ROBUST SEQUENTIAL TESTING[1]

By Pham Xuan Quang[2]

University of California, Berkeley

This paper considers the asymptotic minimax property of the sequential probability ratio test (SPRT) when the given distributions $P_{\pm\epsilon}$ contain a small amount of contamination. Let $\mathscr{P}_{\pm\epsilon}$ be the neighborhoods of $P_{\pm\epsilon}$. Suppose that P_ϵ and $P_{-\epsilon}$ approach each other as $\epsilon \downarrow 0$ and that $\mathscr{P}_{\pm\epsilon}$ shrink at an appropriate rate. We prove (under regularity assumptions) that the SPRT based on the least favorable pair of distributions $(Q^*_{-\epsilon}, Q^*_\epsilon)$ 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 ½.

1. Introduction. Let $\mathscr{M}$ be the set of probability measures on the real line $\mathbb{R}$ and let P_0, P_1 be two distinct elements of $\mathscr{M}$. To discriminate between P_0 and P_1, we may either us 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 $X_1, X_2, \cdots$ 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 $S_n = \sum_{i=1}^n \log[p_1(X_i)/p_0(X_i)]$. 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

$$(1.1) \qquad N = \inf\{n \geq 0: S_n \leq a \text{ or } S_n \geq b\}$$

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 $S_n \geq b$, and accept H (hence reject K) if $S_n \leq a$. We shall denote this test 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, pages 104–110), it holds that

$$(1.2) \qquad \alpha' \leq \alpha \quad \text{and} \quad \beta' \leq \beta \Rightarrow E_0 N \leq E_0 N' \quad \text{and} \quad E_1 N \leq E_1 N'.$$

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 $\mathscr{P}_0$, $\mathscr{P}_1$ of $\mathscr{M}$ with $P_0 \in \mathscr{P}_0$, $P_1 \in \mathscr{P}_1$ and we pretend to know only that $Q \in \mathscr{P}_0 \cup \mathscr{P}_1$. We are thus led to consider the composite testing problem $H(Q \in \mathscr{P}_0)$ vs. $K(Q \in \mathscr{P}_1)$. Huber (1965, 1968, 1981) considered the following models for $\mathscr{P}_j$, $j = 0, 1$:

$$(1.3) \qquad \mathscr{P}_j = \{Q \in \mathscr{M}: Q = (1 - 2\varepsilon_0)P_j + 2\varepsilon_0 H,\ H \in \mathscr{M}\}$$

$$(1.4) \qquad \mathscr{P}_j = \{Q \in \mathscr{M}: d(Q, P_j) \le \varepsilon_0\}$$

where d is any of the following metrics: total variation, Kolmogorov, Levy, Prokhorov. He solved the preceding problem by looking for a "least favorable pair" (Q_0^*, Q_1^*) in $\mathscr{P}_0 \times \mathscr{P}_1$ such that the likelihood ratio test of Q_0^* vs. Q_1^* is precisely the maximin test of $\mathscr{P}_0$ vs. $\mathscr{P}_1$. 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

$$(1.5) \qquad \begin{aligned} \sup\{Q[S_N \ge b]:\ Q \in \mathscr{P}_0\} &= Q_0^*[S_N \ge b] \\ \sup\{Q[S_N \le a]:\ Q \in \mathscr{P}_1\} &= Q_1^*[S_N \le a]. \end{aligned}$$

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:

$$(1.6) \qquad \sup\{E_Q N:\ Q \in \mathscr{P}_j\} = E_{Q_j^*} N \quad j = 0, 1$$

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 $H(Q \in \mathscr{P}_0)$ vs. $K(Q \in \mathscr{P}_1)$ in a sequence of testing problems as follows. In (1.3) replace ε_0, P_0, P_1, Q_0^*, Q_1^*, $\mathscr{P}_0$, $\mathscr{P}_1$ by k_ε, $P_{-\varepsilon}$, P_ε, $Q_{-\varepsilon}^*$, Q_ε^*, $\mathscr{P}_{-\varepsilon}$, $\mathscr{P}_\varepsilon$ where $k > 0$ is fixed and ε decreases to zero. Assume that $P_{-\varepsilon}$ and P_ε approach each other in a smooth way and consider the sequence of testing problems

$$(1.7) \qquad H(Q \in \mathscr{P}_{-\varepsilon}) \quad \text{vs.} \quad K(Q \in \mathscr{P}_\varepsilon).$$

