The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1176–1185

OPERATING CHARACTERISTICS OF SOME SEQUENTIAL DESIGN RULES[1]

BY ROBERT BOHRER

Research Triangle Institute

1. Introduction. The sequential probability ratio test (SPRT) provides an optimum test of a simple hypotheses $\theta = 0$ vs. a simple alternative $\theta = 1$ in the sense of [6]. To assess how good optimum is, the SPRT user must evaluate the operating characteristics (error probabilities and average sample numbers) of the test. Exact evaluations have been carried out only in case the log likelihood ratio is bounded and satisfies

$$(1) \quad \log (g_0/g_1) = kp, \quad p > 0, \quad k \text{ an integer}, \ -m \leqq k \leqq M(\text{say}),$$

for some (p, m, M), where g_θ is the likelihood when θ is the state of nature; see [4] and [5]. Wald assesses the generality of assumption (1), probably too optimistically, at the beginning of Section 4.A of [5].

Using an SPRT, an experimenter is required to perform the same experiment for all observations. Consider, as Abramson has done [1], the more general situation in which either of two experiments $e = 0, 1$ can be used at any trial to generate a random variable with density $f_{\theta,e} = g_{|\theta-e|}$ when experiment e is used and θ is the state of nature.

Let (e_p, X_p) denote the experiment and its random result at trial p. Assume that given e_p, X_p is independent of the past. Note that this is a sequential design analog of random sampling.

An extension of SPRT to this case is the class of sequential design probability ratio tests (SDPRT). Define the log likelihood ratio after the n trials $\mathbf{z}_{2n} = \{(e_p, x_p): 1 \leqq p \leqq n\}$ by

$$L(\mathbf{z}_{2n}) = \sum_{p=1}^{n} \log (f_{1,e_p}(x_p)/f_{0,e_p}(x_p)).$$

For $\mathbf{L} = (L_0, L_1, L_2)$, the SDPRT $\delta = \delta(\mathbf{L})$ proceeds sequentially, taking an $(n + 1)$st observation only if $L_0 < L(\mathbf{z}_{2n}) < L_1$ using $e = 1$ only if $L_2 < L(\mathbf{z}_{2n}) < L_1$; δ stops and decides $\theta = 0$ if $L(\mathbf{z}_{2n}) \leqq L_0$; δ stops and decides $\theta = 1$ if $L(\mathbf{z}_{2n}) \geqq L_1$. Note that there is no loss of generality in the selection of $e = 1$ for large, rather than small, $L(\mathbf{z}_{2n})$, since experiments need only be relabeled for the alternative case.

Rules of this type have been considered in the literature. For example, SPRT which use experiment e at each stage are SDPRT with $L_2 = L_{1-e}$. Each rule of Chernoff's asymptotically optimal sequence (Theorem 2 of [3]) is a SDPRT with $L_2 = 0$. Whittle [7] conjectures that a Bayes rule for the corresponding

Received 12 April 1967; revised 5 January 1968.

[1] The work was supported by the Air Force Office of Scientific Research and the Army Research Office (Durham).

decision problem is a SDPRT, and he shows that this is "approximately true" in Theorem 4.

Exact evaluation of operating characteristics for SDPRT has been accomplished only in the SPRT case and then only when (1) holds. In [7], Whittle suggests the functional form for average risk of a SDPRT "up to boundary conditions," but he makes no explicit evaluations.

The present work uses Whittle's approach to evaluate operating characteristics of any SDPRT in cases where (1) holds. It thus extends the Wald-Girshick evaluation to the general SDPRT case and hence answers in part Cox's call ([7], pg. 389]) for application of the methods of [7]. A method of average risk evaluation is presented in Section 2. Use in evaluating operating characteristics is discussed in Section 3, and the final section considers a simple example.

2. Evaluation of SDPRT average risks. In this section, the problem is formulated from a Bayesian decision theoretic point of view, and the average risk of a SDPRT is evaluated when (1) holds. Succeeding sections discuss and exemplify both classical and Bayesian consequences of this evaluation.

Let $a = 0, 1$ denote the decision that the state of nature is a and suppose the loss $l(\theta)$ is incurred when action $a \neq \theta$ is taken, while no loss results if $a = \theta$. If the cost c_e is paid each time an observation is taken using experiment e, then the risk of SDPRT δ when θ is the state of nature is

$$r(\theta, \delta) = l(\theta)P_\theta + \textstyle\sum_e c_e \mathcal{E}_\theta N_e, \qquad \theta = 0, 1, \tag{2}$$

where $N_e = N_e(\delta)$ is the random number of trials with experiment e used by δ and $P_\theta = P_\theta(\delta)$ is the probability that δ decides the state of nature is $a \neq \theta$.

