The Annals of Statistics
1973, Vol. 1, No. 6, 1126-1138

SOME ASYMPTOTIC ASPECTS OF SEQUENTIAL ANALYSIS[1]

BY ROBERT H. BERK

Rutgers University

The asymptotic behavior is given for the error rates and ASN of the Wald SPRT and of invariant sequential tests. An asymptotic justification of Bhate's conjecture is also provided for invariant sequential tests. Expressions are obtained for the asymptotic relative efficiency of the Wald SPRT as compared with the corresponding best non-sequential test.

1. Introduction. An SPRT of two (simple) hypotheses H_1 and H_2 about a data sequence $X_1, X_2, \cdots$ has a stopping time of the form

$$N = \inf\{n : L_n \notin (-a_1, a_2)\}. \tag{1.1}$$

Here L_n is the log-likelihood ratio for $(X_1, \cdots, X_n)$, compute under the two hypotheses and (a_1, a_2) are prechosen stopping boundaries. One accepts H_1 if $L_N \leqq -a_1$ and H_2 if $L_N \geqq a_2$. In the classical case studied by Wald (1947), $X_1, X_2, \cdots$ are i.i.d. so that $\{L_n\}$ is a simple random walk. More complicated problems give rise to data sequences of dependent elements. For example, the sequential t-test is based on the sequences $X_j = Y_{j+1}/Y_1$, $j = 1, 2, \cdots$. Here $Y_1, Y_2, \cdots$ is the original data sequence, assumed to consist of i.i.d. $N(\mu, \sigma^2)$ observations. In this case L_n is the log-likelihood ratio for the t-statistic or its magnitude based on $Y_1, \cdots, Y_{n+1}$ (computed under two hypotheses of the form $\mu/\sigma = \delta_i$ or $|\mu/\sigma| = \delta_i$, $i = 1, 2$). The structure of L_n in this and other cases is sufficiently complicated so that very few of Wald's elegant results for the i.i.d. case carry over. In fact, only Wald's inequalities for the error rates under the two hypotheses seem to generalize. Termination results have had to be established separately and no close approximations for the power or ASN functions seem to exist, in general.

In this paper we show that, at least asymptotically, one can narrow this gap somewhat. We consider the behavior of the error rates and ASN as a_1 and a_2 become infinite. Results are given for the i.i.d. case (Section 2) and for invariant sequential tests (Section 3). Our considerations also provide an asymptotic justification of "Bhate's conjecture," at least for invariant sequential tests. Finally, in Section 4, we develop expressions for the asymptotic relative efficiency of the Wald SPRT as compared with the corresponding best non-sequential test.

2. The Wald SPRT. We suppose $X, X_1, \cdots$ are i.i.d. with common pdf f_i

Received November 1971; revised April 1973.
[1] Work supported by NSF grant GP-26066.
AMS 1970 *subject classifications.* Primary 62L10, 62E20, 62F20; Secondary 62A05, 62F05, 62G10, 60G40.
Key words and phrases. Sequential analysis, SPRT, invariant sequential test, sequential rank test, Bhate's conjecture, asymptotic relative efficiency.

under hypotheses H_i, $i = 1, 2$. A Wald SPRT of H_1 vs. H_2 gives the stopping time (1.1), where $L_n = \sum_1^n Z_j$, $Z_j = \log[f_2(X_j)/f_1(X_j)]$ and (a_1, a_2) are two positive numbers. (We suppose that Z is finite w.p. 1.) The error rates are given by $\alpha_1 = P_1(L_N \geqq a_2)$ and $\alpha_2 = P_2(L_N \leqq -a_1)$. We shall call (α_1, α_2) the strength of the test. Wald's (1947) inequalities for (α_1, α_2) may be written

$$\alpha_2 \leqq (1 - \alpha_1)e^{-a_1}, \qquad \alpha_1 \leqq (1 - \alpha_2)e^{-a_2}. \tag{2.1}$$

(These inequalities are quite general and do not depend on the i.i.d. structure. In fact, they do not even require that N be finite w.p. 1.)

It is well known [10] that if $X, X_1, \cdots$ are i.i.d. and $P(Z = 0) < 1$, then $Ee^{tN} < \infty$ for some $t > 0$. (Here P and E refer to the actual distribution of X, which need not be either f_1 or f_2.) In particular, then $EN < \infty$. To avoid trivialities, we assume throughout this section that $P(Z = 0) < 1$. Suppose now $a = \min\{a_1, a_2\} \to \infty$. We write $\lim_a$ for $\lim_{a\to\infty}$. Then $\lim_a \alpha_i = 0$ $(i = 1, 2)$ and w.p. 1 $\lim_a N = \infty$ (hence $\lim_a EN = \infty$). The following theorem gives more precise information about the asymptotic behavior of N and EN.

2.1 THEOREM. *Suppose $X, X_1, \cdots$ are* i.i.d. *and $\mu = EZ$ exists. Then if $\mu > 0$,* w.p. 1

$$\lim_a 1_{(L_N \geqq a_2)} = \lim_a P(L_N \geqq a_2) = 1,$$
$$\lim_a N/a_2 = \lim_a EN/a_2 = 1/\mu.$$

If $\mu < 0$, w.p. 1

$$\lim_a 1_{(L_N \leqq -a_1)} = \lim_a P(N_N \leqq -a_1) = 1,$$
$$\lim_a N/a_1 = \lim_a EN/a_1 = -1/\mu.$$

REMARK. We can have $|\mu| = \infty$. The case $\mu = 0$ is covered by Theorem 2.4 below.

