ROBUST SEQUENTIAL TESTING

BY

PHAM XUAN QUANG

TECHNICAL REPORT NO. 206
JANUARY 1985

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
2FFZ459
INGRAM OLKIN, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ROBUST SEQUENTIAL TESTING

Pham Xuan Quang
Department of Statistics
University of California, Berkeley

SUMMARY

This paper considers the asymptotic minimax property of the sequential probability ratio test when the given distributions $P_{\pm \varepsilon}$ contain a small amount of contamination. Let $P_{\pm \varepsilon}$ be the neighborhoods of $P_{\pm \varepsilon}$. Suppose that $P_{\varepsilon}$ and $P_{-\varepsilon}$ approach each other as $\varepsilon \downarrow 0$ and that $P_{\pm \varepsilon}$ shrink at an appropriate rate. We prove (under regularity assumptions) that the SPRT based on the least favorable pair of distributions $(Q^{*}_{-\varepsilon}, Q^{*}_{\varepsilon})$ given by Huber (1965) is asymptotically least favorable for expected sample size and is asymptotically minimax, provided that the limiting maximum error probabilities do not exceed $1/2$.


Key words and phrases. Sequential probability ratio test. Shrinking neighborhoods. Asymptotic minimax.

With partial support from the National Science Foundation Grant GP-31091XI, The Alexander Dallas Bache Grant, and Office of Naval Research Contract N00014-80-C-0163.

*Now at Henkel-France SA, 94250 Gentilly.
1. Introduction.

Let $\mathcal{M}$ be the set of probability measures on the real line $\mathbb{R}$ and let $P_0, P_1$ be two distinct elements of $\mathcal{M}$. To discriminate between $P_0$ and $P_1$, we may either use the likelihood ratio test provided by the Neyman-Pearson lemma, or Wald's sequential probability ratio test.

Let us describe these tests and introduce notations. Let $X_1, X_2, \ldots$ be i.i.d. observations with distribution $Q$. The testing problem is $H(Q = P_0)$ vs. $K(Q = P_1)$. Let $p_0, p_1$ be the respective densities of $P_0, P_1$ with respect to some dominating measure. Put

$$S_n = \sum_{i=1}^{n} \log \frac{p_1(X_i)}{p_0(X_i)}.$$  

For the first test, the sample size is fixed and we reject $H$ if $S_n$ is too large. For the second test, observations are sampled sequentially, we assume that they are available in limitless quantity. Choose 2 numbers $a, b$ with $a < 0 < b$ and define the decisive sample number

$$N = \inf\{n \geq 0 : S_n \leq a \text{ or } S_n \geq b\}$$

with $\inf \emptyset = \infty$. Wald (1947) proved that $N$ is almost surely finite under both $P_0$ and $P_1$. The testing procedure is to stop at stage $N$ and reject $H$ if $S_n \geq b$, accept $H$ (hence reject $K$) if $S_n \leq a$. We shall denote this test $\text{SPRT}(a, b; P_0, P_1)$. The error probabilities are $\alpha = P_0(S_n \geq b)$ and $\beta = P_1(S_n \leq a)$. The average sample numbers are $E_j N$, $j = 0, 1$, where $E_j$ denotes expectation under $P_j$. The SPRT is optimum in the following sense. Consider any other testing procedure with corresponding elements $\alpha', \beta', E_0' N', E_1' N'$ then (cf. Lehmann (1959), pp. 104-110), it holds that
\[
\alpha' \leq \alpha \quad \text{and} \quad \beta' \leq \beta
\]

(1.2)

\[\Rightarrow E_0^N \leq E_0^{N'} \quad \text{and} \quad E_1^N \leq E_1^{N'} .\]

Now suppose that \( P_0 \) and \( P_1 \) are not fully known, as often is the case in practice. To formalize the situation, we introduce two small disjoint subsets \( \mathcal{P}_0, \mathcal{P}_1 \) of \( \mathcal{M} \) with \( P_0 \in \mathcal{P}_0, \ P_1 \in \mathcal{P}_1 \) and we pretend to know only that \( Q \in \mathcal{P}_0 \cup \mathcal{P}_1 \). We are thus led to consider the composite testing problem \( H(Q \in \mathcal{P}_0) \) vs. \( K(Q \in \mathcal{P}_1) \). Huber (1965, 1968, 1981) considered the following models for \( \mathcal{P}_j, \ j = 0, 1: \)

(1.3) \[ \mathcal{P}_j = \{ Q \in \mathcal{M} : Q = (1-2\varepsilon_0)P_j + 2\varepsilon_0 H, \ H \in \mathcal{M} \} \]

(1.4) \[ \mathcal{P}_j = \{ Q \in \mathcal{M} : d(Q, P_j) \leq \varepsilon_0 \} \]

where \( d \) is any of the following metrics: total variation, Kolmogorov-Levy, Prokhorov. He solved the preceeding problem by looking for a "least favorable pair" \((Q_0^*, Q_1^*)\) in \( \mathcal{P}_0 \times \mathcal{P}_1 \) such that the likelihood ratio test of \( Q_0^* \) vs. \( Q_1^* \) is precisely the maximin test of \( P_0 \) vs. \( P_1 \). See also Huber and Strassen (1973) and Rieder (1977).

