SOME PROPERTIES OF GENERALIZED SEQUENTIAL PROBABILITY RATIO TESTS[1]

By J. Kiefer and Lionel Weiss

Cornell University and University of Oregon

0. Introduction and Summary. Generalized sequential probability ratio tests (hereafter abbreviated GSPRT's) for testing between two simple hypotheses have been defined in [1]. The present paper, divided into four sections, discusses certain properties of GSPRT's. In Section 1 it is shown that under certain conditions the distributions of the sample size under the two hypotheses uniquely determine a GSPRT. In the second section, the admissibility of GSPRT's is discussed, admissibility being defined in terms of the probabilities of the two types of error and the distributions of the sample size required to come to a decision; in particular, notwithstanding the result of Section 1, many GSPRT's are inadmissible In Section 3 it is shown that, under certain monotonicity assumptions on the probability ratios, the GSPRT's are a complete class with respect to the probabilities of the two types of error and the average distribution of the sample size over a finite set of other distributions. In Section 4, finer characterizations are given of GSPRT's which minimize the expected sample size under a third distribution satisfying certain monotonicity properties relative to the other two distributions; these characterizations give monotonicity properties of the decision bounds.

1. Uniqueness of certain GSPRT's. In this section we identify a GSPRT with the two sequences of limits characterizing it. Using the same notation as in [1], we assume that the Lebesgue densities f_1 and f_2 satisfy the conditions in Section 2 of [1], even for c equal to zero or infinity (i.e., the probability ratio for any number of observations takes on no single value with positive probability), and are continuous (this last restriction is easily weakened; see also Remark 1 at the end of this section for further generalization).

First we make the transformation $Y_i = F_1(X_i)$. Under H_1, Y_i has a rectangular distribution; under H_2, the density of Y_i will be g (say), where $\int_0^1 g(y)\,dy = 1$.

Next we make the transformation $Z_i = \phi[g(Y_i)]$, where $\phi(u)$ is strictly increasing in u for $u \geqq 0$, $\phi(u)$ takes on no values outside the interval [0, 1], and also

$$\int_{\{y \mid \phi[g(y)] \leqq t\}} dy = t \quad \text{for} \quad 0 \leqq t \leqq 1.$$

Under H_1, the distribution of Z_i is rectangular, while under H_2, the density of Z_i is ϕ^{-1}; we note that $\phi^{-1}(z)$ is strictly increasing in z for $0 \leqq z \leqq 1$ and is 0 elsewhere.

Received December 30, 1955; revised October 2, 1956.

[1] Research sponsored by the Office of Naval Research.

Since we can always transform from $X_1, X_2, \cdots$ to $Z_1, Z_2, \cdots$ and carry out any GSPRT in terms of $Z_1, Z_2, \cdots$, from now on we assume that $f_1(x) = 1$ on $[0, 1]$, and $f_2(x)$ is strictly increasing in x for $0 \leqq x \leqq 1$, with $\int_0^1 f_2(x)\,dx = 1$. We also assume f_2 piecewise differentiable (for the sake of obtaining the g_i below). Any GSPRT is carried out by seeing whether

$$b_m < \prod_{i=1}^{m} f_2(X_i) < a_m, \text{ etc.,}$$

or, defining Q_i as $\log f_2(X_i)$, B_m as $\log b_m$, A_m as $\log a_m$, and W_m as $Q_1 + Q_2 + \cdots + Q_m$, whether $B_m < W_m < A_m$, etc.

Denoting by g_i the density function of Q_1 under H_i, we find that $g_2(q) = e^q \cdot g_1(q)$ identically in q.

Suppose the $m - 1$ pairs $(a_1, b_1), \cdots, (a_{m-1}, b_{m-1})$ are fixed. The joint conditional density function of $(Q_1, Q_2, \cdots, Q_m)$ under H_i, given that sampling continues beyond the $(m - 1)$st observation, is

$$h_i(q_1 \cdots, q_m) = \frac{\prod_{j=1}^{m} g_i(q_j)}{K_i}$$

in the region $\{b_j < w_j < a_j ; j = 1, \cdots, m - 1\}$; $h_i(q_1, \cdots, q_m) = 0$ elsewhere. Here $K_i = P$ {sampling continues beyond $(m - 1)$ observations under H_i}, and we assume K_1 and K_2 are positive. Thus

$$h_2(q_1, \cdots, q_m) = \frac{K_1}{K_2} \cdot e^{w_m} \cdot h_1(q_1, \cdots, q_m).$$

Then, denoting the conditional density of W_m under H_i given that sampling continues beyond $m - 1$ observations by k_i, we have

$$k_2(w) = \frac{K_1}{K_2} \cdot k_1(w) \cdot e^w.$$

Now we make the following *Assumption* A: f_2 is such that $g_i(q) > 0$ for almost all $q(i = 1, 2,)$; thus, if S is any nondegenerate interval, $\int_S g_i(q)\,dq > 0$ for $i = 1, 2$. But this implies that if T is any nondegenerate interval, $\int_T k_i(w)\,dw > 0$ for $i = 1, 2$.

For any given positive numbers C, D, we now show that there is at most one solution (γ, δ) to the two equations

$$\int_\gamma^\delta k_1(w)\,dw = C, \qquad \int_\gamma^\delta k_2(w)\,dw = D.$$

For, given any γ which can be the first element of a solution, let $\delta(\gamma)$ be the uniquely determined value of δ which satisfies the first equation. Then it is easily verified that

$$\int_{\gamma}^{\delta(\gamma)} k_2(w)\, dw = \int_{\gamma}^{\delta(\gamma)} \frac{K_1}{K_2} \cdot k_1(w) \cdot e^w\, dw$$

is strictly increasing in γ. This proves that there is at most one solution (γ, δ) to the two equations.

Let $D_i(n; T)$ denote the probability that a decision is reached after no more than n observations when using the test T and H_i is true. The considerations above yield

THEOREM 1. *If Assumption* A *holds, and if T is a nontruncated* GSPRT, *then there is no* GSPRT *T' different from T and with $D_i(n; T') = D_i(n; T)$ for all n and for $i = 1, 2$. If Assumption* A *holds, and if T is a* GSPRT *truncated after m observations, while T' is another* GSPRT *with $D_i(n; T') = D_i(n; T)$ for all n and for $i = 1, 2$, then T and T' differ only in the terminal decision boundary at stage m.*

REMARK 1. If Assumption A is violated, there are two different GSPRT's, T and T', with $D_i(n; T') = D_i(n; T)$ for all n and for $i = 1, 2$. For we can find a GSPRT T whose first pair of limits is (B_1, A_1), such that for a positive ϵ,

$$\int_{A_1}^{A_1+\epsilon} g_1(q)\, dq = 0.$$

But then T' can be taken as the GSPRT whose first pair of limits is $(B_1, A_1 + \epsilon)$, the other limits agreeing with those for T. The inessential difference between T and T' in such a case as this will be evident to the reader: every sample sequence in a set of probability one under both H_i suffers the same fate under T' as under T. Similarly, in the discrete case (or where Q_1 can take on some constant value with positive probability), there is the aspect of randomization in which two tests with identical $D_i(n; T)$ may differ (e.g., if T and T' both always require at least 3 observations, for some value of $Q_1 + Q_2 + Q_3$, T may stop after 3 observations if $Q_1 < Q_2$, T' if $Q_2 < Q_1$). With these modifications in mind, it is evident that Theorem 1 is of broader validity than its stated form.

