Z. Wahrscheinlichkeitstheorie verw. Gebiete
51, 291–302 (1980)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Structure of Sequential Tests Minimizing an Expected Sample Size

Gary Lorden*

Mathematics Department California Institute of Technology, Pasadena, California 91125, USA

Summary. The testing problem is to decide on the basis of repeated independent observations which of the probability densities f and g is true. Given upper bounds on the probabilities of error, the object is to minimize the expected sample size if the density p is true (allowed to differ from f and g). A characterization of the structure of optimal tests is obtained which is particularly informative in the case where f, g, and p belong to a Koopman-Darmois family. If $p=f$ or g, then the optimal tests are sequential probability ratio tests (SPRT's) and a new proof of the well-known optimality property of these tests is obtained as a corollary.

1. Introduction

Independent and identically distributed random variables $X_1, X_2, \ldots$ are observable one at a time and have probability density f, g, or p with respect to a sigma-finite measure on the space where each X takes values. Assuming f and g are non-equivalent, it is desired to test whether f or g is the true density. The error probabilities of the test, α and β, must satisfy prescribed upper bounds, and the object is to minimize EN, the expected sample size if p is true. This is known as the modified Kiefer-Weiss problem and leads [5] to consideration of the auxiliary problem of finding a test attaining

$$R(u, v) = \inf [EN + u\alpha + v\beta]$$

for given $u, v \geqq 0$, the infimum being taken over the class of all tests.

In the special cases $p=f$ and $p=g$, the solutions of the problem are sequential probability ratio tests (SPRTs) by virtue of the optimality property of these tests (Wald and Wolfowitz [13]). In typical situations, however, such as testing whether a real-valued parameter θ is $\leqq \theta_0$ or $\geqq \theta_1$, it is less critical to minimize the expected sample size for $\theta = \theta_0$ or θ_1 than it is for $\theta_0 < \theta < \theta_1$.

* Research supported by National Science Foundation grant MCS 76-08489

0044-3719/80/0051/0291/$02.40

Rather than choose a particular value of θ for which the expected sample size should be minimized, it is appealing to try to solve the Kiefer-Weiss problem of minimizing the *maximum* over θ of the expected sample size. The latter problem reduces in special cases (e.g. those in [14]) to the modified problem in which a fixed θ (or, equivalently, p) is considered, and the two problems seem to be related more generally (see Remark 1 in Sect. 7).

The present paper investigates the structure of tests attaining $R(u, v)$. Particular attention is paid to the case where f, g, and p belong to a Koopman-Darmois family of densities (Sect. 5), in which more explicit characterizations of optimal tests are obtained, extending results of Kiefer and Weiss [5], Weiss [14], and Lai [6]. The basic theorem is stated in Sect. 2 and is applied in Sect. 3 to give a new proof of the optimality property of the SPRT. This proof avoids some of the technical difficulties in the original argument of Wald and Wolfowitz and in subsequent refinements (e.g. [2, 4, 7, 10]) by considering only one expected sample size at a time. (R. Berk and G. Simons have earlier developed a proof of the optimality property of the SPRT (to appear) which is also of this type.) Additional structural information is obtained in Sect. 4, which is applied to the Koopman-Darmois case in Sect. 5. A numerical example for a symmetric normal case is given in Sect. 6.

2. Basic Structure of Optimal Tests

Tests of f against g are definable on a suitable product space and sequence of sigma-fields, $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots$, allowing for randomization. A test is specified by an extended stopping time, $0 \leqq N \leqq +\infty$, with respect to $\{\mathscr{F}_n\}$, together with an $\mathscr{F}_N$-measurable partition of $\{N<\infty\}$ into $\bar{F}=\{\text{reject } f\}$ and $\bar{G}=\{\text{reject } g\}$.

Let F, G, P denote probability when f, g, p are true and let f_n, g_n, p_n denote the likelihoods (Radon-Nikodym derivatives), for $n=0, 1, \ldots$ observations, e.g. $f_n = f(x_1)\ldots f(x_n)$ for $n \geqq 1$, $f_0 \equiv 1$. Summing over the events $\{(N=n)\cap \bar{F}\}$ and $\{(N=n)\cap \bar{G}\}$, one derives the formulas for error probabilities

$$\alpha = F(\bar{F}\cap(p_N=0)) + \int_{\bar{F}} \frac{f_N}{p_N}\, dP, \qquad \beta = G(\bar{G}\cap(p_N=0)) + \int_{\bar{G}} \frac{g_N}{p_N}\, dP, \tag{1}$$

