The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 979–990

STOPPING TIMES OF SPRTS BASED ON EXCHANGEABLE MODELS[1]

BY ROBERT H. BERK[2]

Columbia University

0. Summary. Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be a stochastic sequence and $\mathscr{P}$ and $\mathscr{Q}$, two composite parametric hypotheses (models) under which the $\mathbf{X}_i$ are i.i.d. We consider SPRTs of $\mathscr{P}$ vs $\mathscr{Q}$ that depend on a sequence of exchangeable densities. Included are SPRTs obtained by the method of weight-functions (Bayesian procedures) and many SPRTs obtained by invariance reduction. Conditions are established under which the stopping time of such a procedure is almost surely finite and has a nontrivial mgf.

The ideas are illustrated using the sequential t-test.

1. Introduction. Let $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \cdots$ be a sequence of abstract random variables and $\mathscr{P}$ and $\mathscr{Q}$, two hypotheses under which the random variables are i.i.d. (independent and identically distributed). We consider SPRTs of $\mathscr{P}$ vs $\mathscr{Q}$ that depend on sequences of exchangeable densities. As discussed below, these can arise in two conceptually distinct ways. Both $\mathscr{P}$ and $\mathscr{Q}$ are dominated parametric families indexed by a parameter θ which ranges in the parameter space Θ. (The parameter spaces for $\mathscr{P}$ and $\mathscr{Q}$ can differ, but for convenience, we denote both by Θ.) We denote the generic density in $\mathscr{P}$ (respectively $\mathscr{Q}$) for $\mathbf{X}$ by $p(\cdot \mid \theta)$ (respectively $q(\cdot \mid \theta)$), taken with respect to some dominating σ-finite measure.

A general prescription for obtaining SPRTs, entailing exchangeable densities, is Wald's method of weight-functions. Let $\mathscr{A}$ be a σ-field of subsets of Θ and P and Q, two measures (weight-functions) on $(\Theta, \mathscr{A})$. Wald (1947) suggested the following type of sequential procedure. At stage n, consider

$$\mathbf{L}_n = p_n(\mathbf{X}_1, \cdots, \mathbf{X}_n)/q_n(\mathbf{X}_1, \cdots, \mathbf{X}_n), \tag{1.1}$$

where

$$p_n(x_1, \cdots, x_n) = \int_\Theta \textstyle\prod_1^n p(x_i \mid \theta)\, dP(\theta) \tag{1.2}$$

and q_n is similarly defined with (q, Q) replacing (p, P). One terminates as soon as $\mathbf{L}_n$ leaves an interval $(A, B) \subset (0, \infty)$. Such a procedure can arise in a Bayesian context if P and Q are priors on Θ. Then but for the initial odds factor, $\mathbf{L}_n$ is the posterior odds favoring $\mathscr{P}$. Hence the procedure amounts to terminating as soon as the posterior odds favoring either hypothesis becomes sufficiently large. The salient fact about the above procedure is that composite hypotheses, under which the data are i.i.d., are replaced by simple hypotheses, under which they are

---

Received August 20, 1968; revised December 10, 1969.

[1] Work supported in part by NIH grants 5F1-MF11, 733–03 and MH14181-01 and NSF grants GP-6008 and GP-7350.

[2] On leave from the University of Michigan.

exchangeable. (For when P and Q are proper (i.e., probability measures), $\{p_n\}$ and $\{q_n\}$ are consistent families of exchangeable densities (in the sense of Kolmogorov). Even when P and/or Q is improper, we may say that $\{\mathbf{L}_n\}$ is a sequence of exchangeable likelihood ratios.)

By assuming an associated group structure for $\mathscr{P}$ and $\mathscr{Q}$ another method of obtaining simple hypotheses (and hence SPRTs) arises: invariance-reduction. (See Lehmann (1959) for the facts about invariance mentioned here.) Let $(\mathscr{X}, \mathscr{B})$ denote the common sample space of the $\mathbf{X}_i$. Suppose both $\mathscr{P}$ and $\mathscr{Q}$ are generated by the same group G. That is, G is a group of one-one bimeasurable transformations of $\mathscr{X}$ onto itself and $\mathscr{P} = \{P_0 g^{-1} : g \in G\}$ (respectively $\mathscr{Q} = \{Q_0 g^{-1} : g \in G\}$), where the choice of $P_0 \in \mathscr{P}$ (respectively $Q_0 \in \mathscr{Q}$) is arbitrary. It then follows that $\mathscr{P}$ and $\mathscr{Q}$ are simple hypotheses about any G-invariant functions of $\mathbf{X}_1, \mathbf{X}_2, \cdots$; in particular, about $\{\mathbf{M}_n\}$, where $\mathbf{M}_n = M_n(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ is a maximally G-invariant function of $\mathbf{X}_1, \cdots, \mathbf{X}_n$. We may thus establish a sequential test based on $\{\mathbf{M}_n\}$ by considering at stage n, $\mathbf{L}_n$, the likelihood ratio of $(\mathbf{M}_1, \cdots, \mathbf{M}_n)$ under $\mathscr{P}$ and $\mathscr{Q}$. The procedure terminates the first time $\mathbf{L}_n$ leaves (A, B). (The arbitrary choice of weight-function is thus replaced by a data reduction.) It should be noted that when $i < n$, $\mathbf{M}_i$ is an invariant function of $\mathbf{X}_1, \cdots, \mathbf{X}_n$, so that $(\mathbf{M}_1, \cdots, \mathbf{M}_{n-1})$ is a function of the maximally invariant $\mathbf{M}_n$. Hence

$$\mathbf{L}_n = p_n(\mathbf{M}_n)/q_n(\mathbf{M}_n), \tag{1.3}$$

where p_n and q_n denote densities for $\mathbf{M}_n$ under $\mathscr{P}$ and $\mathscr{Q}$ respectively. Under a fairly mild condition on G, we may relate (1.1) and (1.3). Specifically, $\mathbf{M}_n$ may be identified with an exchangeable sequence derived from $\mathbf{X}_1, \mathbf{X}_2, \cdots$.

