A Sequential Procedure for Multihypothesis Testing

Carl W. Baum, *Member, IEEE* and Venugopal V. Veeravalli, *Member, IEEE*

***Abstract*—The sequential testing of more than two hypotheses has important applications in direct-sequence spread spectrum signal acquisition, multiple-resolution-element radar, and other areas. A useful sequential test which we term the MSPRT is studied in this paper. The test is shown to be a generalization of the Sequential Probability Ratio Test. Under Bayesian assumptions, it is argued that the MSPRT approximates the much more complicated optimal test when error probabilities are small and expected stopping times are large. Bounds on error probabilities are derived, and asymptotic expressions for the stopping time and error probabilities are given. A design procedure is presented for determining the parameters of the MSPRT. Two examples involving Gaussian densities are included, and comparisons are made between simulation results and asymptotic expressions. Comparisons with Bayesian fixed sample size tests are also made, and it is found that the MSPRT requires two to three times fewer samples on average.**

***Index Terms*—Sequential analysis, hypothesis testing, informational divergence, nonlinear renewal theory.**

I. Introduction

The use of sequential tests for binary hypothesis testing has been well studied, and the properties of the Sequential Probability Ratio Test (SPRT) have been thoroughly investigated in the literature [1]–[3]. The reason for the interest in the SPRT is due mainly to its optimality property; for specified levels of error probabilities, the SPRT is the test with the minimum expected stopping time [1], [4]. Also, in practice, the SPRT outperforms the best fixed sample size test by a very wide margin; typically one-half to one-third the number of samples on average are required.

Although the majority of research in sequential hypothesis testing has been restricted to two hypotheses, there are several situations, particularly in engineering applications, where it is natural to consider more than two hypotheses. The following are some examples.

1) Consider the serial acquisition of direct-sequence spread spectrum signals [5]. The acquisition problem is to determine the true phase of the incoming code sequence. A serial search scheme achieves this goal by testing the correctness of each possible phase serially. If two or more phases (in particular, $M-1$ phases, $M > 2$) are tested at a time, then we have an M-ary hypothesis testing problem—each of $M-1$ of these hypotheses correspond to deciding that a particular phase is correct, and an additional hypothesis corresponds to deciding that none of these $M-1$ phases are correct. A sequential test is of great interest here since the correct phase should be acquired as soon as possible.
2) In the context of multiple-resolution-element radar, we have sequential decision problems with a single null hypothesis and multiple alternative hypotheses [6]. For example, if there are N resolution elements with each element possibly having a target, the number of alternative hypotheses is 2^N.
3) Consider a fault detection problem in a system where there could be more than one possible kind of fault, and the goal is not only to detect the presence of a fault as quickly as possible but also to determine the type of fault. This useful extension of the standard binary quickest change detection problem [7] provides a setting for an M-ary sequential test.
4) In the context of clinical trials, deciding which of several possible medical treatments is the most effective as quickly as possible is a multihypothesis sequential problem.
5) Statistical pattern recognition is a source of multiple hypothesis problems. Fu [8] discusses potential applications of multihypothesis sequential tests in this area.

The problem of sequential testing of multiple hypotheses is considerably more difficult than that of testing two hypotheses. Published work on this problem has taken two approaches. One approach has aimed at finding an optimal multihypothesis sequential test. A recursive solution to the optimization problem in a Bayesian setting has been obtained [9]–[11]. However, this solution is very complex and impractical except in a few special cases.

A second approach has been to extend and generalize the SPRT to the case of more than two hypotheses without much consideration to optimality. Several *ad hoc* tests based primarily on repeated pairwise applications of the SPRT have been proposed and studied. Some exam-

Manuscript received January 25, 1993; revised September 8, 1994. This work was supported in part by the Joint Services Electronics Program under Grant N00014-90-J-1270, and by Motorola, Inc. This paper was presented in part at the 27th Annual Conference on Information Sciences and Systems, Baltimore, MD, Mar. 1993.
C. W. Baum is with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634.
V. V. Veeravalli is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251.
IEEE Log Number 9406217.

ples of these tests are the Sobel–Wald test [12], the Armitage test [13], the Simons test [14], the Lorden test [15], and the m-SPRT [16] (a recent survey of these tests is found in [17]). It should be noted that many of these tests are restricted to the case of three hypotheses.