PROOF. We treat the case $\mu > 0$. Since $\lim_n L_n/n = \mu$ w.p. 1, also $\lim_n L_n = +\infty$ w.p. 1. Thus $L_* = \min_n L_n$ is finite w.p. 1. We then have $1_{(L_N \leqq -a_1)} \leqq 1_{(L^* \leqq -a_1)} \to 0$ w.p. 1 as $a \to \infty$. Thus $\lim_a 1_{(L_N \geqq a_2)} = 1$ and by dominated convergence, $\lim_a P(N_N \geqq a_2) = 1$.

Since w.p. 1 $\lim_a N = \infty$, $\lim_a L_N/N = \mu$ w.p. 1. By the definition of N,

$$L_{N-1} 1_{(L_N \geqq a_2)} < a_2 1_{(L_N \geqq a_2)} \leqq L_N 1_{(L_N \geqq a_2)}.$$

On dividing across by N and letting $a \to \infty$, the extreme terms both approach μ w.p. 1; thus w.p. 1 $\lim_a a_2/N = \mu$ or $\lim_a N/a_2 = 1/\mu$. By Fatou's lemma, $\liminf_a EN/a_2 \geqq 1/\mu$.

Now let $t = \inf\{n : L_n \geqq a_2\}$. Clearly $N \leqq t$. It follows from the results in Siegmund (1967) (for $\mu < \infty$, plus an easy truncation argument if $\mu = \infty$) that under our assumptions, $\lim_a Et/a_2 = 1/\mu$. Hence also $\limsup_a EN/a_2 \leqq 1/\mu$. □

Theorem 2.1 shows that Wald's approximation for the ASN is asymptotically correct. This approximation [11, page 53] applies when $0 < |\mu| < \infty$ and may be written

$$EN \doteq [-a_1 P(L_N \leqq -a_1) + a_2 P(L_N \geqq a_2)]/\mu. \tag{2.2}$$

According to Theorem 2.1, the ratio of the two sides of (2.2) approaches one as $a \to \infty$.

Wald's inequalities (2.1) provide the yet cruder inequalities

$$\alpha_1 \leqq e^{-a_2}, \qquad \alpha_2 \leqq e^{-a_1}. \tag{2.3}$$

The next theorem shows that asymptotically, the inequalities in (2.3) become, in a sense, equalities.

2.2 THEOREM. *Suppose* $X, X_1, \cdots$ *are* i.i.d. *and* $E_i|Z| < \infty$, $i = 1, 2$. *Then*

$$\lim_a a_2^{-1} \log \alpha_1^{-1} = 1 = \lim_a a_1^{-1} \log \alpha_2^{-1}.$$

PROOF. Let $\mu_i = E_i Z$. Necessarily $\mu_1 < 0 < \mu_2$. Wald [11, page 197] gives the following inequality:

$$E_2 L_N = \mu_2 E_2 N \geqq (1 - \alpha_2) \log [(1 - \alpha_2)/\alpha_1] + \alpha_2 \log [\alpha_2/(1 - \alpha_1)]. \tag{2.4}$$

By (2.3), $\alpha_1 = o(1) = \alpha_2$ as $a \to \infty$. Upon dividing across in (2.4) by a_2, we obtain

$$\mu_2 E_2 N/a_2 \geqq a_2^{-1} \log \alpha_1^{-1}[1 + o(1)].$$

From Theorem 2.1, we have that $\lim_a \mu_2 E_2 N/a_2 = 1$; hence $\limsup_a a_2^{-1} \log \alpha_1^{-1} \leqq 1$. By (2.3), $a_2^{-1} \log \alpha_1^{-1} \geqq 1$, so therefore $\lim_a a_2^{-1} \log \alpha_1^{-1} = 1$. The result for α_2 is done similarly. □

REMARK. This result also shows that Wald's approximations for the error rates (obtained by treating the relations in (2.1) as equalities and solving for (α_1, α_2)) are asymptotically correct in the sense of the theorem.

Wald [11, page 50] obtained an approximation for the power curve of the SPRT under the additional assumption that for some (necessarily unique) real number $h \neq 0$, $Ee^{hZ} = 1$. In our notation, the approximation may be written

$$P(L_N \geqq a_2) \doteq (1 - e^{-ha_1})/(e^{ha_2} - e^{-ha_1}).$$

Theorem 2.2, in conjunction with Wald's device of considering the SPRT as being generated by $L_n' = hL_n$ with stopping boundaries

$$(-a_1', a_2') = (-ha_1, ha_2) \quad \text{if} \quad h > 0$$
$$\text{(resp., } (-a_1', a_2') = (ha_2, -ha_1) \text{ if } h < 0)$$

establishes

2.3 COROLLARY. *Suppose* $X, X_1, \cdots$ *are* i.i.d., $E|Z| < \infty$ *and for some* $h \neq 0$, $Ee^{hZ} = 1$. *Then*

$$\lim_a (-ha_2)^{-1} \log P(L_N \geqq a_2) = 1 \quad \text{if} \quad h > 0,$$
$$\lim_a (ha_1)^{-1} \log P(L_N \leqq -a_1) = 1 \quad \text{if} \quad h < 0.$$

PROOF. We recall that L_n' is a log-likelihood ratio for $X_1, \cdots, X_n$ under two i.i.d. distributions, with the true distribution of X in the denominator. Thus Theorem 2.2 applies directly. □

REMARK. Given the existence of h, we have $EhZ < \log Ee^{hZ} = 0$, so that $\mu = EZ \neq 0$ and is opposite in sign to h.

When $EZ = 0$, the asymptotic behavior of EN is substantially different from that given in Theorem 2.1.