THEOREM 1.1. *Suppose there is a positive integer r and a $\mathscr{P} \cup \mathscr{Q}$-null subset, N, of $\mathscr{X}^r$, so that every g in G is determined by its values at the r coordinates of any point of $\mathscr{X}^r - N$. Then, letting $M(x_1, \cdots, x_{r+1})$ denote a maximally G-invariant function of $x_1, \cdots, x_{r+1}$, a maximally G-invariant function of $x_1, \cdots, x_n$, $n > r$, is $M_n(x_1, \cdots, x_n) = (M(x_1, \cdots, x_r, x_{r+1}), \cdots, M(x_1, \cdots, x_r, x_n))$.*

REMARK. In a typical application of this theorem, $\mathscr{X}$ is a vector space and N is the set of points with linearly dependent coordinates.

PROOF. It is clear that M_n is invariant; we show it is maximal. Suppose $(x_1, \cdots, x_n)$ and $(x_1', \cdots, x_n')$ represent two possible outcomes of $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ for which $M_n(x_1, \cdots, x_n) = M_n(x_1', \cdots, x_n')$. Then $M(x_1, \cdots, x_r, x_i) = M(x_1', \cdots, x_r', x_i')$, $i = r+1, \cdots, n$, and since M is maximal, there exist $g_i \in G$ so that $(g_i x_1, \cdots, g_i x_r, g_i x_i) = (x_1', \cdots, x_r', x_i')$, $i = r+1, \cdots, n$. We may assume that $(x_1, \cdots, x_r)$ and $(x_1', \cdots, x_r') \in \mathscr{X}^r - N$; then as $(g_i x_1, \cdots, g_i x_r) = (x_1', \cdots, x_r')$ for all i, $g_{r+1} = \cdots = g_n = g$ (say). Thus $(x_1', \cdots, x_n') = (gx_1, \cdots, gx_n)$, showing that M_n is maximal. □

Letting $\mathbf{Y}_i = M(\mathbf{X}_1, \cdots, \mathbf{X}_r, \mathbf{X}_{r+i})$, $\mathbf{M}_{r+n} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ and the sequential procedure defined by (1.3) is seen to depend on the likelihood ratios for the successive initial segments of the exchangeable sequence $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$.

The following SPRT, the sequential t-test, illustrates invariance-reduction:

Suppose that under P_θ, the $\mathbf{X}_i$ are $N(\delta\theta, \theta^2)$, while under Q_θ, they are $N(\delta'\theta, \theta^2)$, where $0 \leqq \delta' < \delta$ and $\Theta = (-\infty, 0)\cup(0, \infty)$ (or $(0, \infty)$ if $\delta' = 0$). (Thus the hypotheses are $|\mu/\sigma| = \delta$ or δ'.) Both hypotheses are invariant under changes of scale ($\mathbf{X} \to c\mathbf{X}$, $c \neq 0$) and are generated by these transformations. The condition in Theorem 1.1 with $r = 1$ is easily seen to hold provided the origin is removed from $\mathscr{X} = (-\infty, \infty)$. As x_2/x_1 is a maximally invariant function of (x_1, x_2), we see that $(\mathbf{Y}_1, \mathbf{Y}_2, \cdots) = (\mathbf{X}_2/\mathbf{X}_1, \mathbf{X}_3/\mathbf{X}_1, \cdots)$ is a maximally invariant exchangeable sequence guaranteed by the theorem. Thus the hypotheses may be tested sequentially by considering, at stage $n+1$, the likelihood ratio of $\mathbf{M}_{n+1} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ under the hypotheses. (In fact, a further simplification is possible. As $\overline{\mathbf{X}}_n = \sum_1^n \mathbf{X}_i/n$ and $S_n{}^2 = \sum_1^n (\mathbf{X}_i - \overline{\mathbf{X}}_n)^2$ are jointly sufficient for $\mathbf{X}_1, \cdots, \mathbf{X}_n$ (under $\mathscr{P}\cup\mathscr{Q}$), the theory in [4] shows that a maximally invariant function of $(\overline{\mathbf{X}}_n, \mathbf{S}_n)$ is sufficient for $\mathbf{M}_n$. This *sufficient invariant* may be taken to be $|\mathbf{T}_n| = |\overline{\mathbf{X}}_n/\mathbf{S}_n|$; hence the procedure reduces to considering a sequence of likelihood ratios for successive t-statistics.) It is readily checked that other normal theory SPRTs based on invariance-reduction also have this exchangeable structure. These include the sequential χ^2, F and T^2-tests. In these cases, groups of linear transformations are involved and these all satisfy the hypothesis of Theorem 1.1.

The foregoing conceptual link between invariance-reduction and weight-function likelihood ratios is an instance of deFinetti's representation theorem for exchangeable distributions. The theorem, as extended by Hewitt and Savage (1955) shows that in most cases of interest, an exchangeable distribution for a data sequence is a mixture of i.i.d. distributions and is then said to be presentable. In the present instance, conditioned on $\mathbf{X}_1, \cdots, \mathbf{X}_r$, $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$ are i.i.d. and one obtains their marginal exchangeable distribution by mixing with respect to the marginal distribution of $\mathbf{X}_1, \cdots, \mathbf{X}_r$. Thus for the t-test, one mixes the conditional i.i.d. normal distributions, given $\mathbf{X}_1$, of $\mathbf{X}_2/\mathbf{X}_1, \mathbf{X}_3/\mathbf{X}_1, \cdots$. There is another conceptual link of use in the present context. Unlike the preceding, it often involves improper mixing measures. It is well known that the likelihood ratios for certain invariant SPRTs can be obtained by mixing the corresponding joint densities for $\mathbf{X}_1, \mathbf{X}_2, \cdots$ with appropriate measures. This is shown implicitly for the sequential t-test by Wald (1947) and Barnard (1952). The reason for this is expounded by Wijsman (1967a), who shows under general conditions that the probability density for a maximally invariant function of $\mathbf{X}_1, \cdots, \mathbf{X}_n$ is proportional to

$$\int_G \prod_1^n p(\mathbf{X}_i \mid g\theta)\, d\nu(g), \tag{1.4}$$

where θ is a reference point in Θ and ν is absolutely continuous with respect to Haar measure on G. (The omitted factors involve constants and Jacobians, but nothing depending on $\mathscr{P}$.) Typically, ν is improper. Unlike the representation given by Theorem 1.1 and (1.3), (1.4) shows $\mathbf{L}_n$ to be symmetric in $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Thus in treating invariance-reduction SPRTs, one often has the option of using the representation (1.2) for the i.i.d. sequence $\mathbf{X}_1, \mathbf{X}_2, \cdots$ or the exchangeable sequence $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$, with P possibly improper.