Huber (1965) also considered the SPRT\((a, b, Q_0^*, Q_1^*)\) as a sequential candidate for the same testing problem. He proved that \((Q_0^*, Q_1^*)\) are least favorable for errors, namely

\[
\sup\{Q[S_N \geq b] : Q \in \mathcal{P}_0\} = Q_0^*[S_N \geq b] \\
\sup\{Q[S_N \leq a] : Q \in \mathcal{P}_1\} = Q_1^*[S_N \leq a] .
\]

(1.5)
But he also gave an example where \((Q_0^*, Q_1^*)\) are not least favorable for
ASN, i.e., at least one of the following fails:

\[
\text{sup}\{E_Q N : Q \in \rho_j\} = E_{Q_j^* N}, \quad j = 0, 1
\]

where \(E_Q\) denotes expectation under \(Q\).

The main aim of this paper is to prove that (1.6) does hold asymptotically. More precisely, we embed the testing problem \(H(Q \in \rho_0)\) vs. \(K(Q \in \rho_1)\) in

a sequence of testing problems as follows. In (1.3) replace \(\epsilon_0, \rho_0, \rho_1, Q_0^*, Q_1^*, \rho_0, \rho_1\) by \(k_\epsilon, \rho_{-\epsilon}, \rho_\epsilon, Q_{-\epsilon}^*, Q_\epsilon^*, \rho_{-\epsilon}, \rho_\epsilon\) where \(k > 0\) is fixed and \(\epsilon\) decreases to zero. Assume that \(\rho_{-\epsilon}\) and \(\rho_\epsilon\) approach each other in

a smooth way and consider the sequence of testing problems

\[
H(Q \in \rho_{-\epsilon}) \text{ vs. } K(Q \in \rho_\epsilon).
\]

Huber–Carol (1970) treated (1.7) for the variational, fixed-sample

size case. Rieder (1978, 1981a, 1981b) worked with a more general model, and streamlined the regularity conditions. See also Wang (1981). In this paper, we propose to prove that (1.6) holds asymptotically (as \(\epsilon \to 0\)), whenever the limiting maximum error probabilities do not exceed \(1/2\), and hence that the SPRT\((a, b, Q_{-\epsilon}^*, Q_\epsilon^*)\) is asymptotically minimax with respect to expected sample sizes.

2. **Regularity Assumptions.**

Let \(\{\rho_\theta : |\theta| \leq \theta_0\}, \theta_0 > 0,\) be a family of probability measures.

In order to simplify computations, we shall consider the variational model
(1.4) only. The variational metric on $\mathcal{M}$ is $d$ with $d(Q,Q') = \sup\{ |Q'(B) - Q(B)| : \text{all Borel } B \}$. Put $T = \{ \epsilon_n : n = 1,2,\ldots \}$ with $\lim_{n \to \infty} \epsilon_n = 0$. Our working model is then

$$ (2.1) \quad \mathcal{P}_{\pm \epsilon} = \{ Q : d(Q, P_{\pm \epsilon}) \leq k_\epsilon \} $$

where $\epsilon \in T$. Here and in the sequel, plus signs go together, as do minus signs. Denote $E$ the expectation under $P_0$, $E_Q$ the expectation under $Q$, $\vee$ the maximum and $\wedge$ the minimum operators.

The regularity assumptions follow.

$$ (2.2) \quad P_\theta << P_0 \text{ for all } |\theta| \leq \theta_0. \quad \text{Put } p_\theta = dP_\theta/dP_0. $$

$$ (2.3) \quad \text{For each } x \in \mathbb{R}, \log p_\theta(x) \text{ is at least twice continuously differentiable in } \theta. \text{ Let primes denote partial differentiation with respect to } \theta. $$

$$ (2.4) \quad 0 < E|p_\theta'| < \infty $$

$$ (2.5) \quad E \sup_{|\theta| \leq \theta_0} |p_\theta''| < \infty \quad \text{and} \quad E \sup_{|\theta| \leq \theta_0} |(\log p_\theta)'''| < \infty $$

$$ (2.6) \quad E(p_0')^+ > k. $$

Remarks. Assumptions (2.2)-(2.5) are weaker than those given by Cramer (1945, page 500). Exponential families: $P_\theta(dx) = \exp[\theta x - b(\theta)]P_0(dx)$ with $b(0) = b'(0) = 0$, $b''(0) = 1$, are easily seen to satisfy (2.2)-(2.5).
Rieder's sleeker conditions (Rieder, 1978, page 1081) do not seem to be easily amenable to the treatments in section 3.

3. The Least Favorable Pair.

We gather in this section some results of Huber (1965, 1968, 1969) and prove some limits needed in the sequel. The least favorable pair (Q_{-\varepsilon}^*, Q_{\varepsilon}^*) exists, its densities q_{-\varepsilon}^* are given by Huber (1965), page 1757.

Put

\[ Z_\varepsilon = \log(q_{-\varepsilon}^*/q_{\varepsilon}^*) \]

\[ = 2\varepsilon [\gamma' + (\frac{1}{2\varepsilon} \log \frac{p_{\varepsilon}}{p_{-\varepsilon}}) \land \gamma''] \]

(3.1)

\[ S_n = \sum_{i=1}^{n} Z_\varepsilon(X_i) \]

(3.2)

The truncation points \( \gamma' = \gamma'(\varepsilon), \gamma'' = \gamma''(\varepsilon) \) uniquely solve the following equations

\[ E[e^{2\varepsilon\gamma' p_{-\varepsilon} - p_{\varepsilon}}] = k\varepsilon(1 + e^{2\varepsilon\gamma'}) \]

\[ E[e^{-2\varepsilon\gamma'' p_{\varepsilon} - p_{-\varepsilon}}] = k\varepsilon(1 + e^{-2\varepsilon\gamma''}) \]

(3.3)

Moreover, from lemma 2 of Huber (1965):

\[ Q'(S_n > t) \leq Q_{-\varepsilon}^*(S_n > t) \leq Q_{\varepsilon}^*(S_n > t) \leq Q''(S_n > t) \]

(3.4)

for all \( Q' \in \mathcal{P}_{-\varepsilon}, Q'' \in \mathcal{P}_{\varepsilon} \) and \( t \in \mathbb{R} \).
**Lemma 3.1**

(a) There exists $\varepsilon_0$ such that $\gamma'(\varepsilon) < 0 < \gamma''(\varepsilon)$ for all $\varepsilon < \varepsilon_0$.