2.4 THEOREM. *Suppose $X, X_1, \cdots$ are* i.i.d. *with $\sigma^2 = EZ^2 < \infty$ and $EZ = 0$. Let $A = a_1 + a_2$ and for $j > 0$, let*

$$\varphi_j(A) = \max [\sup \{E([Z^+ - u]^j \mid Z^+ \geqq u) : 0 \leqq u \leqq A\}, \sup \{E([Z^- - u]^j \mid Z^- \geqq u) : 0 \leqq u \leqq A\}] .$$

If either $A = O(a)$ or $\varphi_1(A) = o(a)$, then, letting $p = P(L_N \geqq a_2)$ and $\pi = a_1/A$,

$$\lim_a p/\pi = 1 = \lim_a (1 - p)/(1 - \pi) . \tag{2.5}$$

If the above conditions are strengthened to $A = O(a)$ or $\varphi_2(A) = o(a^2)$, then also

$$\lim_a \sigma^2 EN/a_1 a_2 = 1 . \tag{2.6}$$

REMARK. As indicated below, φ_j gives a bound on the jth moment of the "overshoot." If Z is bounded or both Z^+ and Z^- have increasing failure rate distributions, then $\varphi_j(A) = O(1)$.

PROOF. Let

$$\Delta^- = L_N^- - a_1 1_{(L_N \leqq -a_1)} , \qquad \Delta^+ = L_N^+ - a_2 1_{(L_N \geqq a_2)} . \tag{2.7}$$

Thus $\Delta = \Delta^+ + \Delta^-$ is the magnitude of the "overshoot." Upon taking expectations in (2.7) (and noting that by Wald's first lemma, $EL_N = 0$, so that $EL_N^+ = EL_N^- = E|L_N|/2$), we obtain by subtraction

$$a_1(1 - p) - a_2 p = E\Delta^+ - E\Delta^- . \tag{2.8}$$

Upon dividing across by a_1, this last relation is seen to entail

$$|p/\pi - 1| \leqq E\Delta/a_1 . \tag{2.9}$$

A bound for the expected overshoot is given by

$$E\Delta \leqq \varphi_1(A) . \tag{2.10}$$

This is a variant of the bound given by Wald (1947); see, e.g., equations (A.75) and (A.76). Thus if $\varphi_1(A) = o(a)$, (2.9) and (2.10) imply that $p/\pi - 1 = o(1)$ and hence the first relation in (2.5) holds. The second relation in (2.5) follows similarly in this case, upon dividing across in (2.8) by a_2.

Suppose now that $A = O(a)$. We show that this entails

$$E\Delta^2 = o(a^2) . \tag{2.11}$$

We begin by noting that $(L_N - a_2)(L_N + a_1) \geqq 0$, hence $0 \leqq E(L_N^2 + (a_1 - a_2)L_N - a_1 a_2) = \sigma^2 EN - a_1 a_2$ (by Wald's lemmas). Thus

$$\sigma^2 EN/a_1 a_2 \geqq 1 \tag{2.12}$$

and in particular, $\lim_a EN = \infty$. Next we note that $\Delta \leqq |Z_N|$ and it follows from the results of Gundy and Siegmund (1967) that under our conditions,

$$EZ_N^2 = o(EN) . \tag{2.13}$$

Thus to establish (2.11), we need only show that

$$EN = O(a^2) . \tag{2.14}$$

Let $t = \inf\{n : |L_n| \geqq A\}$. Clearly $N \leqq t$. Moreover, $0 \leqq (|L_t| - A)^2 \leqq Z_t^2$, so that

$$0 \leqq \sigma^2 Et + A^2 \leqq 2AE|L_t| + EZ_t^2 \leqq 2A(EL_t^2)^{\frac{1}{2}} + EZ_t^2 = 2A\sigma(Et)^{\frac{1}{2}} + EZ_t^2 ,$$

i.e., $[\sigma(Et)^{\frac{1}{2}} - A]^2 \leqq EZ_t^2 = o(Et)$, which entails $A/\sigma(Et)^{\frac{1}{2}} - 1 = o(1)$. Thus $\lim_a \sigma^2 Et/A^2 = 1$, so $Et = O(A^2) = O(a^2)$ and *a fortiori*, (2.14) holds. (Note that this result for Et is a particular case of (2.6).) As noted above, (2.11) thus holds, which, together with (2.9) entails (2.5).

We consider now EN. Upon squaring the relations in (2.7), adding and taking expectations, we obtain

$$\sigma^2 EN - AE|L_N| + a_2^2 p + a_1^2(1-p) = E\Delta^2 , \tag{2.15}$$

so that

$$\sigma^2 EN/a_1 a_2 - AE|L_N|/a_1 a_2 + (1-\pi)p/\pi + \pi(1-p)/(1-\pi) = E\Delta^2/a_1 a_2 . \tag{2.16}$$

Suppose first that $\varphi_2(A) = o(a^2)$. Then, analogously to (2.10), $E\Delta^2 \leqq \varphi_2(A)$, so that $E\Delta^2 = o(a^2)$. Thus the RHS of (2.16) is $o(1)$. Adding the relations in (2.7) and taking expectations gives $E|L_N| - a_2 p - a_1(1-p) = E\Delta$, hence

$$AE|L_N|/a_1 a_2 - p/\pi - (1-p)/(1-\pi) = AE\Delta/a_1 a_2 \leqq 2E\Delta/a = o(1) . \tag{2.17}$$

Adding, (2.16) and (2.17) yield (2.6).