Assign relative weights or prior probabilities to the hypothesized θ values by

$$\xi = \Pr\{\theta = 1\} = 1 - \Pr\{\theta = 0\};$$

and call $\xi_{\mathbf{z}_{2n}} = \Pr\{\theta = 1 \mid \mathbf{z}_{2n}\} = [1 + (1 - \xi)\xi^{-1}\exp(-L(\mathbf{z}_{2n}))]^{-1}$ the posterior probability (given $\mathbf{z}_{2n}$). Thus, for each fixed ξ in $(0, 1)$, $\xi_{\mathbf{z}_{2n}}$ is an increasing, continuous function of $L(\mathbf{z}_{2n})$.

It is of interest to note that the minimum of the average risk

$$r(\xi, \delta) = \xi r(1, \delta) + (1 - \xi) r(0, \delta)$$

over the class of all sequential design rules δ is called the Bayes risk at ξ, denoted $\rho(\xi)$, and shown in [7] to satisfy

$$\rho(\xi) = \min [l(1)\xi, l(0)(1 - \xi), \min_e [c_e + \textstyle\int_X \rho(\xi_{e,x}) f_{\xi,e}(x)\, d\mu(x)]],$$

where $f_{\xi,e}(x) = \xi f_{1,e}(x) + (1 - \xi) f_{0,e}(x)$. This is an integral equation with at most one solution. These facts will be used in the final section to verify that a SDPRT is a Bayes rule, i.e., one which minimizes average risk. For fixed ξ and $\mathbf{L}$ define $\xi^{(p)} = [1 + (1 - \xi)\xi^{-1}\exp(-L_p)]^{-1}$, for $p = 0, 1, 2$. Then $\delta(\mathbf{L})$ is the rule $\delta(\xi, \mathbf{L})$ which stops and takes $a = 0$ (or 1) as soon as $\xi_{\mathbf{z}_{2n}} < \xi^{(0)}$ (or $> \xi^{(1)}$) and which uses $e = 1$ only in case $\xi^{(2)} < \xi_{\mathbf{z}_{2n}} < \xi^{(1)}$. Here $\mathbf{L} = (L_0, L_1, L_2)$ with $L_0 \leqq L_2 \leqq L_1$.

In general, the average risk $r(\xi) = r(\xi, \delta(\xi, \mathbf{L}))$ of a SDPRT $\delta(\xi, \mathbf{L})$ at prior probability ξ is shown in [7] to satisfy

$$r(\xi) = l(1)\xi \qquad \text{if } \xi \leqq \xi^{(0)}, \tag{3}$$

$$= c_{e_1} + \textstyle\int_X r(\xi_{\mathbf{z}_2}) f_{\xi, e_1}(x_1)\, d\mu(x_1) \qquad \text{if } \xi^{(0)} < \xi < \xi^{(1)}, \tag{4}$$

$$= l(0)(1 - \xi) \qquad \text{if } \xi \geqq \xi^{(1)}. \tag{5}$$

That is, if no samples are taken, then average risk is the product of ξ-probability that the wrong action is taken and the loss for taking that wrong action. If a sample is taken, then average risk is the cost of the sample plus the expected average risk of the rule $\delta(\xi_{\mathbf{z}_2})$ at posterior probability $\xi_{\mathbf{z}_2}$.

If there is at most one solution $r(\cdot)$ to (3)–(5) and if such a solution can be found, then its value at any point ξ' must be the average risk $r(\xi', \delta(\xi', \mathbf{L}))$. Hence the remainder of this section proves generally that a solution to (3)–(5) is unique (Theorems 1 and 2) and then constructs the solution for cases when equation (1) and an easily checked condition are valid (Theorem 3).

THEOREM 1. *If $\delta(\xi)$ terminates with probability one for each θ and each ξ in* $(0, 1)$, *then there is a unique solution to* (3)–(5).

PROOF. If each of r_1 and r_2 is a bounded ξ-function which satisfies (3)–(5), then $d = r_1 - r_2$ is bounded. Induction based on (3)–(5) establishes, for positive integers n, that

$$d(\xi) = \textstyle\int_{S_n} d(\xi_{\mathbf{z}_{2n}}) g_\xi(\mathbf{z}_{2n})\, d\mu^n(\mathbf{x}_n) \qquad \text{if} \quad \xi^{(0)} < \xi < \xi^{(1)} \tag{6}$$

$$= 0 \qquad \text{otherwise}$$

where $g_\xi(\mathbf{z}_{2n}) = \xi \prod_{p=1}^n f_{1, e_p}(x_p) + (1 - \xi) \prod_{p=1}^n f_{0, e_p}(x_p)$, and

$$S_n = \{\mathbf{x}_n : \xi_{\mathbf{z}_{2n}} \varepsilon\ (\xi^{(0)}, \xi^{(1)}),\ 1 \leqq p \leqq n\}.$$

Let $A = \sup |d(\xi)|$; then $A \leqq A p_n$, where $p_n = \Pr_\xi \{S_n\}$ is the probability that δ does not terminate before the nth trial. By the termination assumption, there is an integer n_0 such that if $n > n_0$, then $p_n < 1$. This requires $A = 0$, to establish the theorem.