REMARK 2. If Assumption A holds, one can prove similarly that there is at most one GSPRT having given values for the elements of the following sequences of probabilities:

$$\{P(\text{accepting } H_1 \text{ under } H_2 \text{ at stage } n)\},$$

$$\{P(\text{accepting } H_2 \text{ under } H_1 \text{ at stage } n)\}.$$

Incidentally, it is easy by the methods of [1] or [2] to show that the GSPRT's (and k-decision problem analogues) form a complete class with respect to the generalized risk function consisting of such sequences. A similar remark applies if these sequences and the $D_i(n; T)$ are considered together as the risk function, etc.

2. Questions of admissibility.[2] We have proved in the previous section that under certain conditions there is at most one GSPRT corresponding to any two specified distributions of n (one under each H_i). This does not imply that all GSPRT's are admissible (letting α_i = probability of an incorrect decision under H_i and p_{in}^* = probability that the experiment terminates after more than n observations under H_i, a GSPRT is said to be admissible in this section if there is no second procedure for which all of the numbers α_i and p_{in}^* ($i = 1, 2$; $n = 0, 1, 2, \cdots$) are no greater, and at least one is less, than for the given procedure); as we shall see in a simple example below, the general question of how to characterize the admissible procedures seems quite difficult. Before turning to this example, we make a few remarks on admissibility. Firstly, it is clear that admissibility does not entail any simple monotone character of the constants a_n and b_n: on one hand, putting p_{in} = probability of terminating after n observations under H_i, on considering the minization of

$$\sum_{i=1}^{2} \xi_i \left[\alpha_i W_i + \sum_{n=1}^{\infty} D_{in} p_{in} \right] \text{where } D_{in} = \sum_{1}^{i} C_{jn}, \tag{2.1}$$

where the C_{in} are an increasing (resp., decreasing) sequence of positive numbers, by comparison with the case $C_{in} = 1$, it becomes clear that we may have admissible procedures with $a_n\uparrow$, $b_n\downarrow$ (resp., $a_n\downarrow$, $b_n\uparrow$); similarly, there are admissible procedures with $a_n\uparrow$, $b_n\uparrow$ or $a_n\downarrow$, $b_n\downarrow$. (In the case $C_{in}\uparrow$ (resp., $C_{in}\downarrow$), it is also interesting to note that $a^{(m)} \leqq a_m \leqq b_m \leqq b^{(m)}$ (resp., $a_m \leqq a^{(m)} \leqq b^{(m)} \leqq b_m$), where $a^{(m)}$ and $b^{(m)}$ are the constant bounds for which (2.1) is minimized when C_{in} is replaced by $C_{i,m+1}$ for all n.) On the other hand, by considering $C_{in} = (h_i + k_i) + (-1)^n h_i$, one may obviously obtain admissible procedures for which $a_n = a_{n+2}$, $b_n = b_{n+2}$, $a_{2n} < a_{2n+1} < b_{2n+1} < b_{2n}$ for all n. Other admissible procedures for which the a_n and b_n have no simple monotone character may be constructed similarly.

Secondly, one can sometimes give simple *necessary* conditions for admissibility (a sufficient condition such as that of being the essentially unique procedure which minimizes (2.1) for some choice of the constants ξ_i, W_i, C_{in} will usually be hard to verify). Suppose, as before, that every interval of positive values (but no single value) of $F_2(x)/f_1(x)$ has positive probability under both H_i. Let N be the smallest integer for which $a_N = b_N$ (if no such integer exists, write $N = \infty$). Then a necessary condition for admissibility of a GSPRT is

$$\sup_{n<N+1} a_n \leqq \inf_{n<N+1} b_n. \tag{2.2}$$

To see this, note that for any Bayes solution minimizing (2.1) the constants ξ_i, W_i must satisfy $a_n \leqq (W_1\xi_1/W_2\xi_2) \leqq b_n$ for all $n < N + 1$; thus, any Bayes solution must satisfy (2.2). Since the essentially complete class of procedures which is the closure in the sense of (Wald's) regular convergence (see [3], [2]) of

---

[2] In Section 2 the roles of the symbols a and b are reversed from what they are in the other sections.

this class of Bayes procedures satisfying (2.2) also satisfies (2.2), and since (see Section 1) this class must include all admissible GSPRT's, the necessity of (2.2) is established. As will be seen in the examples below, the condition (2.2) is not in general sufficient for admissibility.

We now give an example which will illustrate how complicated an explicit delimitation of the *admissible* GSPRT's seems to be. Consider all GSPRT's requiring at least 1 and at most 2 observations for testing $H_1:\theta = 1$ against $H_2:\theta = 2$ where

$$f_\theta^*(x) = \begin{cases}\theta e^{-\theta x}, & x \geqq 0\\ 0, & x < 0.\end{cases} \tag{2.3}$$

Let X_1, X_2 be independent with density f_θ^*; write $Y_i = e^{-X_i}$. The hypotheses can then be rewritten as $H_1:Y_i$ have density $f_1(y)$ against $H_2:Y_i$ have density $f_2(y)$, where

$$\begin{aligned} f_1(y) &= \begin{cases}1, \ 0 < y < 1\\ 0 \text{ otherwise}\end{cases}\\ f_2(y) &= \begin{cases}2y, \ 0 < y < 1\\ 0 \text{ otherwise.}\end{cases}\end{aligned} \tag{2.4}$$

Thus, $\frac{1}{2}f_2(y)/f_1(y) = y$, and we may write the general form of the GSPRT as:

$$\begin{cases}\text{If } Y_1 \leqq a \text{ (resp., } \geqq b\text{), stop and accept } H_1 \text{ (resp., } H_2\text{);}\\ \text{If } a < Y_1 < b \text{ and } Y_1Y_2 \leqq k \text{ (resp., } >k\text{), accept } H_1 \text{ (resp., } H_2\text{).}\end{cases} \tag{2.5}$$

Here we may assume a, b, k lie between 0 and 1 inclusive. If $a = b$, k is of no importance. If $a < b$, we may suppose $0 < k < 1$, since $k = 0$ or $k = 1$ is clearly inadmissible (replace b by $b' = a$ or a by $a' = b$, respectively, to obtain better procedures); also, since $Y_1Y_2 \leqq Y_1$ with probability one under both H_1 and H_2, we may suppose $k \leqq a$, since $k > a$ is clearly inadmissible (replace a by $a' = \min(k, b)$ for a better procedure). All procedures with $a = b$ are admissible. To summarize, then, in investigating which tests are admissible, we may eliminate certain trivial cases mentioned above and hereafter assume

$$0 < k \leqq a < b \leqq 1. \tag{2.6}$$

The characteristics of any such procedure are easily computed and may be summarized in the risk vector of any such procedure, which is given by the quadruple

$$\begin{aligned} r(a, b, k) &= \{P_1 \text{ (accept } H_2), P_2 \text{ (accept } H_1), P_1(n = 2), P_2(n = 2)\}\\ &= \left\{1 - a - k \log \frac{b}{a}, a^2 + 2k^2 \log \frac{b}{a}, b - a, b^2 - a^2\right\}.\end{aligned} \tag{2.7}$$

The question of inadmissibility or admissibility of such a procedure is then that of whether or not there exists a test $(\bar{a}, \bar{b}, \bar{k})$ for which all components of $r(\bar{a}, \bar{b}, \bar{k})$ are $\leqq$ the corresponding ones of $r(a, b, k)$, with strict inequality for at least one component. Since no two different tests have identical risk functions (see Section 1), inadmissibility of (a, b, k) is equivalent to (I) the existence of $(\bar{a}, \bar{b}, \bar{k})$ not identical to (a, b, k) and with (II) all components of $r(\bar{a}, \bar{b}, \bar{k}) \leqq$ the corresponding ones of $r(a, b, k)$. The latter condition (II) may be written

$$
(2.8)\qquad
\begin{aligned}
&\text{(a)} && \bar{b} - \bar{a} && \leqq b - a\\
&\text{(b)} && \bar{b}^2 - \bar{a}^2 && \leqq b^2 - a^2\\
&\text{(c)} && \bar{a} + \bar{k}\log(\bar{b}/\bar{a}) && \geqq a + k\log(b/a)\\
&\text{(d)} && \bar{a}^2 + 2\bar{k}^2\log(\bar{b}/\bar{a}) && \leqq a^2 + 2k^2\log(b/a).
\end{aligned}
$$

The possibility that $\bar{b} = \bar{a}$ may be eliminated in all that follows: if $\bar{b} = \bar{a}$, squaring both sides of (c) and comparing with (d) yields $(a + k\log(b/a))^2 \leqq a^2 + 2k^2 \log(b/a)$; i.e., $k\log(b/a) \leqq 2(k - a)$, which is impossible. Thus, we may hereafter assume $\bar{a} < \bar{b}$. Also, $\bar{a} \geqq \bar{k} > 0$ for admissibility. Thus, in particular, $0 < \log(\bar{b}/\bar{a}) < \infty$ in all that follows.