noting that the integrands are well-defined and finite outside P-null sets. To obtain a useful alternative definition of $R(u, v)$, define

$$U_n = u f_n/p_n, \qquad V_n = v g_n/p_n, \qquad n=0, 1, \ldots, \tag{2}$$

and observe that U_n and V_n are a.s. finite under P. By (1),

$$u\alpha + v\beta \geqq \int_{\bar{F}} U_N\, dP + \int_{\bar{G}} V_N\, dP \geqq \int_{\{N<\infty\}} \min(U_N, V_N)\, dP,$$

and equalities hold if $\{N<\infty,\ p_N=0\}$ is empty and $\bar{F}=\{U_N<V_N\}$. Since any test can be modified to satisfy the last two conditions without changing EN, evidently

$$R(u, v) = \inf_N E[N + \min(U_N, V_N)],$$

the infimum begin taken over the class of all extended stopping times such that $P(N<\infty)=1$. (Note however that $F(N<\infty)<1$ and $G(N<\infty)<1$ are permitted.) For $n=1, 2, \ldots$

$$U_n = U_{n-1} \frac{f(x_n)}{p(x_n)} \quad \text{and} \quad V_n = V_{n-1} \frac{g(x_n)}{p(x_n)},$$

and thus $\{(U_n, V_n)\}$ is a stationary Markov sequence and the problem of finding a stopping time attaining $R(u, v)$ reduces in the usual way (e.g. Chapter 5 of [3]) to the determination of the set of (u, v)'s for which it is optimal to take no observations. Defining the "continuation risk",

$$R_1(u, v) = \inf_{N \geqq 1} E[N + \min(U_N, V_N)] = \inf_{N \geqq 1} [EN + u\alpha + v\beta],$$

one obtains by straightforward extension of the results of [3],

$$R(u, v) = \min(u, v, R_1(u, v)) \qquad u, v \geqq 0$$

and the characterization that a test δ attains $R(u, v)$ if and only if for $n=0, 1, \ldots$ it satisfies a.s. P

$$\min(U_n, V_n, R_1(U_n, V_n)) = \begin{cases} U_n & \text{on} \quad (N=n) \cap \bar{F} \\ V_n & \text{on} \quad (N=n) \cap \bar{G} \\ R_1(U_n, V_n) & \text{on} \quad \{N>n\}. \end{cases} \tag{3}$$

Thus, the properties of optimal tests follow from the properties of the sets of points $\{u \leqq R_1(u, v)\}$ and $\{v \leqq R_1(u, v)\}$ in the first quadrant of the (u, v) plane. The structure of these regions is described in the following theorem.

Theorem 1. *There are positive-valued functions $U(\cdot)$ and $V(\cdot)$ on $[0, \infty)$, concave, continuous, and nondecreasing, such that for $u, v \geqq 0$*

$$\operatorname{sgn}(R_1(u, v) - v) = \operatorname{sgn}(V(u) - v) \tag{4}$$

and

$$\operatorname{sgn}(R_1(u, v) - u) = \operatorname{sgn}(U(v) - u). \tag{5}$$

U is bounded above if $P \neq F$, and V is bounded above if $P \neq G$. There is a $w_0 > 0$ such that for $w \geqq 0$

$$\operatorname{sgn}(V(w) - w) = \operatorname{sgn}(w_0 - w) \tag{6}$$

and

$$\operatorname{sgn}(U(w) - w) = \operatorname{sgn}(w_0 - w). \tag{7}$$

Proof. For fixed $u \geqq 0$

$$R_1(u, v) - v = \inf_{N \geqq 1} [EN + u\alpha + v(\beta - 1)],$$

which is evidently concave and continuous in v. Being positive at $v=0$ and negative for large v (considering $N=1$, $\alpha=1$, $\beta=0$), it has a unique zero, v

$=V(u)>0$, and $V(u)$ satisfies (4). Furthermore, in the first quadrant of the (u, v) plane the region $\{v \leqq R_1(u, v)\} = \{v \leqq V(u)\}$ is the intersection of regions of the form $\{v(1-\beta) \leqq EN + u\alpha\}$ and, hence, $V(\cdot)$ is concave, nondecreasing, and continuous on $[0, \infty)$. Similarly, there is a $U(\cdot)$ with the same properties satisfying (5).

