finite, it is convenient to add the points at $\pm\infty$, so that $\mathscr{U} = T$ is taken to be $[-\infty, \infty]$ with the usual topology that makes this a compact space. Take $\mathscr{V} = \{1, 2, \cdots, \infty\}$ and $v_0 = \infty$. Lastly, consulting (3.1.2), define

$$\begin{aligned} l_r(y^r, u, v) &= \textstyle\sum_1^r y_j^2 - 2n(n+r)^{-1}u \sum_1^r y_j \\ &\quad - (n+r)^{-1}(\textstyle\sum_1^r y_j)^2 + n(n+r)^{-1}ru^2 - ra^2 \\ &\qquad\qquad \text{if } |u| < \infty\,, \quad n < \infty\,, \\ l_r(y^r, u, \infty) &= \textstyle\sum_1^r y_j^2 - 2u \sum_1^r y_j + ru^2 - ra^2 \qquad \text{if } |u| < \infty\,, \\ l_r(y^r, \pm\infty, \infty) &= \infty\,. \end{aligned} \tag{3.1.4}$$

The continuity conditions in Theorem 2.2 can then easily be checked, conditions (i)—(iv) are immediate, and N is concluded to be exponentially bounded unless condition (v) is violated. That is, P can only be obstructive if $P\{f(X, u) = 0\} = 1$ for some $u \in T$, with f defined in (3.1.3). Obviously this cannot happen if $|u| = \infty$ so that P is obstructive only if

$$P\{|X - u| = a\} = 1 \qquad \text{for some} \quad u \in R\,. \tag{3.1.5}$$

This is a two-point distribution, therefore X^2 possesses a finite m.g.f. From this and Chernoff's Theorem 1 [2] it follows easily that $(1/n) \sum_1^n (X_i - \bar{X}_n)^2 \to_{\exp} \sigma^2$, where σ^2 is the variance of X under P. This, together with (3.1.1), implies that $L_n/n \to_{\exp} \sigma^2 - a^2$. By Theorem 2.3, N is exponentially bounded unless $\sigma^2 = a^2$. For P given by (3.1.5) this happens if and only if

$$P\{X = u \pm a\} = \tfrac{1}{2}\,, \qquad -\infty < u < \infty\,, \tag{3.1.6}$$

and each such P was shown in [13], Section 4, to be obstructive. Thus, in this example the family of obstructive distributions is completely characterized by (3.1.6).

3.2. *Sufficient conditions for exponential boundedness of N if $||X||$ has finite* m.g.f. The following proposition is essentially Theorem 2.1 in [12], but restated here in slightly different form. Its proof, using Theorem 2.2, is a great simplification over the proof of Theorem 2.1 in [12].

PROPOSITION 3.2.1. *Let $X, X_1, X_2, \cdots$ be* i.i.d. *random vectors with values in R^k (k a positive integer). Let the distribution P of X be such that $E\exp(t||X||) < \infty$ for some $t > 0$ and let $EX = \xi$. Let $\{L_n\}$ be a sequence of statistics and let N be defined by* (2.1). *Suppose there exists a neighborhood W of ξ and a real valued function Φ on W possessing a continuous gradient, such that*

$$|L_n - n\Phi(\bar{X}_n)| < B \qquad \textit{if} \quad \bar{X}_n \in W \quad (n = 1, 2, \cdots) \tag{3.2.1}$$

for some $B < \infty$, *where* $\bar{X}_n = (1/n)\sum_1^n X_i$. *Let* $\Delta = \operatorname{grad}\Phi$ *evaluated at* ξ. *Then* N *is exponentially bounded if*

$$P\{\Delta'(X-\xi)+\Phi(\xi)=0\}<1\,. \tag{3.2.2}$$

PROOF. By (3.2.1) it is permissible to pretend that $L_n = n\Phi(\bar{X}_n)$ whenever $\bar{X}_n \in W$. There is a measurable, real valued function g on $W \times W$, such that $g(x_1, x_2) \to 0$ as $(x_1, x_2) \to (\xi, \xi)$, and $\Phi(x_2) - \Phi(x_1) = \Delta'(x_2 - x_1) + g(x_1, x_2)\|x_2 - x_1\|$ if $x_1, x_2 \in W$. With the help of g one computes

$$\begin{aligned} L_{n+r} - L_n &= (n+r)[\Phi(\bar{X}_{n+r}) - \Phi(\bar{X}_n)] + r\Phi(\bar{X}_n) \\ &= \Delta'(\textstyle\sum_1^r X_{n+j} - r\bar{X}_n) + r\Phi(\bar{X}_n) \\ &\quad + \|\textstyle\sum_1^r X_{n+j} - r\bar{X}_n\| g(\bar{X}_n, (n+r)^{-1}(n\bar{X}_n + \sum_1^r X_{n+j}))\,, \end{aligned} \tag{3.2.3}$$

provided $\bar{X}_n, \bar{X}_{n+r} \in W$. Letting $n \to \infty$ and using $\bar{X}_n \to_{\exp} \xi$ so that $\Phi(\bar{X}_n) \to_{\exp} \Phi(\xi)$, it is seen that the statistic U_n and space $\mathscr{U}$ in Theorem 2.2 are not needed, that V_n can be taken $(n, \bar{X}_n)$, with $v_0 = (\infty, \xi)$, and that

$$f(y) = \Delta'(y-\xi) + \Phi(\xi)\,. \tag{3.2.4}$$

From (3.2.3) the function l_r in Theorem 2.2 follows easily, and the continuity conditions can be verified. By hypothesis, $P\{f(X) = 0\} < 1$, with f defined in (3.2.4), so that (v) of Theorem 2.2 is also verified. □

REMARKS. 1. If $\Phi(\xi) \neq 0$, then Φ needs only to be assumed continuous at ξ in order to reach the same conclusion. This follows from an application of Theorem 2.3 after observing that $\bar{X}_n \to_{\exp} \xi$ implies $L_n/n \to_{\exp} \Phi(\xi)$.