Unlike these *ad hoc* tests, the test that we propose in this paper, the MSPRT, is based on the solution to the Bayesian optimization problem. Unlike the optimal solution, the MSPRT has a simple structure that facilitates implementation. The MSPRT is applicable to any number of hypotheses, and it reduces to the SPRT when there are two hypotheses. Furthermore, like the SPRT, the MSPRT is amenable to an asymptotic analysis using renewal theory.

This paper is organized as follows. After introducing the MSPRT in Section II, we explore its optimality under a Bayesian modeling in Section III. A recursive solution to the general Bayesian optimization problem (akin to that obtained in [9]–[11]) is given first. We then note that the MSPRT has a much simpler structure than the optimal solution, and, furthermore, the MSPRT provides a good approximation to the optimal solution, especially when the cost per observation is small compared to the cost of choosing the wrong hypothesis. Because this cost structure generally corresponds to small error probabilities and large stopping times, we see that the approximation is best for the very applications that can benefit the most by a significant reduction in average sample size.

Performance analysis of the MSPRT is discussed in Sections IV–VI. Just as in the case of the SPRT, it is not possible to obtain exact expressions for the error probabilities and expected stopping time except in special cases. In Section IV, upper bounds on Bayesian and frequentist error probabilities are derived using techniques similar to the ones used by Wald for the SPRT [1]. It is suggested that these bounds can be used as approximations, much like Wald's approximations for the SPRT. In Section V, approximations for the expected stopping time are obtained using elementary renewal theory [3]. In Section VI, improved approximations for the error probabilities are obtained using nonlinear renewal theory [3]. All of the renewal theory approximations are shown to be asymptotically exact as the error probabilities go to zero and the expected stopping time becomes infinite. The approximations for the error probabilities and expected stopping time may be used in the design of the MSPRT for specific applications; methods of design are considered in Section VII. In Section VIII, two examples utilizing the MSPRT are discussed, and comparisons are made between the asymptotic results and simulations. Comparisons are also made between the MSPRT and optimally designed fixed sample size tests. Conclusions and directions for further research are given in Section IX.

II. M-ARY SEQUENTIAL PROBABILITY RATIO TEST

Let $X_1, X_2, \cdots$ be an infinite sequence of random variables, independent and identically distributed (i.i.d.) with density f, and let H_j be the hypothesis that $f = f_j$ for $j = 0, 1, \cdots, M-1$. We assume that $f_k \neq f_j$ almost surely for all $j \neq k$. Stated informally, the problem at hand is to determine the true hypothesis with a desired accuracy as quickly as possible.

Assume that the prior probabilities of the hypotheses are known, and let π_j denote the prior probability of hypothesis H_j for each j. If only n observations are available, the minimum-probability-of-error test uses as the test statistic the following vector of posterior probabilities:

$$\boldsymbol{p}_n = (p_n^0, \cdots, p_n^{M-1})$$

where

$$p_n^j = \mathrm{P}\{H = H_j \mid X_1, \cdots, X_n\}.$$

The minimum-probability-of-error test picks the hypothesis with the largest posterior probability, given the observations [18].

For our problem, the available number of observations is, in theory, infinite, and a compromise must be struck between the number of observations used and the error probability. When the observations arrive periodically over time, this compromise is between the delay in making a decision and the accuracy of that decision. A test that strikes such a compromise is a sequential test, and it consists of a stopping rule and a final decision rule. The stopping rule determines the number of observations that are taken until a decision is made, and the final decision rule chooses one of the M hypotheses as its best estimate of the true hypothesis.