If $P \neq F$, there are tests with stopping times $N \geqq 1$ satisfying $EN < \infty$, $\alpha < 1$, $\beta = 0$, e.g. the test that stops and rejects F at $\tilde{N}$, the first time $p_n > f_n$. If $\tilde{u} = E\tilde{N}/(1-\tilde{\alpha})$, then for all v

$$R_1(\tilde{u}, v) \leqq E\tilde{N} + \tilde{u}\tilde{\alpha} = \tilde{u},$$

whence by (5) $\tilde{u}$ is an upper bound on $U(v)$ for all v. If $P \neq G$, similar reasoning yields an upper bound on $V(u)$.

Since $F \neq G$, either $P \neq F$ or $P \neq G$ – say the latter. Then, since $V(\cdot)$ is bounded above and positive, its concavity and continuity imply that (6) holds for some positive w_0. Relation (7) follows immediately by applying (5), (4), and (6), and the theorem is proved.

3. The Optimality Property of the SPRT

By considering the special cases $P=F$ and $P=G$ of Theorem 1, the following proof of the optimality property of the SPRT is obtained. Like all known proofs, it follows the pattern of Wald and Wolfowitz's original argument, showing that an SPRT solves an auxiliary Bayes problem. The present argument also yields refinements of the optimality property (see Remarks 4–6 of Sect. 7). Note that it is not assumed that competitors of the SPRT have finite expected sample size under F and G. This assumption in [13] was shown to be removable by an additional argument (Burkholder and Wijsman [2]) that is not needed under the present treatment.

Theorem 2. *Let δ and δ' denote tests with stopping times and error probabilities N, α, β and N', α', β', respectively. Suppose δ acts like an SPRT in that for some $0 < B \leqq 1 \leqq A$ $(A \neq B)$ and for $n = 0, 1, \ldots$*

$$\begin{aligned} &(N=n) \cap \bar{F} \subset \{g_n/f_n \geqq A\} \\ &(N=n) \cap \bar{G} \subset \{g_n/f_n \leqq B\} \qquad (8) \\ &(N>n) \subset \{B \leqq g_n/f_n \leqq A\} \qquad \text{a.s. } F, G. \end{aligned}$$

If

$$\alpha' \leqq \alpha \quad \text{and} \quad \beta' \leqq \beta, \qquad (9)$$

then

$$E_F N' \geqq E_F N \quad \text{and} \quad E_G N' \geqq E_G N \qquad (10)$$

and both inequalities in (10) *are strict unless $\alpha' = \alpha$ and $\beta' = \beta$.*

Proof. Consider only the first inequality in (10), since the roles of F and G are interchangeable. Let $P = F$ in Theorem 1 and note that $U_n \equiv u$ for all n and $V_n = v g_n/f_n$. It clearly suffices to show that δ attains $R(u, v)$ for some $u, v > 0$.

The continuous function $V(U(w))/w$ equals one at w_0 by (6) and (7) and goes to zero as w goes to infinity, since $V(\cdot)$ is bounded. Hence, there is a $w_1 > w_0$ such that

$$\frac{B}{A} = \frac{V(U(w_1))}{w_1}.$$

Choose $u = U(w_1)$ and $v = w_1/A$. Then

$$Bv = V(u) \leqq u, \tag{11}$$

the last inequality by (6), since $u \geqq U(w_0) = w_0$. Also,

$$u = U(Av) < Av \tag{12}$$

by (7), since $Av = w_1 > w_0$.

Using (4) and (11),

$$\operatorname{sgn}(R_1(u, w) - w) = \operatorname{sgn}(Bv - w), \qquad w \geqq 0, \tag{13}$$

and, using (5) and (12),

$$\operatorname{sgn}(R_1(u, w) - u) = \operatorname{sgn}(U(w) - U(Av)). \tag{14}$$

By (11)-(14),

$$\min(u, V_n, R_1(u, V_n)) = \begin{cases} u & \text{if } V_n \geqq Av \\ V_n & \text{if } V_n \leqq Bv \\ R_1(u, V_n) & \text{if } Bv \leqq V_n \leqq Av, \end{cases}$$

which combines with (8) to show that δ satisfies (3). Thus, δ attains $R(u, v)$ and the theorem is proved.

4. Properties of the Optimal Stopping Boundaries

Additional information about the functions $U(\cdot)$ and $V(\cdot)$ of Theorem 1 is obtained in the following theorem. The numbers

$$L(P, F) = \exp\left(-\sum_{n=1}^{\infty} n^{-1}[P(p_n \leqq f_n) + F(p_n > f_n)]\right)$$

and

$$L(P, G) = \exp\left(-\sum_{n=1}^{\infty} n^{-1}[P(p_n \leqq g_n) + G(p_n > g_n)]\right)$$

play an important role, slightly different from that of the information numbers $I(P, F) = E\log(p/f)$ and $I(P, G) = E\log(p/g)$. Some basic properties of the L-numbers are described in [9].