Combining (2.8) (c) and (d), we obtain

$$
(2.9)\qquad \left\{\max\left[0, \frac{a - \bar{a} + k\log(b/a)}{\log(\bar{b}/\bar{a})}\right]\right\}^2 \leqq \bar{k}^2 \leqq \frac{a^2 - \bar{a}^2 + 2k^2\log(b/a)}{2\log(\bar{b}/\bar{a})}.
$$

In particular, the right-hand term of (2.9) must be >0. Thus, for a given $\bar{a}, \bar{b}$ with $0 < \bar{a} < \bar{b} \leqq 1$, (2.9) can be satisfied for some $\bar{k}$ with $0 < \bar{k} \leqq \bar{a} < \bar{b}$ if and only if either

$$
(2.10)\qquad \frac{a - \bar{a} + k\log(b/a)}{\log(\bar{b}/\bar{a})} \leqq 0 < \frac{a^2 - \bar{a}^2 + 2k^2\log(b/a)}{2\log(\bar{b}/\bar{a})}
$$

or else

$$
(2.11)\qquad
\begin{cases}
\text{(a)} & 0 < \dfrac{a - \bar{a} + k\log(b/a)}{\log(\bar{b}/\bar{a})} \leqq \bar{a} \quad \text{and}\\[2ex]
\text{(b)} & \left(\dfrac{a - \bar{a} + k\log(b/a)}{\log(\bar{b}/\bar{a})}\right)^2 \leqq \dfrac{a^2 - \bar{a}^2 + 2k^2\log(b/a)}{2\log(\bar{b}/\bar{a})}.
\end{cases}
$$

Equation (2.10) implies

$$
(a + k\log(b/a))^2 \leqq \bar{a}^2 < a^2 + 2k^2\log(b/a),
$$

the extreme members of which give $2a < 2k - k\log(b/a)$, an impossibility. Since the right side of (2.11) (b) is positive, we also see that the first inequality of (2.11) (a) is implied by (2.11) (b): otherwise, we would again have the contradiction $2a \leqq 2k - k\log(b/a)$. Thus, (2.8) (c) and (2.8) (d) may be satisfied for some $\bar{a}, \bar{k}, \bar{b}$ with $0 < \bar{k} \leqq \bar{a} < \bar{b} \leqq 1$ if and only if (2.11) (b) and the second half of (2.11) (a) may be satisfied for some $\bar{a}, \bar{b}$ with $0 < \bar{a} < \bar{b} \leqq 1$. Write

$$
c = \log(b/a), \qquad \bar{c} = \log(\bar{b}/\bar{a}), \lambda = k/a, \qquad \gamma = \bar{a}/a.
$$

Equations (2.8) (a), (b) may be written

$$(2.12)\qquad \begin{aligned} &\text{(a)}\quad \bar{c} \leqq \log\,[1 + (e^{c} - 1)/\gamma],\\ &\text{(b)}\quad \bar{c} \leqq \tfrac{1}{2}\log\,[1 + (e^{2c} - 1)/\gamma^{2}]. \end{aligned}$$

Since $c, \bar{c} > 0$, equation (2.12) (a) implies or is implied by (2.12) (b) according to whether $\gamma \leqq 1$ or $\gamma \geqq 1$. We may write the restriction $\bar{b} \leqq 1$ as

$$(2.13)\qquad \bar{c} \leqq c - \log b\gamma.$$

Equation (2.12) (a) implies (2.13) if $\gamma \leqq 1$, since $b \leqq 1$. If $\gamma > 1$, (2.12) (b) implies or is implied by (2.13) according to whether or not $\gamma < [1 + e^{2c}(b^{-2} - 1)]^{1/2}$.

To summarize, then, (2.11), (2.12), and (2.13) imply that a given (k, a, b) (and hence, (c, λ, b)) is inadmissible if and only if there exist positive numbers $\bar{c}, \gamma$ with either $\bar{c} \neq c$ or $\gamma \neq 1$ (note from (2.9) that $\bar{c} = c, \gamma = 1$ imply $\bar{k} = k$ and hence $(\bar{a}, \bar{b}, \bar{k}) = (a, b, k)$) and satisfying

$$(2.14)\qquad \begin{aligned} &\text{(a)}\quad f_1(\gamma) \leqq \bar{c} \lneqq f_2(\gamma),\\ &\text{(b)}\quad \bar{c} \geqq f_3(\gamma),\\ &\text{(c)}\quad \gamma < \sqrt{1 + 2\lambda^2 c}, \end{aligned}$$

where

$$f_1(\gamma) = (1 + \lambda c - \gamma)/\gamma,$$

$$(2.15)\qquad f_2(\gamma) = \begin{cases} \log\,[1 + (e^{c} - 1)/\gamma] & \text{if} \quad \gamma \leqq 1,\\ \tfrac{1}{2}\log\,[1 + (e^{2c} - 1)/\gamma^{2}] & \text{if} \quad 1 < \gamma < [1 + e^{2c}(b^{-2} - 1)],\\ c - \log b\gamma & \text{if} \quad \gamma \geqq [1 + e^{2c}(b^{-2} - 1)]^{1/2}, \end{cases}$$

$$f_3(\gamma) = 2(1 - \gamma + \lambda c)^2 \,/\, (1 - \gamma^2 + 2\lambda^2 c),$$

and where condition (2.14) (c) merely expresses the positivity of the right side of (2.11) (b).

Suppose $\lambda < 1$ (the case $\lambda = 1$ can be treated easily). It is evident that $f_1(1) = \lambda c < c = f_2(1) = f_3(1)$ and that all points $(\gamma, \bar{c})$ with $|\gamma - 1|$ and $|\bar{c} - c|$ of sufficiently small magnitude and for which $\bar{c} \leqq f_2(\gamma)$, satisfy (2.14) (a) (and, obviously, (2.14)(c)). Hence, primes denoting derivatives, a *necessary* condition for (a, b, k) to be *admissible* is that

$$(2.16)\qquad f_2'(1-) \geqq f_3'(1) \geqq f_2'(1+).$$

Evaluating (2.16) from (2.15), we obtain

$$(2.17)\qquad 1 - e^{-c} \leqq (2\lambda - 1)\,/\,\lambda^2 \leqq \begin{cases} 1 - e^{-2c} & \text{if} \quad b < 1,\\ 1 & \text{if} \quad b = 1. \end{cases}$$

On the other hand, the necessary condition $\sup a_n \leqq \inf b_n$ of (2.2) may be written $1/2 \leqq \lambda \leqq e^c / 2$, or (since $\lambda < 1$)

$$(2.18) \qquad 0 \leqq (2\lambda - 1) / \lambda^2 \leqq \begin{cases} 4e^{-c}(1 - e^{-c}) & \text{if} \quad e^c < 2, \\ 1 & \text{if} \quad e^c \geqq 2. \end{cases}$$

Clearly, (2.18) includes many procedures not included in (2.17) (hence, not in the complement of the set of procedures described by (2.14)). Thus, equation (2.2) is *not sufficient* for admissibility.