THEOREM 2. *If* $\Pr_\theta \{g_0 = g_1\} < 1$ *for each θ, then $\delta(\xi, \mathbf{L})$ terminates with probability one for* $0 < \xi < 1$.

The proof of Theorem 2 follows from a slight modification of the usual Stein argument for the non-design case.

It remains to construct a function $r(\xi)$ which satisfies (3)–(5) when (1) is true. The remainder of this section constructs such a solution under an additional restriction. The construction consists of reducing (3)–(5) to a set of difference equations and solving. The result is summarized in Theorem 3 below.

For convenience, define the continuous, increasing transformation $\zeta = \zeta(\xi) = p^{-1} \log (\xi(1 - \xi)^{-1})$ and the function $R(\zeta) = r(\xi(\zeta))(1 + e^{p\zeta})$. Equation (1) implies that if prior probability has $\zeta(\xi) = (k + \epsilon)$ for some integer k and some ϵ in $(0, 1)$, then for any posterior probability ξ', $\zeta(\xi')$ is in $Z_\epsilon = \{k' + \epsilon : k'$ an integer}, and (3)–(5) defines a separate functional equation for $\{R(\zeta) : \zeta \in Z_\epsilon\}$ for each ϵ in $[0, 1)$.

To describe the functional equation for $R(\zeta)$ on Z_ϵ, note that for ξ such that $\zeta(\xi)$ is in Z_ϵ, $\delta = \delta(\xi, \mathbf{L})$ partitions Z_ϵ as shown schematically in Figure I. That is, for ζ in $E_{e,e}$, if posterior probability is $\xi(\zeta)$ then with probability one δ does not use experiment $1 - e$ for two succeeding trials. With ζ in $E_{e,1-e}$ and posterior probability $\xi(\zeta)$, δ uses experiment e and with non-zero probability takes a next observation using experiment $1 - e$. For ζ in A_a, posterior probability $\xi(\zeta)$ can be attained from some prior probability $\xi = \xi(\zeta')$ with ζ' in $\cup E_{e,e'}$; if attained, sampling is stopped and action a is taken. Note that each A_a contains at most M points and each $E_{e,1-e}$ contains at most m points. Note also that the figure is shown for the case where all $E_{e,e'}$ are non-null but that the succeeding analysis can be carried out formally to provide a solution even when this is not true, e.g., in the SPRT case.

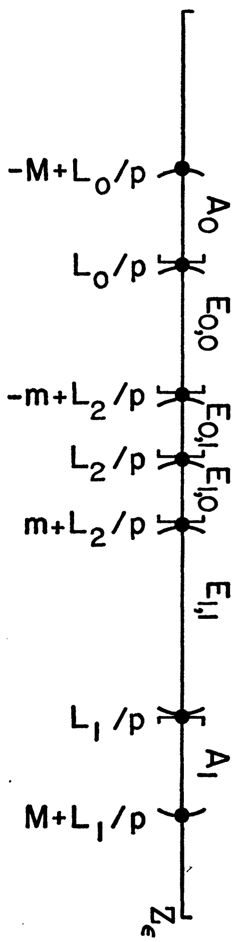

FIG. I

If $r(\xi)$ satisfies (4), then, on $E_{e,e}$, $R(\zeta)$ satisfies

$$\sum_{k=-m}^{M} g_{e,k}R(\zeta - (-1)^e k) = c_e(1 + e^{p\zeta}) \tag{7}$$

where $g_{e,k} = \delta_{0,k} - \Pr_{g_e}\{\log(g_0/g_1) = kp\}$ and $\delta_{0,k}$ is the Kronecker delta. That is, (7) are non-homogeneous, linear difference equations of degree $M + m$ for functions $R_e(\zeta)$ which coincide with $R(\zeta)$ on $E_{e,e}$. A particular solution to (7) is

$$G_e(\zeta) = c_e p\zeta(\Delta_e^{-1} - e^{p\zeta}\Delta_{1-e}^{-1}),$$

where

$$\Delta_s = \int_X (\log g_s g_{1-s}^{-1}) g_s \, d\mu = p \sum_k k \Pr_{g_s}\{\log(g_s/g_{1-s}) = pk\}$$

are the Kullback-Leibler information numbers associated with (g_0, g_1). The sum of G_e and any solution to the homogeneous equation

$$\text{left hand side of} \quad (7) = 0 \tag{8}$$

also satisfies (7). The following lemma determines a class of solutions to (8).