(b) The equation in $c': E(c'-p_{0\varepsilon})^+ = k$ has a unique root which is negative. The equation in $c'': E(p_0^+-c'')^+ = k$ has a unique root which is positive.

(c) Let $c', c''$ be defined as in (b), then $\lim_{\varepsilon \to 0} \gamma'(\varepsilon) = c'$ and $\lim_{\varepsilon \to 0} \gamma''(\varepsilon) = c''$.

(d) $\lim_{\varepsilon \to 0} \varepsilon^{-1}Z_\varepsilon = 2(c' \lor p_{0\varepsilon} \land c'')$ pointwise.

**Proof.** Part (a) comes from the fact that $E(p_{\varepsilon} - p_{-\varepsilon})^+ > 2k\varepsilon \Rightarrow \gamma'(\varepsilon) < 0 < \gamma''(\varepsilon)$, Taylor expansion and dominated convergence.

The proof of (b) proceeds as in Huber (1965), lemma 1. The proof (c) proceeds as in Rieder (1978) pages 1088-1089. Put $B(\theta) = (\log p_\theta)'$, then $(2\varepsilon)^{-1} \log (p_{\varepsilon}/p_{-\varepsilon}) = p_{0\varepsilon} + \frac{1}{4} \varepsilon [B(\theta_1) - B(\theta_2)]$ where $|\theta_1| < \varepsilon$ and $|\theta_2| < \varepsilon$. Hence (d) is immediate.

**Lemma 3.2.** Define $\lambda^* = 2E(c' \lor p_{0\varepsilon} \land c'')^2$ where $c', c''$ are as in lemma 3.1. Then as $\varepsilon \to 0$:

(a) $\varepsilon^{-2}EZ_{\varepsilon} = 0$.

(b) $\varepsilon^{-2}Ep_{\varepsilon}Z_{\varepsilon} + \lambda^* + 2k(c'' - c')$

(c) $\varepsilon^{-2}Ep_{-\varepsilon}Z_{-\varepsilon} + -\lambda^* - 2k(c'' - c')$.

(d) $\varepsilon^{-2}var_{Q\varepsilon}Z_{\varepsilon} + 2\lambda^*$ all $Q \in P_{\varepsilon} \cup P_{-\varepsilon}$.

**Proof.**

(a) Use Taylor expansion on (3.1) to get

$$\varepsilon^{-2}EZ_{\varepsilon} = H_1 + H_2$$
where \[ H_1 = \varepsilon^{-1} \left[ \int I' (\gamma' - p_0'') dP_0 + \int I'' (\gamma'' - p_0') dP_0 \right] \]

\[ H_2 = \frac{\varepsilon}{4} \int J [B(\theta'_1) - B(\theta'_2)] dP_0 \]

\[ I' = \{ e^{2\varepsilon\gamma} p_{-\varepsilon} - p_{\varepsilon} > 0 \}, \quad I'' = \{ e^{-2\varepsilon\gamma} p_{\varepsilon} - p_{-\varepsilon} > 0 \} \]

\[ J = R - (I' \cup I''), \quad |\theta'_1| < \varepsilon \quad \text{and} \quad |\theta'_2| < \varepsilon . \]

Now apply Taylor expansions and dominated convergence to the following equation (which follows from (3.3)):

\[ E[e^{2\varepsilon\gamma} p_{-\varepsilon} - p_{\varepsilon}]^+ - k\varepsilon e^{2\varepsilon\gamma}' \]

\[ = E[e^{-2\varepsilon\gamma} p_{\varepsilon} - p_{-\varepsilon}]^+ - k\varepsilon e^{-2\varepsilon\gamma}'' \]

to see that \( H_1 \to 0 \). Since it is clear that \( H_2 \to 0 \), part (a) is proved.

(b) Again Taylor expansion and dominated convergence leads us to

\[ \varepsilon^{-2} E_{p_{-\varepsilon}} Z + 2E_{p_0} (e' \vee p_0' \wedge c'') \]

Now take into account the equations in lemma 3.1b to get (b).

(c) Let \( I', I'', J \) be as above, let \( q^* = dQ_{\varepsilon}/dP_0 \). It follows from the construction of \( q^* \) in Huber (1965) that

\[ \int I' (q^* - p_{\varepsilon}) dP_0 = \int I'' (p_{\varepsilon} - q^*) dP_0 = k\varepsilon . \]
Hence

\[
E_Q Z^*_\varepsilon = Eq^*_\varepsilon Z^* = 2\varepsilon \gamma' \int_{I^*} q^* d\mathcal{P}_0 + 2\varepsilon \gamma'' \int_{I^*} q^* d\mathcal{P}_0 + \int J \left( \log \frac{p_{-\varepsilon}}{p_{+\varepsilon}} \right) q^* d\mathcal{P}_0 \\
= Ep_{+\varepsilon} Z^* + 2\varepsilon^2 (\gamma' - \gamma'')
\]

Take (b) into account to see that (c) holds.