We shall not consider this example further: it is already evident from (2.14) that, even in simple cases, the delimitation of the admissible procedures can become complicated.

3. Controlling the distribution of the sample size under distributions other than those being tested. In this section we shall characterize (under certain assumptions) an essentially complete class of tests (the risk function is given below) for testing H_0:$f = f_{\theta_0}$ against H_1:$f = f_{\theta_{k+1}}$ sequentially, where the test is based on independent random variables $X_1, X_2, \cdots$ with common density f_θ with respect to some σ-finite measure μ. There are specified values K and $\theta_1, \cdots, \theta_K$, as well as non-negative numbers $a_0, a_1, \cdots, a_{K+1}$ whose sum is unity. The "risk function" of a procedure consists of the vector

$$(3.1) \quad \left\{P_{\theta_0}\{\text{accept } H_1\}, P_{\theta_{K+1}}\{\text{accept } H_0\}, \sum_{i=0}^{K+1} a_i P_{\theta_i}\{n \geqq j\}, j = 1, 2, \cdots\right\},$$

where n is the (chance) number of observations required. We consider only procedures for which $P_{\theta_i}\{n < \infty\} = 1$ for all i. One procedure is said to be at least as good as a second one if each component of (3.1) for the first is no greater than the corresponding component for the second, and the notion of essential completeness is relative to this definition of "as good as."

We assume in this section that the $f_{\theta_i}(0 \leqq i \leqq K + 1)$ are finite everywhere and have the same region of positivity and, writing $p_{im}(x^{(m)}) = \prod_{j=1}^m f_{\theta_i}(x_j)$ with $x^{(m)} = (x_1, \cdots, x_m)$, that the functions $p_{im}(x^{(m)}) / p_{0m}(x^{(m)})$ and $p_{(K+1)m}(x^{(m)}) / p_{im}(x^{(m)})$ are for $1 \leqq i \leqq k$ strictly increasing functions of $p_{(K+1)m}(x^{(m)}) / p_{0m}(x^{(m)})$ on the domain of positivity of p_{0m} (of course, this means that either of the first two ratios can increase as the argument changes from one value to another, only if the last ratio also increases); thus, the results of this section apply to the case $f_\theta(x) = c(\theta)e^{\theta x}h(x)$ with $\theta_0 < \theta_1 < \cdots < \theta_{K+1}$. Our result is

THEOREM 3. *An essentially complete class for the above problem consists of those procedures which at the outset randomize between accepting H_0, accepting H_1, and taking a first observation, and which thereafter are* GSPRT's *for testing H_0 against H_1 (with appropriate randomization rules on the boundaries if μ is not atomless).*

PROOF: A trivial modification of the argument of LeCam [2] (there are two decisions, $K + 2$ states of nature here) shows that an essentially complete class can be obtained by taking the closure in the sense of regular convergence (see [2], [3]) of all Bayes strategies for the problem of minimizing

$$(3.2)\qquad \xi_0 a P_{\theta_0}(\text{accept } H_1) + \xi_{K+1} b P_{\theta_{K+1}}(\text{accept } H_0) + \sum_{i=0}^{K+1} \xi_i E_{\theta_i} C_i(n)$$

for all $a > 0$, $b > 0$, $C_i(m)$ strictly increasing in m and approaching infinity with m if i is such that $a_i > 0$ and $C_i(m) = 0$ if $a_i = 0$ (actually, C need not depend on i here, but must for the considerations of Remark 1 below) and all a priori probability measures $(\xi_0, \xi_1, \cdots, \xi_{K+1})$. Each such Bayes strategy is characterized by an initial randomization of the type described in the statement of the theorem and by a sequence $\{D_{im}\}$ ($i = 0, 1, m = 1, 2, \cdots$) of closed convex subsets of the $(K + 1)$-simplex S (whose elements will be described by $K + 2$ nonnegative barycentric coordinates whose sum is unity) with the property that, Q_i denoting the point of S whose ith barycentric coordinate is unity, $Q_0 \varepsilon \operatorname{int} D_{0m}$ (int denoting interior in the usual topology of S), $Q_{K+1} \varepsilon \operatorname{int} D_{1m}$, these interiors are (for each m) disjoint, and for $1 \leqq j \leqq k$, $Q_j \varepsilon D_{0m} \cap D_{1m}$ and $Q_j \varepsilon \operatorname{int} (D_{0m} \cup D_{1m})$ (see Chapter 4 of [3] or the paragraph following Lemma 4.1 of the present paper for details of arguments yielding these conclusions). The Bayes strategy relative to an a priori probability measure $\xi = (\xi_0, \cdots, \xi_{K+1})$ is then (after some initial randomization as described above) to compute the point $\xi^{(m)} = \xi^{(m)}(x^{(m)})$ of S whose jth component is $\xi_j p_{jm}(x^{(m)}) / \sum_{j=0}^{K+1} \xi_j p_{jm}(x^{(m)})$ ($j = 0, \cdots, K + 1$) after m observations and to accept H_0, accept H_1, or take another observation according to whether $\xi^{(m)} \varepsilon \operatorname{int} D_{0m}$, $\xi^{(m)} \varepsilon \operatorname{int} D_{1m}$, or $\xi^{(m)} \varepsilon S - D_{0m} - D_{1m}$, with some sort of randomization if $\xi^{(m)}$ is in the boundary of one or more D_{im} (under our assumptions, if μ is atomless, randomization is actually unnecessary).

Since the class of procedures described in the theorem is compact and closed in the sense of regular convergence (see [2]), the theorem will be proved if we show each Bayes strategy has the structure enunciated in the theorem. But if this is not true, there are values of ξ, a, b, the functions C_i, and a number $n > 0$ and values $x^{(n)}$, $y^{(n)}$ with $\xi^{(n)}(x^{(n)}) \neq \xi^{(n)}(y^{(n)})$, such that $\xi_0 > 0$ and $\xi_{K+1} > 0$ (otherwise, $P\{n = 0\} = 1$ for any Bayes strategy) and such that

$$(3.3)\qquad \begin{aligned} &\text{(a)} \quad \xi^{(n)}(x^{(n)}) \,\varepsilon\, D_{1n}, \\ &\text{(b)} \quad \xi^{(n)}(y^{(n)}) \not\varepsilon \operatorname{int} D_{1n}, \\ &\text{(c)} \quad p_{(K+1)n}(y^{(n)}) / p_{0n}(y^{(n)}) > p_{(K+1)n}(x^{(n)}) / p_{0n}(x^{(n)}), \end{aligned}$$

or else there is a similar situation for D_{0n} (which is handled similarly). Now, the convex subset of S spanned by $\xi_n(x^{(n)})$, $Q_1, \cdots, Q_{K+1}$ is a subset of D_{1n} which consists of those points $w = (w_0, w_1, \cdots, w_{K+1})$ of S for which

$$(3.4)\qquad w_i t_0 \geqq w_0 t_i \quad \text{for all} \quad i > 0 \quad \text{for which} \quad t_i > 0,$$

where $\xi^{(n)}(x^{(n)}) = (t_0, t_1, \cdots, t_{K+1})$. Hence, writing $\xi^{(n)}(y^{(n)}) = (z_0, z_1, \cdots, z_{K+1})$, (3.3) (b) would imply that $z_i t_0 \leqq z_0 t_i$ for some $i > 0$ for which t_i (hence, ξ_i) > 0.