Suppose next that $A = O(a)$. As shown above, (2.11) then holds, so again the RHS of (2.16) is $o(1)$ and (2.17) holds as well. Again we conclude that (2.6) holds. □

Theorem 2.4 shows that when $EZ = 0$ and $EZ^2 < \infty$, Wald's approximations [11, page 176] for the power and ASN are, under certain conditions, asymptotically exact.

3. Invariant tests. We now suppose that H_1 and H_2 are two invariant composite hypotheses about an i.i.d. data sequence $X, X_1, X_2, \cdots$, both generated by the same group G. That is, G is a group of 1–1 bimeasurable transformations acting on range X and under H_i, the distribution of X belongs to $\mathscr{P}_i = GP_i$, $i = 1, 2$. (P_i can be any distribution in $\mathscr{P}_i$.) For further elaboration, see [7]. Let $\mathscr{I}_n$ denote the G-invariant subsets of $\mathscr{F}_n = \mathscr{B}(X_1, \cdots, X_n)$. ($G$ acts coordinate-wise on each X_i.) For $P_i \in \mathscr{P}_i$, we let P_{in} denote the induced product measure on $\mathscr{F}_n$ and Q_{in} denotes the restriction of P_{in} to $\mathscr{I}_n$. (Note that because G generates $\mathscr{P}_i$, every distribution in $\mathscr{P}_i$ gives the same Q_{in}.) We suppose there exists P_i in $\mathscr{P}_i$, $i = 1, 2$ so that $P_{2n} \equiv P_{1n}$ for all n. Then the likelihood ratio under H_1 and H_2 of the G-maximal invariant for $(X_1, \cdots, X_n)$ is

$$\Lambda_n = dQ_{2n}/dQ_{1n} = E_1(dP_{2n}/dP_{1n} \mid \mathscr{I}_n) .$$

Since $\{\Lambda_n\}$ is a sequence of likelihood ratios on increasing σ-fields ($\mathscr{I}_n \subset \mathscr{I}_{n+1}$), an SPRT of H_1 vs. H_2 can be based on $\{\Lambda_n\}$. Letting $L_n = \log \Lambda_n$, one can use the procedure described in Section 1. Wald's inequalities (2.1) for the error rates remain valid. Of course, as L_n need not be a sum of i.i.d. random variables, Wald's approximation for the ASN is not available. Under certain conditions, we develop an asymptotic expression for the ASN. As above, P denotes the actual distribution of X and need not be in either hypothesis. The following result does not use the invariance structure; thus it applies to any SPRT satisfying the hypotheses. These hypotheses are satisfied for many invariant SPRTs (e.g., the sequential t-test) and for many (other) SPRTs obtained by Wald's method of weight functions. Verification of the hypotheses will be discussed elsewhere. An application to a sequential rank-test is given below.

3.1 THEOREM. *Suppose that* w.p. 1, $L_n/n \to \mu \in (0, \infty]$. *Then* w.p. 1

(i) $$\lim_a 1_{(L_N \geqq a_2)} = \lim_a P(L_N \geqq a_2) = 1 ,$$
$$\lim_a N/a_2 = 1/\mu .$$

If also, for some $\nu \in (0, \mu)$, *the "large-deviation" probabilities* $p_n = P(L_n/n < \nu)$ *satisfy* $\lim_n np_n = 0$ *and* $\sum_n p_n < \infty$, *then also*

(ii) $$\lim_a EN/a_2 = 1/\mu .$$

Analogous statements hold if $L_n/n \to \mu < 0$.

PROOF. The argument in Thorem 2.1 carries over verbatim to establish (i). To prove (ii), it then suffices to show that N/a_2 is uniformly integrable. For this, we need only show that

$$\sup_{a>1} \{nP(N > na_2) + \textstyle\sum_{k>n} P(N > ka_2)\} = o(1) \quad \text{as} \quad n \to \infty . \tag{3.1}$$

It is sufficient to establish (3.1) when a_2 ranges in the positive integers (since N does not exceed the stopping time obtained by replacing a_2 with $\{a_2\} = \inf\{n : n \geqq a_2\}$). For any integer $n > 1/\nu$, letting $s = na_2$, we have

$$P(N > na_2) \leqq P(L_s < a_2) = P(L_s/s < 1/n) \leqq p_s . \tag{3.2}$$

It follows from the hypothesis that $b_n = \sup_{k \geqq n} kp_k \downarrow 0$ as $n \to \infty$. Thus (3.2) entails

$$nP(N > na_2) \leqq np_s \leqq b_n \tag{3.3}$$

and also that

$$\textstyle\sum_{k>n} P(N > ka_2) \leqq \sum_{k>s} p_k \leqq \sum_{k>n} p_k \downarrow 0 \quad \text{as} \quad n \uparrow \infty . \tag{3.4}$$

Together, (3.3) and (3.4) entail (3.1). □

We establish next an analog of Theorem 2.2 for invariant SPRTs. We choose $P_i \in \mathscr{P}_i$, $i = 1, 2$ and let $Z = dP_2/dP_1(X)$.

3.2 THEOREM. *Suppose that for* $i = 1, 2$, $P_i(L_n/n \to \mu_i) = 1$, *where* $-\infty < \mu_1 < 0 < \mu_2 < \infty$ *and that for some* $\nu_i \in (0, |\mu_i|)$, $p_n = P_1(L_n/n > -\nu_1) + P_2(L_n/n < \nu_2)$ *satisfies* $\lim_n np_n = 0$ *and* $\sum_n p_n < \infty$. *Suppose also that* (P_1, P_2) *can be chosen so that* $E_i|Z| < \infty$, $i = 1, 2$ *and that invariance and almost-invariance are equivalent for* $\mathscr{P}_i$, $i = 1, 2$. *Then*

$$\lim_a a_1^{-1} \log \alpha_2^{-1} = 1 = \lim_a a_2^{-1} \log \alpha_1^{-1} .$$

REMARK. Conditions for the equivalence of invariance and almost-invariance are given in [3].