(d) Since \( p_{+\varepsilon} + p_0 = 1 \), Scheffé's theorem implies \( d(p_{+\varepsilon}, p_0) \rightarrow 0 \).

Moreover \( d(Q, P_{\varepsilon}) \leq k\varepsilon \) if \( Q \in \mathcal{P}_{\varepsilon} \), \( d(Q, P_{-\varepsilon}) \leq k\varepsilon \) if \( Q \in \mathcal{P}_{-\varepsilon} \), hence \( d(Q, P_0) \rightarrow 0 \) uniformly in \( Q \in \mathcal{P}_{\varepsilon} \cup \mathcal{P}_{-\varepsilon} \). Refer to (3.1) and put for simplicity \( G_{\varepsilon} = (2\varepsilon)^{-2} Z^2_{\varepsilon} \), then \( G_{\varepsilon} \) is bounded (say by M) and

\[ G_{\varepsilon} \rightarrow (c' \lor p_0 \land c'')^2 \] .

We deduce that \( \int G_{\varepsilon} d\mathcal{P}_0 \rightarrow \lambda^*/2 \) by dominated convergence. Consider \( \varepsilon^{-2} E Q Z^2_{\varepsilon} = \int G_{\varepsilon} dQ = \int G_{\varepsilon} d(Q-P_0) + \int G_{\varepsilon} d\mathcal{P}_0 \). We have the first integral on the right converging to zero, since

\[
\left| \int G_{\varepsilon} d(Q-P_0) \right| \leq \left| \int G_{\varepsilon} d(Q-P_0)^+ \right| + \left| \int G_{\varepsilon} d(Q-P_0)^- \right| \\
\leq 2Md(Q,P_0)
\]

Thus (d) is proved.

4. Approximation by Brownian Motion.

We shall approximate the random walk \( S_n \) on \( \mathbb{R}^2 \)-scale by a Brownian motion process with drift. For computational reasons, consider also the SPRT \((a,b,Q^*_{-\varepsilon},Q^*_{\varepsilon})\) curtailed at some \( K > 0 \):

If \( N \leq K \) then either \( S_n < a \) or \( S_n > b \). Accept \( H \) in the first case, reject \( H \) in the second case. If \( N > K \), stop sampling at stage \([K]\)
and do not decide, where \([K]\) denotes the largest integer not exceeding \(K\).

Let \(C[0,\infty)\) and \(C[0,K]\) be the sets of all continuous functions on \([0,\infty)\) and on \([0,K]\) respectively. For \(x \in \mathbb{R}^+\) and \(\omega \in C[0,\infty)\), define the following stopping times:

\[
\begin{align*}
\tau_x(\omega) &= \inf\{t \geq 0 : \omega(t) = x\} \\
\tau(\omega) &= \tau_{ab}(\omega) = \min\{\tau_a(\omega), \tau_b(\omega)\} \\
\tau^K_x(\omega) &= \min\{\tau_x(\omega), K\} \\
\tau^K(\omega) &= \min\{\tau(\omega), K\}
\end{align*}
\]

Let \(A, R, C\) be the regions of acceptance, rejection and no-decision of the SPRT with barriers \(a, b\). They are subsets of \(C[0,\infty)\):

\[
\begin{align*}
A &= \{\tau < \infty \text{ and } \tau = \tau_a\} \\
R &= \{\tau < \infty \text{ and } \tau = \tau_b\} \\
C &= \{\tau = \infty\}
\end{align*}
\]

Let \(A^K, R^K, C^K\) be the corresponding regions of the SPRT curtailed at \(K\). For example \(\omega \in A^K\) means \(\tau^K(\omega) < K\) and \(\tau^K(\omega) = \tau^K_a(\omega)\), or \(\tau^K(\omega) = K\) and \(\omega(K) = a\). Endow \(C[0,K]\) with the uniform topology and for \(E \subset C[0,K]\), denote \(\partial E\) the boundary of \(E\).

Let \(\{W(t) : 0 \leq t < \infty\}\) be a standard Brownian motion process on some space \((\Omega, \mathcal{F}, P)\), namely \(W(0) = 0\), and if \(0 < t_1 < t_2 < \ldots < t_m\)
then \( W(t_1), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1}) \) are mutually independent random variables which are normally distributed with mean zero and respective variances \( t_1, t_2 - t_1, \ldots, t_m - t_{m-1} \). Let \( \lambda \) be a given real number. Put

\[
B_\lambda(t) = (2\lambda)^{1/2}W(t) + \lambda t
\]

where \( \lambda^* \) is given by 3.2. Thus \( B_{\lambda} \) is a Brownian motion with variance \( 2\lambda^* \) and drift \( \lambda \) per unit time. We shall also consider \( B_{\lambda} \) as the Wiener measure induced by \( \{B_\lambda(t) : 0 \leq t \leq K\} \) on \( C[0,K] \).

**Theorem 4.1.** The mappings \( \tau^K_a, \tau^K_b, \tau^K \) are continuous a.e. \( [B_{\lambda}] \).

**Proof.** The proof is inspired from Breslow (1969). Let \( U \) be the set of all \( \omega \in C[0,K] \) such that

\[
\begin{align*}
(\text{i}) & \quad \omega(0) = 0 \\
(\text{ii}) & \quad \limsup_{h \to 0} h^{-1}[\omega(t+h) - \omega(t)] = +\infty \\
(\text{iii}) & \quad \liminf_{h \to 0} h^{-1}[\omega(t+h) - \omega(t)] = -\infty.
\end{align*}
\]

Dvoretzky, Erdős, Kakutani (1961) proved that \( P[W \in U] = 1 \). It results that \( P[B_{\lambda} \in U] = 1 \).