Thus, for that i (since also $\xi_0 > 0$), we have

$$p_{in}(y^{(n)}) \,/\, p_{0n}(y^{(n)}) \leqq p_{in}(x^{(n)}) \,/\, p_{0n}(x^{(n)}). \tag{3.5}$$

Since (3.3) (c) implies (by the assumption of this section) the negation of (3.5), we obtain a contradiction and the theorem is proved.

REMARKS: 1. Essentially the same proof works to show that the essentially complete class of Theorem 3 is essentially complete for the more general problem where the components of the risk function are

$$P_{\theta_0}\{\text{accept } H_1\}, P_{\theta_{K+1}}\{\text{accept } H_0\}, \sum_{i=0}^{K+1} a_{ir} P_{\theta_i}\{n \geqq j\} (r = 1, \cdots, S; j = 1, 2, \cdots)$$

the a_{ir} being given non-negative constants. One can also treat the case where the finite linear combination in (3.1) is to be replaced by $\int P_\theta\{n \geqq j\}\, dA(\theta)$ where A is a probability measure on a suitable family $\{P_\theta\}$ of probability measures. These considerations have obvious applications to practical problems where the a_i (or A) represent the probability distribution of the process parameter.

2. Theorem 3 can also be proved using the method of [1]; in fact, a proof of Theorem 3 can be obtained essentially by going through the proof in [1] and replacing P_1 and P_2 there by $\sum a_i P_{\theta_i}$ and making other obvious similar alterations.

3. In cases like those of Lemma 4.2 below other than (3), the assumptions of the present section are not satisfied; however, such cases can also be treated here with only minor modifications of the above analysis.

4. Procedures minimizing En at a third point, etc. Let f_{-1}, f_0, f_1 be three densities with respect to a σ-finite measure μ. We assume no two of the f_i are identical almost everywhere (μ). Let $X_1, X_2, \cdots$ be independently and identically distributed random variables with common density f with respect to μ. It is desired to test between the hypotheses $H_{-1}: f = f_{-1}$ and $H_1: f = f_1$. Let $\alpha_i(\delta)$ denote the probability that the procedure δ terminates with an incorrect decision when H_i is true ($i = \pm 1$). Let α_i^* be specified numbers satisfying $0 < \alpha_i^* < 1$ ($i = \pm 1$). Let $A_0(\delta)$ denote the expected value of n (the number of observations which have been taken at termination) when $f = f_0$ and δ is used. Our purpose here is to characterize procedures δ which, among all procedures satisfying

$$\alpha_i(\delta) \leqq \alpha_i^* (i = \pm 1), \tag{4.1}$$

minimize $A_0(\delta)$. Under suitable assumptions (those of Sections 3 and 4 differ), the class of procedures delimited in Theorem 3 will evidently contain the procedures which do this, but we shall obtain here a much finer characterization of them. For the remainder of this section we shall term such procedures "optimum." To avoid trivialities, we hereafter assume $\alpha_1^* + \alpha_2^* < 1$.

We first note a fairly obvious property of optimum procedures. Let Γ be the set of points in three-space of the form $(\alpha_{-1}(\delta), \alpha_1(\delta), A_0(\delta))$ for all possible δ

(not merely those satisfying (4.1)). Since one can randomize between two procedures at the outset, Γ is clearly convex. (The existence of points (a, b, c) with $c < \infty$ for any $a, b > 0$ follows from consideration of fixed-sample-size procedures. A convex combination of points giving positive weight to a point with $c = \infty$ will itself have $c = \infty$.) For any procedure δ satisfying (4.1) and with strict inequality for either $i = 1$ or $i = -1$, we may (by randomizing between δ and a procedure requiring no observations) obtain a δ' satisfying (4.1) and for which $A_0(\delta') < A_0(\delta)$; we may therefore restrict our search for optimum procedures to those δ for which equality holds in (4.1). Among the class of all such procedures there exists one minimizing $A_0(\delta)$, a consequence of Theorem 3.1 of Wald [3]. Let $e(\alpha_{-1}^*, \alpha_1^*) = \min_\delta A_0(\delta)$, the minimum being taken subject to (4.1) (with equality). For all $\epsilon > 0$ with $\epsilon < \min(\alpha_{-1}^*, \alpha_1^*)$ we have (recalling $\alpha_1^* + \alpha_{-1}^* < 1$) that $e(\alpha_{-1}^* - \epsilon, \alpha_1^* - \epsilon) > e(\alpha_{-1}^*, \alpha_1^*)$; for otherwise, if equality held, a randomization of the type noted parenthetically above would produce a δ satisfying (4.1) and for which $A_0(\delta) < e(\alpha_{-1}^*, \alpha_1^*)$, a contradiction. Since $e(\alpha_{-1}^* - \epsilon, \alpha_1^* - \epsilon) > e(\alpha_{-1}^*, \alpha_1^*)$, and since for any value $e > 0$ the points $(0, 1, e)$ and $(1, 0, e)$ are clearly in Γ, it is clear that Γ can not be supported at $(\alpha_{-1}^*, \alpha_1^*, e(\alpha_{-1}^*, \alpha_1^*))$ by a plane any of whose direction cosines is zero. Since Γ obviously can be supported at this point by a plane with non-negative direction cosines, we have

LEMMA 4.1. *Any optimum procedure must, for some positive ξ_1, ξ_{-1}, ξ_0, minimize*

$$\xi_1\alpha_1(\delta) + \xi_{-1}\alpha_{-1}(\delta) + \xi_0 A_0(\delta) \tag{4.2}$$

among all procedures δ. (Conversely, any procedure minimizing (4.2) *for some positive ξ_i's is obviously optimum for some α_i^*'s.)*

Thus, necessary conditions on optimum procedures may be found by characterizing "Bayes solutions" which minimize (for a given ξ_1, ξ_{-1}, ξ_0, all positive, and whose sum we may take to be unity) the "integrated risk" (4.2). Results like Theorems 4.8, 4.9, and 4.10 of Wald [3] (see also [4]) are easily seen to be valid in the present case (with the two values of the loss function and cost of experimentation altered from their unit values in (4.2) if desired). To summarize what we need of these results, *all* procedures minimizing (4.2) for all possible $\xi = (\xi_1, \xi_2, \xi_3)$ with $\xi_i \geqq 0$, $\sum \xi_i = 1$, are characterized in the 2-simplex in barycentric coordinates by two closed convex regions C_{-1} and C_1 as follows: after m observations ($m = 0, 1, 2, \cdots$) compute the a posteriori probability measure $\xi^{(m)}$ for the given a priori measure $\xi = \xi^{(0)}$ and the observed values of $X_1, \cdots, X_m$. Accept H_{-1}, accept H_1, or take another observation according to whether $\xi^{(m)}$ lies in the interior of C_{-1}, C_1, or the complement of $C_{-1} \cup C_1$; on the boundaries between regions, a Bayes solution may randomize in any way (depending on $X_1, \cdots, X_m$ if desired) between (or among) appropriate actions. We now describe the C_i. Let V_i be the point where $\xi_i = 1 (i = 0, \pm 1)$. A point ξ of C_i will be called an *interior* or *boundary* point of C_i according to whether or not *every* Bayes solution with respect to ξ immediately accepts H_i

with probability one. Clearly (see p. 121 of [3]) V_i is in the boundary of C_i for $i = \pm 1$, and a line segment V_0P (of positive length) of the line $\xi_1 = \xi_2$ is the intersection $C_1 \cap C_2$; the curve V_0PV_i is the boundary of C_i; on the segment V_0P, except at the point P where one may randomize among accepting H_1 or H_{-1} or taking another observation, a Bayes solution must stop with probability one and randomize between accepting H_1 or H_{-1} (analogous to Theorem 4.9 of [3]). Of course, a necessary condition for a Bayes solution minimizing (4.2) to be optimum (for some α_i^*) is that it stop with probability one whenever $\xi^{(m)} = V_1$ or V_{-1}; *we hereafter consider only Bayes solutions of this nature.*

In order to obtain a more detailed characterization, we now introduce certain assumptions. Write $x^{(m)} = (x_1, \cdots, x_m)$ and $p_{im}(x^{(m)}) = f_i(x_1)f_i(x_2) \cdots f_i(x_m)$ for $m > 0$ and $=1$ for $m = 0$.