Before proving the theorem, we establish

3.3 LEMMA. *Let* $\{U_a\}$ *be a collection of nonnegative uniformly integrable random variables, all measurable with respect to a* σ*-field* $\mathscr{F}$. *Let* $\{\mathscr{F}_a\}$ *be a similarly indexed system of sub-*σ*-fields of* $\mathscr{F}$ *and let* $V_a = E(U_a \mid \mathscr{F}_a)$. *Then* $\{V_a\}$ *is uniformly integrable.*

PROOF. Since $EV_a = EU_a$, $\sup_a EV_a = \sup_a EU_a = b < \infty$. For $x > 0$, $\int_{(V_a > x)} V_a = \int_{(V_a > x)} U_a$. Since for every a, $P(V_a > x) \leqq b/x$, it follows that $\sup_a \int_{(V_a > x)} V_a = \sup_a \int_{(V_a > x)} U_a \to 0$ as $x \to \infty$. $\square$

PROOF OF THEOREM 3.2. We treat α_2. From (2.1), we see that $\lim_a \alpha_1 = \lim_a \alpha_2 = 0$ and in fact,

$$\liminf_a a_1^{-1} \log \alpha_2^{-1} \geqq 1 .$$

It is also true that

$$\begin{aligned} E_1 L_N^- &\geqq -E_1 L_N \\ &\geqq (1 - \alpha_1) \log [(1 - \alpha_1)/\alpha_2] + \alpha_1 \log [\alpha_1/(1 - \alpha_2)] \\ &= \log \alpha_2^{-1} [1 + o(1)] . \end{aligned} \tag{3.5}$$

Thus

$$E_1 L_N^-/a_1 \geqq a_1^{-1} \log \alpha_2^{-1} [1 + o(1)]$$

and the theorm will be established if we show that

$$\lim_a E_1 L_N^-/a_1 = 1 . \tag{3.6}$$

By hypothesis, $P_1(\lim_n L_n/n = \mu_1 < 0) = 1$, so also $P_1(\lim_n L_n^-/n = -\mu_1) = 1$. Since $P_1(\lim_a N = \infty) = 1$, $P_1(\lim_a L_N^-/N = -\mu_1) = 1$. By Theorem 3.1, $P_1(\lim_a N/a_1 = -1/\mu_1) = 1$ and hence

$$P_1(\lim_a L_N^-/a_1 = 1) = 1 . \tag{3.7}$$

In view of (3.7), to establish (3.6), it is enough to show that L_N^-/a_1 is uniformly integrable.

Let, $\mathscr{S}_n \subset \mathscr{F}_n$ denote the sufficient σ-field of all sets invariant under permutations of $(X_1, \cdots, X_n)$. If X is real-valued, $\mathscr{S}_n$ is generated by the order-statistic obtained from $(X_1, \cdots, X_n)$. Let $\mathscr{J}_n = \mathscr{T}_n \cap \mathscr{S}_n$. It follows from Theorem 3.2 of [7] that under any $P_1 \in \mathscr{P}_1$, $\mathscr{T}_n$ and $\mathscr{S}_n$ are conditionally independent given $\mathscr{J}_n$. (Theorem 3.2 of [7] requires two conditions. First: that $\mathscr{S}_n$ be equivariant, which in this case is immediate. Second: that invariance and almost-invariance be equivalent for $\mathscr{P}_1$, which we have assumed to be true.)

Since $R_n = dP_{2n}/dP_{1n} = \exp\{\sum_1^n Z_j\}$ is symmetric in $(X_1, \cdots, X_n)$, so is $\Lambda_n = E_1(R_n \mid \mathscr{T}_n)$. Thus $\Lambda_n = E_1(R_n \mid \mathscr{J}_n)$. We then have

$$\begin{aligned} L_n = \log \Lambda_n &\geqq E_1(\log R_n \mid \mathscr{J}_n) \\ &= E_1(\textstyle\sum_1^n Z_j \mid \mathscr{J}_n) = nE_1(Z_1 \mid \mathscr{J}_n) . \end{aligned}$$

(The last equality follows by symmetry: $E_1(Z_j \mid \mathscr{J}_n) = E_1(Z_1 \mid \mathscr{J}_n), j = 1, \cdots, n.$). Thus

$$\begin{aligned} L_n^- &\leqq n[E_1(Z_1 \mid \mathscr{J}_n)]^- \leqq nE_1(Z_1^- \mid \mathscr{J}_n) \\ &= E_1(s_n \mid \mathscr{J}_n) , \end{aligned} \tag{3.8}$$

where $s_n = Z_1^- + \cdots + Z_n^-$. Letting

$$V_n = E_1(Z_1^- \mid \mathscr{J}_n) = E_1(s_n/n \mid \mathscr{J}_n) ,$$

$L_n^- \leqq nV_n$, so also

$$L_N^- \leqq NV_N . \tag{3.9}$$

We complete the proof by showing that NV_N/a_1 is uniformly integrable.

