The Annals of Statistics
1992, Vol. 20, No. 1, 555–561

TWO-SIDED SEQUENTIAL TESTS

BY LAWRENCE D. BROWN AND EITAN GREENSHTEIN

Cornell University and Bar Ilan University

Let X_i be i.i.d. $X_i \sim F_\theta$. For some parametric families $\{F_\theta\}$, we describe a monotonicity property of Bayes sequential procedures for the decision problem H_0: $\theta = 0$ versus H_1: $\theta \neq 0$. A surprising counterexample is given in the case where F_θ is $N(\theta, 1)$.

Introduction and preliminaries. Let $X_1, X_2, \dots, X_m$, $m \leq \infty$, be i.i.d. with $X_i \sim F_\theta$, where F_θ is a one-parameter exponential family. Assume X_i are canonical observations and θ the canonical parameter. We will consider the two-sided sequential testing problem H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$. For recent discussion of this problem see Emerson and Fleming (1989) and Jennison and Turnbull (1990).

The action space in such a problem is a pair (N, τ), where $N = 1, 2, \dots, m$ is the stopping time and τ is the terminal decision, is 0 or 1. The loss function is denoted $\tilde{L}(\theta, (N, \tau)) = c(N - 1) + L(\theta, \tau)$. Here c represents the cost of one observation; the cost of the first observation is 0 and it is always taken.

In this paper we assume the following type of loss function:

(1) $L(\theta, 0)$ is nonincreasing for $\theta < \theta_0$ and nondecreasing for $\theta \geq \theta_0$, $L(\theta, 1)$ is nondecreasing for $\theta < \theta_0$ and nonincreasing for $\theta \geq \theta_0$.

We will consider only procedures based on $S_n = X_1 + \cdots + X_n$, as in Brown, Cohen and Strawderman (1979) (to be referred to in the sequel as B.C.S.). This can be justified by sufficiency and transitivity of S_n. A procedure Δ consists of a set of nonnegative functions $\delta_{in}(s_n)$ defined for every s_n such that $\sum_{i=0}^{2} \delta_{in}(s_n) = 1$. The quantities $\delta_{in}(s_n)$, $i = 0, 1, 2$, represent, respectively, the conditional probability of accepting H_0, accepting H_1 and taking another observation when n observations have been taken and $S_n = s_n$. Such a procedure Δ implicitly defines the stopping rule, N.

Define the risk function $R(\theta, \Delta) = E_\theta L(\theta, \Delta)$. The Bayes risk for a prior $\pi(\theta)$ is

$$r(\pi, \Delta) = \int R(\theta, \Delta)\, d\pi(\theta).$$

Let R be the real line and η some additional point. Denote $\overline{R} = R \cup \{\eta\}$. Map to the event $N = n_0$, $S_1 = s_1, \dots, S_{n_0} = s_{n_0}$, the point $(s_1, \dots, s_{n_0},$

Received March 1990; revised February 1991.

Research supported in part by NSF Grant DMS-88-09016.

AMS 1980 *subject classifications*. Primary 62L10; secondary 62C99.

Key words and phrases. Sequential testing, total-positivity, complete class, monotone procedures.

$\eta, \eta, \dots)$ in $\overline{R}_1 \times \overline{R}_2 \times \cdots \times \overline{R}_m$. Thus, if a point has nth coordinate η, this means that sampling has stopped before stage η. This mapping induces a measure on $\overline{R}_1 \times \cdots \times \overline{R}_m$, under a parameter θ and a procedure Δ, denote it by $H_{\theta\Delta}$. Denote $H_{\pi\Delta}$ the measure defined by: $H_{\pi,\Delta}(dx) = \int H_{\theta\Delta}(dx)\, d\pi(\theta)$.

DEFINITION 1. A sequential procedure Δ for the preceding problem is said to be monotone under the prior π if for every n, there exist numbers $-\infty \le a_1^n \le a_2^n \le a_3^n \le a_4^n \le \infty$ such that: For almost every real value s_n under $H_{\pi\Delta}$, $\delta_{0n}(s_n) = 0$ if $s_n < a_2^n$ or $s_n > a_3^n$; $\delta_{1n}(s_n) = 0$ if $a_1^n < s_n < a_4^n$; $\delta_{2n}(s_n) = 0$ if $s_n < a_1^n$, $s_n > a_4^n$ or $a_2^n < s_n < a_3^n$. Certain obvious randomizations are allowed when $s_n = a_i^n$, $i = 1, 2, 3, 4$.