LEMMA. *Consider the difference equation*

$$\sum_{k=a}^{b} g_k R(\zeta + k) = 0 \tag{9}$$

where $g_a \neq 0$. *If* β *is a root of multiplicity* S *for the polynomial* $\sum_{k=0}^{b-a} g_{k+a}x^k$, *then*

$$T(\zeta) = \zeta^s\beta^\zeta, \qquad 0 \leqq s < S, \tag{10}$$

is a solution to (9)

PROOF. Since $g_a \neq 0$, it follows that $\beta \neq 0$ and that

$$\sum_{k=a}^{b} g_k\beta^{\zeta+k} = \beta^{\zeta+a}\sum_{k=0}^{b-a} g_{k+a}\beta^k = 0,$$

i.e., (9) holds for $p = 0$. For $0 < s < S$, suppose (9) holds for non-negative integers less than s. Then

$$\begin{aligned} \sum_{k=a}^{b} g_k(\zeta + k)^s\beta^{\zeta+k} &= \beta^{a-s}[(d^s/dx^s)\sum_{k=0}^{b-a} g_{k+a}x^k]_{x=\beta} \\ &= \beta^{a-s}[(d^s/dx^s)(x - \beta)^S P(x)]_{x=\beta}, \end{aligned} \tag{11}$$

for some polynomial $P(x)$. For $s < S$, $(x - \beta)$ factors each term of the derivative in (11). Hence, the derivative is zero at β, i.e., $T(\zeta)$ satisfies (9). The lemma thus follows by induction.

COROLLARY. *If* $\{T_k(\zeta): 1 \leqq k \leqq b - a\}$ *is the set of functions defined by* (9). *and* (10) *and* $\mathbf{v}$ *is any* $(b - a) \times 1$ *complex vector, then* $\sum_{k=1}^{b-a} v_k T_k(\zeta)$ *satisfies* (9).

For each e, define $\{T_{e,k}(\zeta): 1 \leqq k \leqq M + m\}$, corresponding to (7), as in the corollary. Then for any complex $(M + m) \times 1$ vector $\mathbf{a}_e$,

$$R_e(\zeta, \mathbf{a}_e) = G_e(\zeta) + \sum_{k=1}^{M+m} a_{e,k}T_{e,k}(\zeta) \tag{12}$$

satisfies (7) on $E_{e,e}$. Define

$$R(\zeta) = R(\zeta, \mathbf{a}_0, \mathbf{a}_1) = R_e(\zeta, \mathbf{a}_e), \qquad \zeta \text{ in } A_e \cup E_{e,e} \cup E_{e,1-e}. \tag{13}$$

Suppose $(\mathbf{a}_0, \mathbf{a}_1)$ can be found such that R satisfies the "boundary conditions"

$$(14)\quad R(\zeta) = l(1)e^{p\zeta} \quad \text{on} \quad A_0, \qquad R(\zeta) = l(0) \quad \text{on} \quad A_1,$$
$$\text{and (7) on } E_{1,0} \cup E_{0,1}.$$

Then $r(\xi(\zeta)) = R(\zeta)/(1 + e^{p\zeta})$ satisfies (3)–(5) and by Theorems 1 and 2 is the average risk of $\delta(\xi, \mathbf{L})$ at $\xi = \xi(\zeta)$.

As noted above, (14) is a system of at most $2(M + m)$ conditions on the $\mathbf{a}_e$ each of which, by (12), is a linear equation in the $2(M + m)$ unknowns $\{a_{e,k}\}$. If this system is consistent, then any solution $\{\alpha_{e,k}\}$ defines the solution to (3)–(5), viz.,

$$r(\xi(\zeta)) = R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)/(1 + e^{p\zeta}).$$

The form of these boundary conditions on $E_{0,1} \cup E_{1,0}$ is simplified by the following lemma.