2. If (3.2.2) is not satisfied, i.e., $f(X) = \Delta'(X-\xi) + \Phi(\xi) = 0$ with probability one, then by taking expectation on both sides of this equation one sees that $\Phi(\xi) = 0$ and therefore $P\{\Delta'(X-\xi) = 0\} = 1$. If, in addition, Φ has continuous second partial derivatives in a neighborhood of ξ, then it was proved in [14] that P is obstructive. (For that result the only moment condition on P is that $E\|X\|^2$ be finite.) An application of this remark will be made in the next subsection.

3.3. *Sequential F-test for the general univariate linear hypothesis.* Let $X, X_1, X_2, \cdots$ be i.i.d. with values in R^k, $k \geqq 2$. Write $x_1, \cdots, x_k$ for the components of X, $x_{ij}, \cdots, x_{kj}$ for the components of X_j, $j = 1, 2, \cdots$. The model specifies $x_1, \cdots, x_k$ to be independently normal with common variance σ^2. Integers q, s are given, with $1 \leqq q \leqq s < k$. Put $Ex_i = \mu_i$ and $\gamma = \sum_1^q \mu_i^2/\sigma^2$. The problem in its canonical form is to test one value of γ against another, say γ_1 versus γ_2, with $\gamma_1 \neq \gamma_2$ given nonnegative numbers. It is convenient to introduce the following shortened notation:

$$\textstyle\sum_1 = \sum_{i=1}^q\,, \qquad \sum_2 = \sum_{i=q+1}^s\,, \qquad \sum_3 = \sum_{i=s+1}^k\,, \qquad \sum = \sum_{i=1}^k\,, \tag{3.3.1}$$

and $\bar{x}_{in} = (1/n)\sum_{j=1}^n x_{ij}$. In this notation the usual F-statistic at the nth sampling stage is in 1-1 correspondence with

$$Y_n = n\textstyle\sum_1 \bar{x}_{in}^2 / \sum_{j=1}^n \{\sum_1 x_{ij}^2 + \sum_2 (x_{ij} - \bar{x}_{in})^2 + \sum_3 x_{ij}^2\} \tag{3.3.2}$$

(see, e.g., [4], proof of Theorem 7). The sequential F-test is the sequential probability ratio test based on the sequence $\{Y_n\}$. Lai [4], Section 5, showed that for the study of asymptotic properties of the stopping time N it is permissible to pretend that the log probability ratio is

$$L_n = n(Y_n - \beta) + \nu \log n \,, \tag{3.3.3}$$

in which β and ν are constants depending on the model, and $0 < \beta < 1$. The constant ν depends also on (γ_1, γ_2), and $\nu = 0$ if $q = 1$ or if both γ_1 and γ_2 are > 0 (see [4]). Lai [4] proved termination with probability one and obtained asymptotic results on the moments of N.

In this subsection it will be shown that N is exponentially bounded under P unless

$$P\{\textstyle\sum_1 (x_i - \beta^{-1}\mu_i)^2 + \sum_2 (x_i - \mu_i)^2 + \sum_3 x_i^2 = (\beta^{-2} - \beta^{-1}) \sum_1 \mu_i^2\} = 1 \,. \tag{3.3.4}$$

Before doing that, a short discussion will be given about distributions satisfying (3.3.4). First of all it has to be ascertained that there are such distributions. It is convenient to put (3.3.4) in a different form. Let $S(b, r)$ be the $(k-1)$-sphere with center b and radius r; then (3.3.4) is equivalent to

$$P\{X \in S(b, r)\} = 1 \tag{3.3.5}$$

in which

$$\begin{aligned} b_i &= \beta^{-1}\mu_i \,, & i &= 1, \cdots, q \,, \\ &= \mu_i \,, & i &= q+1, \cdots, s \,, \\ &= 0 \,, & i &= s+1, \cdots, k \,, \end{aligned} \tag{3.3.6}$$

$$r^2 = (\beta^{-2} - \beta^{-1}) \textstyle\sum_1 \mu_i^2 \,. \tag{3.3.7}$$

Observe that the distribution (3.3.5) is bounded, so that $||X||^2$ has a finite m.g.f. Let $\mu = EX$ and let $B(b, r)$ be the closed k-ball with center b and radius r. Any P satisfying (3.3.5) obviously has $\mu \in B(b, r)$. Conversely, it is easy to see that given any point $y \in B(b, r)$ it is possible to find a distribution P supported on $S(b, r)$ that has $\mu = y$ (in fact, many such P's if y is in the interior of $B(b, r)$). Expressing the condition $\mu \in B(b, r)$, using (3.3.6) and (3.3.7), leads to

$$\textstyle\sum_3 \mu_i^2 \leqq (\beta^{-1} - 1) \sum_1 \mu_i^2 \,. \tag{3.3.8}$$

Thus, P satisfying (3.3.5) exists if and only if (3.3.8) is fulfilled.

The next question is whether every P satisfying (3.3.5) is obstructive. This is indeed true if $\nu = 0$ in (3.3.3), for then (3.2.1) is satisfied, with Φ having continuous second partial derivatives, and obstructiveness is concluded from Remark 2 following Proposition 3.2.1. If $\nu \neq 0$ (which is the more usual situation) the question whether P is obstructive reduces essentially to a problem of random walk between tilted and widening absorbing barriers. This has been solved recently by Portnoy [7], and by Lai and Wijsman [5]. The result is that

P is obstructive also in this case. Therefore, P is obstructive if and only if it satisfies (3.3.5).