ASSUMPTION A. For each m and $x^{(m)}$, $y^{(m)}$, if

$$p_{1m}(x^{(m)})p_{-1m}(y^{(m)}) \geqq p_{1m}(y^{(m)})p_{-1m}(x^{(m)}), \tag{4.3}$$

then

$$p_{0m}(x^{(m)})p_{-1m}(y^{(m)}) \geqq p_{0m}(y^{(m)})p_{-1m}(x^{(m)}) \tag{4.4}$$

and

$$p_{1m}(x^{(m)})p_{0m}(y^{(m)}) \geqq p_{1m}(y^{(m)})p_{0m}(x^{(m)}); \tag{4.5}$$

and strict inequality in (4.3) with both sides positive implies strict inequality in (4.4) and (4.5).

ASSUMPTION B. For each x_1, if

$$f_1(x_1) \geqq f_{-1}(x_1), \tag{4.6}$$

then

$$f_0(x_1) \geqq f_{-1}(x_1); \tag{4.7}$$

and, if

$$f_{-1}(x_1) \geqq f_1(x_1), \tag{4.8}$$

then

$$f_0(x_1) \geqq f_1(x_1). \tag{4.9}$$

ASSUMPTION C:

$$\limsup_{m\to\infty}\ \sup_{x^{(m)}} \frac{\min[p_{1m}(x^{(m)}), p_{-1m}(x^{(m)})]}{p_{0m}(x^{(m)})} = 0 \tag{4.10}$$

where the supremum is taken over those $x^{(m)}$ for which the denominator is positive.

Of course, if there is a value z_1 for which $f_{-1}(z_1) = f_1(z_1) = f_0(z_1) > 0$ (which will usually not be so), then Assumption B follows from Assumption A. Assumptions A and B are related to the monotone likelihood ratio assumption which occurs elsewhere in certain fixed-sample-size problems in statistics (e.g., [5]), but are not quite in that form, which (for a parametric class: put $\theta_j = 0, \pm 1$ in our case) states (for $n = 1$) $f_{\theta_2}(x)f_{\theta_1}(y) \geqq f_{\theta_1}(x)f_{\theta_2}(y)$ if $\theta_2 \geqq \theta_1$, $x \geqq y$. In Assumption A the $x^{(m)}$ are not necessarily simply-ordered (although a simple ordering of certain equivalence classes for the purpose of the present discussion will often be obvious), unlike the monotone likelihood ratio case; and, at least for fixed n, it is easy to see by examples that neither of Assumption B and the monotone assumption implies the other.

Before proceeding to the consequences of Assumptions A, B, and C, we note that they will be satisfied in many important cases (see also Remark 12 below):

LEMMA 4.2. *If* $f_i(x) = f(x, \theta_i)$ $(i = 0, \pm 1)$ *with* $\theta_{-1} < \theta_0 < \theta_1$ *(the inequalities may be reversed), then Assumptions* A, B, *and* C *are satisfied if* $f(x, \theta)$ *is (for example) of any of the following forms*:

$$(1)\quad f(x, \theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & x \leqq \theta \end{cases}$$

$(-\infty < \theta < \infty, \mu =$ *Lebesgue measure*);

$$(2)\quad f(x, \theta) = \begin{cases} \theta^{-1}, & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

$(0 < \theta < \infty, \mu =$ *Lebesgue measure*);

$$(3)\quad f(x, \theta) = r(\theta)e^{\theta x} \text{ (Koopman-Darmois)}$$

(μ *any* σ-*finite measure not giving all measure to one point*, $r^{-1}(\theta) = \int e^{\theta x}\, d\mu(x)$, θ *any value for which* $r^{-1}(\theta) < \infty$).

We remark that the case $f(x, \theta) = \phi(\theta)e^{g(\theta)t(x)}$ with g strictly monotone can be reduced to case (3).

PROOF: Cases (1) and (2) are easy to verify directly (note that the last part of Assumption A is vacuous here). In case (3), we have $\prod_{i=1}^{n} f(x_i, \theta) / f(y_i, \theta) = e^{\theta z}$ where $z = \sum_1^n (x_i - y_i)$; Assumption A follows at once. Next, we note (differentiating under the integral sign where necessary, which is easily justified for any $\theta \varepsilon L$, where L is the interior of the interval of values θ for which $r^{-1}(\theta) < \infty$) that, for $\theta \varepsilon L$, we have $d^2 \log r(\theta) / d\theta^2 = (E_\theta X)^2 - E_\theta X^2 < 0$, where $E_\theta g(X)$ denotes the expected value of $g(X)$ when X has density $f(x, \theta)$. Hence, $-\log r(\theta)$ is *strictly* convex over the interval of θ for which $r^{-1}(\theta) < \infty$.

Putting $f_i(x) = f(x, \theta_i)$ with $\theta_{-1} < \theta_0 < \theta_1$, equation (4.6) is equivalent to

$$\frac{1}{\theta_1 - \theta_{-1}} \log \frac{r(\theta_1)}{r(\theta_{-1})} \geqq -x_1, \tag{4.11}$$

while (4.7) is equivalent to the expression obtained from (4.11) by substituting θ_0 for θ_1. Hence, we will have shown that (4.6) implies (4.7) if we show that $q(\theta) = (\theta - \theta_{-1})^{-1} \log r(\theta) / r(\theta_{-1})$ is monotonically nonincreasing in θ for $\theta > \theta_{-1}$. Thus, it suffices to show, for $\theta > \theta_{-1}$, that $b(\theta) \leqq 0$, where $b(\theta) = (\theta - \theta_{-1})^2\, dq(\theta) / d\theta = [-\log r(\theta) / r(\theta_{-1})] + (\theta - \theta_{-1})\, d \log r(\theta) / d\theta$. Since

$b(\theta_{-1}) = 0$, it suffices to show, for $\theta > \theta_{-1}$, that $0 \geqq db(\theta) / d\theta$. Since $db(\theta) / d\theta = (\theta - \theta_{-1}) d^2 \log r(\theta) / d\theta^2$, the desired result follows from that of the previous paragraph. The proof that (4.8) implies (4.9) is similar (or may be obtained from the preceding argument by replacing θ and x by $-\theta$ and $-x$).

Finally, let ρ be defined by $\theta_0 = \rho\theta_1 + (1 - \rho)\theta_{-1}$. Clearly, $0 < \rho < 1$. Because of the strict convexity of $-\log r(\theta)$ we have, for some number h with $0 < h < 1$ and for all x,

A few remarks can be made about restricted product problems in general. They are mainly consequences of the fact that the risk function of $\delta = (\delta', \delta'')$ is given by

$$R_\delta(\theta) = R_{\delta'}(\theta) + R_{\delta''}(\theta) \tag{6.2}$$

and are also valid for the slightly more general case

$$R_\delta(\theta) = \rho R_{\delta'}(\theta) + (1 - \rho)R_{\delta''}(\theta).$$

(i) *Bayes solutions.* We mention first the following result, which was previously noted by Duncan [3]. Let δ_0' and δ_0'' be Bayes solutions of two component problems with respect to a common a priori distribution λ, and suppose that (δ_0', δ_0'') is compatible with the given set of restrictions. Then (δ_0', δ_0'') is a Bayes solution with respect to λ for the restricted product problem. For let (δ_1', δ_1'') be any other compatible procedure. Then

$$\int R_{\delta_0'}(\theta)\, d\lambda(\theta) \leqq \int R_{\delta_1'}(\theta)\, d\lambda(\theta); \qquad \int R_{\delta_0''}(\theta)\, d\lambda(\theta) \leqq \int R_{\delta_1''}(\theta)\, d\lambda(\theta),$$

and the result follows from (6.2).

