TWO-SIDED SEQUENTIAL TESTS

by

L.D. Brown¹ and E. Greenshtein
Purdue University

Technical Report #90-03

Department of Statistics
Purdue University

January 1990

¹ Research supported in part by NSF DMS 8809016
TWO-SIDED SEQUENTIAL TESTS
by
L.D. Brown and E. Greenshtein

Abstract

Let $X_i$ be i.i.d. $X_i \sim F_\theta$. For some parametric families $\{F_\theta\}$, we characterize Bayes sequential procedures for the decision problem $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$. A surprising counter example 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. $X_i \sim F_\theta$ where $F_\theta$ is an exponential family. Assume $X_i$ are canonical observations, and $\theta$ the canonical parameter. We will consider the sequential testing problem $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

First we will formulate a general sequential testing problem. In this problem there exists a parameter set $\Theta$, two subsets $\Theta_0, \Theta_1 \subseteq \Theta$, and the hypotheses are $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$.

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

When $\Theta \subseteq \mathbb{R}$, two important cases are:

1. $L(\theta, 0)$ is nondecreasing and $L(\theta, 1)$ is nonincreasing. This condition on $L(\theta, \cdot)$ suits the one sided testing problem $H_0: \theta < \theta_0$ $H_1: \theta \geq \theta_0$.

2. $L(\theta, 0)$ 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 $\theta \geq \theta_0$. This condition suits the two sided problem $H_0: \theta = \theta_0$ vs. $H_2: \theta \neq \theta_0$.

As in Brown, Cohen and Strawderman [1] (to be referred to in the sequel as B.C.S.), we consider only procedures based on $S_n = X_1 + \ldots + X_n$. A procedure $\Delta$ consists of

Research was supported in part by NSF DMS 8809016.


Key words and phrases: Sequential Testing; Total–Positivity.
a set of non negative functions $\delta_i(s_n)$ defined for every $s_n$ such that $\sum_{i=0}^{2} \delta_i(s_n) = 1$. The quantities $\delta_i(s_n)$ 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: $R(\theta, \Delta) = E_\theta L(\theta, \Delta)$. 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, $\eta$ some additional point, denote $\overline{R} = R \cup \{\eta\}$. Map to the event $N = n_0$, $S_1 = s_1, \ldots, S_{n_0} = s_{n_0}$, the point $(s_1, \ldots, s_{n_0}, \eta, \eta, \ldots)$ in $\overline{R}_1 \times \overline{R}_2 \times \ldots \times \overline{R}_m$. This mapping induces a measure on $\overline{R}_1 \times \ldots \times \overline{R}_m$, under a parameter $\theta$ and a procedure $\Delta$, denote it by $H_{\theta, \Delta}$. Denote $H_{\pi, \Delta}$ the measure defined by: $H_{\pi, \Delta}(dx) = \int H_{\theta}(dx) d\pi(\theta)$.

**Definition 1:** A sequential procedure $\Delta$ of a one sided hypothesis testing is said to be monotone under the prior $\pi$ if for every $n$, there exist numbers $-\infty < a_1^n \leq a_2^n \leq \infty$ such that: For almost every real value $s_n$ under $H_{\pi, \Delta}$, $\delta_{1n}(s_n) = 0$ if $s_n > a_1^n$; $\delta_{2n}(s_n) = 0$ if $s_n < a_1^n$ or $s_n > a_2^n$; $\delta_{1n}(s_n) = 0$ if $s_n < a_2^n$. If $s_n = a_1^n$ then $\delta_{1n}(s_n) = 0$ but $\delta_{0n}(s_n)$ and $\delta_{2n}(s_n)$ are arbitrary. If $s_n = a_2^n$ then $\delta_{0n}(s_n) = 0$ but $\delta_{1n}(s_n)$ and $\delta_{2n}(s_n)$ are arbitrary.

**Definition 2:** A sequential procedure $\Delta$ of a two sided hypothesis testing is said to be monotone under the prior $\pi$ if for every $n$, there exist numbers $-\infty < a_1^n \leq a_2^n \leq a_3^n \leq a_4^n \leq \infty$ 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 $a_1^n < s_n < a_2^n$; $\delta_{2n}(s_n) = 0$ if $s_n < a_1^n$, $s_n > a_3^n$ 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 shown by Sobel [7] under the above condition, and was later generalized by B.C.S., that in the one sided testing problem every Bayes procedure is monotone. It was also pointed out by 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.
Section 1: Main Theorem

Assume $X_i$ are i.i.d. normal with unknown mean $\theta$. For two sided sequential testing it seems plausible that every Bayes procedure is monotone without assuming further symmetry. This conjecture is implicit in B.C.S. The example in the next section will show the last conjecture to be false. In view of the example, the following Theorem 1 seems to give the "best" possible general characterization 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 [6], [2].