Theorem 3. *The* w_0 *in Theorem* 1 *satisfies*

$$1 \leqq w_0 \leqq [1 - F(f < g) - G(f \geqq g)]^{-1}, \tag{15}$$

with equality on the right if $\min(f, g) \leqq p$ a.s. P. *Letting* $\bar{U} = \lim U(v)$ *as* $v \to \infty$,

$$\bar{U} \leqq L(P, F)^{-1}, \tag{16}$$

with equality if $P(g=0)=0$. *A similar result holds for* $\bar{V} = \lim V(u)$ *as* $u \to \infty$.

Remark. The assumption $\min(f, g) \leqq p$ a.s. P which is Assumption B of Kiefer and Weiss (*op. cit.*), is satisfied whenever optimal tests are truncated, i.e. for some n take at most n observations. In fact, it is easy to see using Theorem 1 that all optimal tests are truncated if and only if $\min(f, g) \leqq (1-\varepsilon) p$ a.s. P for some $\varepsilon > 0$.

Proof. Considering one-observation tests, clearly

$$1 \leqq R_1(w_0, w_0) \leqq 1 + E \min(U_1, V_1) = 1 + w_0 [F(f < g) + G(f \geqq g)]. \tag{17}$$

Since $R_1(w_0, w_0) = w_0$ by Theorem 1, (15) follows immediately. In case $\min(f, g) \leqq p$, $R_1(w_0, w_0)$ is attained by stopping after the first observation, since $\min(U_1, V_1)$ is at most w_0, and therefore equality actually holds in the second inequality in (17) and, accordingly, (15).

If $P = F$, then $L(P, F) = 0$ and (16) is trivial. If $P \neq F$, the upper bound on $U(v)$ obtained in the proof of Theorem 1 is

$$\bar{U} \leqq \tilde{u} = \frac{E\tilde{N}}{F(\tilde{N} = \infty)}, \tag{18}$$

where $\tilde{N}$ = first time $p_n > f_n$. Using Corollary 2 of Spitzer [12], if $f, p > 0$ a.s. F, P, then

$$E\tilde{N} = \sum_{n=0}^{\infty} P(\tilde{N} > n) = \exp\left(\sum_{n=1}^{\infty} n^{-1} P(\log(p_n/f_n) \leqq 0)\right),$$

and by virtue of (4.7) of [12]

$$\begin{aligned} F(\tilde{N} = \infty) &= F(\log(p_n/f_n) \leqq 0 \text{ for } n \geqq 1) \\ &= \exp\left(-\sum_{n=1}^{\infty} n^{-1} F(\log(p_n/f_n) > 0)\right), \end{aligned}$$

which combine with (18) to yield (16). In the non-absolutely continuous case where $P(f=0) > 0$ or $F(p=0) > 0$, Spitzer's results must be applied with greater care, but the resulting expressions for $E\tilde{N}$ and $F(\tilde{N} = \infty)$ are the same, as shown in the remark on p. 5 of [9].

To prove that equality holds in (16) if $P(g=0)=0$, consider first the case where $P \neq F$. Fix $u = \bar{U}$ and let $N^* = N^*(v)$ denote the rule that attains $R_1(u, v)$ by stopping when $R_1(U_n, V_n)$ exceeds $\min(U_n, V_n)$. At almost every (P) point in the sample space, the following hold for sufficiently large v:

$$U_n < V_n = v g_n / p_n \quad \text{for all } n \leqq \tilde{N}$$

and, since $\tilde{N}$ is the first time $U_n < \bar{U}$,

$$U_{\tilde{N}} < U(v) \leqq \bar{U} \leqq \min_{n < \tilde{N}} U_n,$$

which imply

$$N^* + \min(U_{N^*}, V_{N^*}) = \tilde{N} + U_{\tilde{N}}.$$

Therefore, by Fatou's Lemma and (5)

$$E\tilde{N} + \bar{U}\tilde{\alpha} = E(\tilde{N} + U_{\tilde{N}}) \leqq \lim_{v\to\infty} R_1(\bar{U}, v) \leqq \bar{U},$$

from which the reverse inequality in (18) follows immediately.