Lemma. *$R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)$ satisfies (14) on $E_{0,1} \cup E_{1,0}$ only in case*

$$(15)\qquad R_0(\zeta, \boldsymbol{\alpha}_0) = R_1(\zeta, \boldsymbol{\alpha}_1) \quad \textit{for each} \quad \zeta \quad \textit{in} \quad E_{0,1} \cup E_{1,0}.$$

Proof. If $R_0(\zeta, \boldsymbol{\alpha}_0) \equiv R_1(\zeta, \boldsymbol{\alpha}_1)$ on $E_{0,1} \cup E_{1,0}$, then (7) is

$$\textstyle\sum_{k=-m}^{M} g_{e,k} R(\zeta - (-1)^e k, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1) = \sum_{k=-m}^{M} g_{e,k} R_e(\zeta - (-1)^e k, \boldsymbol{\alpha}_e) = c_e(1 + e^{p\zeta})$$

on $E_{e,1-e}$, i.e., R satisfies (7).

If $R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)$ satisfies (14) on $E_{1,0} \cup E_{0,1}$ then

$$(16)\qquad R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1) - R_e(\zeta, \boldsymbol{\alpha}_e) \quad \text{satisfies (8) on} \quad E_{e,1-e}.$$

By referring to Figure I, note that at the smallest point ζ^* in $E_{0,1}$(16) says $R_0(\zeta^* + m, \boldsymbol{\alpha}_0) = R_1(\zeta^* + m, \boldsymbol{\alpha}_1)$, i.e., the lemma is true at the smallest point in $E_{1,0}$. If true at all $\zeta^* + s$ with $m \leqq s \leqq s_1 \leqq 2m$, then (16) says $R_0(\zeta^* + s_1, \boldsymbol{\alpha}_0) = R_1(\zeta^* + s_1, \boldsymbol{\alpha}_1)$, i.e., the lemma is true on $E_{0,1}$. Proof for ζ in $E_{1,0}$ is similar and omitted.

Theorem 3 summarizes this development of a solution to (3)–(5).

Theorem 3. *Define $R_e(\zeta, \boldsymbol{\alpha}_e)$ by (12) and $R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)$ by (13). If $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ satisfy the linear system $R_0(\zeta, \boldsymbol{\alpha}_0) = l(1)e^{p\zeta}$ on A_0, $R_1(\zeta, \boldsymbol{\alpha}_1) = l(0)$ on A_1, $R_0(\zeta, \boldsymbol{\alpha}_0) = R_1(\zeta, \boldsymbol{\alpha}_1)$ on $E_{0,1} \cup E_{1,0}$, then $r(\xi(\zeta)) = R(\zeta, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)/(1 + e^{p\zeta})$ satisfies (3)–(5).*

The system in Theorem 3 is consistent if the matrix of $\{\alpha_{e,k}\}$ coefficients has full rank. This rank condition can be made more explicit as follows.

Corollary. *If $Z \subset Z_\epsilon$ let $T_e(Z)$ denote an $(m + M)$ column matrix with rows $\mathbf{T}_e(\zeta)$, the kth element of which is $T_{e,k}(\zeta)$, one row for each ζ in Z with rows arranged in order of increasing ζ. The matrix of $\{\alpha_{e,k}\}$ coefficients is then equivalent to*

$$T = \begin{bmatrix} T_0(A_0) & [0] \\ T_0(E_{0,1}) & -T_1(E_{0,1}) \\ -T_0(E_{1,0}) & T_1(E_{1,0}) \\ [0] & T_1(A_1) \end{bmatrix}.$$

Hence, if T has full rank, then the result of Theorem 3 is true.

3. Evaluation of operating characteristics. Consider any SDPRT δ $(\mathbf{L})$. For fixed $(l(\theta), c_e)$ and any ξ at which the average risk

$$r(\xi, \delta(\mathbf{L})) = r(\xi, \delta(\xi, \mathbf{L}))$$

can be evaluated by Theorem 3, the evaluation is

$$r(\xi, \delta(\xi, \mathbf{L})) = \xi r(1, \delta(\mathbf{L})) + (1 - \xi) r(0, \delta(\mathbf{L})). \tag{17}$$

If there are two such ξ, then (17) and Theorem 3 give two linear equations from which to determine

$$r(\theta, \delta(\mathbf{L})) = l\,(\theta) P_\theta + \textstyle\sum_e c_e E_\theta N_e\,. \tag{18}$$

Evaluation of (18) for three independent vectors $(l(\theta), c_0, c_1)$ gives linear equations from which error probability P_θ and average sampling $\mathcal{E}_\theta N_e$ with experiment e can be determined.

4. A binomial example. The following binomial example illustrates the use, as well as the complexity, of risk evaluation by Theorem 3.

Suppose that g_s is the density of a binomial random variable with success probability p_s where for some $p > 0$

$$p_0 = (e^{2p} - 1)/(e^{3p} - 1), \qquad p_1 = e^p p_0\,.$$

Then (1) holds, since $p^{-1} \log\,(g_0/g_1)$ is either 2 or -1. Suppose also that costs are symmetric in the sense that $l(0) = l\,(1) = 1$ and $c_0 = c_1 = c$. Note that this is perhaps the simplest, non-trivial case possible. Approximate evaluations of SDPRT operating characteristics for this case were derived in [2].