In what follows, restrict attention to those \( \omega \in U \). Let \( \omega_n \in C[0,K] \) and \( \omega_n \to \omega \) in the uniform norm, then it is required to prove that

\[
\tau^K_a(\omega_n) = \theta_n \to \theta = \tau^K_a(\omega).
\]

In the case \( 0 < \theta < K \), for all \( \delta \) satisfying \( 0 < \theta - \delta < \theta + \delta < K \), there exists \( \tau^K_{\delta}, \theta < \tau^K_{\delta} < \theta + \delta \) such that \( \omega(\tau^K_{\delta}) < a \). Since \( \omega(0) = 0 \),

10
we have that $\omega(\theta - \delta) \geq a'$ > a. Since $\omega_n(t) + \omega(t)$ uniformly in t,
we get for all large $n$:

$$\omega_n(t_\delta) < a \text{ and } \omega_n(t) > a \text{ all } 0 \leq t \leq \theta - \delta.$$ 

Hence $\theta - \delta < \theta_n < \theta + \delta$ and thus $\theta_n = \theta$.

In the case $\theta = K$, either $\omega(K) = a$ or $\omega$ does not hit a
on $[0,K]$. It remains to consider the second case since $\{\omega = a\}$ is
$B^\lambda$-null. Now $\omega_n + \omega$ so $\omega_n$ does not hit a on $[0,K]$ for all large
$n$, and for those $n$, $\theta_n = K$ so we are done.

**Corollary 4.1.** $A^K, R^K$ and $C^K$ are $B^\lambda$-continuity sets.

**Proof.** Consider $A^K$. Let $\omega \in A^K$. So there exist $\{\omega_n\} \subset A^K$ and
$\{\xi_n\} \subset (A^K)^c$ such that $\omega_n + \omega$ and $\xi_n + \omega$. Suppose $\omega$ does not hit
a on $[0,K]$, then $\omega_n$ does not hit a on $[0,K]$ for all large $n$,
contradicting $\omega_n \in A^K$.

Since $\{\omega = a\}$ is $B^\lambda$-null, we may assume that $r^K_a(\omega) < K$. The
argument presented in theorem 4.1 implies that $\xi_n$ hits a on $[0,K]$
for all large $n$, contradicting $\xi_n \notin A^K$, unless perhaps $\omega \in C[0,K] - U$.
Similar proof for $R^K$. Finally remark that $C^K \subset A^K \cup R^K$.

Define $X_\varepsilon$ as the polygonal line joining the points with successive
coordinates

$$(4.4) \quad (0,0), (\varepsilon^2, S_1), \ldots, (j\varepsilon^2, S_j), \ldots$$

then $X_\varepsilon$ is a random element of $C[0,\infty)$. 

11
Theorem 4.2. Let \( \{Q_\varepsilon\} \) be a sequence in \( P_\varepsilon \cup P_{-\varepsilon} \) such that \( E_{Q_\varepsilon} Z_\varepsilon \) has a limit in \( \varepsilon^2 \)-scale, say

\[
\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} E_{Q_\varepsilon} Z_\varepsilon = \lambda
\]

then \( X_\varepsilon \) weakly converges to \( B_\lambda \) on \( C[0,K] \) and

\[
\mathcal{L}[\tau(X_\varepsilon) | Q_\varepsilon] \to \mathcal{L}[\tau(B_\lambda)]
\]

Proof. Put \( \mu = E_{Q_\varepsilon} Z_\varepsilon, \sigma^2 = \text{var}_{Q_\varepsilon} Z_\varepsilon, \xi_1 = \varepsilon \sigma^{-1}[Z_\varepsilon(X_\varepsilon) - \mu], \xi_i = \sum_{i=1}^j \xi_i \),

Remark that \( |\xi_i| \leq 4\varepsilon^2 \sigma^{-1}[\max(|c'|,c'') + 1] \) for all small \( \varepsilon \),

where \( c', c'' \) are given by 3.1.5. As \( \varepsilon \to 0 \), we have \( \mu/\varepsilon^2 + \lambda \) by (4.5) and \( \sigma^2/\varepsilon^2 + 2\lambda \) by lemma 3.2.d.

Let \( Y_\varepsilon \) be the polygonal line joining the points

\[
(0,0), (\varepsilon^2, \xi_1), \ldots, (j\varepsilon^2, \xi_j), \ldots
\]

then \( Y_\varepsilon \) is a random element of \( C[0,\infty) \). Also define the following random element \( Y'_\varepsilon \) of \( D[0,K] \) (cf. Billingsley (1968), page 109):

\[
Y'_\varepsilon(t) = \xi_j
\]

for \( j\varepsilon^2 \leq t < (j+1)\varepsilon^2 \), \( j = 0,1,\ldots,n \), where \( n \) is defined by \( n\varepsilon^2 \leq K < (n+1)\varepsilon^2 \). Then Donsker's theorem on \( D[0,K] \) (cf. Billingsley (1968), page 137) immediately applies and we get \( Y'_\varepsilon \Rightarrow W \), where \( W \)
is an extension on $D[0,K]$ of standard Brownian motion on $C[0,K]$.