It was proven in B.C.S. that in a two-sided testing problem, if the distributions, loss function and prior distribution are all symmetric, then every Bayes procedure is monotone. In this work we will investigate what happens when symmetry is not assumed. The following conjecture is also implicit in B.C.S.: When the X_i are i.i.d. normal with mean θ, every Bayes procedure is monotone. In Section 1 we show this conjecture is false. In Section 2 we describe a (weaker) complete class property.

1. Counterexample. We consider the two-sided hypothesis testing problem H_0: $\theta = 0$ versus H_1: $\theta \ne 0$ for the mean of a normal distribution. The example will be of a Bayes procedure that decides H_1 for values of X_1 belonging to three disjoint intervals. This contradicts the monotonicity conjecture expressed in the previous section.

Let $\theta_1 < \theta_0 = 0 < \theta_2$. Denote $\Theta_0 = \{0\}$ and $\Theta_1 = \{\theta_1, \theta_2\}$. Consider a two-stage testing problem (i.e., $m = 2$) with

$$\tilde{L}(\theta, (N, 0)) = \begin{cases} c(N-1) + 1, & \text{if } \theta \in \theta_1, \\ c(N-1), & \text{if } \theta \in \theta_0, \end{cases}$$

$$\tilde{L}(\theta, (N, 1)) = \begin{cases} c(N-1), & \text{if } \theta \in \theta_1, \\ c(N-1) + 1, & \text{if } \theta \in \theta_0. \end{cases}$$

Let the prior give mass (π_1, π_0, π_2) to the points $\theta_1, 0, \theta_2$, respectively. Denote by x_1^* the unique value such that $\operatorname{Max}_{x_1} P(\Theta_0 | x_1) = P(\Theta_0 | x_1^*)$. Here $P(\Theta_0 | x_1)$ denotes the posterior probability of Θ_0 given $X_1 = x_1$. [The uniqueness of x_1^* follows by showing, similarly to Lemma 1 (in Section 2), that $\rho_0(s) - W$ changes sign twice at most in the strong sense for every W and if there are two sign changes the function is first positive.]

Suppose

(i) $$P(\Theta_0 | x_i^*) = \tfrac{1}{2}.$$

Let $\rho_\tau(x_1) = 1 - P(\Theta_\tau | x_1)$ and let $\beta_1(s)$ denote the conditional Bayes risk of a procedure which takes two observations, conditional on $S_1 = s$. Let

$$\Delta(x_1) = \mathrm{Min}(\rho_1(x_1), \rho_0(x_1)) - (\beta_1(x_1) - c).$$

In our case by (i), $\rho_1(x_1) \le \rho_0(x_1)$ and $\Delta(x_1)$ can be written as

$$\Delta(x_1) = P(\Theta_0|x_1) - E\big(\mathrm{Min}(P(\Theta_0|x_1 + X_2), P(\Theta_1|x_1 + X_2))\big|x_1\big).$$

Notice that $\Delta(x)$ does not depend on c and that the Bayes action conditional on $X_1 = x_1$ is to take one more observation if and only if $\Delta(x_1) - c \ge 0$. Let x_1^{**} be the unique value such that $\mathrm{Max}_{x_1} \Delta(x_1) = \Delta(x_1^{**})$. [It can be shown that x_1^{**} is unique by writing $\rho_0(s) - \beta_1(s) - W$ similar to (i) in Lemma 3, and showing it has at most two sign changes for every real number W. Notice that this method applies only when the horizon is of size 2; it seems that the analogous function $\rho_0(s) - \beta_1^n(s) - W$, as defined in Section 2, can have more than one local maximum when $n > 2$.] Suppose that

$$\text{(ii)} \qquad x_1^{*} \ne x_1^{**}.$$

We now show that when (i) and (ii) are satisfied for some $\theta_1, \theta_2, \pi_1, \pi_2$, then a counterexample can be constructed. Define a new problem with $\tilde{\theta}_1$, $\tilde{\theta}_2$, $\tilde{\pi}_1$, $\tilde{\pi}_0$, $\tilde{\pi}_2$, $\tilde{c}$, where $\theta_i = \tilde{\theta}_i$, $\tilde{\pi}_1 = \pi_1 - (\varepsilon/2)$, $\tilde{\pi}_0 = \pi_0 + \varepsilon$, $\tilde{\pi}_2 = \pi_2 - (\varepsilon/2)$.