Let $\mathscr{J}_N = \{\bigcup_n A_n(N = n) : A_n \in \mathscr{J}_n\}$. $\mathscr{J}_N$ is the stopped σ-field for the sequence $\{\mathscr{J}_n\}$. We show that

$$NV_N = E_1(s_N \mid \mathscr{J}_N) . \tag{3.10}$$

From Berk (1969), Proposition 2.2, we obtain

$$\begin{aligned} E_1(s_N \mid \mathscr{J}_N) &= \textstyle\sum_n \{E_1(s_N 1_{(N=n)} \mid \mathscr{J}_n)/P_1(N = n \mid \mathscr{J}_n)\} 1_{(N=n)} \\ &= \textstyle\sum_n \{E_1(s_n 1_{(N=n)} \mid \mathscr{J}_n)/P_1(N = n \mid \mathscr{J}_n)\} 1_{(N=n)} \\ &= \textstyle\sum_n E_1(s_n \mid \mathscr{J}_n) 1_{(N=n)} = \sum_n nV_n 1_{(N=n)} = NV_N , \end{aligned}$$

where we use the conditional independence of $\mathscr{S}_n$ and $\mathscr{T}_n$ to obtain the third equality (note that $(N = n) \in \mathscr{T}_n$). From (3.10), we then have

$$NV_N/a_1 = E_1(s_N/a_1 \mid \mathscr{J}_N) . \tag{3.11}$$

We note that s_N/a_1 is uniformly integrable. (For $0 \leqq s_N/a_1 = (s_N/N)(N/a_1) \to -E_1 Z^-/\mu_1[P_1]$ and $E_1 s_N/a_1 = E_1 N E_1 Z^-/a_1 \to -E_1 Z^-/\mu_1$ by Theorem 3.1). Thus we see from (3.11) and Lemma 3.3 that NV_N/a_1 is also uniformly integrable. This, in conjunction with (3.7) and (3.9) establishes (3.6). □

We do not present an analog of Theorem 2.4, but simply note following easily established fact.

3.4 THEOREM. *If* $P(L_n/n \to 0) = 1$, *then* w.p. 1, $\lim_a N/a = \lim_a EN/a = \infty$.

PROOF. We have $|L_N| \geqq a$. The result follows upon noting that $P(\lim_a L_N/N = 0) = 1$. Hence $P(\lim_a N/a = \infty) = 1$ and then, by Fatou, $\lim_a EN/a = \infty$. $\square$

As one application of the foregoing results, we mention a class of two-sample sequential rank-tests discussed by Berk and Savage [4]; see also [7]. At stage n, independent samples $(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_n)$ are available, from which is obtained the rank-order statistic $R_n = (R_{n1}, \cdots, R_{nn})$. (The coordinates of R_n are the ranks of $Y_1, \cdots, Y_n$ among $\{X_1, \cdots, X_n, Y_1, \cdots, Y_n\}$.) Let F and G denote the respective distributions of X and Y. To test the (null) hypotheses $H_1: G = F$, the following class of SPRTs has been proposed: Choose an alternative hypothesis $H_2: G = \varphi(F)$, where φ is a given df on $[0, 1]$. The distribution of R_n is then determined under both hypotheses and an SPRT can be based on $L_n = \log[P_{2n}(R_n)/P_{1n}(R_n)]$, where $P_{in}(r) = P_i(R_n = r)$ is computed under H_i, $i = 1, 2$. As shown in [4], under certain conditions on φ, w.p. 1 L_n/n converges to a limit μ and in fact, the following large-deviation result holds: For all $\varepsilon > 0$, there is a $c > 0$ and $\rho < 1$ so that $P(|L_n/n - \mu| > \varepsilon) \leqq c\rho^n$. Here P denotes the actual distribution of (X, Y), which need not be given by either hypothesis. (In [4], it is shown that μ is the difference between two Kullback–Leibler information numbers.) When $\mu \neq 0$, the large-deviation result for L_n/n obviously implies the conditions of Theorem 3.1, so the limiting behavior of EN given there applies to this class of sequential rank-tests. Letting μ_i denote the limiting value of L_n/n under H_i, it can be verified that $-\infty < \mu_1 < 0 < \mu_2 < \infty$ and the conditions of Theorem 3.2 are satisfied as well for these SPRTs.

The preceding provides an asymptotic justification of Bhate's "conjecture" for invariant sequential tests. One version of this conjecture is as follows. By "neglecting" the overshoot, we have the approximate equality

$$E_1 L_N \doteq -(1 - \alpha_1)a_1 + \alpha_1 a_2 . \tag{3.12}$$

Letting

$$\lambda_1(n) = E_1 L_n ,$$

Bhate (unpublished; see [12], e.g.) conjectured that a reasonable approximation is given by

$$E_1 L_N \doteq \lambda_1(E_1 N) , \tag{3.13}$$

(where $\lambda_1(\cdot)$ is supposed extended to R in a convenient manner). This yields an approximation n_1 to $E_1 N$, obtained by equating the right-hand expressions in (3.12) and (3.13).

3.5 THEOREM. *Under the conditions of Theorem* 3.2, *Bhate's approximation is asymptotically correct. That is,*

$$\lim_a n_i/E_i N = 1 , \qquad i = 1, 2 .$$

PROOF. We give the argument for n_1. First we show that $\lim_n \lambda_1(n)/n = \lim_n E_1 L_n/n = \mu_1 < 0$. Since $P_1(L_n/n \to \mu_1) = 1$, we see that $P_1(L_n^+/n \to 0) = 1$ and

$P_1(L_n^-/n \to -\mu_1) = 1$. We note that for $x > 0$, $P_1(L_n^+ > x) = P_1(\Lambda_n > e^x) \leqq e^{-x}$ since $E_1 \Lambda_n = 1$. Thus $\{L_n^+\}$ is uniformly integrable and, *a fortiori*, so is $\{L_n^+/n\}$; thus $E_1 L_n^+/n \to 0$. We see from (3.8) that $L_n^-/n \leqq E_1(Z_1^- \mid \mathscr{F}_n)$, so that also $\{L_n^-/n\}$ is uniformly integrable. Thus $E_1 L_n^-/n \to -\mu_1$ and hence $\lim_n \lambda_1(n)/n = \mu_1$.

