Two-Sided Sequential Tests
Author(s): Lawrence D. Brown and Eitan Greenshtein
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/2242177
Accessed: 25/03/2010 15:48

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ims.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to The Annals of Statistics.
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 $F_\theta$ is $N(\theta, 1)$.

Introduction and preliminaries. Let $X_1, X_2, \ldots, X_m$, $m \leq \infty$, be i.i.d. with $X_i \sim F_\theta$, where $F_\theta$ is a one-parameter exponential family. Assume $X_i$ are canonical observations and $\theta$ 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 $(N, \tau)$, where $N = 1, 2, \ldots, m$ is the stopping time and $\tau$ is the terminal decision, is 0 or 1. The loss function is denoted $\bar{L}(\theta, (N, \tau)) = c(N - 1) + L(\theta, \tau)$. 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:

$$L(\theta, 0) \text{ is nonincreasing for } \theta < \theta_0 \text{ and nondecreasing for } \theta \geq \theta_0,$$

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

(1)

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 $\Delta$ consists of a set of nonnegative functions $\delta_{i\theta}(s_n)$ defined for every $s_n$ such that $\Sigma_{i=0}^2 \delta_{i\theta}(s_n) = 1$. The quantities $\delta_{i\theta}(s_n), i = 0, 1, 2$, 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 $\Delta$ implicitly defines the stopping rule, $N$.

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

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

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

Received March 1990; revised February 1991.
Research supported in part by NSF Grant DMS-88-09016.
Key words and phrases. Sequential testing, total-positivity, complete class, monotone procedures.

555
\( \eta, \eta, \ldots \) in \( \bar{R}_1 \times \bar{R}_2 \times \cdots \times \bar{R}_m \). Thus, if a point has \( n \)th coordinate \( \eta \), this means that sampling has stopped before stage \( \eta \). This mapping induces a measure on \( \bar{R}_1 \times \cdots \times \bar{R}_m \), under a parameter \( \theta \) and a procedure \( \Delta \), denote it by \( H_{\theta \Delta} \). Denote \( H_{\pi \Delta} \) the measure defined by: \( H_{\pi \Delta}(dx) = \int \theta \Delta(dx) d\pi(\theta) \).

**Definition 1.** A sequential procedure \( \Delta \) for the preceding problem is said to be monotone under the prior \( \pi \) if for every \( n \), there exist numbers \(-\infty \leq a_1^n \leq a_2^n \leq a_3^n \leq a_4^n \leq \infty \) 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 \( \theta \), 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 \( H_1: \theta \neq 0 \) 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

\[
\mathcal{L}(\theta, (N, 0)) = \begin{cases} 
c(N - 1) + 1, & \text{if } \theta \in \Theta_1, \\
c(N - 1), & \text{if } \theta \in \Theta_0,
\end{cases}
\]

\[
\mathcal{L}(\theta, (N, 1)) = \begin{cases} 
c(N - 1), & \text{if } \theta \in \Theta_1, \\
c(N - 1) + 1, & \text{if } \theta \in \Theta_0.
\end{cases}
\]

Let the prior give mass \( (\pi_1, \pi_0, \pi_2) \) to the points \( \theta_1, 0, \theta_2 \), respectively. Denote by \( x_1^* \) the unique value such that \( \max_{x_1} P(\Theta_0|x_1^*) = P(\Theta_0|x_1^*) \). Here \( P(\Theta_0|x_1) \) denotes the posterior probability of \( \Theta_0 \) 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) \( P(\Theta_0|x_1^*) = \frac{1}{2} \).