Define $E_0 = \{x_1 | P(\Theta_0|x_1) \ge P(\Theta_1|x_1)\}$. E_0 is an interval by Lemma 1 of Section 2. Choose $\varepsilon > 0$ small enough so that $\tilde{x}^{**} \notin E_0$. Such an $\varepsilon > 0$ must exist by (ii) and by continuity considerations. Also by continuity consideration, there exists an $\varepsilon_1 > 0$ such that $\Delta(\tilde{x}_1^{**}) - \varepsilon_1 > \Delta(x_1)$ for every $x_1 \in E_0$. Take $\tilde{c} = \Delta(x_1^{**}) - \varepsilon_1$. In our new problem there are two disjoint separated intervals containing $\tilde{x}_1^{*}$ and $\tilde{x}_1^{**}$, respectively, such that: The Bayes procedure decides H_0 if x_1 is in the interval containing $\tilde{x}_1^{*}$; it takes one more observation if x_1 is in the interval containing $\tilde{x}_2^{*}$ and it decides H_1 otherwise. Hence there are three separate intervals where the Bayes procedure decides H_1 for values x_1 in these intervals. Such a procedure is not monotone.

A numerical example is the following: Take $\theta_1 = -1$, $\theta_0 = 0$, $\theta_2 = 2$. Straightforward calculations show that in order to get

$$x_1^{*} = 0 \quad \text{and} \quad P(\Theta_0|x_1^{*}) = 1/2,$$

we should choose $\pi_1 = 0.219$, $\pi_0 = 0.240$, $\pi_2 = 0.539$. Using numerical integration we get

$$\Delta(x_1^{*}) = \Delta(0) = 0.259, \qquad \Delta(0.2) = 0.266;$$

that is, $x_1^{*} \ne x_1^{**}$. Table 1 shows some further values of $P(\Theta_0|x_1)$ and $\Delta(x_1)$.

Remarks. (i) As noted before, when the size of the horizon is greater than 2, it seems there can be more than one local maximum to $\Delta(\cdot)$. If so, examples can be given with more than three disjoint intervals where H_1 is accepted by a Bayes procedure. (ii) In principle, the above example leaves unsettled the monotonicity conjecture for possibly open-ended procedures. However, in light of the preceding results, we believe that the monotonicity conjecture is false in this case also. (iii) In constructing the counterexample we have used the fact that the cost of the first observation is zero, and hence the first observation is

TABLE 1

x_1	$P(\Theta_0 \mid x_1)$	$\Delta(x_1)$
-0.4	0.466	0.206
-0.3	0.480	0.223
-0.1	0.497	0.250
$x_1^* = 0.0$	0.500	0.259
0.1	0.497	0.265
$x_1^{**} \approx 0.2$	0.489	0.266
0.3	0.475	0.263
0.4	0.457	0.256
0.5	0.433	0.245

always taken. It is not clear whether a counterexample can be given when the cost is c per observation including the first one.

2. Complete class theorem. Assume X_i are i.i.d. normal with unknown mean θ. As we have shown, for two-sided sequential testing it is not true that every Bayes procedure is monotone (unless one assumes further symmetry). In view of this, the following Theorem 1 seems to give the best possible general monotonicity statement for the class of Bayes procedures in such a case.

We will first review some facts about total-positivity which will be needed for the proof of Theorem 1. Some references on this subject are Karlin (1968) and Brown, Johnstone and MacGibbon (1981).

DEFINITION 1. The function $\varphi(x)$: $R \to R$ changes signs at most n times if and only if there exist $-\infty = a_0 < a_1 < \cdots < a_n < a_{n+1} = \infty$ such that $\varphi(x)$ preserves its sign on (a_i, a_{i+1}), $i = 0, \dots, n$, that is, it is either nonnegative or nonpositive on each interval.

Let $\{G_\theta: \theta \in \theta \subseteq R\}$ be a family of distributions on the real line.