It remains only to prove that if $P = F$ and $P(g=0)=0$, then $\bar{U} = +\infty$, i.e. $U(\cdot)$ is not bounded above. This follows from (5) upon noting that for all choices of $u > 0$, $U_n \equiv u$ and

$$R_1(u, v) \geqq E \min\left[1+u, \min_{1 \leqq n < 1+u} \{n + \min(u, v g_n/p_n)\}\right] \to 1+u \tag{19}$$

as $v \to \infty$, since the g_n's are almost surely positive.

5. The Koopman-Darmois Case

If p, f, and g belong to a Koopman-Darmois family of densities,

$$f_\theta(x) = \exp(\psi(\theta)\, T(x) - b(\theta)),$$

and if $\psi(\theta(p))$ is strictly between $\psi(\theta(f))$ and $\psi(\theta(g))$, then

$$a \log(p/f) + b \log(p/g) \equiv 1 \qquad \text{for some } a, b > 0 \tag{20}$$

and it is helpful to transform from the first quadrant of the (u, v) plane to the (t, s) plane by

$$t(u, v) = a \log u + b \log v, \qquad s(u, v) = \log(v/u). \tag{21}$$

The image of the sequence $\{(U_n, V_n)\}$, where $U_0 = u$ and $V_0 = v$ are specified, is found from (20) to be

$$t_n = a \log u + b \log v - n, \qquad S_n = \log(v/u) + \log(g_n/f_n), \tag{22}$$

so that the representation $\{(t_n, S_n)\}$ is the left-to-right reversal of the usual plot of log-likelihood ratio vs. time. To transform the optimality characterization of Theorem 1, it is only necessary to characterize the images under (21) of the stopping regions, $\{u \leqq U(v)\}$, $\{v \leqq V(u)\}$, and their boundaries.

Theorem 4. *Suppose that f, g, and p satisfy* (20). *Then there are unique functions $A(\cdot)$ and $B(\cdot)$ such that if (t, s) denotes the image of (u, v) under* (21)

$$\operatorname{sgn}(s - A(t)) = \operatorname{sgn}(U(v) - u) \tag{23}$$

and

$$\operatorname{sgn}(s - B(t)) = \operatorname{sgn}(v - V(u)). \tag{24}$$

Moreover, the functions $A(\cdot)$ *and* $B(\cdot)$ *satisfy*

(i) $A(t)\uparrow$ *and* $B(t)\downarrow$, *strictly and continuously.*

(ii) $\operatorname{sgn} A(t)=\operatorname{sgn}(t-(a+b)\log[1-F(f<g)-G(f\geqq g)]^{-1})=-\operatorname{sgn} B(t)$.

(iii) $A(t)-b^{-1}t\downarrow$ *and has limit* $b^{-1}(a+b)\log L(P,F)$ *as* $t\to\infty$.

(iv) $B(t)+a^{-1}t\uparrow$ *and has limit* $-a^{-1}(a+b)\log L(P,G)$ *as* $t\to\infty$.

Proof. Define $A(\cdot)$ parametrically in terms of $v>0$ by

$$x=a\log U(v)+b\log v,\qquad A(x)=\log(v/U(v)),\tag{25}$$

noting that the relation between v and x is strictly increasing and continuous. Thus, $A(\cdot)$ is well-defined and, since the continuous function $U(\cdot)$ is concave and positive, $A(\cdot)$ is strictly increasing and continuous, as claimed in (i). To prove (23), let x be as in (25) and note that

$$\operatorname{sgn}(s-A(x))=\operatorname{sgn}(\log(U(v)/u))=\operatorname{sgn}(U(v)-u)\tag{26}$$

and also

$$\operatorname{sgn}(A(x)-A(t))=\operatorname{sgn}(x-t)=\operatorname{sgn}(U(v)-u),\tag{27}$$

which combine to yield (23). Similarly, $B(\cdot)$ is defined in terms of u and $V(u)$, is strictly decreasing and continuous, and satisfies (24). The uniqueness of $A(\cdot)$ and $B(\cdot)$ is clear from (23) and (24), which imply (ii) also, using (6) and (7) and the evaluation of w_0 in Theorem 3. By (25),

$$A(x)-b^{-1}x=-b^{-1}(a+b)\log U(v)\downarrow\quad\text{in } v,\tag{28}$$

so that the left-hand side is decreasing in x, and Theorem 3 evaluates the limit, proving (iii). Relation (iv) is similar and the proof is complete.