By applying Theorem 2.2, exponential boundedness of N under P will now shown provided P does not satisfy (3.3.4). This excludes in particular those P for which $x_1, \cdots, x_q$ and $x_{s+1}, \cdots, x_k$ are degenerate at 0, and $x_{q+1}, \cdots, x_s$ are degenerate at any constants (such P would also leave Y_n in (3.3.2) undefined). The following simplifying notation will be employed:

$$S_i = \sum_{j=1}^n x_{ij}\,, \qquad Z = \sum_{j=1}^n \sum_{i=1}^k x_{ij}^2 \tag{3.3.9}$$

$$T_i = \sum_{j=1}^r x_{i,n+j}\,, \qquad W = \sum_{j=1}^r \sum_{i=1}^k x_{i,n+j}^2\,. \tag{3.3.10}$$

On the left-hand sides of (3.3.9) and (3.3.10) the dependence on n has been suppressed. In the notation (3.3.9) and (3.3.1), (3.3.2) can be written

$$Y_n = \textstyle\sum_1 S_i^2/(nZ - \sum_2 S_i^2)\,. \tag{3.3.11}$$

In (3.3.11), replace n by $n + r$ and use (3.3.10):

$$Y_{n+r} = \textstyle\sum_1 (S_i + T_i)^2/[(n+r)(Z+W) - \sum_2 (S_i + T_i)^2]\,. \tag{3.3.12}$$

From (3.3.3) compute $L_{n+r} - L_n = r(Y_n - \beta) + \nu \log(1 + rn^{-1}) + (n + r) \times (Y_{n+r} - Y_n)$, and in the last term substitute (3.3.11) and (3.3.12). The result is

$$\begin{aligned} L_{n+r} &- L_n \\ &= r(Y_n - \beta) + \nu \log(1 + rn^{-1}) \\ &\quad + [Z + W - (n+r)^{-1} \textstyle\sum_2 (S_i + T_i)^2]^{-1}\{2 \sum_1 S_i T_i \\ &\quad + \textstyle\sum_1 T_i^2 - Y_n[rZ + (n+r)W - 2\sum_2 S_i T_i - \sum_2 T_i^2]\}\,. \end{aligned} \tag{3.3.13}$$

From this expression it can be deduced how the statistics U_n, V_n, and the functions l_r, f in Theorem 2.2 ought to be taken. Define

$$U_n = (nZ^{-1}, S_1 Z^{-1}, \cdots, S_s Z^{-1}) \tag{3.3.14}$$

with values $u = (u_0, u_1, \cdots, u_s) \in R^{s+1}$, and

$$V_n = (n, Y_n, Z^{-1}) \tag{3.3.15}$$

with values $v = (v_1, v_2, v_3) \in \mathscr{V} = \{1, 2, \cdots, \infty\} \times [0, 1] \times [0, \infty]$ (observe that $v_1 = n$). The limit point v_0 in Theorem 2.2 is

$$v_0 = (\infty, \beta, 0)\,. \tag{3.3.16}$$

First condition (ii) in Theorem 2.2 will be checked. Taking a product neighborhood $G = G_1 \times G_2 \times G_3$ of v_0 given in (3.3.16), it is sufficient to verify that $A_n \cap [n \notin G_1]$, $A_n \cap [Y_n \notin G_2]$, and $A_n \cap [Z^{-1} \notin G_3]$ are all three exponentially bounded. The first and third of these assertions are even true without the intersection with A_n since $n \to \infty$ and $Z \to_{\exp} \infty$ by [11], using $P(\|X\|^2 = 0) < 1$. In the second, let $\varepsilon > 0$ be arbitrary and $G_2 = (\beta - \varepsilon, \beta + \varepsilon)$. Further, choose integer n_0 such that $n\varepsilon - |\nu| \log n \geqq d/2$ if $n > n_0$. Using (3.3.3), (2.1), and (2.2), if $n > n_0$ and $Y_n \notin G_2$, then $|L_n| \geqq d/2$ so that the event A_n does not happen. In other words, $A_n \cap [Y_n \notin G_2]$ is empty for $n > n_0$.

The space $\mathscr{U}$ and subset T in Theorem 2.2 need to be defined, and condition (i) checked. By (3.3.9) and an application of Schwarz' inequality it is found that $\sum_{i=1}^s (S_i Z^{-1})^2 \leqq nZ^{-1}$. Thus, the values u of U_n defined in (3.3.14) satisfy $\sum_1^s u_i^2 \leqq u_0$. Therefore, define $\mathscr{U} = \{u \in R^{s+1}\colon \sum_1^s u_i^2 \leqq u_0\}$. Denote $\tau^2 = E\|X\|^2$ and observe that $0 < \tau^2 \leqq \infty$. Choose any real number a such that $0 < a < \tau^2$. If the distribution of $\|X\|^2$ is bounded under P, then by Chernoff's Theorem 1 [2], $[Z_n/n < a]$ is exponentially bounded. This is easily seen to be true also for unbounded distribution, using a truncation argument. Define

$$T = \{u \in R^{s+1}\colon \textstyle\sum_1^s u_i^2 \leqq u_0 \leqq a^{-1}\}\,, \tag{3.3.17}$$

then T is a compact subset of $\mathscr{U}$ and $[U_n \notin T] = [Z_n/n < a]$ is exponentially bounded so that condition (i) is satisfied.

In order to write down the function l_r, in (3.3.13) U_n has to be replaced by u, V_n by v, and $X_{n+1}, \cdots, X_{n+r}$ by the (nonrandom) vectors $y_1, \cdots, y_r$. The latter substitution amounts to replacing T_i by $\sum_{j=1}^r y_{ij}$ $(i = 1, \cdots, s)$ and W by $\sum_{j=1}^r \|y_j\|^2$. Then letting $v \to v_0$ (defined in (3.3.16)) furnishes the function f:

$$\begin{aligned} f(y, u) = {} & (1 - u_0^{-1} \textstyle\sum_2 u_i^2)^{-1} \\ & \times [2 \textstyle\sum_1 u_i y_i - \beta(1 + u_0 \sum y_i^2 - 2 \sum_2 u_i y_i)]\,, \end{aligned} \tag{3.3.18}$$