In Section 2 we consider the termination of the SPRTs discussed above. Section

3 gives (stronger) conditions that guarantee the existence of a non-trivial mgf for the stopping time. Section 4 illustrates the general theory with the sequential t-test and Section 5 presents some further examples to indicate the limitations of the theory discussed here.

2. Termination. We consider the almost sure termination of SPRTs based on exchangeable likelihood ratios, as given by (1.1) and (1.2). Many parametric SPRTs that have been proposed are of this ilk. Unless specified otherwise, we assume that the data sequence, $\mathbf{X}_1, \mathbf{X}_2, \cdots$ is i.i.d. F will denote the true (joint) distribution of the data sequence. For invariance-reduction SPRTs, we may also consider the reduced data sequence, $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$, whose exchangeable joint distribution will be denoted by F_I. Clearly F_I is just the restriction of F to the σ-field generated by $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$. Expectations, if not otherwise specified, are under F.

Termination of parametric SPRTs has been considered by many writers. Some have dealt with specific procedures, others with a class of procedures. Among the former, we mention the termination proofs given by David and Kruskal (1956), Jackson and Bradley (1961) and Ray (1957). Among the latter, Wald (1947) provided a general result when $\mathscr{P}$ and $\mathscr{Q}$ are simple, showing termination under any i.i.d. F for which the successive increments to the log likelihood ratio do not almost surely vanish. Wirjosudirdjo (1961) considered SPRTs depending on a real (sufficient) statistic and real parameter and established termination when the model obtains. Ifram (1965) obtained improved conditions, again assuming F to be in the model. Wijsman (1967b) showed that SPRTs based on multivariate normal models terminate for a large class of F. The present development sets forth a general approach to the parametric termination problem and gives unified results for many of the SPRTs that have been proposed. The methods used here differ from those of previous writers, although the results, when they overlap, are comparable. (Establishing termination for the sequential t-test, whether by the methods of David and Kruskal, Wijsman or the present writer, seems to entail the assumption $E\mathbf{X}^2 < \infty$.) In the work of Ifram, Wijsman and Wirjosudirdjo, as well as in the present effort, termination is obtained as a consequence of certain results about the asymptotic behavior of $\mathbf{L}_n$. (Results used in other termination proofs may be interpreted in this way too.) In this area, one frequently encounters "exceptional points" (i.e., F's) that entail added difficulty. The present approach avoids this problem; correspondingly, one learns slightly less about the behavior of $\mathbf{L}_n$. However, the extra detail seems unnecessary for the application to sequential analysis. (This point is further discussed after Theorem 2.5.)

An important preliminary for our termination result is the material in Berk (1970). We adopt the definitions and notation of that paper, indicating references to it with an asterisk. (See especially, Definition 1.1*.) Treating P as a prior, let $\mathbf{P}_n$ denote the formal posterior distribution on $(\Theta, \mathscr{A})$, given $\mathbf{X}_1, \cdots, \mathbf{X}_n$. (See Equation (1.1)*.) We shall say

DEFINITION 2.1. $\{\mathbf{P}_n\}$ is consistent at $\theta_p \in \Theta$, $[F]$ if $F(\mathbf{P}_n$ converges weakly to $P^*) = 1$, where P^* is the probability distribution on Θ degenerate at θ_p.

REMARK. In order that weak convergence have meaning, Θ must be a topological space and $\mathscr{A}$ must contain the open sets. (See the paragraph preceding Definition 3.5*.) Sufficient conditions for $\{\mathbf{P}_n\}$ to be consistent are given in Berk (1970). We also say

DEFINITION 2.2. p remains bounded if, whenever $P\sup\prod_1^n p(x_i \mid \theta) < \infty$, then for all integers $k > n$ and $x_{n+1}, \cdots, x_k$, $P\sup\prod_1^k p(x_i \mid \theta) < \infty$. Here, $P\sup$ means $P\operatorname{ess\,sup}$ over Θ; see Definition 1.1 (h)* and the discussion following Lemma 2.2*. Let τ be the first integer $n \geqq 1$ for which $P\sup\prod_1^n p(\mathbf{X}_i \mid \theta) < \infty$. τ is a possibly infinite stopping time on $\mathbf{X}_1, \mathbf{X}_2, \cdots$. Let $\mathbf{T}$ be the last time $P\sup\prod_1^n p(\mathbf{X}_i \mid \theta) = +\infty$; we take $\mathbf{T} = 0$ if the expressions are finite for all n. $\mathbf{T}$ is a possibly infinite reverse stopping time on $\mathbf{X}_1, \mathbf{X}_2, \cdots$ and $\tau \leqq \mathbf{T}+1$, with equality holding if p remains bounded. We call τ the time that p becomes bounded.

Following Definition 2.4*, we say a random variable $\mathbf{Y}$ is exponentially bounded if $E\exp(r|\mathbf{Y}|) < \infty$ for some $r > 0$.

LEMMA 2.3. *Suppose p remains bounded. Then* $\mathbf{T}$ *is exponentially bounded* $\Leftrightarrow$ *for some integer $s \geqq 1$,*

$$F(P\sup\prod_1^s p(\mathbf{X}_i \mid \theta) < \infty) > 0.$$

PROOF. If $\mathbf{T}$ is exponentially bounded, then $F(\mathbf{T} < \infty) = 1$; hence for some $s \geqq 1$, $F(P\sup\prod_1^s p(\mathbf{X}_i \mid \theta) < \infty) > 0$. Conversely, suppose $F(P\sup\prod_1^s p(\mathbf{X}_i \mid \theta) < \infty) = \beta > 0$. Then, since p remains bounded, $F(\mathbf{T} > ns) \leqq F(P\sup\prod_{ks+1}^{ks+s} p(\mathbf{X}_i \mid \theta) = \infty$, $k = 0, 1, \cdots, n-1) = (1-\beta)^n$, showing that $\mathbf{T}$ is exponentially bounded. □

We collect the assumptions to be made.

ASSUMPTIONS 2.4. (a) Θ *is a Hausdorff space and* $\mathscr{A}$ *contains the open sets.*

(b) $p(\cdot \mid \cdot)$ *is jointly measurable in* (x, θ) *and for almost every* $x[F]$, *is continuous in* θ.

(c) *For all* $\theta, \theta' \in \Theta$, $F(p(\mathbf{X} \mid \theta) = q(\mathbf{X} \mid \theta')) < 1$.

(d) $\mathbf{P}_n$ *is consistent at* $\theta_p \in \Theta$, $[F]$.