Definition 1: The function $\varphi(x): R \rightarrow R$ changes signs at most $n$ times if and only if there exist $-\infty = a_0 < a_1 < \ldots < a_n < a_{n+1} = \infty$ such that $\varphi(x)$ preserves its sign on $(a_i, a_{i+1}) i = 0, \ldots, n$ i.e. it is either non negative or non positive.

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

Definition 2: $\{G_\theta\}$ is $TP_n$ if for any function $\varphi(x)$, that changes signs 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 $-\infty < a_1 \leq \ldots \leq a_{n-1} < \infty$ as in Definition 1 such that $h(\theta)$ 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 $d\nu^{n+1}_{s_n} = d\nu^{n+1}(s_{n+1}|S_n = s_n, \pi(\theta))$, the conditional distribution of $S_{n+1}$ given $S_n = s_n$ and prior $\pi(\theta)$.

In the sequel we will require $\nu^{n+1}_{s_n}$ to be $STP_3$ with respect to the parameter $s_n$. Denote $F^{n}_{\theta}(ds)$ the distribution of $s_n$ under $\theta$, $F^{n}_{\theta}(ds_n|S_{n-1} = s_{n-1})$ its conditional distribution conditional on $s_{n-1} = s_{n-1}$.

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

3
Proof: See [4].

Proposition 2: In the cases where $\{F_\theta\}$ are: Binomial $\theta = P$, Exponential $\theta = \lambda$, $\lambda^{-1}$ its expectation, Poisson $\theta = \lambda$ its expectation, Geometric $\theta = P$, Normal $\theta = \mu$ its expectation, $dF_{\theta_0}^{n+1}(S_{n+1}|S_n = s_n)$ is $STP_\infty$ in $s_n$ (i.e. $STP_k$ for every $k = 1, 2, \ldots$) for every $n$.

Proof: Immediate from chapters 7, 8 in [6].

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 $\Delta = \{\delta_n\}$ is of the following type: There exist numbers $a_2^n \leq a_3^n$ 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 $H_{\pi\Delta}$.

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

$$\rho_\tau(s_n) = \int L(\theta, \tau) d\pi(\theta|S_n = s_n), \quad \tau = 0, 1.$$ 

Here $\pi(\theta|s_n)$ 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. Suppose $X_1, \ldots, X_n$ have already been observed. 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 $S_n = s_n$. Then:

$$\beta_{n+k}^{n+k}(s_n) = \int \min(c + \beta_{n+1}^{n+k}(s), c + \rho_0^{n+1}(s), c + \rho_1^{n+1}(s)) d\nu^{n+1}(s|s_n).$$

Of course, $\beta_{n+k-1}^{n+k}(s_n) = \int \min(c + \rho_0^{n+1}(s), c + \rho_1^{n+1}(s)) d\nu^{n+1}(s|s_n)$.

Lemma 1: For every real number $W$, $\rho_1^n(s) - \rho_0^n(s) - W$ changes signs 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 (2) stated in the introduction, the proof follows as in Karlin [5].

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

Proof: Sobel [7].
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:

(i)

$$\beta_n^{n+k}(s_n) - \rho_0^n(s_n) + W = \beta_n^{n+k}(s_n) - \int \rho_0^{n+1}(s) d\nu^{n+1}(s|s_n) = \int \text{Min}(c + \beta_{n+1}^{n+k}(s) - \rho_0^{n+1}(s) + W, c + W, c + \rho_0^{n+1}(s) - \rho_0^{n+1}(s) + W) d\nu^{n+1}(s|s_n).$$

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 3. 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 $\nu^{n+1}_{s_n}$.

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 [3]. Define $\beta_n^{\infty}(s_n)$ 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 Theorem 4.4 and 4.7 of [3] that: $\beta_n^M(s_n) \xrightarrow{M \to \infty} \beta_n^{\infty}(s_n)$. 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 $\nu^{n+1}_{s_n}$ 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.