in which now $y_1, \cdots, y_k$ are the components of the k-vector y. The continuity conditions on l_r and f, as well as (iv) of Theorem 2.2, are easily checked. Condition (5) is also fulfilled unless $P\{f(X, u) = 0\} = 1$, with f of (3.3.18), i.e., unless

$$P\{2 \textstyle\sum_1 u_i x_i - \beta - \beta u_0 \sum x_i^2 + 2\beta \sum_2 u_i x_i = 0\} = 1 \tag{3.3.19}$$

for some $u \in T$, with T defined in (3.3.17). In (3.3.19) u_0 cannot be $= 0$ since otherwise by (3.3.17) $u_1 = \cdots = u_s = 0$, whereas $\beta \neq 0$, contradicting (3.3.19). Hence (3.3.19) restricts X to a $(k - 1)$ sphere so that P is bounded. Therefore $\|X\|^2$ has finite m.g.f. and Chernoff's Theorem 1 [2] implies that U_n (defined in (3.3.14)) converges exponentially to $u^0 = (\tau^{-2}, \mu_1\tau^{-2}, \cdots, \mu_s\tau^{-2})$. For any $u \neq u^0$ there is a compact neighborhood T_0 of u^0 in $\mathscr{U}$ such that $u \notin T_0$. Since $[U_n \notin T_0]$ is exponentially bounded, one can replace T by T_0 in the application of Theorem 2.2. Then P can be obstructive only if (3.3.19) holds for some $u \in T_0$. Taking the intersection of all such T_0 it follows that P can be obstructive only if (3.3.19) holds with $u = u^0$, i.e.,

$$P\{\textstyle\sum x_i^2 - 2\beta^{-1} \sum_1 \mu_i x_i - 2 \sum_2 \mu_i x_i + \tau^2 = 0\} = 1\,. \tag{3.3.20}$$

Taking expectation on both sides of the equality inside the curly brackets in (3.3.20) shows $\tau^2 = \beta^{-1} \sum_1 \mu_i^2 + \sum_2 \mu_i^2$, and substitution of this into (3.3.20) yields (3.3.4).

4. Nonparametric applications. Tests depending on ranks can often conveniently be expressed in terms of empirical distribution functions. Usually the empirical distribution function F_n of a sample $X_1, \cdots, X_n$ is taken to be right continuous, i.e., $nF_n(z) = $ (number of X's $\leqq z$) $= \sum_1^n I_{(-\infty, z]}(X_i)$, where for any set A, I_A denotes its indicator. The theoretical distribution function F is then also right continuously defined. If possible occurrence of ties between the observations is dealt with by assigning midranks, then it may be more convenient to define both empirical and theoretical distribution functions in a way that constitutes an average between right and left continuity:

$$nF_n(z) = \tfrac{1}{2} \textstyle\sum_1^n [I_{(-\infty,z]}(X_i) + I_{(-\infty,z)}(X_i)] , \tag{4.1}$$

$$F(z) = \tfrac{1}{2}[P(X \leqq z) + P(X < z)] , \tag{4.2}$$

if X has distribution P and distribution function F.

In the sequel two results on exponential convergence of empirical distribution functions will be needed and are stated below as Lemmas 4.1 and 4.2. The first one follows from Dvoretzky, Kiefer and Wolfowitz [3], equation (2.6). It is also implied by Theorem 1 in Sethuraman [9], but that theorem is stronger than necessary for our purpose. Both lemmas are valid no matter how F_n and F are defined, as long as they are defined in the same manner. That is, both right continuous, or left continuous, or a fixed mixture of the two as in (4.1) and (4.2).

LEMMA 4.1. *Let* $X_1, X_2, \cdots$ *be* i.i.d. *real valued random variables with distribution function* F, *and let* F_n *be the empirical distribution function of* $X_1, \cdots, X_n$. *Then* $\sup\{|F_n(x) - F(x)| : x \in R\} \to_{\exp} 0$.

Now let $Y, Y_1, Y_2, \cdots$ be another sequence of i.i.d. real random variables with arbitrary common distribution. Since $F(Y)$ is a bounded random variable it follows from Chernoff's Theorem 1 [2] that $n^{-1} \sum_1^n F(Y_i) \to_{\exp} EF(Y)$. Combining this with Lemma 4.1 proves the next lemma.

LEMMA 4.2. *Let* $X_1, X_2, \cdots$ *be* i.i.d. *real random variables with distribution function* F, *let* F_n *be the empirical distribution of* $X_1, \cdots, X_n$, *and let* $Y, Y_1, Y_2, \cdots$ *be* i.i.d. *real random variables. If* $\{c_n\}$ *is a sequence of members such that* $nc_n \to 1$ *as* $n \to \infty$, *then* $c_n \sum_{i=1}^n F_n(Y_i) \to_{\exp} EF(Y)$.

A generalization of Lemma 4.1 will be needed in subsection 4.2. Suppose for definiteness that the F_n and F are right continuously defined so that $F_n(x) = (1/n) \sum_1^n I_{(-\infty,x]}(X_i)$ and $F(x) = \int I_{(-\infty,x]}(z)F(dz)$. The generalization consists in replacing $I_{(-\infty,x]}$ by a function $f(x, \cdot)$ that is monotonic in x.

LEMMA 4.3. *Let* $Z, Z_1, Z_2, \cdots$ *be a sequence of* i.i.d. *random variables taking values in a measurable space* $(\mathbf{Z}, \mathrm{A})$ *and let* P *be the distribution of* Z. *Furthermore, let* I *be a real interval (possibly infinite) and let* f *be a bounded, real valued function on* $I \times \mathbf{Z}$ *such that* $f(x, \cdot)$ *is* A*-measurable* $\forall\, x \in I$ *and* $f(\cdot, z)$ *is nondecreasing* $\forall\, z \in \mathbf{Z}$ *or nonincreasing* $\forall\, z \in \mathbf{Z}$. *Define*

$$G_n(x) = \frac{1}{n} \textstyle\sum_1^n f(x, Z_i)$$

$$G(x) = \int f(x, z)P(dz)$$

$$H_n = \sup\{|G_n(x) - G(x)| : x \in I\} .$$

Then $H_n \to_{\exp} 0$.