Let \( \rho_1(x_1) = 1 - P(\Theta_1|x_1) \) and let \( \beta_1(s) \) denote the conditional Bayes risk of a
procedure which takes two observations, conditional on \( S_1 = s \). Let
\[
\Delta(x_1) = \text{Min}(\rho_1(x_1), \rho_0(x_1)) - (\beta_1(x_1) - c).
\]

In our case by (i), \( \rho_1(x_1) \leq \rho_0(x_1) \) and \( \Delta(x_1) \) can be written as
\[
\Delta(x_1) = P(\Theta_0|x_1) - E(\text{Min}(\tilde{P}(\Theta_0|x_1 + X_2), P(\Theta_1|x_1 + X_2))|x_1).
\]
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 \( \Delta(x_1) - c \geq 0 \). Let \( x_1^{**} \) be the unique value such that \( \Delta(x_1) = \Delta(x_1^{**}) \). [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
\[
(i)
\]

\[
\begin{align*}
\hat{x}_1^* \neq x_1^{**}.
\end{align*}
\]

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 \( \hat{\theta}_1, \hat{\theta}_2, \hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_0, \hat{\pi}_2, \hat{\epsilon}, \) where \( \hat{\theta}_1 = \hat{\theta}_2, \hat{\pi}_1 = \pi_1 - (\epsilon/2), \hat{\pi}_0 = \pi_0 + \epsilon, \hat{\pi}_2 = \pi_2 - (\epsilon/2) \).

Define \( E_0 = \{x_1|P(\Theta_0|x_1) \geq P(\Theta_1|x_1)\} \). \( E_0 \) is an interval by Lemma 1 of Section 2. Choose \( \epsilon > 0 \) small enough so that \( \hat{x}_1^{**} \not\in E_0 \). Such an \( \epsilon > 0 \) must exist by (ii) and by continuity considerations. Also by continuity consideration, there exists an \( \epsilon_1 > 0 \) such that \( \Delta(\hat{x}_1^{**}) - \epsilon_1 > \Delta(x_1) \) for every \( x_1 \in E_0 \). Take \( \hat{c} = \Delta(\hat{x}_1^{**}) - \epsilon_1 \). In our new problem there are two disjoint separated intervals containing \( \hat{x}_1^* \) and \( \hat{x}_1^{**} \), respectively, such that: The Bayes procedure decides \( H_0 \) if \( x_1 \) is in the interval containing \( \hat{x}_1^* \); it takes one more observation if \( x_1 \) is in the interval containing \( \hat{x}_2^* \) 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
\[
\begin{align*}
x_1^* = 0 \text{ and } P(\Theta_0|x_1^*) = 1/2,
\end{align*}
\]
we should choose \( \pi_1 = 0.219, \pi_0 = 0.240, \pi_2 = 0.539 \). Using numerical integration we get
\[
\Delta(x_1^*) = 0.259, \quad \Delta(0) = 0.266;
\]
that is, \( x_1^* \neq x_1^{**} \). 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
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 $\theta$. 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)

**Definition 1.** The function $\varphi(x): R \rightarrow R$ 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 $(a_i, a_{i+1})$, $i = 0, \ldots, n$, 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 $\varphi$. $(G_\theta)$ is
STP$_n$ if, in addition, for any $\varphi$ 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 $a_0 = -\infty < a_1 \leq \cdots \leq a_{n-1} < a_n = \infty$ such that $h(\theta)$ preserves its
sign on $(a_i, a_{i+1})$ and $h(\theta)$ can be zero only at $a_i$, $i = 1, \ldots, n - 1$.

Suppose $X_i \sim F_\theta$ are i.i.d. and $\pi(\theta)$ is a prior distribution on $\theta$. Denote
by $\nu^{n+1}(ds_{n+1}|s_n)$, 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 $\text{STP}_3$ with respect to the parameter $s_n$. Denote by $F^n_\theta(ds)$ the distribution of $s_n$ under $\theta$, and let $F^{n+1}_\theta(ds_{n+1}|s_n)$ be its conditional distribution conditional upon $S_n = s_n$.