For any positive integer I define $\delta_I(\xi)$ as the SDPRT $\delta(\xi, \mathbf{L})$ with $L_0 = p[\zeta(\xi) - I]$, $L_1 = p[\zeta(\xi) + I]$, $L_2 = p\zeta(\xi)$. As an application of Theorem 3, Result 1 evaluates the average risk of $\delta_I(\xi)$ at those ξ such that $\zeta(\xi)$ is in Z_0, i.e., for $\xi_k = 1/(1 + e^{pk})$.

RESULT 1.

$$\begin{aligned} r(\xi_k\,, \delta_I(\xi_k)) &= r(k) \qquad \textit{if}\ \ k \geqq 0 \\ &= r(-k) \qquad \textit{if}\ \ k < 0, \end{aligned}$$

where

$$\begin{aligned} r(k) &= \xi_k[H(k) + e^{pk} A_1 + A_2 + (-(1 + e^p))^k A_3], \\ H(k) &= cpk(\Delta_1^{-1} - e^{pk}\Delta_0^{-1}), \end{aligned}$$

$$\begin{aligned} A_3(I, c) = A_3 = [cp(p_0 + p_1)(\Delta_0^{-1} + \Delta_1^{-1}) + (p_1 - p_0)(F_1 - F_2)] \\ \cdot[1 + (p_1 - p_0)(G_2 - G_1)]^{-1}, \end{aligned}$$

$$\begin{aligned} A_s &= F_s + G_s A_3 \qquad (s = 1, 2), \\ F_1 &= e^{-Ip}(H(I) - H(I + 1))/(e^p - 1), \\ G_1 &= (2 + e^p) e^{-Ip}(-(1 + e^p))^I/(e^p - 1), \end{aligned}$$

$$F_2 = 1 + (H(I + 1) - e^p H(I))/(e^p - 1),$$

$$G_2 = -(2e^p + 1)(-(1 + e^p))^I/(e^p - 1).$$

In fact, for some experimental costs c there is a "stopping value" I such that $\delta_I(\xi)$ is Bayes.

RESULT 2. *If, in the notation of Result* 1, (c, I) *satisfy*

$$(19) \qquad (p_1 - p_0)(F_2 - F_1) = cp(p_0 + p_1)(\Delta_0^{-1} + \Delta_1^{-1}),$$

then $\delta_I(\xi_k)$ *is Bayes at* ξ_k *for* $-(I + 1) \leqq k \leqq (I + 1)$.

PROOF. Condition (19) is equivalent to $A_3(I, c) = 0$, in which case

$$r(\xi_k) = \xi_k[H(k) + e^{pk}A_1 + A_2].$$

It is verified directly that this function is the solution to the Bayes risk equation

$$(20) \quad \rho(\xi) = \min\,[\xi,\, 1 - \xi,\, c + \min_e \textstyle\sum_{x=0}^{1} \rho(\xi_{e,x}) f_{\xi,e}(x)] \qquad \text{for} \quad -(I + 1) \leqq k \leqq I + 1.$$

See the proof of Result 4 for details.

Next note that $\delta_I(\frac{1}{2})$ is the Chernoff sequential design rule which stops sampling as soon as $|L(\mathbf{z}_{2n})| \geqq Ip$. Result 3 shows a minimax property of these rules for the present examples.

RESULT 3. *If the Chernoff rule* $\delta_I(\frac{1}{2})$ *is Bayes at* $\xi = \frac{1}{2}$ *when cost per trial is c, then it is minimax.*

PROOF. The result follows from the fact that $\rho(\xi)$ is symmetric about $\xi = \frac{1}{2}$. This is true since any risk which can be attained at ξ, say by a rule δ, can also be attained at $1 - \xi$, viz., by δ' which, as a function of observations, stops whenever δ stops but which chooses experiments and actions exactly oppositely. Since ρ is a concave function symmetric about $\frac{1}{2}$, it is maximum there, i.e., $\frac{1}{2}$ is the least-favorable prior probability. Hence, if the Chernoff rule is Bayes at $\xi = \frac{1}{2}$, it is minimax.

Condition (19) specifies $c = c_I$ as a decreasing function of I. Hence Result 2 states that a SDPRT is Bayes at ξ_k for countably many c. This is extended by Result 4.