(ii) *Minimax procedures.* Let δ_0' and δ_0'' be minimax solutions of the component problems, and suppose that the same sequence of *a priori* distributions $\{\lambda_n\}$ is least favorable for both so that

$$\sup R_{\delta_0'}(\theta) = \lim \int R_{\delta_{\lambda_n}'}(\theta)\, d\lambda_n(\theta),$$

$$\sup R_{\delta_0''}(\theta) = \lim \int R_{\delta_{\lambda_n}''}(\theta)\, d\lambda_n(\theta),$$

where δ_{λ_n}' and δ_{λ_n}'' denote any Bayes solutions with respect to λ_n. Then if (δ_0', δ_0'') is compatible with the given set of restrictions, the procedure $\delta_0 = (\delta_0', \delta_0'')$ is a minimax solution of the restricted product problem.

To prove this we note as a consequence of the minimax property that

$$\sup [R_{\delta_0'}(\theta) + R_{\delta_0''}(\theta)] \leqq \lim \int [R_{\delta_{\lambda_n}'}(\theta) + R_{\delta_{\lambda_n}''}(\theta)]\, d\lambda_n(\theta),$$

CORRECTION

Page 14, formula (4.12) through page 17, line 23 should be exchanged with page 70, line 8 through page 72, next-to-last line.

while also

$$\sup [R_{\delta_0'}(\theta) + R_{\delta_0''}(\theta)] \geqq \int [R_{\delta_0'}(\theta) + R_{\delta_0''}(\theta)] \, d\lambda_n(\theta)$$

$$\geqq \int [R_{\delta'_{\lambda_n}}(\theta) + R_{\delta''_{\lambda_n}}(\theta)] \, d\lambda_n(\theta).$$

Hence

$$\sup [R_{\delta_0'}(\theta) + R_{\delta_0''}(\theta)] = \lim \int [R_{\delta'_{\lambda_n}}(\theta) + R_{\delta''_{\lambda_n}}(\theta)] \, d\lambda_n(\theta),$$

which is a sufficient condition for δ_0 to be minimax.

(iii) *Procedures with uniformly minimum risk.* Let $\mathcal{C}'$ and $\mathcal{C}''$ be classes of decision procedures for the two component problems, and suppose that within these classes the procedures δ_0' and δ_0'' respectively have uniformly minimum risk. Let $\mathcal{C}$ be the class of compatible pairs (δ', δ'') with $\delta' \varepsilon \mathcal{C}'$ and $\delta'' \varepsilon \mathcal{C}''$. Then if (δ_0', δ_0'') is compatible and hence belongs to $\mathcal{C}$, the procedure $\delta_0 = (\delta_0', \delta_0'')$ uniformly minimizes the risk within the class $\mathcal{C}$. This follows from the fact that if δ_i', δ_i'' $(i = 1, 2)$ are such that $R_{\delta_1'}(\theta) \leqq R_{\delta_2'}(\theta)$ and $R_{\delta_1''}(\theta) \leqq R_{\delta_2''}(\theta)$ for all θ, and if $\delta_i = (\delta_i', \delta_i'')$ are both compatible, then $R_{\delta_1}(\theta) \leqq R_{\delta_2}(\theta)$ for all θ. This result again extends immediately to the case of infinite products.

(iv) *Unbiasedness.* In an earlier paper the author defined a decision procedure δ to be unbiased if it satisfies

$$E_\theta W(\theta', \delta(X)) \geqq E_\theta W(\theta, \delta(X)) \tag{6.3}$$

for all θ, θ'. For the type of problem with which we are concerned this means roughly that on the average the actual decision is closer to the correct decision than to any false one. In this sense the condition is an expression of the requirement that the decision procedure should not favor any one parameter value, or any subset, at the expense of all others, but that it should be impartial towards the various values the parameter can take on. Without some such restriction minimization of the risk will not lead to acceptable results since the procedure that without regard to the data takes the constant decision $d{:}\theta \varepsilon \Omega_i$ clearly minimizes the risk for $\theta \varepsilon \Omega_i$. As was shown in [9], condition (6.3) reduces to the usual condition of unbiasedness in the case of hypothesis testing and point estimation for suitable loss functions.

If δ' and δ'' are unbiased, it follows from addition of the associated inequalities (6.3) that the same is true for the product procedure $\delta = (\delta', \delta'')$. More generally consider products of a family of decision problems with decision spaces $\mathfrak{D}_\gamma$ and loss functions W_γ, $\gamma \varepsilon \Gamma$, where the risk function of a product procedure δ with components δ_γ is given by

$$R_\delta(\theta) = \int R_{\delta_\gamma}(\theta) \, d\mu(\gamma).$$

Then again the unbiasedness of each δ_γ implies that of δ.

The converse, that unbiasedness of a product of two procedures implies that of the components, is not true in general. However, it does hold for example if $\theta = (\xi, \eta)$, $W'(\theta, d') = U(\xi, d')$, $W''(\theta, d'') = V(\eta, d'')$, and if the parameter space Ω is such that given any points ξ and ξ' in the projection of Ω onto the ξ-axis, there exist two points in Ω with abscissae ξ and ξ' respectively and with common ordinate η, and if the corresponding condition holds with ξ and η interchanged. For putting $\theta' = (\xi', \eta)$ an $\theta = (\xi, \eta)$, we have

$$E_{\xi,\eta}U(\xi', \delta'(X)) + E_{\xi,\eta}V(\eta, \delta''(X)) \geqq E_{\xi,\eta}U(\xi, \delta'(X)) + E_{\xi,\eta}V(\eta, \delta''(X)),$$

and hence

$$E_{\xi,\eta}U(\xi', \delta'(X)) \geqq E_{\xi,\eta}U(\xi, \delta'(X)).$$

Therefore δ' is unbiased, and analogously also δ''. The above condition on Ω is satisfied in particular when Ω is a direct product, but also for example in a trinomial (or more generally multinomial) situation with $\xi = p_1$, $\eta = p_2$ and Ω the triangle defined by $0 \leqq p_1, p_2$ and $p_1 + p_2 \leqq 1$.

If this condition holds, and if δ_0', δ_0'' are unbiased procedures with uniformly minimum risk for the component problems and (δ_0', δ_0'') is compatible, it follows from (iii) that (δ_0', δ_0'') is unbiased with uniformly minimum risk for the product problem. In the next section we shall give a much weaker condition on the structure of the decision problem, for which a similar conclusion holds.