Theorem 4 along with Theorem 1 characterizes optimal tests as almost surely continuing only if $B(t_n)\leqq S_n\leqq A(t_n)$, rejecting f only if $S_n\geqq A(t_n)$, and rejecting g only if $S_n\leqq B(t_n)$. Thus, the optimal tests are generalized sequential probability ratio tests (GSPRT's), which was already established by Kiefer and Weiss (op. cit.) for a slightly larger class of problems. They also proved the monotonicity property in (i) and established (ii) without determining explicitly the "truncation point", $t=(a+b)\log[1-F(f<g)-G(f\geqq g)]^{-1}$, where the boundaries cross. (Weiss [14] found the truncation point, however, in the symmetric normal and binomial cases.) Note that the values of $A(t)$ and $B(t)$ to the left of the truncation point are immaterial, since optimality requires stopping there and choosing between f and g according to the sign of S_n. Remark 7 in Sect. 7 indicates that the Koopman-Darmois case is nearly the most general circumstance in which the optimal tests are GSPRT's.

Remark 3 in Sect. 7 shows that the monotonicity in (iii) and (iv) is strict.

6. An Example

In the Koopman-Darmois case, one can calculate the stopping boundaries, $A(t)$ and $B(t)$, along a sequence $t=t_0$, t_0+1, $t_0+2, \ldots$ by the method of "backward induction" ([13]) based upon the relation

$$R_1(u, v) = 1 + ER(U_1, V_1). \tag{29}$$

Using the transformation from $(u, v) \to (t, s)$ of the preceding section, define versions of R and R_1 in terms of s and t, $R^*(t, s)$ and $R_1^*(t, s)$, so that (29) becomes

$$R_1^*(t, s) = 1 + ER^*(t-1, s + \log(g/f)). \tag{30}$$

By (5) and (23)

$$\operatorname{sgn}(R_1^*(t, s) - u(t, s)) = \operatorname{sgn}(s - A(t)). \tag{31}$$

Thus, $A(t)$ is calculated by finding the s where $R_1^*(t, s) - u(t, s)$ changes sign and, similarly, $B(t)$ is the change point of $R_1^*(t, s) - v(t, s)$. Then $R^*(t, s) = R_1^*(t, s)$ for $B(t) \leqq s \leqq A(t)$, $= u(t, s)$ for $s \geqq A(t)$, $= v(t, s)$ for $s \leqq B(t)$, and the induction proceeds to $t+1$ after calculating $R^*(t, s)$ on a suitable grid of s-values. For an initial $t \leqq (a+b) \log w_0$, $R^*(t, s) = u(t, s)$ for $s \geqq 0$, $= v(t, s)$ for $s \leqq 0$.

This scheme was carried out on a computer for the case where f, g, and p are normal densities with variance one and means -0.1, 0.1, and 0, respectively. In this case, $a = b = 100$, $w_0 = 12.554$, $\bar{U} = \bar{V} = 211.74$, and $\log(g/f) = 0.2X$, where X denotes an observation with density f, g, or p. The following tabulated data describe the optimal boundaries $A(t)$, $B(t)$ ($B(t) = -A(t)$) and characteristics of optimal tests (the ones starting from $s = 0$ and the indicated t).

t	$\bar{A}(t)$	$A(t)$	EN	α, β
506	−5.65	0	0	0.500
606	−4.65	0.09	1.50	0.445
706	−3.65	0.22	3.02	0.418
806	−2.65	0.43	7.17	0.371
906	−1.65	0.74	16.21	0.310
1006	−0.65	1.18	32.86	0.241
1106	0.35	1.74	58.74	0.175
1206	1.35	2.41	93.72	0.121
1306	2.35	3.16	136.62	0.0810
1406	3.35	3.98	186.04	0.0527
1506	4.35	4.84	240.68	0.0338
1606	5.35	5.74	299.54	0.0214
1706	6.35	6.66	361.82	0.0134
1806	7.35	7.59	426.93	0.00838

For large-sample considerations, it is natural to use $\bar{A}(t)$ and $\bar{B}(t)$ as stopping boundaries instead of $A(t)$ and $B(t)$, which are a great deal harder to compute. It was shown in [8] that this approach attains $R(u, v)$ to within $o(1)$ as $\max(u, v) \to \infty$, assuming that $\log(p/f)$ and $\log(p/g)$ have finite second moments under P and $p > 0$ a.s. F, G. This leads to the theorem in [8] that a 2-SPRT, i.e. a test that rejects F the first time $f_n/p_n \leqq A$ and rejects G the first time $g_n/p_n \leqq B$

(rejecting either if both inequalities hold), minimizes EN to within $o(1)$ as $\min(A, B) \to 0$, among all tests with the same or smaller error probabilities. In the normal mean example, calculations reported in [8] showed that with $A = B$ the 2-SPRT's minimize EN to within 1%.