PROOF (the essence of the proof is contained in the union of [6], page 20, and [9], proof of Theorem 1). The random variable $f(x, Z)$ is bounded and has therefore a finite moment generating function. The same is true for $f(x+, Z)$ and $f(x-, Z)$. By Theorem 1 of Chernoff [2],

$$G_n(x) \to_{\exp} G(x) \qquad \text{for every} \quad x \in I\,, \tag{4.3}$$

and likewise with x replaced by $x+$, and by $x-$. Let $\varepsilon > 0$ be arbitrary. Since f is bounded and monotone (in the same direction for every z), the same is true for G. This is used to construct a finite set of points, say $x_1 < \cdots < x_m$, such that $|G(x_{j+1}-) - G(x_j+)| < \varepsilon/2$ for all j, i.e., G changes by less than $\varepsilon/2$ between any two successive x_j. Denote by $\{y_k : k \in K\}$ the set of "points" $\{x_j-, x_j, x_j+ : j = 1, \cdots, m\}$ and define the random variable

$$M_n = \max\{|G_n(y_k) - G(y_k)| : k \in K\}\,.$$

Since K is a finite set, (4.3) implies $M_n \to_{\exp} 0$, i.e.,

$$P(M_n > \varepsilon/2) < c\rho^n\,, \qquad n = 1, 2, \cdots \tag{4.4}$$

for some $c > 0$, $\rho < 1$. From the fact that G_n and G are monotonic in the same direction it can easily be established that $|G_n(x) - G(x)| < M_n + \varepsilon/2$ for every $x \in I$. Taking the sup over all $x \in I$ yields $H_n \leqq M_n + \varepsilon/2$. Then use (4.4) to obtain $P(H_n > \varepsilon) < c\rho^n$, $n = 1, 2, \cdots$. □

The next and final lemma will be needed in subsection 4.1 and is also of interest in its own right. As in Lemmas 4.1 and 4.2, the distribution functions F and G in Lemma 4.4 may be right or left continuous, or a mixture of these, as long as they are defined in the same way. The proof will be given in a way that is valid for any definition.

LEMMA 4.4. *Let X and Y be real valued random variables with joint distribution P and distribution function F, G, respectively. Then $P\{F(Y) = G(X)\} = 1$ if and only if $P(X = Y) = 1$.*

PROOF. The "if" part is of course trivial. To prove the "only if" part it will be shown first that $F = G$. The proof will be by contradiction. If $F(x) = G(x)$ for every x that is a continuity point of both F and G, then $F(x) = G(x)$ for all x. Suppose then, on the contrary, that there is a continuity point x of both F and G such that $F(x) \neq G(x)$. Without loss of generality it may be assumed that $F(x) > G(x)$. Then

$$F(x-) = F(x) = F(x+)\,, \tag{4.5}$$

$$G(x) < F(x)\,. \tag{4.6}$$

Using (4.6) and the first of the equalities (4.5) there is $y < x$ with

$$G(x) < F(y)\,. \tag{4.7}$$

The implication

$$[X \leqq x] \Rightarrow [G(X) \leqq G(x)] \tag{4.8}$$

is immediate from the monotonicity of G, and the implication

$$[F(Y) \leqq G(x)] \Rightarrow [Y < y] \tag{4.9}$$

follows from (4.7) and the monotonicity of F. By hypothesis, $F(Y)$ and $G(X)$ have the same distribution, which implies

$$P\{F(Y) \leqq G(x)\} = P\{G(X) \leqq G(x)\} . \tag{4.10}$$

There results the following string of inequalities:

$$\begin{aligned} F(x) &= P(X \leqq x) \quad \text{by (4.5)} \\ &\leqq P\{G(X) \leqq G(x)\} \quad \text{by (4.8)} \\ &= P\{F(Y) \leqq G(x)\} \quad \text{by (4.10)} \\ &\leqq P(Y < y) \quad \text{by (4.9)} \\ &\leqq P(Y < x) \quad \text{since} \quad y < x \\ &\leqq G(x) \\ &< F(x) \quad \text{by (4.6)} , \end{aligned}$$

which is a contradiction. Hence $F = G$, and replacing in the hypothesis G by F it follows that

$$P\{F(X) = F(Y)\} = 1 . \tag{4.11}$$

Now let A be the complement of the support of the distribution on R defined by the distribution function F; that is, $P(X \in A) = 0$, and A is the largest open subset of R with that property (A may be empty, of course). Clearly, A is a disjoint union of open intervals on each of which F is constant. The following characterization of A is useful:

(4.12) $x \in A$ if and only if there exists $y \in A$ with $y \neq x$ and $F(x) = F(y)$.

From (4.12) follows

$$\{(x, y) \in R^2 : x \neq y,\ F(x) = F(y)\} \subset A \times A . \tag{4.13}$$

Hence

$$\begin{aligned} P(X \neq Y) &= P\{X \neq Y, F(X) = F(Y)\} \quad \text{using (4.11)} \\ &\leqq P(X \in A, Y \in A) \quad \text{by (4.13)} \\ &\leqq P(X \in A) = 0 . \end{aligned}$$