RESULT 4. *A* SDPRT *is a Bayes rule at* ξ_k *for uncountably many c.*

PROOF. From Result 2, there is a Bayes rule at ξ_k for countably many c, viz., if $I > 0$, then $\delta_I(\xi_k)$ is Bayes at cost c_I such that $A_3(I, c_I) = 0$. Result 4 follows by showing that if $|c - c_I|$ is sufficiently small and cost is c, then $R_c(\xi_k) = r(\xi_k, \delta_I(\xi_k))$ satisfies (18), i.e., that $\delta_I(\xi_k)$ is Bayes at ξ_k for an interval of cost values including c_I. By symmetry, (18) needs be verified only for non-negative integers k. Note that

$$|r(\xi, \delta_I(\xi), c) - r(\xi, \delta_I(\xi), c_I)| = |c - c_I|\, \mathcal{E}_\xi N(\delta_I(\xi)) < M\,|c - c_I|,$$

where $M = \max_{0 \leqq k \leqq I} \mathcal{E}_{\xi_k} N(\delta_I(\xi_k))$. From Result 1,

$$R_{c_I}(\xi_{I-1})/\min\,(\xi_{I-1}, 1 - \xi_{I-1}) = 1 - (1 - e^{-p})c_I p(\Delta_1^{-1} + e^{Ip}\Delta_0^{-1}) < 1,$$

for $I \geqq 0$, so concavity of the Bayes risk proves, for $|c - c_I|$ sufficiently small, that $R_c(\xi_k) < \min(\xi_k, 1 - \xi_k)$ for $0 \leqq k \leqq I - 1$, i.e., R_c satisfies the Bayes "optimum stopping" condition. The "optimum experiment" condition at ξ_k can be written in terms of $s(\xi_k) = R_c(\xi_k)/\xi_k$, viz.,

$$D_{k,c} = c/\xi_k + (1 - p_1)s(\xi_{k-2}) + p_1 s(\xi_{k+1}) - s(\xi_k) \geqq 0,$$

for $0 \leqq k \leqq I - 1$. By symmetry, $D_{0,c} = 0$ while by Result 2, if $A_3 = 0$ then

$$D_{1,c_I} = D_{1,c_I} - D_{0,c_I} = (e^p - 1)[c_I + (1 - p_1)(F_2 - F_1)] > 0,$$

since $(F_2 - F_1) = cp(p_0 + p_1)(\Delta_1^{-1} - \Delta_0^{-1})/(p_1 - p_0) > 0$; and

$$D_{k,c_I} = c_I(\Delta_0 - \Delta_1)(e^{pk}\Delta_0^{-1} - \Delta_1^{-1}) > 0 \quad \text{for} \quad k \geqq 2.$$

In summary, $D_{0,c} = 0$ and $D_{k,c} = D_{k,c_I} + (c - c_I)f_{k,I}$, $-I \leqq k \leqq I$, where $f_{k,I}$ is bounded for $-I \leqq k \leqq I$; i.e., if $|c - c_1|$ is sufficiently small, then $D_{k,c} \geqq 0$ for $-I \leqq k \leqq I$, and R_c satisfies the "optimum experiment" condition at each ξ_k. Hence R_c satisfies (20) and $\delta_I(\xi_k)$ is Bayes at ξ_k when cost is c.

COMPUTATIONS. Table I presents exact evaluations of Chernoff-rule operating characteristics for some binomial situations wherein these SDPRT are minimax. The first three columns define the problems evaluated by specifying success probabilities and cost per experiment. The fourth column gives the "stopping value" I, such that the Chernoff rule stops as soon as $|L(\mathbf{z}_{2n})| > Ip$. The next six columns evaluate average losses and average sampling per experiment for the minimax rule. The penultimate column gives the average risk, $.5(r(0, \delta_I(\frac{1}{2})) + r(1, \delta_I(\frac{1}{2})))$ of this "best" SDPRT. This can be compared with the corresponding average risk, tabulated in the final column, of the SPRT δ_I^0 which uses $e = 0$ for each trial and which stops as soon as $|L(\mathbf{z}_{2n})| > Ip$.

For the relatively large experimental costs (relatively small average sample numbers) considered, the SDPRT is seen to give little improvement over the more easily applied SPRT. This improvement can be compared with the asymptotic case wherein experimental cost approaches zero. In this case, the ratio of risk for the minimax rule $\delta_I(\frac{1}{2})$ to that of δ_I^0 approaches $R = 2\Delta_1/(\Delta_0 + \Delta_1)$, which is evaluated below for the cases considered:

$p =$	.25	.50	.75
$R =$	.9586	.9184	.8806.