(e) τ_p, *the time p becomes bounded, is* a.s. $[F]$ *finite.*

Corresponding assumptions are made for (q, Q, F).

We let $\tau = \max(\tau_p, \tau_q)$. It follows from Assumption 2.4e that $F(\tau < \infty) = 1$.

Let $\mathbf{N}$ be the stopping time of the SPRT: $\mathbf{N}$ is the first integer $n \geqq 1$ for which $\mathbf{L}_n \notin (A, B)$; $\mathbf{N} = \infty$ if no such n exists. Our main termination result is

THEOREM 2.5. *Under Assumptions* 2.4, $F(\mathbf{N} < \infty) = 1$.

PROOF. The consistency of $\mathbf{P}_n$ entails that eventually, $0 < \int\prod_1^n p(\mathbf{X}_i \mid \theta)\,dP = p_n(\mathbf{X}_1, \cdots, \mathbf{X}_n) < \infty$. (See Equations (1.1)* and (1.2) and the discussion at the beginning of Section 3*.) Fixing an integer $k \geqq 1$ and working with $\mathbf{X}_{k+1}, \mathbf{X}_{k+2}, \cdots$, let $\mathbf{S}$ be the first $n \geqq 1$ so that $0 < p_n(\mathbf{X}_{k+1}, \cdots, \mathbf{X}_{k+n})\, q_n(\mathbf{X}_{k+1}, \cdots, \mathbf{X}_{k+n}) < \infty$. $F(\mathbf{S} < \infty) = 1$. Then for $n > k$ and on $(\mathbf{S} < n-k)$, we may formally write

$$p_n(\mathbf{X}_1, \cdots, \mathbf{X}_n) = p_{n-k}(\mathbf{X}_{k+1}, \cdots, \mathbf{X}_n)p(\mathbf{X}_1, \cdots, \mathbf{X}_k \mid \mathbf{X}_{k+1}, \cdots, \mathbf{X}_n),$$

the conditional density being formally defined by this relation. Upon taking the ratio with the corresponding expression for q_n, we obtain the (self-explanatory) decomposition: on $(\mathbf{S} < n-k)$

$$\mathbf{L}_n = \mathbf{L}_{nk}\mathbf{R}_{nk}. \tag{2.1}$$

Let $\mathbf{L} = \limsup \mathbf{L}_n$. Since $\mathbf{L}_n$ is symmetric in $\mathbf{X}_1, \cdots, \mathbf{X}_n$, $\mathbf{L}$ is symmetric in $\mathbf{X}_1, \mathbf{X}_2, \cdots$ (invariant under finite permutations of the sequence). By the Hewitt–Savage (1955) zero-one law, $\mathbf{L}$ is a.s. a constant, say L. We note too that the behavior of $\mathbf{L}_{nk}$ is the same as that of $\mathbf{L}_n$ (for k fixed), hence $\limsup_n \mathbf{L}_{nk} = L[F]$.

We now show that $\mathbf{R}_{nk}$ converges $[F]$ as $n \to \infty$ by showing a similar fact for its numerator and denominator. Treating P as a formal prior on $(\Theta, \mathscr{A})$, under the model $\mathscr{P}$, $\mathbf{X}_1, \mathbf{X}_2, \cdots$ are conditionally (on θ) i.i.d. Thus, letting $\mathbf{P}_n'$ denote the posterior distribution on Θ given $\mathbf{X}_{k+1}, \cdots, \mathbf{X}_n$, on $(\mathbf{S} \nless n-k)$,

$$\begin{aligned} p(\mathbf{X}_1, \cdots, \mathbf{X}_k \mid \mathbf{X}_{k+1}, \cdots, \mathbf{X}_n) &= \int p(\mathbf{X}_1, \cdots, \mathbf{X}_k \mid \theta, \mathbf{X}_{k+1}, \cdots, \mathbf{X}_n)\, d\mathbf{P}_n'(\theta) \\ &= \int \textstyle\prod_1^k p(\mathbf{X}_i \mid \theta)\, d\mathbf{P}_n'. \end{aligned} \tag{2.2}$$

Since $\mathbf{P}_n$ is consistent at θ_p, so is $\mathbf{P}_n'$. On $(\boldsymbol{\tau} \leqq k)$, the integrand in (2.2) is continuous and essentially bounded (it can be replaced by $\max\{\prod_1^k p(\mathbf{X}_i \mid \theta), P\sup \prod_1^k p(\mathbf{X}_i \mid \theta)\}$ without changing the value of the integral, since $\mathbf{P}_n' \ll P$). Thus by weak convergence, on $(\boldsymbol{\tau} \leqq k)$, as $n \to \infty$,

$$p(\mathbf{X}_1, \cdots, \mathbf{X}_k \mid \mathbf{X}_{k+1}, \cdots, \mathbf{X}_n) \to \textstyle\prod_1^k p(\mathbf{X}_i \mid \theta_p) \quad [F]. \tag{2.3}$$

Thus on $(\boldsymbol{\tau} \leqq k)$, $\mathbf{R}_{nk} \to \mathbf{R}_k = \prod_1^k p(\mathbf{X}_i \mid \theta_p)/q(\mathbf{X}_i \mid \theta_q)[F]$.

From (2.1) we see that on $(\boldsymbol{\tau} \leqq k)$, $L = L\mathbf{R}_k$, so if we assume $0 < L < \infty$, it follows that $\mathbf{R}_k = 1$ on $(\boldsymbol{\tau} \leqq k)$. Since $F(\boldsymbol{\tau} < \infty) = 1$, this implies $F(\mathbf{R}_k \to 1) = 1$. We show this leads to a contradiction. For let $\mathbf{Z}_k = \ln[p(\mathbf{X}_k \mid \theta_p)/q(\mathbf{X}_k \mid \theta_q)]$. $\mathbf{Z}_1, \mathbf{Z}_2, \cdots$ are i.i.d. and $\mathbf{R}_k \to 1[F] \equiv \sum_1^k \mathbf{Z}_i \to 0[F]$. But then $\sum_2^k \mathbf{Z}_i \to 0[F]$, which implies $\mathbf{Z}_1 \equiv 0[F]$. This contradicts Assumption 2.4(c), which entails $F(\mathbf{Z}_1 = 0) < 1$. We conclude that $L = 0$ or ∞, which implies $F(\mathbf{N} < \infty) = 1$. □

REMARK. Since $\mathbf{L}$ is a.s. constant, it is only an apparent weakening to write the above conclusion as $F(\mathbf{L} = 0 \text{ or } \infty) = 1$. The above technique shows too that $\liminf \mathbf{L}_n$ is either 0 or ∞. Hence either $F(\mathbf{L}_n \to 0) = 1$ or $F(\mathbf{L} \to \infty) = 1$, or $F(\liminf \mathbf{L}_n = 0 \text{ and } \limsup \mathbf{L}_n = \infty) = 1$. For our purposes, it does not matter which of these three types of divergence (of $\ln \mathbf{L}_n$) occurs. In certain cases, previous writers have given a more explicit characterization of the behavior of $\mathbf{L}_n$ corresponding to a given F. See, e.g., Ifram (1965), Wijsman (1967b) and Wirjosudirdjo (1961).