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 $\varepsilon \to 0$), whenever the limiting maximum error probabilities do not exceed ½, and hence that the SPRT$(a, b, Q_{-\varepsilon}^*, Q_\varepsilon^*)$ is asymptotically minimax with respect to expected sample sizes.

2. Regularity assumptions.

Let $\{P_\theta: |\theta| \le \theta_0\}$, $\theta_0 > 0$, be a family of probability measures. In order to simplify computations, we shall consider the variational model (1.4) only. The variational metric on $\mathscr{M}$ is d with $d(Q, Q') =$

$\sup\{|Q'(B) - Q(B)|: \text{ all Borel } B\}$. Put $T = \{\varepsilon_n: n = 1, 2, \cdots\}$ with $\lim_{n\to\infty}\varepsilon_n = 0$. Our working model is then

$$(2.1) \qquad \mathscr{P}_{\pm\varepsilon} = \{Q: d(Q, P_{\pm\varepsilon}) \leq k_\varepsilon\}$$

where $\varepsilon \in T$. 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.

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

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

$$(2.4) \qquad 0 < E\,|P_0'| < \infty.$$

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

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

REMARKS. Assumptions (2.2)–(2.5) are weaker than those given by Cramér (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 $q_{\pm\varepsilon}^*$ are given by Huber (1965, page 1757). Put

$$(3.1) \qquad Z_\varepsilon = \log(q_\varepsilon^*/q_{-\varepsilon}^*) = 2\varepsilon[\gamma' \vee (1/2\varepsilon \log p_\varepsilon/p_{-\varepsilon}) \wedge \gamma'']$$

$$(3.2) \qquad S_n = \sum_{i=1}^n Z_\varepsilon(X_i).$$

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

$$(3.3) \qquad \begin{aligned} E[e^{2\varepsilon\gamma'}p_{-\varepsilon} - p_\varepsilon]^+ &= k\varepsilon(1 + e^{2\varepsilon\gamma'}) \\ E[e^{-2\varepsilon\gamma''}p_\varepsilon - p_{-\varepsilon}]^+ &= k\varepsilon(1 + e^{-2\varepsilon\gamma''}). \end{aligned}$$

Moreover, from Lemma 2 of Huber (1965):

$$(3.4) \qquad Q'(S_n > t) \leq Q_{-\varepsilon}^*(S_n > t) \leq Q_\varepsilon^*(S_n > t) \leq Q''(S_n > t)$$

for all $Q' \in \mathscr{P}_{-\varepsilon}$, $Q'' \in \mathscr{P}_\varepsilon$ and $t \in R$.

LEMMA 3.1.

(a) *There exists ε_0 such that $\gamma'(\varepsilon) < 0 < \gamma''(\varepsilon)$ for all $\varepsilon < \varepsilon_0$.*

(b) *The equation in c': $E(c' - p_0')^+ = k$ 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) $\lim_{\varepsilon \to 0} \varepsilon^{-1} Z_\varepsilon = 2(c' \vee p_0' \wedge c'')$ *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 of (c) proceeds as in Rieder (1978, pages 1088–1089). Put $B(\theta) = (\log p_\theta)''$, then $(2\varepsilon)^{-1} \log(p_\varepsilon / p_{-\varepsilon}) = p_0' + \frac{1}{4}\varepsilon[B(\theta_1) - B(\theta_2)]$ where $|\theta_1| < \varepsilon$ and $|\theta_2| < \varepsilon$. Hence (d) is immediate.