7. Unbiasedness. We now return to the multiple decision problems of Section 2, which were obtained as restricted products of the problems of testing $H_\gamma: \theta \varepsilon \omega_\gamma$, $\gamma \varepsilon \Gamma$. The losses resulting from false rejection and acceptance are assumed to be a_γ and b_γ respectively, so that the risk for the testing problem is

$$(7.1) \qquad R_{\varphi_\gamma}(\theta) = \begin{matrix} a_\gamma E_\theta \varphi_\gamma(X) & \text{for} & \theta \varepsilon \omega_\gamma \\ b_\gamma E_\theta[1 - \varphi_\gamma(X)] & \text{for} & \theta \varepsilon \omega_\gamma^{-1}, \end{matrix}$$

which may be written in a single formula as

$$(7.2) \quad R_{\varphi_\gamma}(\theta) = \tfrac{1}{2}(x_\gamma + 1)a_\gamma E_\theta \varphi_\gamma(X) - \tfrac{1}{2}(x_\gamma - 1)b_\gamma E_\theta \varphi_\gamma^{-1}(X) \quad \text{for} \quad \theta \varepsilon \omega_\gamma^{x_\gamma}.$$

The risk of the prodı ɔt procedure is therefore

$$(7.3) \qquad R_\psi(\theta) = E_\theta \textstyle\sum_\gamma [\frac{1}{2}(x_{i\gamma} + 1)a_\gamma\varphi_\gamma(X) - \frac{1}{2}(x_{i\gamma} - 1)b_\gamma\varphi_\gamma^{-1}(X)],$$

when $\theta \varepsilon \Omega_i = \bigcap_\gamma \omega_\gamma^{x_{i\gamma}}$ and the x's are defined as in (2.1).

The purpose of the present and following sections is to prove that all of the procedures described in Sections 3 to 5 are unbiased, and among all unbiased procedures possess uniformly minimum risk, when the loss function is given by (2.4). The result is independent of the weight function μ, provided in the case with infinite Γ, μ is equivalent to Lebesgue measure in the sense of mutual absolute continuity. It is however valid only within the class of procedures the risk of which is finite for the chosen μ.

5. The methods used herein (especially the geometric type argument of

CORRECTION

Page 14, formula (4.12) through page 17, line 23 should be exchanged with page 70, line 8 through page 72, next-to-last line.

Theorem 3 and Lemma 4.3) may also be usefully applied to obtain structure theorems in other sequential decision problems under suitable regularity conditions. For example, the stopping rule for any Bayes solution (risk = expected loss + En) for the k-decision problem of choosing which of $\theta_1 < \theta_2 < \cdots < \theta_k$ is the true parameter value when the X_i have the density $f(x, \theta_j)$ of (3) of Lemma 4.2 for some j is easily seen for large n to approach that which says to stop if and only if each of $k - 1$ certain SPRT's of θ_j against θ_{j+1} $(1 \leqq j < k)$ says to stop.

6. For practical use, our results may be put into more convenient form. For example, if f_j is for some $\Delta > 0$ the normal density with mean Δj and unit variance, our results say that there are constants $A_1 > A_2 > \cdots > A_{N-1} > 1$ such that the (essentially unique) procedure with type I and type II errors α which minimizes $A_0(\delta)$ stops the first time $|\sum_1^n X_i| \geqq A_n$ (making the appropriate decision) and never takes more than N observations (assuming a first observation is taken with probability one). Similar characterizations in the space of the range of the sufficient statistic may be made in other cases of Lemma 4.2.

7. For given α_i^*, it is interesting to consider $A_0(\delta^*)$ where δ^* is the SPRT with $\alpha_i(\delta^*) = \alpha_i^*$ (which minimizes $E_j n$ for $j = \pm 1$). Let $M = M(\alpha_{-1}^*, \alpha_1^*)$ be the smallest integer such that (4.1) may be satisfied by a fixed-sample-size procedure δ requiring M observations. It is easy to give examples where $A_0(\delta^*) < M$ (e.g., let f_0 be close to f_{-1} or f_1) and where $M < A_0(\delta^*)$ (in the example of Remark 6 above, as $\alpha \to 0$, $A_0(\delta^*)$ is of order $(\log \alpha)^2 > M(\alpha, \alpha)$). It would be interesting to obtain useful inequalities and limiting formulas for $e(\alpha_{-1}, \alpha_1)$, as well as $e(\alpha_{-1}, \alpha_1) / A_0(\delta^*)$ and $e(\alpha_{-1}, \alpha_1) / M(\alpha_{-1}, \alpha_1)$, analogous to those which can be obtained in sequential analysis [6]. Of course, if for each $\xi^{(0)}$ one has a knowledge of an upper bound on N, one can compute the procedures of Theorem 4 (for all α_i^*) by "working backwards" as in [3], [4]. Without investigating these topics further, we mention an interesting suggestion of Wolfowitz (who is also to be thanked for suggesting the problem of this section): There is in Case (3) of Lemma 4.2 with μ equivalent to Lebesgue measure a one-parameter family C of tests of the form "stop the first time there is a violation of the inequality $h_1 + Sn < \sum_1^n X_i < h_2 + Sn$ (h_1, h_2, S constants)" and which satisfy $\alpha_i = \alpha_i^*$ $(i = \pm 1)$. One of these other than the unique SPRT δ^* of f_1 against f_{-1} which is a member of C may minimize $A_0(\delta)$ among members of C and may reduce $A_0(\delta)$ considerably from its values for δ^*. Investigation now being undertaken shows that this improvement may be appreciable in practical examples and can often be achieved without modifying $E_i n$ greatly for $i = \pm 1$. We also remark that truncated SPRT's will often be much better than untruncated SPRT'S (e.g., in the example of Remark 6 for α small) in making $A_0(\delta)$ small subject to (4.1); some data on this are available in (e.g.) [7]. These remarks apply also to 9 and 10 below.

8. Our results may be extended in an obvious fashion to consideration of minimizing $A_0(\delta)$ subject to (4.1) for continuous time processes [8].

9. One may obtain a result similar to that of our Theorem 4 for the problem of minimizing subject to (4.1) a (probability) average of $E_\theta n$ over a set of θ between θ_{-1} and θ_1 in the cases of Lemma 4.2 (see also Section 3). This corresponds to the practical situation where θ may be thought of as having a known probability distribution (e.g., certain industrial problems).

10. In any of the cases of Lemma 4.2, one can obtain results on the problem of minimizing $\sup_{\theta_{-1} \leqq \theta \leqq \theta_1} E_\theta n$ subject to (4.1). This can be done by obvious application of the Bayes technique, using Remark 9. In some cases it will be easy to guess at a value θ_0 ($\theta_{-1} \leqq \theta_0 \leqq \theta_1$) such that a procedure minimizing $A_0(\delta)$ subject to (4.1) has its maximum $E_\theta n$ at $\theta = \theta_0$. This procedure will then clearly minimize $\sup_\theta E_\theta n$.

11. Results like those of Theorem 4 and Remarks 9 and 10 can also be obtained if a restriction of the type $E_i n \leqq c_i (i = \pm 1)$ is imposed in addition to (4.1).

12. Lemma 4.2 can easily be extended to include many other cases, e.g., many cases arising in simple fashion from those of Lemma 4.2. For example, Lemma 4.2 also holds for $f(x, \theta) = (t + 1)x^t\theta^{-t-1}$ if $0 < x < \theta$ (and $= 0$ otherwise), where $t > -1$.

REFERENCES

[1] L. Weiss, "Testing one simple hypothesis against another," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 273–281.

[2] L. LeCam, "Note on a theorem of Lionel Weiss," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 791–794.

[3] A. Wald, *Statistical Decision Functions*, John Wiley & Sons, 1950.

[4] A. Wald and J. Wolfowitz, "Bayes solutions of sequential decision problems," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 82–89.

[5] H. Rubin, "A complete class of decision procedures for distributions with monotone likelihood ratio," (abstract) *Ann. Math. Stat.*, Vol. 22 (1951), p. 608.

[6] A. Wald, *Sequential Analysis*, John Wiley & Sons.

[7] R. E. Bechhofer, J. Kiefer, and M. Sobel, "A Sequential Multiple Decision Procedure for Certain Identification and Ranking Problems," to be published.

[8] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Sequential decision problems for processes with continuous time parameter. Testing hypotheses," *Ann. Math. Stat.*, Vol. 24 (1953) pp. 254–264.