Theorem 2.5 applies directly to weight-function SPRTs, including those invariance-reduction SPRTs for which the representation (1.4) holds. Of course Assumptions 2.4 must be verified; we illustrate such a verification in Section 4. Working still with the model (1.2), Theorem 2.5 can be extended to sequences that are actually exchangeable as follows. Let $\mathscr{F}$ denote the class of i.i.d. distributions for the sequence for which Theorem 2.5 holds. Thus for $F \in \mathscr{F}$, $F(\mathbf{L} = 0 \text{ or } \infty) = 1$.

Then if F' is a mixture of elements of $\mathscr{F}$, $F'(\mathbf{L} = 0 \text{ or } \infty) = 1$. Thus the theorem holds for any exchangeable distribution that is a mixture of elements a.s. in $\mathscr{F}$. This remark applies in particular to the representation (1.3) when Theorem 1.1 holds, for then $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$ are exchangeable with distribution F_I. The desired conclusion follows if F_I is a mixture of elements a.s. in the appropriate $\mathscr{F}$.

3. Exponential boundedness. We establish conditions under which $\mathbf{N}$ is exponentially bounded (has a non-trivial moment generating function). Of relevance here are the large-deviation results of Section 5*.

DEFINITION 3.1. $\{\mathbf{P}_n\}$ is exponentially bounded if for all $\delta > 0$ and $\varepsilon > 0$, $\{(\mathbf{P}_n A_\delta' > \varepsilon)\}$ is exponentially bounded.

REMARK. The A_δ, defined in 1.1(i)*, are, under Assumption 3.2(e) below, essentially a nested neighborhood system at θ_p. The meaning of exponential boundedness for a sequence of events is given in 2.4*. A' denotes $\Theta - A$. Theorems 5.1* and 5.3* give sufficient conditions for $\{\mathbf{P}_n\}$ to be exponentially bounded. We strengthen Assumption 2.4(a–d) to

ASSUMPTIONS 3.2. (a), (b) *and* (c) *as in* 2.4.

(d) $\{\mathbf{P}_n\}$ *is exponentially bounded.*

(e) *The* A_δ *form a weak base at* $\theta_p \in \Theta$. (*See Definition* 3.5*.) *Similar assumptions hold for* q.

REMARK. Corollary 3.6* shows that 3.2 (d, e) are a strengthening of 2.4(d).

Let t be a positive integer and let $\mathbf{X}^1$ denote $\mathbf{X}_1, \cdots, \mathbf{X}_t$, $\mathbf{X}^2$ denote $\mathbf{X}_{t+1}, \cdots, \mathbf{X}_{2t}$, etc. We use notation such as $p(\mathbf{X}^1 \mid \theta) = \prod_1^t p(\mathbf{X}_i \mid \theta)$. The reasoning that leads to (2.2) shows that

$$p(\mathbf{X}^n \mid \mathbf{X}^1, \cdots, \mathbf{X}^{n-1}) = \int p(\mathbf{X}^n \mid \theta)\, d\mathbf{P}_{(n-1)t}(\theta) \quad [F]. \tag{3.1}$$

By weak convergence, we would expect this to converge to $p(\mathbf{X}^n \mid \theta_p)$, except that the integrand may not be bounded. Nevertheless, because of the special nature of this integrand, we can establish

LEMMA 3.3. *Let* $\mathbf{Z}_n = \ln [p(\mathbf{X}^n \mid \mathbf{X}^1, \cdots, \mathbf{X}^{n-1})/p(\mathbf{X}^n \mid \theta_p)]$, $r(\mathbf{X}^n, \theta) = p(\mathbf{X}^n \mid \theta)/p(\mathbf{X}^n \mid \theta_p)$ *and let* $B_n = B_{n,\delta,\varepsilon}$ *denote the event* $(P \sup \{|\ln r(\mathbf{X}^n, \theta)| : \theta \in A_\delta\} < \varepsilon/2)$. *Then if* $\{\mathbf{P}_n\}$ *is exponentially bounded, for all* $\varepsilon > 0$, $\{(B_n, |\mathbf{Z}_n| > \varepsilon)\}$ *is exponentially bounded.*

PROOF.

$$|Z_n| = |\ln \int r(\mathbf{X}^n, \theta)\, d\mathbf{P}_{(n-1)t}(\theta)| = |\ln \int_{A_\delta} r(\mathbf{X}^n, \theta)\, d\mathbf{P}_{(n-1)t} + \ln (1 + \int_{A_\delta'} d\mathbf{P}_{nt} / \int_{A_\delta} d\mathbf{P}_{nt})|$$
$$\leqq \varepsilon/2 + |\ln \mathbf{P}_{(n-1)t} A_\delta| + |\ln \mathbf{P}_{nt} A_\delta|$$

on B_n. Let $\varepsilon' = 1 - e^{-\varepsilon/4}$. Then $\mathbf{P}_{(n-1)t} A_\delta' < \varepsilon'$ and $P_{nt} A_\delta' < \varepsilon'$ imply that $|\ln \mathbf{P}_{(n-1)t} A_\delta| < \varepsilon/4$ and $|\ln \mathbf{P}_{nt} A_\delta| < \varepsilon/4$. Hence

$$(B_n, |Z_n| > \varepsilon) \subset (\mathbf{P}_{(n-1)t} A_\delta' > \varepsilon') \cup (\mathbf{P}_{nt} A_\delta' > \varepsilon')$$

and both events on the right are exponentially bounded. □

The main result of this section is

THEOREM 3.4. $\mathbf{N}$ *is exponentially bounded under any* i.i.d. F *for which Assumptions* 3.2 *hold.*