DEFINITION 2. $\{G_\theta\}$ is TP_n if for any function $\varphi(x)$, that changes sign at most $n-1$ times, $h(\theta) = E_\theta \varphi(x)$ changes signs at most $n-1$ times, and if it does change sign $n-1$ times, then it does so in the same order as φ. $\{G_\theta\}$ is STP_n if, in addition, for any φ as above which is not identically zero, the function $h(\theta)$ changes sign at most $n-1$ times in the stronger sense that there are $a_0 = -\infty < a_1 \le \cdots \le a_{n-1} < \infty = a_n$ such that $h(\theta)$ preserves its sign on (a_i, a_{i+1}) and $h(\theta)$ can be zero only at a_i, $i = 1, \dots, n-1$.

Suppose $X_i \sim F_\theta$ are i.i.d. and $\pi(\theta)$ is a prior distribution on θ. Denote by $\nu^{n+1}(ds_{n+1} \mid s_n)$, the conditional distribution of S_{n+1} given $S_n = s_n$ and prior $\pi(\theta)$.

In the sequel we will require $\nu^{n+1}(ds_{n+1}|s_n)$ to be STP_3 with respect to the parameter s_n. Denote by $F_\theta^n(ds)$ the distribution of s_n under θ, and let $F_\theta^{n+1}(ds_{n+1}|s_n)$ be its conditional distribution conditional upon $S_n = s_n$.

PROPOSITION 1. *Suppose for some $\theta_0 \in \theta$, $F_{\theta_0}^{n+1}(ds_{n+1}|S_n = s_n)$ is $(S)TP_k$ with respect to the parameter s_n. Then for every $\pi(\theta)$, $\nu^{n+1}(ds_{n+1}|s_n)$ is $(S)TP_k$.*

PROOF. See Greenshtein (1990). □

PROPOSITION 2. *$F_{\theta_0}^{n+1}(ds_{n+1}|S_n = s_n)$ is STP_∞ (i.e., STP_k for $k = 1, 2, \dots$) in s_n for every n when $\{F_\theta\}$ is any one of the following: binomial $\theta = p$; exponential $\theta = \lambda$ when λ^{-1} is its expectation; Poisson $\theta = \lambda$, where λ is its expectation; geometric $\theta = p$; normal $\theta = \mu$, where μ is its expectation.*

PROOF. Immediate from Chapters 7 and 8 in Karlin (1968). □

THEOREM 1. *Consider a two-sided sequential testing problem. Assume for every $\pi(\theta)$, $\nu^{n+1}(ds|s_n)$ is STP_3 in the parameter s_n. Then every Bayes procedure $\Delta = \{\delta_{in}\}$ is of the following type: There exist numbers $a_2^n \le a_3^n$ such that $\delta_{0n}(s_n) = 1$ if $s_n \in (a_2^n, a_3^n)$ and $\delta_{0n}(s_n) = 0$ if $s_n \notin [a_2^n, a_3^n]$ for almost every s_n under $H_{\pi\Delta}$.*

INTERPRETATION. The theorem says that the Bayes procedure stops and accepts H_0 whenever $S_n \in (a_2^n, a_3^n)$. It cannot continue sampling for any such S_n. It may also accept if $S_n = a_2^n$ or a_3^n. For values of S_n outside of $[a_2^n, a_3^n]$, the procedure might either continue sampling or stop and decide H_1, but it cannot stop and accept H_0.

Before proving the theorem, some further lemmas and notation are needed. Let

$$\rho_\tau^n(s_n) = \int L(\theta, \tau)\pi^n(d\theta|s_n), \qquad \tau = 0, 1.$$

Here $\pi^n(d\theta|s_n)$ denotes the posterior distribution given $S_n = s_n$. Assume a finite horizon where $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+k}$ are the available observations and $X_1, \dots, X_n$ have already been observed $(n + k = m)$. Denote by $\beta_n^{n+k}(s)$, the conditional additional Bayes risk of a procedure that takes at least one more observation and proceeds optimally, conditional on $S_n = s_n$. (The definitions of ρ and β are extensions of the definitions in Section 1.) Then,

$$\beta_n^{n+k}(s_n) = \int \mathrm{Min}\big(c + \beta_{n+1}^{n+k}(s), c + \rho_0^{n+1}(s), c + \rho_1^{n+1}(s)\big)\nu^{n+1}(ds|s_n).$$

Of course,

$$\beta_n^{n+1}(s_n) = \int \operatorname{Min}\left(c + \rho_0^{n+1}(s), c + \rho_1^{n+1}(s)\right)\nu^{n+1}(ds|s_n).$$

LEMMA 1. *For every real number W, $\rho_1^n(s) - \rho_0^n(s) - W$ changes sign at most twice. If there are two sign changes, then it is first negative. Moreover the function is zero only at its crossing points.*

PROOF. From condition (1) stated in the Introduction, the proof follows as in Karlin (1955). □

LEMMA 2. $\rho_\tau^n(s_n) = \int \rho_\tau^{n+1}(s)\nu^{n+1}(ds|s_n)$, $\tau = 0, 1$.