LEMMA 3.2. *Define $\lambda^* = 2E(c' \vee p_0' \wedge c'')^2$ where c', c'' are as in Lemma 3.1. Then as $\varepsilon \to 0$:*

(a) $\varepsilon^{-2} E Z_\varepsilon \to 0$.
(b) $\varepsilon^{-2} E p_\varepsilon Z_\varepsilon \to \lambda^* + 2k(c'' - c')$
 $\varepsilon^{-2} E p_{-\varepsilon} Z_\varepsilon \to -\lambda^* - 2k(c'' - c')$.
(c) $\varepsilon^{-2} E_{Q_\varepsilon^*} Z_\varepsilon \to \lambda^*$
 $\varepsilon^{-2} E_{Q_{-\varepsilon}^*} Z_\varepsilon \to -\lambda^*$.
(d) $\varepsilon^{-2} \mathrm{var}_Q Z_\varepsilon \to 2\lambda^*$ *all $Q \in \mathscr{P}_\varepsilon \cup \mathscr{P}_{-\varepsilon}$.*

PROOF.

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

$$\varepsilon^{-2} E Z_\varepsilon = H_1 + H_2$$

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

$$\varepsilon^{-2} E p_\varepsilon Z_\varepsilon \to 2E p_0'(c' \vee p_0' \wedge c'').$$

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

(c) Let I', I'', J be as above, let $q_\varepsilon^* = dQ_\varepsilon^*/dP_0$. It follows from the construction

of q^* in Huber (1965) that

$$\int_{I'} (q_\varepsilon^* - p_\varepsilon) \, dP_0 = \int_{I''} (p_\varepsilon - q_\varepsilon^*) \, dP_0 = k\varepsilon.$$

Hence

$$E_{Q_\varepsilon^*} Z_\varepsilon = Eq_\varepsilon^* Z_\varepsilon = 2\varepsilon\gamma' \int_{I'} q_\varepsilon^* \, dP_0 + 2\varepsilon\gamma'' \int_{I''} q_\varepsilon^* \, dP_0 + \int_J \left(\log \frac{p_\varepsilon}{p_{-\varepsilon}} \right) q_\varepsilon^* \, dP_0$$

$$= Ep_\varepsilon Z_\varepsilon + 2k\varepsilon^2(\gamma' - \gamma'').$$

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

(d) Since $p_{\pm\varepsilon} \to p_0 = 1$, Scheffé's theorem implies $d(P_{\pm\varepsilon}, P_0) \to 0$. Moreover $d(Q, P_\varepsilon) \le k\varepsilon$ if $Q \in \mathscr{P}_\varepsilon$, $d(Q, P_{-\varepsilon}) \le k\varepsilon$ if $Q \in \mathscr{P}_{-\varepsilon}$, hence $d(Q, P_0) \to 0$ uniformly in $Q \in \mathscr{P}_\varepsilon \cup \mathscr{P}_{-\varepsilon}$. Refer to (3.1) and put for simplicity $G_\varepsilon = (2\varepsilon)^{-2} Z_\varepsilon^2$, then G_ε is bounded (say by M) and $G_\varepsilon \to (c' \vee p_0' \wedge c'')^2$. We deduce that $\int G_\varepsilon \, dP_0 \to \lambda^*/2$ by dominated convergence. Consider $\varepsilon^{-2} E_Q Z_\varepsilon^2 = \int G_\varepsilon \, dQ = \int G_\varepsilon \, d(Q - P_0) + \int G_\varepsilon \, dP_0$. We have the first integral on the right converging to zero, since

$$\left| \int G_\varepsilon \, d(Q - P_0) \right| \le \left| \int G_\varepsilon \, d(Q - P_0)^+ \right| + \left| \int G_\varepsilon \, d(Q - P_0)^- \right|$$

$$\le 2Md(Q, P_0).$$

Thus (d) is proved.