PROOF. The following proof is motivated by the development found in Sethuraman (1967). The final steps incorporate a suggestion due to R. A. Wijsman. Let $a = \ln B/A$, $\mathbf{R}_k = \ln[p(\mathbf{X}^k \mid \mathbf{X}^1, \cdots, \mathbf{X}^{k-1})/q(\mathbf{X}^k \mid \mathbf{X}^1, \cdots, \mathbf{X}^{k-1})]$ and $\mathbf{S}_k = \ln[p(\mathbf{X}^k \mid \theta_p)/q(\mathbf{X}^k \mid \theta_q)]$. Let $A_k = (|\mathbf{R}_k| < a)$ and let B_k be the intersection of the sets $B_{k,\delta,\varepsilon}$ of Lemma 3.3, one for p and one for q. We note that as $\delta \downarrow 0$, $PB_k' \downarrow 0$. (To wit: Since the A_δ form a weak base at θ_p and $\ln r(\mathbf{X}^k, \cdot)$ is a.s. continuous at θ_p, as $\delta \to 0$, $P\sup\{|\ln r(\mathbf{X}^k, \theta)| : \theta \in A_\delta\} \to |\ln r(\mathbf{X}^k, \theta_p)| = 0[F]$. A similar fact holds for q.) Let $C_k = (|\mathbf{S}_k| < a+\varepsilon) \cup B_k'$ and $D_k = (B_k, |\mathbf{R}_k - \mathbf{S}_k| > \varepsilon)$. Then for $\varepsilon > 0$,

$$A_k \subset C_k \cup D_k \tag{3.2}$$

and the C_k are independent. Hence

$$(\mathbf{N} > 2nt) \subset \textstyle\bigcap_n^{2n} A_k \subset (\bigcap_n^{2n} C_k) \cup (\bigcup_n^{2n} D_k). \tag{3.3}$$

Since $\mathbf{S}_k$ is a sum of t i.i.d. random variables, and $F(\ln[p(\mathbf{X} \mid \theta_p)/q(\mathbf{X} \mid \theta_q)] = 0) < 1$, for t sufficiently large, $F(|\mathbf{S}_k| < a+\varepsilon) < 1-2\alpha < 1$. For δ sufficiently small, $FB_k' < \alpha$, hence $FC_k < 1-\alpha$. Also, by Lemma 3.3, $FD_k < c\rho^k$ for some $\rho < 1$. It follows from (3.3) that $F(\mathbf{N} > 2nt) < (1-\alpha)^n + \sum_n^{2n} c\rho^k < (1-\alpha)^n + c\rho^n/(1-\rho)$. Thus $\mathbf{N}$ is exponentially bounded. □

As discussed above, this result can be applied to many invariance-reduction SPRTs as well as weight-function SPRTs, to show that for some $s > 0$, $E\exp(s\mathbf{N}) < \infty$. The result does not seem to extend directly to exchangeable sequences, for one then obtains the result that $\mathbf{N}$ is conditionally exponentially bounded. This need not imply that $\mathbf{N}$ is exponentially bounded.

4. The sequential *t*-test. The sequential t-test will serve to illustrate the foregoing. We have $p(x \mid \theta) = \exp\{-(x/\theta - \delta)^2/2\}/|\theta|(2\pi)^{\frac{1}{2}}$ for some $\delta > 0$ and, for convenience, assume $\mathcal{Q}$ corresponds to $\delta = 0$. The weight-function here is $dP = d\theta/|\theta|$. We consider Assumptions 2.4. Conditions (a) and (b) hold. Since $p(\cdot \mid \theta)$ and $q(\cdot \mid \theta')$ are normal densities with different means, (c) holds unless $\mathbf{X}$ is degenerate. (If $\mathcal{Q}$ also corresponds to a $\delta' > 0$, it may happen that the means coincide, but then the variances will differ. It then seems necessary to assume that $\mathbf{X}$ can take on at least three values.) Verification of (d) proceeds essentially as in example I of Section 6*. Choosing $p^*(x) = (2\pi)^{\frac{1}{2}}\exp\{\delta^2/2\}$, $l(x \mid \theta) = -x^2/2\theta^2 + \delta x/\theta - \ln|\theta|$. The relevant expectations in 1.1(f)* certainly exist if $E\mathbf{X}^2 < \infty$. $\mathcal{P}$ is an exponential model, so Corollary 4.2* applies. Replacing δx by α_1 and x^2 by α_2 in $l(x \mid \theta)$, we obtain $l(\alpha, \theta) = -\alpha_2/2\theta^2 + \alpha_1/\theta - \ln|\theta|$, which has a finite maximum if $\alpha_2 > 0$. Thus (see Section 4*) $D = (\alpha_2 > 0)$. Letting $a_1(x) = \delta x$ and $a_2(x) = x^2$, $F(\mathbf{a}_1 \in \bar{D}) = 1$ while $\alpha_F \in D^0$ unless $\mathbf{X}$ is degenerate at zero. Condition 4.1(ii)* holds if $E\mathbf{X}^2 < \infty$. Referring to Definition 3.1*, we see that P becomes proper as soon as $\mathbf{X}_n \neq 0$. ($\mathbf{S}$ is the waiting time until the first non-zero observation.) It may be seen directly (or from Proposition 4.3*) that $F(\mathbf{S} < \infty) = 1$ unless $\mathbf{X}$ is degenerate at zero. Thus if $\mathbf{X}$ is not degenerate and $E\mathbf{X}^2 < \infty$, we may verify the hypothesis of Corollary 4.2*; then $F(\mathbf{P}_n A_\delta \to 1) = 1$.

Turning to A_δ, let $\mu = E\mathbf{X}$ and $\nu = E\mathbf{X}^2$. Since P is equivalent to Lebesgue measure, θ_p is the point where $\lambda(\theta) = El(\mathbf{X} \mid \theta)$ attains its maximum. $\theta_p = [\pm(\delta^2\mu^2 + 4\nu)^{\frac{1}{2}} - \delta\mu]/2$ (choose the root having the same sign as μ). Moreover, it is easily seen that the A_δ become a nested set of intervals decreasing to θ_p and hence form a weak base there. By Corollary 3.6*, $\mathbf{P}_n$ converges weakly to $P^*[F]$. Condition 2.4(e) is also seen to hold since τ is also the waiting time until the first non-zero observation. Hence Theorem 2.5 applies, and we conclude for the sequential t-test that if $\mathbf{X}$ is not degenerate and $E\mathbf{X}^2 < \infty$, then $F(\mathbf{N} < \infty) = 1$. As mentioned above, this particular result has been obtained by other writers using different methods.