Note that, if θ is true, then experiment θ is the more informative one, in the sense of Kullback and Leibler exploited in [3]. In this regard, it is of interest to compare the $\mathcal{E}_\theta N_\theta$ and $\mathcal{E}_\theta N_{1-\theta}$ columns and to note that, on average, sampling with the more informative experiment predominates, regardless of the true state of nature.

Acknowledgment. I thank W. Hoeffding for useful insights. For help in making the paper more correct and comprehensible I thank H. Chernoff and a referee.

TABLE I

Operating Characteristics of Some Minimax Sequential Design Rules with Binomial Random Vavirables

	p	c	I	P_0	$\mathcal{E}_0 N_1$	$\mathcal{E}_0 N_0$	P_1	$\mathcal{E}_1 N_1$	$\mathcal{E}_1 N_0$	Risks	
										SDPRT	SPRT
	.25	.02409	2	.26046	1.76385	1.02438	.44857	1.31555	1.76407	.42518	.42531
	.25	.01672	3	.34717	3.91000	1.36473	.26513	1.75278	3.39083	.39324	.39359
	.25	.01250	4	.23079	5.71394	2.95121	.27066	4.36502	4.75453	.36186	.36249
	.25	.00972	5	.21905	8.62610	3.84631	.20082	5.93649	6.48614	.33094	.33185
$p_0 = .581$ $p_1 = .746$	.25	.00773	6	.16550	11.37518	5.33966	.17385	9.13000	8.00891	.30059	.30177
	.25	.00623	7	.14011	14.79394	6.45737	.13639	11.76938	9.57397	.27097	.27239
	.25	.00506	8	.10960	18.12115	7.70395	.11158	15.28744	11.00445	.24234	.24393
	.25	.00411	9	.08871	21.81714	8.73483	.08834	18.61157	12.30896	.21494	.21664
	.25	.00335	10	.06998	25.62451	9.71628	.07045	22.33533	13.33154	.18901	.19078
	.50	.04311	2	.15366	1.67093	.82466	.31135	1.35966	1.67117	.35164	.35250
	.50	.02738	3	.19357	3.69736	1.03880	.13252	1.71271	2.81608	.28989	.29189
	.50	.01810	4	.08837	5.00639	1.87717	.11365	4.06352	3.52998	.23203	.23485
	.50	.01204	5	.07100	7.26250	2.20718	.06115	5.15224	4.25978	.17971	.18293
$p_0 = .491$ $p_1 = .814$	.50	.00793	6	.03824	8.95798	2.66125	.04210	7.49283	4.69861	.13461	.13776
	.50	.00515	7	.02592	11.14838	2.89590	.02438	9.07858	5.02960	.09768	.10052
	.50	.00330	8	.01496	12.98193	3.10622	.01554	11.23605	5.27625	.06897	.07127
	.50	.00208	9	.00949	15.04440	3.23089	.00924	13.03925	5.44180	.04757	.04935
	.50	.00130	10	.00564	16.92639	3.32384	.00572	15.07428	5.55457	.03219	.03352
	.75	.05382	2	.08385	1.55343	.63718	.20442	1.34887	1.55305	.28116	.28338
	.75	.02938	3	.09914	3.36391	.75327	.05941	1.59476	2.28796	.19679	.20075
	.75	.01583	4	.02981	4.21266	1.14764	.04334	3.57220	2.60847	.12791	.13204
	.75	.00823	5	.02076	5.95070	1.24937	.01634	4.26947	2.86128	.07756	.08097
$p_0 = .410$ $p_1 = .868$	.75	.00414	6	.00782	6.96970	1.37168	.00923	6.01251	3.00546	.04448	.04687
	.75	.00203	7	.00438	8.52093	1.41497	.00392	7.01520	3.05395	.02448	.02599
	.75	.00098	8	.00186	9.67694	1.44798	.00200	8.53513	3.06464	.01307	.01400
	.75	.00047	9	.00095	10.91470	1.46195	.00089	9.69409	3.30847	.00687	.00739
	.75	.00022	10	.00042	12.17493	1.47059	.00044	11.09650	3.40063	.00357	.00384

REFERENCES

[1] Abramson, L. R. (1966). Asymptotic sequential design of experiments with two random variables. *J. Roy. Statist. Soc. Ser. B* **28** 73–87.

[2] Bohrer, Robert (1966). On Bayes sequential design with two random variables. *Biometrika* **53** 469–75.

[3] Chernoff, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–70.

[4] Girshick, M. A. (1946). Contributions to the theory of sequential analysis I. *Ann. Math. Statist.* **17** 123–143.

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

[6] Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math Statist.* **19** 326–39.

[7] Whittle, P. (1965). Some general results in sequential design (with discussion). *J. Roy. Statis. Soc. Ser. B* **27** 371–94.