4. Approximation by Brownian motion.

We shall approximate the random walk S_n on ε^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 $N \le K$ then either $S_n \le a$ or $S_n \ge b$. 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 $[0, K]$, respectively. For $x \in \mathbb{R}^+$ and $\omega \in C[0, \infty)$, define the following stopping times:

$$\tau_x(\omega) = \inf\{t \ge 0 \colon \omega(t) = x\}$$

$$\tau(\omega) = \tau_{ab}(\omega) = \min\{\tau_a(\omega), \tau_b(\omega)\}$$

(4.1)

$$\tau_x^K(\omega) = \min\{\tau_x(\omega), K\}$$

$$\tau^K(\omega) = \min\{\tau(\omega), K\}.$$

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

(4.2)

$$A = \{\tau < \infty \text{ and } \tau = \tau_a\}$$

$$R = \{\tau < \infty \text{ and } \tau = \tau_b\}$$

$$C = \{\tau = \infty\}.$$

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 $\tau^K(\omega) = \tau_a^K(\omega)$, 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 ∂E the boundary of E.

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

$$(4.3) \qquad B_\lambda(t) = (2\lambda^*)^{1/2} W(t) + \lambda t$$

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

THEOREM 4.1. *The mappings τ_a^K, τ_b^K, τ_K are continuous a.e. $[B_\lambda]$.*

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

   (i) $\omega(0) = 0$
   (ii) $\lim \sup_{h \downarrow 0} h^{-1}[\omega(t + h) - \omega(t)] = +\infty$
   (iii) $\lim \inf_{h \downarrow 0} h^{-1}[\omega(t + h) - \omega(t)] = -\infty$.

Since Dvoretzky, Erdös, and Kakutani (1960) proved that $P[W \in U] = 1$, it follows 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 $\tau_a^K(\omega_n) = \theta_n \to \theta = \tau_a^K(\omega)$.

In the case $0 < \theta < K$, for all δ satisfying $0 < \theta - \delta < \theta + \delta < K$, there exists t_δ, $\theta < t_\delta < \theta + \delta$ such that $\omega(t_\delta) < a$. Since $\omega(0) = 0$, we have that $\omega(\theta - \delta) \ge a' > a$. Since $\omega_n(t) \to \omega(t)$ uniformly in t, we get for all large n:

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

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

In the case $\theta = K$, either $\omega(K) = a$ or ω does not hit a on $[0, K]$. It remains to consider the second case since $\{\omega = a\}$ is B_λ-null. Now $\omega_n \to \omega$ so ω_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_λ-continuity sets.*

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

Since $\{\omega = a\}$ is B_λ-null, we may assume that $\tau_a^K(\omega) < K$. The argument presented in theorem 4.1 implies that ξ_n hits a on $[0, K]$ for all large n,

contradicting $\xi_n \notin A^K$, unless perhaps $\omega \in C[0, K] - U$. The proof for ∂R^K is similar. Finally remark that $\partial C^K \subset \partial A^K \cup \partial R^K$.

Define X_ϵ as the polygonal line joining the points with successive coordinates

(4.4) $(0, 0), (\epsilon^2, S_1), \cdots, (j\epsilon^2, S_j), \cdots$

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

THEOREM 4.2.   *Let $\{Q_\epsilon\}$ be a sequence in $\mathscr{P}_\epsilon \cup \mathscr{P}_{-\epsilon}$ such that $E_{Q_\epsilon} Z_\epsilon$ has a limit in ϵ^2-scale, say*

(4.5) $\lim_{\epsilon \to 0} (1/\epsilon^2) E_{Q_\epsilon} Z_\epsilon = \lambda$

then X_ϵ weakly converges to B_λ on $C[0, K]$ and

(4.6) $\mathscr{L}[\tau(X_\epsilon) \mid Q_\epsilon] \to \mathscr{L}[\tau(B_\lambda)].$

PROOF.   Put $\mu = E_{Q_\epsilon} Z_\epsilon$, $\sigma^2 = \mathrm{var}_{Q_\epsilon} Z_\epsilon$, $\xi_i = \epsilon \sigma^{-1}[Z_\epsilon(X_i) - \mu]$, $\zeta_i = \sum_{i=1}^{j} \xi_i$.