PROOF. See Sobel (1952). □

LEMMA 3. $\beta_n^{n+k}(s) - \rho_0^n(s) + W$, *changes sign at most twice for every real number W. If there are two sign changes, it is first negative. Moreover the function is zero only at its crossing points.*

PROOF. The proof is by induction on the number of remaining observations. The general induction step is as follows:

$$\begin{aligned}
&\beta_n^{n+k}(s_n) - \rho_0^n(s_n) + W \\
&\quad = \beta_n^{n+k}(s_n) - \int \rho_0^{n+1}(s)\nu^{n+1}(ds|s_n) + W \\
&\quad = \int \operatorname{Min}\left(c + \beta_{n+1}^{n+k}(s) - \rho_0^{n+1}(s) + W, c + w,\right. \\
&\qquad\qquad \left. c + \rho_1^{n+1}(s) - \rho_0^{n+1}(s) + W\right)\nu^{n+1}(ds|s_n).
\end{aligned} \tag{i}$$

All the functions in the brackets change signs twice at most and in the right order. The last fact is true by the induction hypothesis and by Lemma 1. Hence the Min of the three functions changes signs twice at most and if it does, it is first negative. The desired conclusion follows now by STP_3 of $\nu^{n+1}(\cdot|s_n)$. □

PROOF OF THE THEOREM. For the finite horizon case, the proof follows from Lemmas 1 and 3 letting $W = 0$. For the infinite horizon we proceed as in Chow, Robbins and Siegmund (1971). Define $\beta_n^\infty(s_n)$ to be the additional risk of a procedure that takes at least one more observation and proceeds optimally conditional on $S_n = s_n$. For the M truncated problem we get by their Theorem 4.4 and 4.7 that $\beta_n^M(s_n) \to_{M\to\infty} \beta_n^\infty(s_n)$. Thus $\beta_n^\infty(s_n) - \rho_0^n(s_n)$ has at most two sign changes and if there are two sign changes, it is first negative. Now (i) holds replacing $\beta_n^{n+k}(s_n)$ by $\beta_n^\infty(s_n)$. Thus, using STP_3 of $\nu_{s_n}^{n+1}$, we conclude

$\beta_n^{\infty} - \rho_0^n(s_n)$ is zero only at its crossing points. Now the conclusion follows as in the finite horizon case. □

Acknowledgments. The authors would like to thank Dr. Fu-Shing Hsieh for helping with the numerical work. The second author would like to thank Professor Yosi Rinott for a helpful discussion concerning total-positivity.

REFERENCES

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1979). Monotonicity of Bayes sequential tests. *Ann. Statist.* **7** 1222–1230.

BROWN, L. D., JOHNSTONE, I. M. and MACGIBBON, K. B. (1981). Variation diminishing transformations. *J. Amer. Statist. Assoc.* **76** 824–832.

CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations*. Houghton Mifflin, Boston.

EMERSON, S. S. and FLEMING, T. R. (1989). Symmetric group sequential test designs. *Biometrics* **45** 905–923.

GREENSHTEIN, E. (1990). On some decision theory problems in sequential analysis. Ph.D. dissertation, Cornell Univ.

JENNISON, C. and TURNBULL, B. W. (1990). Statistical approaches to monitoring of medical trials: A review and commentary. *Statist. Sci.* **5** 299–317.

KARLIN, S. (1955). Decision theory for polya type distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 115–128. Univ. California Press, Berkeley.

KARLIN, S. (1968). *Total Positivity*. Stanford Univ. Press.

SOBEL, M. (1952). An essentially complete class of decision function for certain standard sequential problems. *Ann. Math. Statist.* **23** 319–337.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

DEPARTMENT OF MATHEMATICS
BAR ILAN UNIVERSITY
RAMAT GAN 52900
ISRAEL