Therefore, $P(X = Y) = 1$. □

4.1. *Sequential two-sample Wilcoxon test.* With a slight change from the general notation used in Section 2, let $(X, Y), (X_1, Y_1), \cdots$ be i.i.d. random vectors in R^2 and let F, G be the distribution functions of X, Y, respectively. The Wilcoxon test is based, at any sampling stage n, on the sum of the ranks of the Y's in the combined sample. It is not strictly a probability ratio test but can be considered the limit, as $\Delta \to 0$, of the probability ratio test based on the ranks of the Y's for testing the hypothesis $G = F$ against the alternative $G = F + \Delta(F^2 - F)$, or, alternatively, G is F shifted to the right by Δ if F is the logistic. (Also, it is to be understood that under the model, X and Y are independent.) Possible ties between the X's and Y's will be dealt with by assigning midranks. Consequently, definitions (4.1) and (4.2) are to be adopted for F_n and F, with similar definitions for G_n and G. As a consequence, the following useful equation is valid, no matter what ties occur:

$$\sum_1^n [F_n(X_i) + G_n(X_i) + F_n(Y_i) + G_n(Y_i)] = 2n \tag{4.1.1}$$

(for a proof consider $(F_n + G_n)/2$ as the empirical distribution function of the X's and Y's combined). The log probability ratio based on the ranks of the Y's at the nth stage (or rather its limit as $\Delta \to 0$) is

$$L_n = 2n(2n+1)^{-1} \sum_1^n [F_n(Y_i) + G_n(Y_i) + (2n)^{-1}] - n\,. \tag{4.1.2}$$

The test stops according to (2.1), and exponential boundedness of N is to be investigated with no assumptions on the distribution P of (X, Y) (in particular, it is not assumed that X and Y are independent).

Suppose X and Y were equal under P:

$$P(X = Y) = 1\,. \tag{4.1.3}$$

Then from (4.1.2) and (4.1.1) it follows that $P(L_n = 0, n = 1, 2, \cdots) = 1$ so that P is obstructive as was to be expected (worse: the test never stops). It will be shown now that the distributions (4.1.3) are the only obstructive ones. For this purpose L_n may be replaced by another statistic that differs from it by a uniformly bounded amount. It is convenient to replace (4.1.2) by

$$L_n = \sum_1^n [F_n(Y_i) + G_n(Y_i)] - n\,. \tag{4.1.4}$$

Corollary 2.2 will be applied now. In that corollary $\mathscr{X} = R^2$; due to our change in notation, X^n in the corollary is to be replaced by (X^n, Y^n) and y^r by (x^r, y^r). Furthermore, the set S_0 can be taken to be R^2. Next, $L_{n+r} - L_n$ has to be computed and the function f identified. Put $X_{n+j} = x_j$, $Y_{n+j} = y_j$, $j = 1, \cdots, r$, then from (4.1.4) one obtains

$$\begin{aligned} L_{n+r}(X^n, Y^n; x^r, y^r) - L_n \\ = -r + \sum_1^r [F_{n+r}(y_j) + G_{n+r}(y_j)] \\ + \sum_1^n [F_{n+r}(Y_i) - F_n(Y_i)] + \sum_1^n [G_{n+r}(Y_i) - G_n(Y_i)]\,. \end{aligned} \tag{4.1.5}$$

The third and fourth terms on the right-hand side in (4.1.5) are dealt with as follows. First (4.1) yields

$$(n + r)F_{n+r}(z) = nF_n(z) + \tfrac{1}{2} \sum_1^r [I_{(-\infty,z]}(x_j) + I_{(-\infty,z)}(x_j)] \tag{4.1.6}$$

and a similar equation with F replaced by G, x by y. Replace z by Y_i and sum over $i = 1, \cdots, n$ to obtain

$$(n + r) \sum_1^n F_{n+r}(Y_i) = n \sum_1^n F_n(Y_i) + nr - n \sum_1^r G_n(x_j)\,, \tag{4.1.7}$$

and a similar equation with F replaced by G, x by y. Substitution of (4.1.7) into (4.1.5) leads to

$$\begin{aligned} &L_{n+r}(X^n, Y^n; x^r, y^r) - L_n \\ (4.1.8) \qquad &= \sum_1^r [F_{n+r}(y_j) + G_{n+r}(y_j)] - n(n+r)^{-1} \sum_1^r [G_n(x_j) + G_n(y_j)] \\ &\quad - r(n+r)^{-1} \sum_1^n [F_n(Y_i) + G_n(Y_i)] + r(n-r)(n+r)^{-1} . \end{aligned}$$

By letting $n \to \infty$ the function f is discovered:

$$(4.1.9) \qquad f(x, y) = F(y) - G(x) - E[F(Y) + G(Y)] + 1 .$$

In the present notation the random quantity K_n defined in (2.10) reads

$$(4.1.10) \qquad K_n = \sup |L_{n+r}(X^n, Y^n; x^r, y^r) - L_n - \textstyle\sum_1^r f(x_j, y_j)| ,$$

where the sup is taken over all real values of x_j and y_j, $j = 1, \cdots, r$. After substituting (4.1.8) and (4.1.9) into (4.1.10) it is obvious that K_n is measurable. It remains to show that $K_n \to_{\exp} 0$. This follows from the four convergence relations below.

$$(4.1.11) \qquad \sup_{y \in R} |F_{n+r}(y) - F(y)| \to_{\exp} 0 ,$$

$$(4.1.12) \qquad \sup_{y \in R} |G_{n+r}(y) - n(n+r)^{-1}G_n(y)| \to_{\exp} 0 ,$$

$$(4.1.13) \qquad \sup_{x \in R} |n(n+r)^{-1}G_n(x) - G(x)| \to_{\exp} 0 ,$$

$$(4.1.14) \qquad (n+r)^{-1} \textstyle\sum_1^n [F_n(Y_i) + G_n(Y_i)] - E[F(Y) + G(Y)] \to_{\exp} 0 ,$$

all four for fixed r, as $n \to \infty$. Of these four, (4.1.11)—(4.1.13) follow from Lemma 4.1 (after adding a few trivial steps), and (4.1.14) follows from Lemma 4.2.