Remark that $|\xi_i| \leq 4\epsilon^2 \sigma^{-1}[\max(|c'|, c'') + 1]$ for all small ϵ, where c', c'' are given by 3.1(b). As $\epsilon \to 0$, we have $\mu/\epsilon^2 \to \lambda$ by (4.5) and $\sigma^2/\epsilon^2 \to 2\lambda^*$ by Lemma 3.2(d).

Let Y_ϵ be the polygonal line joining the points

(4.7) $(0, 0), (\epsilon^2, \zeta_1), \cdots, (j\epsilon^2, \zeta_j) \cdots$

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

(4.8) $Y_\epsilon'(t) = \zeta_j$

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

$$| Y_\epsilon'(t) - Y_\epsilon(t) | \leq \max_{1 \leq j \leq n} |\xi_i|$$

$$\leq 2\epsilon^2 \sigma^{-1}[\max(|c'|, c'') + 1] \to 0.$$

Notice that

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

hence $X_\epsilon \Rightarrow B_\lambda$ by Slutsky's theorem.

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

(4.9) $\sup_{0 < \epsilon \leq \epsilon_0} Q_\epsilon(N > m\epsilon^{-2}) \leq 2\rho^m$

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

$$(4.10) \qquad g(\lambda, a, b) = \begin{cases} \dfrac{a(1 - e^{-b'}) + b(e^{-a'} - 1)}{\lambda(e^{-a'} - e^{-b'})} & \lambda \neq 0 \\[3mm] |a| \, b/(2\lambda^*) & \lambda = 0 \end{cases}$$

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

COROLLARY 4.2. *As $\varepsilon \to 0$, assuming that (4.5) holds:*

(a) $Q_\varepsilon(S_N \le a) \to (1 - e^{-b'})/(e^{-a'} - e^{-b'})$
(b) $Q_\varepsilon(S_N \ge b) \to (1 - e^{-a'})/(e^{-b'} - e^{-a'})$
(c) $\varepsilon^2 E_{Q_\varepsilon} N \to g(\lambda, a, b)$.

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

$$Q_\varepsilon(S_N \le a) \to P[\tau(B_\lambda) = \tau_a(B_\lambda)]$$

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^*_ε) are

$$(4.11) \qquad \begin{aligned} \alpha_\varepsilon &= \sup\{Q(S_N \ge b) \colon Q \in \mathscr{P}_{-\varepsilon}\} \\ \beta_\varepsilon &= \sup\{Q(S_n \le a) \colon Q \in \mathscr{P}_\varepsilon\}. \end{aligned}$$

COROLLARY 4.3. *The limiting maximum error probabilities of the SPRT(a, b, $Q^*_{-\varepsilon}$, Q^*_ε) 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 follows that $\alpha_\varepsilon = Q^*_{-\varepsilon}(S_N \ge b)$ and $\beta_\varepsilon = Q^*_\varepsilon(S_N \le a)$. Now it suffices to replace Q by Q^* in Corollary 4.2(a), then use Lemma 3.2(c) to get 4.3(a). The argument for 4.3(b) is similar.

Note that α, β are the error probabilities of a Brownian SPRT (see Dvoretzky, Kiefer and Wolfowitz, 1953).

5. Asymptotic minimaxity of the SPRT. At stage ε, the SPRT(a, b, $Q^*_{-\varepsilon}$, Q^*_ε) has decisive sample number N and maximum error probabilities given

by (4.11). Let α'_ε, β'_ε, N' be the corresponding elements of a rival procedure for testing (1.7). Say that condition (C) holds whenever the following two implications are satisfied:

(C1) $\lambda \geq \lambda^* \Rightarrow g(\lambda, a, b) \leq g(\lambda^*, a, b)$

(C2) $\lambda \leq -\lambda^* \Rightarrow g(\lambda, a, b) \leq g(-\lambda^*, a, b)$.

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

(5.1) $\lim_{\varepsilon \to 0}(\sup\{E_Q N: Q \in \mathscr{P}_{\pm\varepsilon}\}/E_{Q^*_{\pm\varepsilon}} N) = 1$.