Section 2: Counter Example

We consider the two sided hypothesis testing problem $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$ for the mean of a normal distribution. The example will be of a Bayes procedure that accepts $H_1$ 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 $\Theta_0 = \{0\}$, $\Theta_1 = \{\theta_1, \theta_2\}$. Consider the following two stage testing problem. $X_1, X_2$ are i.i.d. $N(\theta, 1)$ $\theta \in \Theta_0 \cup \Theta_1$. Let the prior give mass $(\pi_1, \pi_0, \pi_2)$ to the points $\theta_1, 0, \theta_2$ respectively. Let

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

$$
\tilde{L}(\theta, (N, 1)) = \begin{cases} 
  c \cdot (N - 1) & \text{if } \theta \in \Theta_1 \\
  c \cdot (N - 1) + 1 & \text{if } \theta \in \Theta_0
\end{cases}
$$

Denote by $x_1^*$ the unique value such that $\text{Max}_{x_1} P(\Theta_0 | x_1) = P(\Theta_0 | x_1^*)$. Here $P(\Theta_0 | x_1)$ is the posterior probability of $\Theta_0$ given $X_1 = x_1$. The uniqueness of $x_1$ follows by showing, similarly to Lemma 1, 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) $P(\Theta_0 | x_1^*) = \frac{1}{2}$

Denote: $\Delta(x_1) = \text{Min}(\rho_1(x_1), \rho_0(x_1)) - \beta_1^2(x_1) - c$.

In our case by (i) $\rho_1^1(x_1) \leq \rho_0^1(x_1)$ and $\Delta(x_1)$ can be written as:

(ii) $\Delta(x_1) = P(\Theta_0 | x_1) - E(\text{Min}(P(\Theta_0 | X_1 + X_2), P(\Theta_1 | X_1 + X_2) | x_1)$

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 iff $\Delta(x_1) - c \geq 0$. Let $x_1^{**}$ be the unique value such that $\text{Max}_{x_1} \Delta(x_1) = \Delta(x_1^{**})$. It can be shown that $x_1^{**}$ is unique by writing $\rho_0(s) - \beta_1^2(s) - W$ 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 function $\rho_0(s) - \beta_1^n(s) - W$ can have more than one local maximum when $n > 2$.

Suppose that:

(iii) $x_1^* \neq x_1^{**}$

We will show now that when (i) and (iii) are satisfied for some $\theta_1, \theta_2, \pi_1, \pi_2$, then a counterexample can be constructed. Define a new problem with $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\pi}_1, \tilde{\pi}_0, \tilde{\pi}_2, \tilde{\epsilon}$. Assume $\theta_i = \tilde{\theta}_i$, $\tilde{\pi}_1 = \pi_1 - \frac{\epsilon}{2}$, $\tilde{\pi}_0 = \pi_0 + \epsilon$, $\tilde{\pi}_2 = \pi_2 - \frac{\epsilon}{2}$. For $\epsilon$ small enough by continuity considerations

$$
\tilde{x}_1^{**} \notin \{ x_1 | P(\Theta_0 | x_1) > P(\Theta_1 | x_1) \} = E_0
$$
$E_0$ is an interval by Lemma 1. By continuity consideration there exists $\varepsilon_1$ such that:
$\Delta(\tilde{x}_1^{**}) - \varepsilon_1 > \Delta(x_1)$ for every $x_1 \in E_0$. Take $\bar{c} = \Delta(x_1^{**}) - \varepsilon_1$. In our new problem there are two separate intervals containing $\tilde{x}_1$ and $\tilde{x}_1^{**}$, such that the Bayes procedure respectively accepts $H_0$, take one more observation, conditional on $x_1$ belonging to these intervals and accept $H_1$ otherwise. Hence there are three separate intervals where the Bayes procedure accepts $H_1$ 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 $x_1^* = 0$ and $P(\Theta_0|x_1^*) = \frac{1}{2}$, we should choose $\pi_1 = 0.219$ $\pi_0 = 0.240$ $\pi_2 = 0.539$. Using numerical integration we get:

$$\Delta(x_1^*) = \Delta(0) = 0.259, \quad \Delta(0.2) = 0.266;$$

i.e. $x_1^* \neq x_1^{**}$. Table 1 shows some further values of $P(\theta_0|x_1)$ and $\Delta(x_1)$.

| $x_1$  | $P(\Theta_0|x_1)$ | $\Delta(x_1)$ |
|--------|-------------------|---------------|
| -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         |

**REMARKS:**

1. 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.

2. 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.
3. In constructing the counter example we have used the fact that the cost of the first observation is zero, and hence first observation is always taken. It is not clear whether a counter example can be given when the cost is $c$ per observation including the first one.

Acknowledgement

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

References


L.D. Brown  
Department of Mathematics  
Cornell University  
Ithaca, NY 14853

E. Greenshtein  
Department of Statistics  
Purdue University  
West Lafayette, IN 47907