Since in the present situation S can be taken as any subset of R^2 which has positive probability and on which $f \neq 0$, the conclusion of Corollary 2.2 applies unless $P\{f(X, Y) = 0\} = 1$, i.e., unless

$$(4.1.15) \qquad P\{F(Y) - G(X) = E[F(Y) + G(Y)] - 1\} = 1 .$$

On the other hand, from (4.1.4) and an application of Lemma 4.2 it is seen that

$$(4.1.16) \qquad L_n/n \to_{\exp} E[F(Y) + G(Y)] - 1 ,$$

so that by Theorem 2.3 N is exponentially bounded unless the right-hand side of (4.1.16) equals 0. Combining this with (4.1.15) it follows that N is exponentially bounded unless

$$(4.1.17) \qquad P\{F(Y) - G(X) = 0\} = 1 .$$

By Lemma 4.4 (4.1.17) is equivalent to (4.1.3). Thus, P is obstructive if and only if (4.1.3) holds.

4.2. *Sequential rank-order test based on Lehmann alternatives.* Let (X, Y), (X_1, Y_1), $\cdots$ be i.i.d. random vectors, with X and Y real valued, having joint distribution P and marginal distribution functions F, G, respectively. It will simplify the exposition if all distribution functions are taken to be right continuous (but see Remark 1 at the end of this subsection). For the sequential probability ratio test based on ranks for testing $G = F$ against $G = F^A$ (and assuming in the model X and Y to be independent), for some $A > 0$ with $A \neq 1$, Savage and Sethuraman [8] obtained partial results for the exponential boundedness of N, and Sethuraman [10] obtained a nearly complete solution of the form: N is exponentially bounded unless $P\{f(X, Y) = 0\} = 1$. The function f in this statement is given in (4.2.12) below. It will be outlined in this subsection how Sethuraman's result follows rather easily from an application of Corollary 2.2, and how the function f is discovered in the course of that application.

Let F_n and G_n be the empirical distribution function of the X's and Y's respectively, at the nth stage. Following Sethuraman [10] define

$$W_n = F_n + AG_n \,, \tag{4.2.1}$$

$$W = F + AG \,, \tag{4.2.2}$$

$$W(x, y; z) = I_{(-\infty,z]}(x) + AI_{(-\infty,z]}(y) \,. \tag{4.2.3}$$

From [10], equation 13, it is seen that

$$L_n = \textstyle\sum_1^n [-\log W_n(X_i)W_n(Y_i) + \log 4A - 2] + \frac{1}{2}\log n \tag{4.2.4}$$

differs from the log probability ratio by a uniformly bounded amount. Writing (4.2.4) down with n replaced by $n + r$, setting $X_{n+j} = x_j$, $Y_{n+j} = y_j$, $j = 1, \cdots, r$, and taking the difference with (4.2.4) yields

$$\begin{aligned} L_{n+r}&(X^n, Y^n; x^r, y^r) - L_n \\ &= \textstyle\sum_1^r [-\log W_{n+r}(x_j)W_{n+r}(y_j)] + r(\log 4A - 2) \\ &\quad + \textstyle\sum_1^n [\log W_n(X_i)W_n(Y_i) - \log W_{n+r}(X_i)W_{n+r}(Y_i)] \\ &\quad + \tfrac{1}{2}\log(1 + rn^{-1}) \,. \end{aligned} \tag{4.2.5}$$

The last term on the right-hand side in (4.2.5) converges to 0 as $n \to \infty$ so that it may be ignored when verifying the exponential convergence of K_n to 0 in (2.10) (as in subsection 4.1, in (2.10) X^n is to be replaced by (X^n, Y^n), y^r by (x^r, y^r)).

In order to treat the first term on the right-hand side in (4.2.5), it should be realized that although Lemma 4.1 implies $\sup|W_{n+r}(x) - W(x)| \to_{\text{exp}} 0$ (sup over all real x), it cannot be concluded that $\sup|\log W_{n+r} - \log W(x)| \to_{\text{exp}} 0$ since $W(x)$ is not bounded away from 0 (and, in fact, could equal 0 for some x). This points to the fact that the set S_0 in Corollary 2.2 cannot be taken as the whole of R^2. In order to find a subset that can serve as S_0 first observe that $P\{W(\min(X, Y)) > 0\} = 1$. As a result, if $P\{f(X, Y) \neq 0\} > 0$, then also $P\{f(X, Y) \neq 0, W(\min(X, Y)) > 0\} > 0$. Hence there exists $w_0 > 0$ such that $P\{f(X, Y) \neq 0, W(\min(X, Y)) \geqq w_0\} > 0$. Define z_0 by $z_0 = \inf\{z \in R : W(z) \geqq w_0\}$, then $W(z) \geqq w_0$ if and only if $z \geqq z_0$. Now take $S_0 = \{(x, y) \in R^2 : \min(x, y) \geqq z_0\}$ and $S = \{(x, y) \in S_0 : f(x, y) \neq 0\}$, then $S \subset S_0$, $P\{(X, Y) \in S\} > 0$, and $f \neq 0$ on S, as required by the hypothesis of Corollary 2.2. In (2.10) the supremum is now over $x_j \geqq z_0$, $y_j \geqq z_0$, $j = 1, \cdots, r$. As a result, $W(x_j)$ and $W(y_j)$ are bounded below by $w_0 > 0$; moreover, by (4.2.2) they are bounded above by $1 + A$. This, together with Lemma 4.1, shows that

$$\sup \{|\log W_{n+r}(x_j)W_{n+r}(y_j) - \log W(x_j)W(y_j)|\} \to_{\exp} 0 . \tag{4.2.6}$$

In order to treat the third term on the right-hand side in (4.2.5), first observe that $(n + r)F_{n+r}(z) = nF_n(z) + \sum_1^r I_{(-\infty,z]}(X_{n+j})$, and similarly with F replaced by G, X by Y. Now put $X_{n+j} = x_j$, $Y_{n+j} = y_j$ and use (4.2.1) and (4.2.3) to obtain

$$(n + r)W_{n+r}(z) = nW_n(z) + \sum_1^r W(x_j, y_j; z) . \tag{4.2.7}$$

Replacing in (4.2.7) z by X_i, taking log, and summing over i from 1 to n gives

$$\begin{aligned} &\sum_1^n [\log W_{n+r}(X_i) - \log W_n(X_i)] \\ &\quad = n \log n(n + r)^{-1} \\ &\qquad + \sum_1^n \log [1 + (nW_n(X_i))^{-1} \textstyle\sum_{j=1}^r W(x_j, y_j; X_i)] . \end{aligned} \tag{4.2.8}$$