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

(5.2) $\alpha'_\varepsilon \leq \alpha_\varepsilon$ *and* $\beta'_\varepsilon \leq \beta_\varepsilon$ *imply* $\limsup_{\varepsilon \to 0} \dfrac{\sup\{E_Q N: Q \in \mathscr{P}_{\pm\varepsilon}\}}{\sup\{E_Q N': Q \in \mathscr{P}_{\pm\varepsilon}\}} \leq 1$.

Recall that $+$ signs go together, as do $-$ signs.
To prove Theorem 5.1, we need

LEMMA 5.1. *As* ε *countably decreases to zero, whatever the sequence* $\{Q_\varepsilon\}$ *with* $Q_\varepsilon \in \mathscr{P}_\varepsilon$ *or with* $Q_\varepsilon \in \mathscr{P}_{-\varepsilon}$, *the sequence* $\{\varepsilon^{-2} E_{Q_\varepsilon} N\}$ *is bounded.*

PROOF.   Consider the case $Q_\varepsilon \in \mathscr{P}_\varepsilon$. 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 λ^*. Next since $d(Q_\varepsilon, \mathscr{P}_\varepsilon) \leq k\varepsilon$ and $(1/2\varepsilon)|Z_\varepsilon| \leq \gamma'' - \gamma'$ according to (3.1) and Lemma 3.1(a), we have

$$\varepsilon^{-2} E_{Q_\varepsilon} Z_\varepsilon = \varepsilon^{-2} \int Z_\varepsilon \, dP_\varepsilon + \frac{2}{\varepsilon} \int \frac{1}{2\varepsilon} Z_\varepsilon d(Q_\varepsilon - P_\varepsilon)$$

$$\leq \varepsilon^{-2} E_{P_\varepsilon} Z_\varepsilon + 2k(\gamma'' - \gamma')$$

$$\to \lambda^* + 4k(c'' - c')$$

(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 ε.
(a) Let L be the left-hand side of (5.1) then $L \geq 1$ since $Q^*_\varepsilon \in \mathscr{P}_\varepsilon$. Consider for a contradiction a subsequence $\{\varepsilon'\} \subset T$ such that

$$1 < L = \lim_{\varepsilon' \to 0}(\sup_{\mathscr{P}_{\varepsilon'}} E_Q N)/E_{Q^*_\varepsilon} N$$

where $\sup_{\mathscr{P}_{\varepsilon'}} E_Q N = \sup\{E_Q N: Q \in \mathscr{P}_{\varepsilon'}\}$.

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

$$(\sup_{\mathscr{S}_{\varepsilon}'} E_Q N)/E_{Q_{\varepsilon}^*} N > L - \delta.$$

Limit attention to these ε'. For each integer m there exists $\varepsilon'' = \varepsilon'(m)$ such that $\varepsilon'' \to 0$ as $m \to \infty$ and such that

$$\varepsilon''^2 E_{Q_{\varepsilon''}} N + 1/m > \sup_{\mathscr{S}_{\varepsilon''}} \varepsilon''^2 E_{Q_{\varepsilon''}} N > (L - \delta)\varepsilon''^2 E_{Q_{\varepsilon''}^*} N.$$

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

$$g(\lambda, a, b) \geq (L - \delta)g(\lambda^*, a, b) > g(\lambda^*, a, b)$$

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

(b) The optimality property of the SPRT$(a, b, Q_{-\varepsilon}^*, Q_{\varepsilon}^*)$ leads to

$$\alpha_{\varepsilon}' \leq \alpha_{\varepsilon} \quad \text{and} \quad \beta_{\varepsilon}' \leq \beta_{\varepsilon} \Rightarrow E_{Q_{\varepsilon}^*} N \leq E_{Q_{\varepsilon}^*} N'.$$

Hence

$$E_{Q_{\varepsilon}^*} N \leq \sup_{\mathscr{S}_{\varepsilon}'} E_Q N'$$

$$(\sup_{\mathscr{S}_{\varepsilon}'} E_Q N)/(\sup_{\mathscr{S}_{\varepsilon}'} E_Q N') \leq (\sup_{\mathscr{S}_{\varepsilon}'} E_Q N)/E_{Q_{\varepsilon}^*} N$$