It follows that $Y_\varepsilon \to W$ on $C[0,K]$ since

$$|Y'_\varepsilon(t) - Y_\varepsilon(t)| \leq \max_{1 \leq j \leq n} |\xi_j|$$

$$\leq 2\varepsilon^2 \sigma^{-1} \max(|c'|, c'') + 1$$

$$\to 0.$$ 

Notice that

$$X_\varepsilon(t) = \varepsilon^{-1} \sigma Y_\varepsilon(t) + \varepsilon^{-2} \mu t$$

hence $X_\varepsilon \to B_\lambda$ by Slutsky's theorem.

For (4.6) to hold, we need to prove first that $\mathbb{L}[\tau(X_\varepsilon)|Q_\varepsilon]$ is uniformly integrable. According to Bickel and Yahav (1968), it suffices to prove the following

$$\sup_{0 < \varepsilon \leq \varepsilon_0} Q_\varepsilon(N > m \varepsilon^{-2}) \leq 2 \rho^m$$

for some $\rho \in (0,1)$ and all large $m$, since $0 < \varepsilon^2 N - \tau(X_\varepsilon) < \varepsilon^2$.

But (4.9) is a straightforward application of Stein's technique (1946).

Now (4.6) is a consequence of theorem 4.1, Mann-Wald theorem and standard truncation techniques.
Put

\[
g(\lambda, a, b) = \begin{cases} 
\frac{a(1-e^{-b'}) + b(e^{-a'}-1)}{\lambda(e^{-a'} - e^{-b'})} & \lambda \neq 0 \\
|a|b/(2\lambda^*) & \lambda = 0
\end{cases}
\]

where \(a' = \lambda a/\lambda^*\) and \(b' = \lambda b/\lambda^*\) then

Corollary 4.2. As \(\varepsilon \to 0\), assuming that (4.5) holds:

(a) \(Q_{\varepsilon}(S_N \leq a) \to (1 - e^{-b'})/(e^{-a'} - e^{-b'})\)

(b) \(Q_{\varepsilon}(S_N \geq b) \to (1 - e^{-a'})/(e^{-b'} - e^{-a'})\)

(c) \(\varepsilon^2 E_Q g(\lambda, a, b)\)

**Proof.** A straightforward truncation argument using Theorem 4.1 and Corollary 4.1 leads us to

\[Q_{\varepsilon}(S_N \leq a) \to P[\tau(B_{\lambda}) = \tau_a(B_{\lambda})]\]

and the above limit has been computed by Anderson (1960), page 175, thus parts (a) and (b) are proved. Next we get from Shepp (1967) that \(EW(\tau) = 0\), where \(\tau = \tau(B_{\lambda})\) for short. Then using (4.3):

\[\lambda E(\tau) = EB_{\lambda}(\tau) = aP(\tau = \tau_a) + bP(\tau = \tau_b)\]

But \(P(\tau = T_{a})\) and \(P(\tau = T_{b})\) are the right sides of (a) and (b) respectively, so (c) is proved for case \(\lambda \neq 0\). Freedman (1971), page 71, gave directly the result for case \(\lambda = 0\).
According to (1.5), the maximum error probabilities of the SPRT\((a, b, Q^*_\varepsilon, Q^*_\varepsilon)\) are

\[
\alpha_\varepsilon = \sup\{Q(S_N \geq b) : Q \in \mathcal{P}_{-\varepsilon}\}
\]

\[
\beta_\varepsilon = \sup\{Q(S_N \leq a) : Q \in \mathcal{P}_{\varepsilon}\}.
\]

Corollary 4.3. The limiting maximum error probabilities of the SPRT\((a, b, Q^*_\varepsilon, Q^*_\varepsilon)\) are given by

(a) \(\alpha = \lim_{\varepsilon \to 0} \alpha_\varepsilon = (1 - e^{a})/(e^b - e^a)\)

(b) \(\beta = \lim_{\varepsilon \to 0} \beta_\varepsilon = (1 - e^{-b})/(e^{-a} - e^{-b})\).

Proof. Huber (1965) proved that (3.4) holds when \(n\) is replaced by \(N\).

It results that \(\alpha_\varepsilon = Q^*_\varepsilon(S_N \geq b)\) and \(\beta_\varepsilon = Q^*_\varepsilon(S_N \leq a)\). Now it suffices to replace \(Q\) by \(Q^*_\varepsilon\) in Corollary 4.2.a, then use Lemma 3.2.c to get 4.3.a. Similar argument for 4.3.b.

Remark that \(\alpha, \beta\) are the error probabilities of a Brownian SPRT.

See Dvoretzky, Kiefer and Wolfowitz (1953).

5. Asymptotic Minimality of the SPRT.

At stage \(\varepsilon\), the SPRT\((a, b, Q^*_\varepsilon, Q^*_\varepsilon)\) has decisive sample number \(N\) and maximum error probabilities given by (4.11). Let \(\alpha'_\varepsilon, \beta'_\varepsilon, N'_\varepsilon\) be the corresponding elements of a rival procedure for testing (1.7). Say that condition (C) holds whenever the following 2 implications are satisfied:

\[
(C1) \quad \lambda \geq \lambda^* \implies g(\lambda, a, b) \leq g(\lambda^*, a, b)
\]
(C2) \[ \lambda < - \lambda^* \implies g(\lambda, a, b) \leq g(-\lambda^*, a, b) \].