Turning to Assumptions 3.2, we need only discuss 3.2(d). Referring to Theorem 5.1*, we see that $\Delta = D$, and if $\mathbf{X}$ is not degenerate and $\mathbf{X}^2$ is exponentially bounded, the conditions of that theorem are satisfied. Hence under this more restrictive assumption on $\mathbf{X}$, we have that $\mathbf{N}$ is exponentially bounded. This result is slightly stronger than that implied by Wijsman's (1968) results for the sequential t-test since we have no exceptional points to worry about. Wijsman (1968) gives additional references to the exponential boundedness problem for $\mathbf{N}$.

5. Further examples. We discuss two more examples to indicate further the scope of the assumptions made in Section 2. The first is an SPRT that does not terminate w.p.1. It is shown below that Assumption 2.4(c) fails for this procedure. Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be an exchangeable Gaussian process with all parameters known, except for the mean μ. We are interested in testing the hypotheses $H' : \mu \leqq 0$ vs $K' : \mu > 0$. By rescaling, if necessary, we may assume that $\operatorname{Var} X_1 = 1 + \delta$, $\operatorname{Cov}(X_1, X_2) = \delta$, where $\delta > 0$ is known. As a test of the above hypotheses (or as a procedure in its own right), we consider an SPRT based on $\mathbf{X}_1, \mathbf{X}_2, \cdots$ of $H : \mu = 0$ vs $K : \mu = \nu$, where ν is a specified positive number. We may consider the process $\mathbf{X}_1, \mathbf{X}_2, \cdots$ as arising as follows: Let $\mathbf{Y}, \mathbf{Y}_1, \mathbf{Y}_2, \cdots$ be independent normal variables; $\mathbf{Y} \sim N(\mu, \delta)$ and $\mathbf{Y}_i \sim N(0, 1)$ for all i. Then $\{\mathbf{Y}_i + \mathbf{Y}\}$ has the same joint distribution as $\{\mathbf{X}_i\}$ and we may replace the latter by the former. Letting P_μ denote the joint distribution of $\mathbf{Y}, \mathbf{Y}_1, \mathbf{Y}_2, \cdots$ when μ obtains, we may consider the above-mentioned SPRT as a test of whether P_μ is P_0 or P_ν, based on the reduced data sequence $\mathbf{Y}_1 + \mathbf{Y}, \mathbf{Y}_2 + \mathbf{Y}, \cdots$. We argue that for all μ and ν, P_μ and P_ν are equivalent measures on the reduced sequence. Indeed, they are equivalent on the sequence $\mathbf{Y}, \mathbf{Y}_1, \cdots$ since only the distribution of the first coordinate can differ. Then the restrictions of P_μ and P_ν to the (σ-field generated by the) reduced sequence are surely equivalent as well. Letting $\mathbf{L}_n$ be the likelihood ratio under $\mu = \nu$ and $\mu = 0$ of $(\mathbf{Y}_1 + \mathbf{Y}, \cdots, \mathbf{Y}_n + \mathbf{Y})$, it follows from the martingale convergence theorem that for $\mu = 0$ and ν, $P_\mu(\mathbf{L}_n \to \mathbf{L}) = 1$, where $\mathbf{L}$ is the corresponding likelihood ratio for the entire sequence $\mathbf{Y}_1 + \mathbf{Y}, \mathbf{Y}_2 + \mathbf{Y}, \cdots$. (For specificity, we suppose that the ν-measure is in the numerator of the likelihood ratio.) $\mathbf{L}$ is a non-degenerate random variable, so that already this behavior of $\mathbf{L}_n$ is not as described in Section 2 (see the remark after the proof of Theorem 2.5). One might suspect that this test does not terminate w.p.1. This is indeed the case, as we show next.

A straightforward computation shows that

$$\ln \mathbf{L}_n = nv(\overline{\mathbf{Y}}_n + \mathbf{Y} - v/2)/(1+n\delta), \qquad \text{where} \quad \overline{\mathbf{Y}}_n = \textstyle\sum_1^n \mathbf{Y}_i/n.$$

For given $a > 0$, let $\mathbf{N}$ be the first $n \geqq 1$ so that $|\ln \mathbf{L}_n| \geqq a$, or be $+\infty$ if no such n occurs. Then

$$P_\mu(\mathbf{N} = \infty) \geqq P_\mu(|\overline{\mathbf{Y}}_n + \mathbf{Y} - v/2| < b, n = 1, 2, \cdots), \tag{5.1}$$

where $b = \delta a/v$. To evaluate (5.1), consider the related probability $p(z) = \Pr(|\overline{\mathbf{Y}}_n + z| < b,\ n = 1, 2, \cdots)$. If $|z| < b$, $p(z) > 0$. This follows readily from the fact that for all $c > 0$, $\Pr(|\overline{\mathbf{Y}}_n| < c,\ n = 1, 2, \cdots) \geqq \Pr(|\mathbf{W}(t)| < c,\ 0 \leqq t \leqq 1) > 0$, where $\mathbf{W}(t)$ is a standard Wiener process. (Note that the joint distributions of $\{\overline{\mathbf{Y}}_n\}$ and $\{\mathbf{W}(1/n)\}$ are the same.) Evidently, (5.1) implies that $P_\mu(\mathbf{N} = \infty \mid \mathbf{Y} = y) \geqq p(y - v/2)$. Since $p(y-v/2) > 0$ if $|y - v/2| < b$, and since $P_\mu(|\mathbf{Y} - v/2| < b) > 0$, it follows that $P_\mu(N = \infty) > 0$. Thus SPRTs based on $\mathbf{L}_n$ do not terminate w.p.1 under any μ when the model holds.