The first term on the right-hand side in (4.2.8) converges to $-r$ as $n \to \infty$. In the spirit of approximating $\log (1 + \varepsilon)$ by ε if ε is small, and W_n by W, it may be expected that the second term on the right-hand side in (4.2.8) may be replaced by

$$\sum_1^n (nW(X_i))^{-1} \sum_{j=1}^r W(x_j, y_j; X_i) \tag{4.2.9}$$

in the sense that the supremum (over $x_j, y_j \geqq z_0$) of the absolute difference of the two expressions converges exponentially to 0. This is indeed not hard to prove (Lemma 4.1 is used again). Consulting (4.2.3) it is seen that (4.2.9) contains expressions of the form $n^{-1} \sum_1^n (W(X_i))^{-1}I_{(-\infty,X_i]}(x)$, with x being an x_j or y_j. Its exponential convergence is governed by

$$\begin{aligned} \sup \{|n^{-1} &\textstyle\sum_1^n (W(X_i))^{-1}I_{(-\infty,X_i]}(x) \\ &- \int (W(z))^{-1}I_{(-\infty,z]}(x)F(dz)| : x \geqq z_0\} \to_{\exp} 0 . \end{aligned} \tag{4.2.10}$$

This follows from Lemma 4.3 by taking in that lemma $f(x, z) = (W(z))^{-1}I_{(-\infty,z]}(x)$, which is bounded for $z \geqq z_0$ and nonincreasing as a function of x for each z. Therefore, (4.2.9) may be replaced by

$$\sum_{j=1}^r \int (W(z))^{-1}W(x_j, y_j; z)F(dz) . \tag{4.2.11}$$

In the third term on the right-hand side in (4.2.5), the terms involving the Y_i are treated similarly, and the only change in (4.2.11) will be to replace F by G. Putting everything together it is seen that

$$\begin{aligned} f(x, y) = &-\log W(x)W(y) + \log 4A \\ &- \int (W(z))^{-1}W(x, y; z)(F(dz) + G(dz)) , \end{aligned} \tag{4.2.12}$$

in agreement with [8] and [10]. According to the conclusion of Corollary 2.2 N is exponentially bounded under P unless $P\{f(X, Y) = 0\} = 1$.

REMARKS. 1. Ties between the X's and Y's can be dealt with by assigning midranks, as in subsection 4.1. Thus, the empirical and theoretical distribution functions should have been defined in the manner of (4.1) and (4.2) in order to cover arbitrary F and G. This amounts to replacing indicators of the form $I_{(-\infty,z]}$ by $\frac{1}{2}(I_{(-\infty,z]} + I_{(-\infty,z)})$. If care is taken to choose $w_0 > 0$ in such a way that z_0 is a continuity point of W, then it is still true that $W(z) \geqq w_0$ if and only if $z \geqq z_0$. It is easy to verify that with this change in definition of the distribution functions all equations remain unchanged, and the conclusion is therefore also the same. The proof could have been given in that way but would have been slightly more cumbersome, although essentially the same.

2. In this example the conclusion does not say whether distributions P for which $P\{f(X, Y) = 0\} = 1$ are obstructive, or whether such distributions even exist. Positive results on these questions have been obtained, but will be presented elsewhere [17].

Acknowledgment. I would like to thank the Associate Editor and the referee for valuable suggestions concerning the presentation of this paper.

REFERENCES

[1] BERK, R. H. (1970). Stopping times of SPRTS based on exchangeable models. *Ann. Math. Statist.* **41** 979–990.

[2] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[3] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[4] LAI, TZE LEUNG (1975). Termination, moments and exponential boundedness of the stopping rule for certain invariant sequential probability ratio tests. *Ann. Statist.* **3** 581–598.

[5] LAI, T. L. and WIJSMAN, R. A. (1976). On the first exit time of a random walk from the stopping bounds $f(n) \pm cg(n)$, with application to obstructive distributions in sequential tests. *IMS Bull.* **5** 259.

[6] LOÈVE, MICHEL (1963). ***Probability Theory***, 3rd ed. Van Nostrand, Princeton.

[7] PORTNOY, STEPHEN L. (1976). Probability bounds on first exits through tilted moving boundaries. *IMS Bull.* **5** 209.

[8] SAVAGE, I. RICHARD and SETHURAMAN, J. (1966). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.

[9] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316. (Corrections (1970), *Ann. Math. Statist.* **41** 1376–1380.)

[10] SETHURAMAN, J. (1970). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives II. *Ann. Math. Statist.* **41** 1322–1333.

[11] STEIN, C. (1946). A note on cumulative sums. *Ann. Math. Statist.* **17** 498–499.

[12] WIJSMAN, R. A. (1971). Exponentially bounded stopping time of sequential probability ratio tests for composite hypotheses. *Ann. Math. Statist.* **42** 1859–1869.

[13] WIJSMAN, R. A. (1972). Examples of exponentially bounded stopping time of invariant sequential probability ratio tests when the model may be false. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 109–128, Univ. of California Press.

[14] WIJSMAN, R. A. (1972). A theorem on obstructive distributions. *Ann. Math. Statist.* **43** 1709–1715.

[15] WIJSMAN, R. A. (1973). Stopping time in sequential samples from multivariate exponential families. *Proc. Third Internat. Symp. Multivariate Anal.* 171–179, Academic Press, New York.

[16] WIJSMAN, R. A. (1975). Exponentially bounded stopping time of the sequential t-test. *Ann. Statist.* **3** 1006–1010.

[17] WIJSMAN, R. A. (1977). Obstructive distributions in a sequential rank-order test based on Lehmann alternatives. *Statistical Decision Theory and Related Topics* II, Purdue Univ., Academic Press, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801