7. Remarks and Refinements

1. In the Koopman-Darmois case of Sect. 5 there is a natural way of searching for Kiefer-Weiss solutions. Assuming that $A(t)$ and $B(t)$ have been determined for $t = t^*, t^*+1, \ldots, t^*+k$, say, where $t^* \geqq t_0 > t^* - 1$, fix t at one of these values, say $t = t_1$. If there is an s_1 between $A(t_1)$ and $B(t_1)$ such that the optimal test starting from (t_1, s_1) has its maximum expected sample size when $\theta = \theta(p)$, then this test evidently minimizes the maximum expected sample size among all tests with the same or smaller α and β. It is plausible that such an s_1 can be found because one expects that for s near one boundary the maximizing θ should be larger than $\theta(p)$ and for s near the other boundary it should be smaller. If the location of the maximizing θ varies continuously with s, then there should be an s_1 where $\theta(p)$ is maximizing. This suggests that solutions of the Kiefer-Weiss problem are to be found among the solutions of the modified Kiefer-Weiss problem in the Koopman-Darmois case.

2. In case $\min(f, g) < p$, the results of Sect. 4 can be extended to show that as $u, v \downarrow w_0$

$$R(u, v) = 1 + uF(f \leqq g) + vG(f > g) + o(v - w_0),$$

which leads to the conclusion that $U(\cdot)$ has right-hand derivative at w_0 equal to $G(f > g)/F(f > g)$. A similar formula holds for $V(\cdot)$, and in the Koopman-Darmois case these yield expressions for the right-hand derivatives of the boundaries at the truncation point, t_0:

$$A'(t_0) = \frac{F(f > g) - G(f > g)}{bF(f > g) + aG(f > g)},$$

$$B'(t_0) = -\frac{G(g > f) - F(g > f)}{aG(g > f) + bF(g > f)}.$$

For the example of Sect. 6, the derivatives are ± 0.0008, whereas the slopes of the asymptotes are ± 0.01. Anderson [1] found that for 5% error probabilities the best choice of slopes for straight-line boundaries is about ± 0.0066, which yields boundaries reasonably close to the optimum ones over the range needed. (As pointed out in Sect. 6, however, using the asymptotes themselves as boundaries is highly efficient.)

3. It is straightforward to show that the monotonicity of $U(\cdot)$ and $V(\cdot)$ in Theorem 1 is strict under stronger assumptions, e.g. if $P(f \geqq p > g > 0) > 0$ then U is strictly increasing. The argument is based on noting that for every $v > w_0$ there is an n such that $\{f \geqq p > (v/w_0)^{1/n} g > 0\}$ has positive P-probability, and the occurrence of this event on each of the first n observations would cause a certain

test optimal for $u=U(v)$ to reject G, showing that this test has a positive β. It follows that $U(v)$ is larger than $U(v')$ for all $v'<v$ and, since v can be chosen arbitrarily large, U, being concave, must be strictly increasing.

4. In Theorem 2, the error probability conditions, (9), are needed in the proof only to establish that $u\alpha'+v\beta'\leqq u\alpha+v\beta$, which follows alternatively from (11) and (12) if it is assumed that

$$\lambda\alpha'+\beta'\leqq\lambda\alpha+\beta \qquad \text{for } \lambda=A,B. \tag{32}$$

Interchanging the roles of f and g does not change this condition. (Simons' [11] conditions $\alpha'/(1-\beta')\leqq\alpha/(1-\beta)$ and $\beta'/(1-\alpha')\leqq\beta/(1-\alpha)$ are also sufficient, being stronger than (32) by virtue of Wald's bounds, $\alpha/(1-\beta)\leqq A^{-1}$ and $\beta/(1-\alpha)\leqq B$.) Furthermore, at least one of the inequalities in (10) is strict unless δ' satisfies (8). To see this, note that if neither inequality is strict then δ' attains $R_1(u,v)$ and, hence, by (3) and (13) the events $(N'>n)\cap\{g_n<Bf_n\}$ and $(N'=n)\cap\bar{G}\cap\{g_n>Bf_n\}$ are F-null. It clearly follows that the first event is G-null, and the second is, too, because its intersection with $\{f_n=0\}$ is G-null (otherwise, β' could be reduced without changing α' or N'). Similarly, the conditions related to the A-inequality hold a.s. F, G, so that δ' satisfies (8).

5. Both inequalities in (10) are, in fact, strict, except in rather special circumstances. If, say, the expected sample sizes under F are equal, then either

(i) $g\geqq f$ a.s. G, *or*

(ii) δ' *satisfies* (8) a.s. F.