The equation for n_1 is

$$\lambda_1(n_1) = -a_1[1 + o(1)] \tag{3.14}$$

so clearly $\lim_a n_1 = \infty$ and therefore

$$\lim_a \lambda_1(n_1)/n_1 = \mu_1 . \tag{3.15}$$

From (3.14) we see that $\lim_a \lambda_1(n_1)/a_1 = -1$, hence by Theorem 3.1 and (3.15),

$$\lim_a n_1/a_1 = -1/\mu_1 = \lim_a E_1 N/a_1 . \tag{3.16}$$

It follows that $\lim_a n_1/E_1 N = 1$. □

REMARK 1. Since $\lambda_1(n) \sim n\mu_1$, a modified Bhate approximation would replace $\lambda(E_1 N)$ by $\mu_1 E_1 N$ in (3.13). Similarly, the RHS of (3.12) can be replaced by the no-overshoot approximation

$$(1 - \alpha_1) \log [\alpha_2/(1 - \alpha_1)] + \alpha_1 \log [(1 - \alpha_2)/\alpha_1] , \qquad \text{with} \quad (\alpha_1, \alpha_2)$$

then being replaced by appropriate asymptotic expressions; (e^{-a_2}, e^{-a_1}), e.g. Since the approximations obtained for $E_i N$ all, apparently, have only an asymptotic justification, it seems simplest to use the asymptotic expression given by Theorem 3.1: $E_i N \doteq a_i/|\mu_i|$. One does have to determine μ_i for this, but the necessity of inverting $\lambda_i(\cdot)$ is avoided. Numerical investigation of these approximations would be desirable.

REMARK 2. The proof of Theorem 3.2 shows that $\lim_a E_1 L_N/a_1 = -1$. Since $\lim_a \lambda_1(E_1 N)/E_1 N = \mu_1$, it follows from (3.16) that

$$\lim_a E_1 L_N/\lambda_1(E_1 N) = 1 .$$

That is, Bhate's conjecture (3.13) is asymptotically correct.

4. Asymptotic efficiency of Wald SPRT. We consider again the Wald SPRT for testing two simple hypotheses about i.i.d. data. The theorems of Section 2 allow us to obtain the asymptotic relative efficiency of such tests, as compared with the best non-sequential tests of the same strength. Theorem 2.2 gives the asymptotic behavior of (α_1, α_2) as $a \to \infty$, while Theorems 2.1 and 2.4 give the corresponding behavior of EN. To effect the comparison, we need a corresponding (asymptotic) expression for the sample size required by the best non-sequential test of strength (α_1, α_2). In making the computation, we assume that

$$\lim_a [(\log \alpha_2)/(\log \alpha_1)] = \lambda , \tag{4.1}$$

$0 < \lambda < \infty$, or, in view of Theorem 2.2, that $\lim_a a_1/a_2 = \lambda$.

For a sample size n, the most powerful test of H_1 vs. H_2 rejects H_1 if $L_n > c_n$. We must choose (n, c_n) so that asymptotically, the non-sequential test has strength

(α_1, α_2). That is, we must have

$$\text{(4.2)} \qquad \text{(a)} \quad P_1(L_n > c_n) \doteq \alpha_1 , \qquad \text{(b)} \quad P_2(L_n \leqq c_n) \doteq \alpha_2 ,$$

in the sense that the ratios of the corresponding logarithms tend to unity. (Thus (n, c_n) depends on (a_1, a_2).)

We argue that $c_n = O(n)$, or more exactly, that we may achieve (4.2) by choosing $c_n = n\zeta$, where ζ is a real number depending on H_1, H_2 and λ. The reason for this is to be found in Chernoff's (1952) large-deviation result for a series of i.i.d. summands. Let

$$c_1(t) = \log E_1 e^{tZ} .$$

Since Z is a log-likelihood ratio, $c_1(t) < \infty$, at least for $0 \leqq t \leqq 1$. Chernoff's theorem then says that for $z > E_1 Z$,

$$\text{(4.3)} \qquad \lim_n n^{-1} \log P_1(L_n > nz) = -k_1(z) ,$$

where

$$\text{(4.4)} \qquad k_1(z) = \sup_{-\infty<t<\infty} \{tz - c_1(t)\} ,$$

Similarly, if $z < E_2 Z$,

$$\text{(4.5)} \qquad \lim_n n^{-1} \log P_2(L_n \leqq nz) = -k_2(z) .$$

In view of (4.1), (4.3) and (4.5), we see that a non-sequential test asymptotically of strength (α_1, α_2) is obtained by choosing $c_n = n\zeta(\lambda)$, where $\mu_1 < \zeta(\lambda) < \mu_2$ is the unique soultion (see below) of

$$\text{(4.6)} \qquad k_2(z)/k_1(z) = \lambda$$

and n is choosen so that (4.2a) holds. Thus (4.2) and (4.3) give for n the relation $nk_1(\zeta) = \log \alpha_1^{-1}$ or, in view of Theorem 2.2,

$$\text{(4.7)} \qquad nk_1(\zeta) = a_2 .$$

Hence the sample size required to asymptotically obtain strength (α_1, α_2) is

$$\text{(4.8)} \qquad \nu(a, \lambda) = a_2/k_1(\zeta)$$

(and the corresponding critical value for L_ν is $\zeta(\lambda)a_2/k_1(\zeta)$).