In Section 6, we shall investigate the validity of (C). The main result of this paper is

**Theorem 5.1.** Assume that (C) holds, then for

(a) \((Q_{-\varepsilon}^*, Q_\varepsilon^*)\) are asymptotically least favorable expected sample sizes, namely:

\[
(5.1) \quad \lim_{\varepsilon \to 0} \sup_{Q} \frac{\mathbb{E}_Q N}{Q_{\pm \varepsilon}^* N} = 1
\]

(b) The \(\text{SPRT}(a, b, Q_{-\varepsilon}^*, Q_\varepsilon^*)\) is asymptotically minimax in the sense that

\[
(5.2) \quad \alpha_{\varepsilon}' \leq \alpha_{\varepsilon} \quad \text{and} \quad \beta_{\varepsilon}' \leq \beta_{\varepsilon} \implies \limsup_{\varepsilon \to 0} \sup_{Q} \frac{\mathbb{E}_Q N}{\mathbb{E}_Q N'} < 1.
\]

Recall that + signs go together, as do − signs.

To prove Theorem 5.1, we need

**Lemma 5.1.** As \(\varepsilon\) countably decreases to zero, whatever the sequence \(\{Q_\varepsilon\}\) with \(Q_\varepsilon \in \mathcal{P}_\varepsilon\) or with \(Q_\varepsilon \in \mathcal{P}_{-\varepsilon}\), the sequence \(\varepsilon^{-2} \mathbb{E}_{Q_\varepsilon} N\) is bounded.

**Proof.** Consider the case \(Q_\varepsilon \in \mathcal{P}_\varepsilon\). It follows from (3.4) that
\[ \lambda^* = \lim_{\varepsilon \to 0} \varepsilon^{-2} E_{Q_{\varepsilon}^*} Z_{\varepsilon} \leq \liminf_{\varepsilon \to 0} \varepsilon^{-2} E_{Q_{\varepsilon}} Z_{\varepsilon} \]

So \( \{\varepsilon^{-2} E_{Q_{\varepsilon}} N\} \) is bounded from below by \( \lambda^* \). Next since \( d(Q_{\varepsilon}, P_{\varepsilon}) \leq k\varepsilon \) and \( \frac{1}{Z_{\varepsilon}} \left| Z_{\varepsilon} \right| \leq \gamma'' - \gamma' \) according to (3.1) and Lemma 3.1.a, we have

\[
\varepsilon^{-2} E_{Q_{\varepsilon}} Z_{\varepsilon} = \varepsilon^{-2} \int Z_{\varepsilon} dP_{\varepsilon} + \frac{Z_{\varepsilon}}{Z_{\varepsilon}} \cdot 2 \frac{1}{Z_{\varepsilon}} \varepsilon d(Q_{\varepsilon} - P_{\varepsilon}) \\
\leq \varepsilon^{-2} E_{P_{\varepsilon}} Z_{\varepsilon} + 2k(\gamma'' - \gamma') \\
+ \lambda^* + 4k(c'' - c')
\]

(cf. Lemma 3.2.b). So \( \{\varepsilon^{-2} E_{Q_{\varepsilon}} N\} \) is bounded from above by \( \lambda^* + 4k(c'' - c') \).

**Proof of Theorem 5.1.** Limit attention to the cases of + signs in (5.1) and (5.2). Recall that \( N \) depends on \( \varepsilon \).

(a) Let \( L \) be the left side of (5.1) then \( L \geq 1 \) since \( Q_{\varepsilon}^* \in P_{\varepsilon} \).

Consider for a contradiction a subsequence \( \{\varepsilon'\} \subset T \) such that

\[ 1 < L = \lim_{\varepsilon' \to 0} \frac{\sup_{E_{Q_{\varepsilon}^*} N}}{\sup_{E_{Q_{\varepsilon}^*} N}} \]

where \( \sup_{E_{Q_{\varepsilon}^*} N} = \sup\{E_{Q_{\varepsilon}^*} N : Q_{\varepsilon} \in P_{\varepsilon}'\} \).

Pick \( \delta > 0 \) such that \( L - \delta > 1 \). There exists \( \varepsilon(\delta) \) such that, for all \( \varepsilon' < \varepsilon(\delta) \):

\[ (\sup_{E_{Q_{\varepsilon}^*} N})/E_{Q_{\varepsilon}^*} N > L - \delta \]

Limit attention to these \( \varepsilon' \). For each integer \( m \) there exists \( \varepsilon'' = \varepsilon'(m) \) such that \( \varepsilon'' \to 0 \) as \( m \to \infty \) and such that
\[ \epsilon^{2E_{Q_e} N + \frac{1}{m}} > \sup_{\rho_e} \epsilon^{2E_{Q_e} N} \]
\[ > (L-\delta)\epsilon^{2E_{Q_e} N} \]

According to Lemma 5.1, by passing to subsequences if necessary, we may assume that \((\epsilon^n)^{-2E_{Q_e^n Z_e^n}}\) has a limit, say \(\lambda\). It follows from the proof of Lemma 5.1 that \(\lambda' = \lambda^*\). Corollary 4.2.c applies and we get, as \(m \to \infty\):

\[ g(\lambda, a, b) > (L-\delta)g(\lambda^*, a, b) > g(\lambda^*, a, b) \]

but this result contradicts condition \((C1)\) and part (a) is proved.

(b) The optimality property of the SPRT\((a, b, Q_e, Q_e^*)\) leads to

\[ \alpha_e' \leq \alpha_e \quad \text{and} \quad \beta_e' \leq \beta_e \quad \Rightarrow \quad E_{Q_e^* N} \leq E_{Q_e^* N'} \]

Hence

\[ E_{Q_e^* N} \leq \sup_{\rho_e} E_{Q_e N'} \]

\[ (\sup_{\rho_e} E_{Q_e N})/(\sup_{\rho_e} E_{Q_e N'}) \leq (\sup_{\rho_e} E_{Q_e N})/(\sup_{\rho_e} E_{Q_e^* N}) \]

Now part (b) of the theorem is a direct consequence of part (a).