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

6. Admissible choices of (α, β). 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 $\{(\alpha, \beta): \alpha < 0, \beta > 0, \alpha + \beta < 1\}$, and we get

$$a = -\log(1 - \alpha)/\beta, \quad b = \log(1 - \beta)/\alpha.$$

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

THEOREM 6.1. (α, β) is admissible for all α, β satisfying $0 < \alpha \leq \frac{1}{2}, 0 < \beta \leq \frac{1}{2}, \alpha + \beta < 1$.

PROOF. Put $r = b/|a|$, $x = \lambda |a|/\lambda^*$ and

$$h_r(x) = \frac{re^x + e^{-rx} - (r + 1)}{x(e^x - e^{-rx})} \quad \text{for} \quad x \neq 0$$

and $h_r(0) = r/2$, then (4.10) implies

$$g(\lambda, a, b) = (a^2/\lambda^*)h_r(x).$$

Hence condition (C) can be rephrased as follows:

(6.1)
$$x \geq a \Rightarrow h_r(x) \leq h_r(-a)$$

$$x \leq a \Rightarrow h_r(x) \leq h_r(a).$$

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 $u(t) = (1 - e^{-rt})/(e^t - e^{-rt})$, 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) $(\partial/\partial x)|_{x=t(r)} h_r(x) \leq 0$.

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

(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

$$H(y) = y(\log(1 + y)/(1 - y))^2 + (2y + \log(1 + y)/(1 - y))\log(1 - y^2).$$

Now we prove (e), using series expansion of H in powers of y. We have directly computed the first coefficients:

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

where

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

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.

REFERENCES

ANDERSON, T. W. (1960). A modification of the SPRT to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

BICKEL, P. J. and YAHAV, J. A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.* **39** 442–456.

BILLINGLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BRESLOW, N. (1969). On large sample sequential analysis with application to survivorship data. *J. Appl. Probab.* **6** 261–274.

CRAMÉR, H. (1945). *Mathematical Methods of Statistics.* Princeton Univ. Press.

DVORETZKY, A., ERDÖS, P. and KAKUTANI, S. (1960). Nonincrease everywhere of Brownian motion process. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2** 103–116. Univ. California Press.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1953). Sequential decision problems for processes with continuous time parameter testing hypotheses. *Ann. Math. Statist.* **24** 254–264.

FREEDMAN, D. A. (1971). *Brownian Motion and Diffusion.* Holden-Day, San Francisco.

HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. verw. Gebiete* **10** 269–278.

HUBER, P. J. (1969). Théorie de l'inférence statistique robuste. Lecture notes. Univ. de Montréal.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

HUBER, P. J. and STRASSEN, V. (1973). Minimax test and the Neyman-Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

HUBER-CAROL, C. (1970). Etude asymptotique de tests robustes. Ph.D. dissertation, Swiss Federal Institute of Technology.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

RIEDER, H. (1977). Least favorable pairs for special capacities. *Ann. Statist.* **5** 909–921.

RIEDER, H. (1978). A robust asymptotic model. *Ann. Statist.* **6** 1080–1094.

RIEDER, H. (1981a). Robustness of one- and two-sample rank tests against gross errors. *Ann. Statist.* **9** 245–265.

RIEDER, H. (1981b). On local asymptotic minimaxity and admissibility in robust estimation. *Ann. Statist.* **9** 266–277.

SHEPP, L. A. (1967). A first passage problem for the Wiener process. *Ann. Math. Statist.* **38** 1912–1914.

STEIN, C. (1946). A note on cumulative sums. *Ann. Math. Statist.* **17** 498–499.

WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

WANG, P. C. C. (1981). Robust asymptotic tests of statistical hypotheses involving nuisance parameters. *Ann. Statist.* **9** 1096–1106.

7 RUE JEAN MACE, APT 134
94120 FONTENAY SOUS BOIS
FRANCE