**Proposition 1.** Suppose for some $\theta_0 \in \Theta$, $F^{n+1}_{\theta_0|s_n}(ds_{n+1}|S_n = s_n)$ is $(S)\text{TP}_k$ with respect to the parameter $s_n$. Then for every $\pi(\theta)$, $\nu^{n+1}(ds_{n+1}|s_n)$ is $(S)\text{TP}_k$.

**Proof.** See Greenshtein (1990).

**Proposition 2.** $F^{n+1}_{\theta_0|s_n}(ds_{n+1}|S_n = s_n)$ is $\text{STP}_n$ (i.e., $\text{STP}_k$ for $k = 1, 2, \ldots$) in $s_n$ for every $n$ when $\{F_\theta\}$ is any one of the following: binomial $\theta = p$; exponential $\theta = \lambda$ when $\lambda^{-1}$ is its expectation; Poisson $\theta = \lambda$, where $\lambda$ is its expectation; geometric $\theta = p$; normal $\theta = \mu$, where $\mu$ 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 $\text{STP}_3$ in the parameter $s_n$. Then every Bayes procedure $\Delta = \{\delta_{n}\}$ is of the following type: There exist numbers $a_2^n \leq a_3^n$ such that $\delta_0^n(s_n) = 1$ if $s_n \in (a_2^n, a_3^n)$ and $\delta_0^n(s_n) = 0$ 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

$$\rho^n_s(s_n) = \int L(\theta, \tau) \pi^n(d\theta|s_n), \quad \tau = 0, 1.$$  

Here $\pi^n(d\theta|s_n)$ denotes the posterior distribution given $S_n = s_n$. Assume a finite horizon where $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+k}$ are the available observations and $X_1, \ldots, X_n$ have already been observed ($n + k = m$). Denote by $\beta^{n+k}_n(s)$, 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 $\rho$ and $\beta$ are extensions of the definitions in Section 1.) Then,

$$\beta^{n+k}_n(s_n) = \int \min(c + \beta^{n+k}_{n+1}(s), c + \rho^{n+1}_0(s), c + \rho^{n+1}_1(s)) \nu^{n+1}(ds|s_n).$$
Of course,
\[
\beta_{n+1}^n(s_n) = \int \min(c + \rho_0^{n+1}(s), c + \rho_1^{n+1}(s)) \nu^{n+1}(ds|s_n).
\]

**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). \( \Box \)

**Lemma 2.** \( \rho_\tau^n(s_n) = \int \rho_1^{n+1}(s) \nu^{n+1}(ds|s_n), \tau = 0, 1. \)

**Proof.** See Sobel (1952). \( \Box \)

**Lemma 3.** \( \beta_{n+1}^n(s) - \rho_0^n(s) + W \), 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:
\[
\beta_{n+1}^n(s_n) - \rho_0^n(s_n) + W = \beta_{n+1}^n(s_n) - \int \rho_0^{n+1}(s) \nu^{n+1}(ds|s_n) + W
\]

(i)
\[
= \int \min(c + \beta_{n+1}^n(s) - \rho_0^{n+1}(s) + W, c + w, c + \rho_1^{n+1}(s) - \rho_0^{n+1}(s) + W) \nu^{n+1}(ds|s_n).
\]

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 of \( \nu^{n+1}(\cdot|s_n) \). \( \Box \)

**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 \( \beta_n^n(s_n) \) 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 \( \beta_n^M(s_n) \to \beta_n^n(s_n) \). Thus \( \beta_n^n(s_n) - \rho_0^n(s_n) \) has at most two sign changes and if there are two sign changes, it is first negative. Now (i) holds replacing \( \beta_{n+1}^n(s_n) \) by \( \beta_n^n(s_n) \). Thus, using STP of \( \nu^{n+1}_{s_n} \), we conclude
$\beta_n^p - \rho_n^p(s_n)$ 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**


**Department of Mathematics**

**Cornell University**

**Ithaca, New York 14853**

**Department of Mathematics**

**Bar Ilan University**

**Ramat Gan 52900**

**Israel**