6. **Admissible Choices of \((a, \beta)\).**

Theorem 5.1 hinges on the validity of condition \((C)\). Corollary 4.3 shows that there exists a one-to-one correspondence between \(\{(a, b) : a < 0 < b\}\)
and \( \{ (a, \beta) : a < 0, \beta > 0, a + \beta < 1 \} \), and we get

\[
a = -\log(1-a)/\beta, \quad b = \log(1-\beta)/a.
\]

Let us say that \((a, \beta)\), or \((a, b)\), is admissible whenever \((C)\) holds. It turns out that \((-b, b)\) is admissible whenever \(b > 0\), or equivalently, \((a, a)\) is admissible for all \(a \in (0, \frac{1}{2})\).

**Theorem 6.1.** \((a, \beta)\) is admissible for all \(a, \beta\) satisfying \(0 < a \leq \frac{1}{2}, 0 < \beta \leq \frac{1}{2}, a + \beta < 1\).

**Proof.** Put \(r = b/|a|, x = \lambda |a|/\lambda^*\) and

\[
h_r(x) = \frac{re^x + e^{-rx} - (r+1)}{x(e^x - e^{-rx})}
\]

for \(x \neq 0\) and \(h_r(0) = \frac{x}{2}\), then (4.10) implies

\[
g(\lambda, a, b) = \frac{a^2}{\lambda^*} h_r(x).
\]

Hence condition \((C)\) can be rephrased as follows:

\[
x \geq a \implies h_r(x) \leq h_r(-a)
\]

\[
x \leq a \implies h_r(x) \leq h_r(a).
\]

The function \(h_r\) is positive unimodal. This can be seen by investigating the number of roots of the equation \(h_r(x) = c\) with unknown \(x\) and varying \(c\). We skip the details. Let \(x(r)\) be the value of \(x\) for
which \( h_r(x) \) attains its maximum. An examination of (6.1) shows that
the following procedure gives us an admissible \((a,b)\): (1) Choose
\( r > 0 \), (2) Compute \( x(r) \), (3) Choose \( a < 0 \) satisfying \( |a| \geq |x(r)| \),
and finally, (4) Compute \( b = r|a| \).

Since \( r = 1 \) implies that \( h_r \) is even and hence \( x(r) = 0 \), the
pair \((-b,b)\) is admissible for all \( b > 0 \). Now assume without loss of
generality that \( |a| < b \). Thus \( r > 1, x_r > 0 \) and \( 0 < \alpha < \beta \leq \frac{1}{2} \).
Put \( u(t) = \frac{1-e^{-rt}}{(e^t-e^{-rt})} \), then \( u \) is strictly decreasing for
\( 0 < t < \infty \) and \( u(|a|) = \beta \). Thus the critical inequality in the above
procedure \( |a| \geq |x(r)| \), is equivalent to \( \beta \leq u[x(r)] \). Thus Theorem 6.1
is proved if \( u[x(r)] \geq \frac{1}{2} \), or equivalently, if \( e^{x(r)} + e^{-rx(r)} \leq 2 \).
Algebraic manipulations show that the last inequality is successively
equivalent to each of the following statements (a)-(e):

(a) Let \( t(r) \) be the unique root of the equation \( e^t + e^{-rt} = 2 \).

Then \( x(r) \leq t(r) \).

(b) \( \frac{\partial}{\partial x}|_{x = t(r)} h_r(x) \leq 0 \).

(c) \( U[t(r)] \leq 0 \) where \( \frac{\partial}{\partial x} h_r(x) = \frac{U(x)}{V(x)} \), \( V(x) = x^2(e^x - e^{-rx})^2 \).

(d) Let \( t \) solve \( tr = -\log(2-e^t) \), then \( T(t) \leq 0 \) where

\[
T(t) = (r+1)t[(r+1)(e^t-1)-(r-1)]-2(r-1)(e^t-1)
\]

(e) \( H(y) \leq 0 \) for \( 0 < y < 1 \) where

\[
H(y) = y(\log \frac{1+y}{1-y})^2 + (2y + \log \frac{1+y}{1-y}) \log(1-y^2)
\]
Now we prove (e), using series expansion of $H$ in powers of $y$.

We have directly computed the first coefficients:

$$H(y) = -\frac{1}{90} y^7 - \frac{1}{63} y^9 + \sum_{n=5}^{\infty} \frac{1}{2n} \left( c_n - 1 \right) y^{2n+1}$$

where

$$c_n = \sum_{i=2}^{n} \frac{1}{i(2i-1)} + \sum_{i=1}^{n} \frac{1}{i(2n+1-2i)}$$

We have, for $n \geq 5$

$$\sum_{i=2}^{n} \frac{1}{i(2i-1)} \leq \sum_{i=1}^{5} \frac{1}{i(2i-1)} + \int_{5}^{n} \frac{dx}{x(2x-1)} \leq 0.3967$$

$$\sum_{i=1}^{n} \frac{1}{i(2n+1-2i)} \leq \frac{1}{2n+1} \log [n(2n-1) + 2.0513]$$

$$c_n \leq 0.9292$$

and so statement (e) is proved.

7. **Acknowledgements.**

This paper is part of a thesis submitted in partial requirements for the Ph.D. degree at the University of California, Berkeley. My sincere thanks go to Professor P. J. Bickel for support and encouragements, and to the referees for their abundant notes and suggestions. The revision was carried out when I was visiting Stanford University and I should like to thank Professor D. O. Siegmund for his hospitality.
REFERENCES