Regarding a solution of (4.6), we note first that

$$c_2(t) = \log \textstyle\int [f_2(x)/f_1(x)]^t f_2(x)\, dx = c_1(t+1) ,$$

from which it follows that

$$\text{(4.9)} \qquad k_2(z) = k_1(z) - z .$$

Moreover, when $z = \mu_1 < 0$, $t\mu_1 - c_1(t)$ is maximum at $t = 0$ (since $\dot{c}_1(0) = \mu_1$, where $\dot{c}_1(t) = dc_1(t)/dt$); hence $k_1(\mu_1) = 0$. Similarly, $k_2(\mu_2) = 0$, so that $k_1(\mu_2) = \mu_2 > 0$. In view of (4.9), (4.6) becomes

$$(1 - \lambda)k_1(z) = z ,$$

which has the solution $\zeta = 0$ if $\lambda = 1$. Otherwise (4.6) gives the equation

$$k_1(z) = z/(1 - \lambda) . \tag{4.10}$$

Since, as shown by Chernoff (1952), if $z \geqq \mu_1$, $k_1(z) = \sup_{t>0} \{tz - c_1(t)\}$, it follows that k_1 is convex in z and increasing for $z > \mu_1$. Since also $k_1(\mu_1) = 0$, $k_1(\mu_2) = \mu_2$ and $0 < \lambda < \infty$ entails $(1 - \lambda)^{-1} \notin (0, 1)$, it follows that the curves defined by the two sides of (4.10) intersect in a single point, whose abscissa, $\zeta(\lambda)$ (say) is in (μ_1, μ_2) and is thus the unique solution of (4.6). In view of (4.10), (4.8) becomes

$$\begin{aligned} \nu(a, \lambda) &= a_2(1 - \lambda)/\zeta(\lambda) , & \lambda \neq 1 \\ &= a_2/\mathscr{K} , & \lambda = 1 . \end{aligned} \tag{4.11}$$

where, as shown by Chernoff (1952),

$$\mathscr{K} = k_1(0) = k_2(0) = -\log \inf_{0<t<1} \textstyle\int f_1^t(x) f_2^{1-t}(x)\, dx .$$

The asymptotic efficiency of the SPRT relative to the corresponding best non-sequential test is now obtained via (4.11) and the theorems of Section 2. If $EZ = \mu \neq 0$, then

$$\begin{aligned} EN &\sim a_2/\mu , & \mu > 0 \\ &\sim -a_1/\mu , & \mu < 0 , \end{aligned}$$

while the corresponding non-sequential test requires $\nu(a, \lambda)$ observations. Hence the ARE of the SPRT is given by

$$\begin{aligned} \text{ARE} = \lim_a \nu(a, \lambda)/EN &= (1 - \lambda)\mu/\zeta(\lambda) , & \mu > 0 , \quad \lambda \neq 1 \\ &= -\mu(1 - \lambda)/\lambda\zeta(\lambda) , & \mu < 0 , \quad \lambda \neq 1 \\ &= |\mu|/\mathscr{K} , & \mu \neq 0 , \quad \lambda = 1 . \end{aligned}$$

If $EZ = 0$, we see from Theorem 2.4 that $EN \sim a_1 a_2$, so then $\lim_a \nu(a, \lambda)/EN = 0$. That is, the SPRT has ARE zero under distributions for which $EZ = 0$. This phenomenon is well known and has been pointed out explicitly by Bechhofer (1960) in the normal case. Similar results (qualitatively) are given by Sakaguchi (1967) for exponential models. However, his formulas appear to be in error, due to his misapproximating the large-deviation probabilities in (4.3) and (4.5) by using the central limit theorem. Other notions of asymptotic efficiency for sequential tests have been considered, notably a Pitman approach, in which the error rates do not tend to zero. See, e.g., Sakaguchi (1967).

5. Acknowledgment. The author is grateful to Professor R. A. Wijsman for a careful reading of an earlier draft of this paper that resulted in the elimination of some serious errors and to Professor W. J. Hall for helpful comments.

REFERENCES

[1] BECHHOFER, R. (1960). A note on the limiting relative efficiency of the Wald sequential probability ratio test. *J. Amer. Statist. Assoc.* **55** 660–663.

[2] Berk, R. H. (1969). Strong consistency of certain sequential estimators. *Ann. Math. Statist.* **40** 1492–1495. Correction (1971) **42** 1135–1137.

[3] Berk, R. H. and Bickel, P. J. (1968). On invariance and almost invariance. *Ann. Math. Statist.* **39** 1573–1576.

[4] Berk, R. H. and Savage, I. R. (1968). The information in a rank-order and the stopping time of some associated SPRTs. *Ann. Math. Statist.* **39** 1661–1674.

[5] 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.

[6] Gundy, R. F. and Siegmund, D. (1967). On a stopping rule and the central limit theorem. *Ann. Math. Statist.* **38** 1915–1917.

[7] 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.

[8] Sakaguchi, M. (1967). Relative efficiency of the Wald SPRT and the Chernoff information number. *Kōdai Math. Sem. Rep.* **19** 138–146.

[9] Siegmund, D. (1967). Some one sided stopping rules. *Ann. Math. Statist.* **38** 1641–1646.

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

[11] Wald, A. (1947). *Sequential Analysis.* Wiley, New York.

[12] Wetherill, G. B. (1966). *Sequential Methods in Statistics.* Methuen, London.

Statistics Center
Rutgers University
New Brunswick, N.J. 08903