(In case (i) holds, $\beta=0$ for all SPRT's with $B<1$.) If (i) does not hold, then $\{f>g>0\}$ has positive probability under G, hence also under F, and by Remark 3 U is strictly increasing. Thus, (14) can be strengthened to

$$\operatorname{sgn}(R_1(u,w)-u)=\operatorname{sgn}(w-A\,v),$$

which combines with (13) to show that δ' can attain $R_1(u,v)$ only if it obeys the same (A,B)-prescription as δ, a.s. F.

It is, however, possible that δ' does *not* satisfy (8) a.s. G. This can occur only if $G(f=0)>0$ and if (with positive probability) δ' continues rather than rejecting f when $f_n=0$. (Note that this "indecisiveness" does not affect α', β' or $E_F N'$.) In the absolutely continuous case where $F(g=0)=0$ and $G(f=0)=0$, one concludes that both inequalities in (10) are strict unless δ' satisfies (8). The only tests other than SPRT's and "indecisive" SPRT's that can minimize one of the expected sample sizes are those in case (i) (or its analog with f and g interchanged), where it is easy to see that the tests that are optimal (in the sense of minimizing $E_F N$ among tests with $\beta=0$ and with the same or smaller α) are those tests that reject g (a.s. F) if and only if $g_n=0$. Among these tests, however, only the SPRT's are optimal for $E_G N$ (by Remark 4).

6. If two SPRT's have the same α and β, then by Theorem 1 their expected sample sizes match and by Remark 4 they satisfy (8) for the same (A,B) pairs. (This uniqueness result is essentially Theorem 5 of [15].) The same conclusion holds if two SPRT's have at least one match between error probabilities and at least one between expected sample sizes, since by the last line of Theorem 1 both error probabilities must match.

7. Under mild regularity conditions it is easily shown that the optimal tests are GSPRT's only if g/f is a sufficient statistic for $\{F, G, P\}$, whence p belongs to the Koopman-Darmois family generated by f and g (and lies between them). In the case of discrete measures, for instance, it is easily seen that the value of g/f must determine the value of p/f, or else there is an initial point (u_0, v_0) from which the same value of g/f can yield (U_1, V_1) belonging to the (strict) stopping region or, alternatively, the continuation region, depending upon the value of p/f.

References

1. Anderson, T.W.: A modification of the sequential probability ratio test to reduce the sample size. Ann. Math. Statist. **31**, 165–197 (1960)
2. Burkholder, D.L., Wijsman, R.A.: Optimum properties and admissibility of sequential tests. Ann. Math. Statist. **34**, 1–17 (1963)
3. Chow, Y.S., Robbins, H., Siegmund, D.: Great Expectations: The Theory of Optimal Stopping. Boston: Houghton-Mifflin 1971
4. Ferguson, T.S.: Mathematical Statistics, a Decision Theoretic Approach. New York: Academic Press 1967
5. Kiefer, J., Weiss, L.: Some properties of generalized sequential probability ratio tests. Ann. Math. Statist. **28**, 57–75 (1957)
6. Lai, T.L.: Optimal stopping and sequential tests which minimize the maximum expected sample size. Ann. Statist. **1**, 659–673 (1973)
7. Lehmann, E.L.: Testing Statistical Hypotheses. New York: Wiley 1959
8. Lorden, G.: 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size. Ann. Statist. **4**, 281–291 (1976)
9. Lorden, G.: Nearly-optimal sequential tests for finitely many parameter values. Ann. Statist. **5**, 1–21 (1977)
10. Matthes, T.K.: On the optimality of sequential probability ratio tests. Ann. Math. Statist. **34**, 18–21 (1963)
11. Simons, G.: An improved statement of optimality for sequential probability ratio tests. Ann. Statist. **4**, 1240–1243 (1976)
12. Spitzer, F.: A combinatorial lemma and its application to probability theory. Trans. Amer. Math. Soc. **82**, 332–339 (1956)
13. Wald, A., Wolfowitz, J.: Optimum character of the sequential probability ratio test. Ann. Math. Statist. **19**, 326–339 (1948)
14. Weiss, L.: On sequential tests which minimize the maximum expected sample size. J. Amer. Statist. Assoc. **57**, 551–566 (1962)
15. Wijsman, R.A.: Existence, uniqueness, and monotonicity of sequential probability ratio tests. Ann. Math. Statist. **34**, 1541–1548 (1963)

Received May 18, 1978