This problem is superficially different from those discussed in Section 1, for H_0 and H_v are *simple* hypotheses about a sequence $\{\mathbf{X}_i\}$, which here is *not* i.i.d. However, the $\mathbf{X}_i$ are exchangeable and, in fact, are conditionally i.i.d. given $\mathbf{Y}$. Thus this problem does, in fact, entail sequences of exchangeable densities. Moreover, we may identify $p(\cdot \mid \cdot)$ (respectively, $q(\cdot \mid \cdot)$) with the conditional density of $\mathbf{X}_1$ given $\mathbf{Y}$ under H_0 (respectively, H_v) and P and Q, with the distributions of $\mathbf{Y}$ under H_0 and H_v. Here the families p and q are the same normal translation family (with unit variance), so that Assumption 2.4(c) is patently false. What distinguishes the (simple) hypotheses $\mathscr{P} \equiv H_0$ and $\mathscr{Q} = H_v$ here is solely the difference between P and Q. This example suggests that when p and q coincide, we should not expect the SPRT to terminate a.s., unless P and Q have disjoint supports. Thus when $\delta = 0$, the (degenerate) distributions of $\mathbf{Y}$ under H_0 and H_v do have disjoint supports and the separation required by 2.4(c) holds. (We have then, in fact, the ordinary Wald SPRT for a normal mean based on independent observations.)

We consider next weight-function SPRTs of $H: \mu = 0$ vs $K: \mu > 0$ for the mean of i.i.d. normal variable with unit variance. Condition 2.4(c) is formally satisfied by the densities in H and K. However, 2.4(d) typically fails when F is the $N(0, 1)$ distribution: If the weight-function P on $\Theta = (0, \infty)$ has 0 in its (topological) support, one gets no convergence of $\mathbf{P}_n$ on Θ, since $\mathbf{P}_n$ converges to the point-mass at zero. Adjoining zero to Θ overcomes this difficulty but violates 2.4(c). Thus if zero is in the support of P, the above procedures do not seem to be amenable to our method, when the null hypothesis obtains. Nevertheless, we show that these SPRTs do terminate under the null hypothesis. For simplicity, we suppose P is proper. Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ be the data sequence, i.i.d. normal variables with unit variance. Let F_μ denote their (joint) distribution when the common mean is μ. For $\mu \neq 0$, $F_\mu \perp F_0$, since $F_\mu(\overline{\mathbf{X}}_n \to 0)$ is zero unless $\mu = 0$ (and then the probability is one). If F_P is the P-mixture of $\{F_\mu : \mu > 0\}$, it follows that $F_P(\bar{\mathbf{X}}_n \to 0) = 0$, hence that $F_P \perp F_0$ too. Let $F_0{}^n$, etc. denote the restriction of F_0 to the σ-field generated by $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$. Then the likelihood ratio for $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ is $\mathbf{L}_n = dF_P{}^n/dF_0{}^n$. Let $\lambda = (F_0 + F_P)/2$. We note that $F_0 \ll \lambda$, $F_P \ll \lambda$ and that the martingale convergence theorem implies that $\lambda(dF_0{}^n/d\lambda^n \to dF_0/d\lambda) = 1$. This convergence thus holds a.s. $[F_0]$ and a.s. $[F_P]$. Moreover, the orthogonality of F_0 and F_P implies that $F_0(dF_0/d\lambda = 2) = 1$ and $F_P(dF_0/d\lambda = 0) = 1$. We note too that $dF_0{}^n/d\lambda^n = 2/(1 + \mathbf{L}_n)$; hence it follows that $F_0(\mathbf{L}_n \to 0) = 1$ and $F_P(\mathbf{L}_n \to \infty) = 1$. Thus an SPRT based on $\mathbf{L}_n$ terminates a.s. when F_0 obtains and also when F_P obtains. This last fact implies that the test terminates w.p.1 under almost every $F_\mu[P]$. (But the method of Section 2 establishes this termination for *every* $\mu > 0$. It is when F_0 obtains, that Assumptions 2.4 do not hold.)

The method of Section 3 cannot be used either, of course, when F_0 obtains. In fact, it is conjectured that $E_0 \mathbf{N} = \infty$ unless zero is not in the support of P. The reason for this conjecture and some additional insight into the behavior of this SPRT is provided by Lemma 3.3. There, we see that the nth increment to the log likelihood ratio is approximately $\ln[p(\mathbf{X}_n \mid \theta_p)/q(\mathbf{X}_n \mid \theta_q)]$. For the procedure under discussion q is the $N(0, 1)$ density and $p(x \mid \theta)$ are the $N(\mu, 1)$ densities. When F_0 obtains, $\theta_p = 0$ too, so that the successive increments to the log likelihood ratio become essentially zero. In this light, it is even surprising that the procedure terminates a.s. under F_0.

Dr. J. Yahav observes that when F_μ obtains, $E_\mu N = O(1/\mu^2)$, asymptotically. Hence it is also reasonable to conjecture that under F_P, $E_P N < \infty \Leftrightarrow \int dP(\mu)/\mu^2 < \infty$.

6. Acknowledgment. The author is indebted to Prof. R. A. Wijsman for a thorough reading of an earlier version of this paper that resulted in the elimination of some serious errors.

REFERENCES

[1] Barnard, G. A. (1952). The frequency justification of certain sequential tests. *Biometrika* **39** 144–150.

[2] Berk, R. H. (1970). Consistency a posteriori. *Ann. Math. Statist.* **41** 894–906

[3] David, H. T. and Kruskal, W. H. (1956). The WAGR sequential t-test reaches a decision with probability one. *Ann. Math. Statist.* **27** 797–805.

[4] Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

[5] Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

[6] Ifram, A. F. (1965). On the asymptotic behavior of densities with applications to sequential analysis. *Ann. Math. Statist.* **36** 615–637.

[7] Jackson, J. E. and Bradley, R. A. (1961). Sequential χ^2- and T^2-tests. *Ann. Math. Statist.* **32** 1063–1077.

[8] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[9] Ray, W. D. (1957). A proof that the sequential probability ratio test of the general linear hypothesis terminates with probability unity. *Ann. Math. Statist.* **28** 521–523.

[10] Sethuraman, J. (1967). Stopping time of a rank order sequential probability ratio test based on Lehmann alternatives II. Technical Report No. 4/67, Indian Statistical Institute.

[11] WIJSMAN, R. A. (1967a). Cross-sections of orbits and their application to densities of maximal invariants. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press **1** 389–400.

[12] WIJSMAN, R. A. (1967b). General proof of termination with probability one of invariant sequential probability ratio tests based on multivariate normal observations. *Ann. Math. Statist.* **38** 8–24.

[13] WIJSMAN, R. A. (1968). Bounds on the sample size distribution for a class of invariant sequential probability ratio tests. *Ann. Math. Statist.* **39** 1048–1056.

[14] WIRJOSUDIRDJO, S. (1961). Limiting behavior of a sequence of density ratios. Ph.D. dissertation, Univ. of Illinois.