In this paper, we consider one such sequential test, which we call the M-ary sequential probability ratio test (MSPRT). The stopping time N_A and final decision δ for the MSPRT can be described as follows:

$$N_A = \text{first } n \geq 1 \text{ such that } p_n^k > \frac{1}{1 + A_k} \text{ for at least one } k$$

$$\delta = H_m, \text{ where } m = \arg\max_j p_{N_A}^j.$$

Using Bayes' rule, the posterior probabilities can be written as

$$p_n^j = \frac{\pi_j \prod_{i=1}^{n} f_j(X_i)}{\sum_{k=0}^{M-1} \pi_k \left(\prod_{i=1}^{n} f_k(X_i) \right)}$$

and thus, an equivalent description of the MSPRT is

$$N_A = \text{first } n \geq 1 \text{ such that } \frac{\pi_k \prod_{i=1}^{n} f_k(X_i)}{\sum_{j=0}^{M-1} \pi_j \left(\prod_{i=1}^{n} f_j(X_i) \right)} > \frac{1}{1 + A_k} \text{ for at least one } k$$

$$\delta = H_m, \quad \text{where } m = \arg\max_j \left(\pi_j \prod_{i=1}^{n} f_j(X_i) \right).$$

The parameters A_j are assumed to be positive, and as we shall see in the following sections, the typical design values are all less than one. If each parameter A_j is indeed less than one, then the condition $p_j^n > 1/(1+A_j)$ can be satisfied by at *most* one value of j (since the posterior probabilities must sum to one). In this case, the MSPRT takes the form

$$N_A = \text{first } n \geq 1 \text{ such that } \frac{\pi_k \prod_{i=1}^{n} f_k(X_i)}{\sum_{j=0}^{M-1} \pi_j \left(\prod_{i=1}^{n} f_j(X_i) \right)} > \frac{1}{1+A_k} \text{ for some } k$$

$$\delta = H_k.$$

For comparison, the SPRT (for sequential testing of two hypotheses H_0 and H_1) with parameters A' and B' is defined [1] as

$$N_{(A',B')} = \text{first } n \geq 1 \text{ such that } L_n = \frac{\prod_{i=1}^{n} f_1(X_i)}{\prod_{i=1}^{n} f_0(X_i)} \notin [A', B']$$

$$\delta = H_1 \text{ if } L_n > B' \text{ and } \delta = H_0 \text{ if } L_n < A'.$$

It is straightforward to show that the MSPRT with $M = 2$ is identical to the SPRT with parameters $\pi_0 A_0/\pi_1$ and $\pi_0/(\pi_1 A_1)$. Note that for the SPRT, the prior probabilities can be incorporated into the parameters A' and B'. For the MSPRT, however, the priors cannot be absorbed into the thresholds $A_0, A_1, \cdots, A_{M-1}$ unless $M = 2$.

III. Bayesian Optimality

In this section we investigate the MSPRT in the context of designing optimal sequential hypothesis tests under a Bayesian framework. Towards this end, we first provide a formulation for determining the best (in the Bayesian sense) sequential test for determining the true hypothesis. A set of admissible sequential tests is defined, and the performance of each admissible sequential test is quantified through the use of a cost assignment.

An admissible sequential test γ is defined as follows: Let $\mathscr{X}_n$ be the σ-field generated by $X_1, X_2, \cdots, X_n$. The stopping time N of γ is a $\{\mathscr{X}_n, n = 1, 2, \cdots\}$-stopping time, and the final decision δ of γ is measurable with respect to $\mathscr{X}_N$. The set of all admissible sequential tests is denoted by Γ.

For notational convenience, we let H denote the random variable that represents the true hypothesis. Then H takes the value H_j with probability π_j. To avoid any trivialities, we assume throughout that the priors π_j are all nonzero.

For the cost assignment, we assume that each time step taken by γ costs a positive amount c, and that decision errors in the final decision δ are penalized through a *uniform* cost function $W(\delta; H)$, that is,

$$W(H_j, H_j) = 0$$

$$W(H_j, H_k) = 1 \text{ for } j \neq k.$$

The Bayesian optimization problem is then formulated as

$$\min_{\gamma \in \Gamma} \mathrm{E}\{cN + W(\delta; H)\}. \tag{1}$$

A recursive solution to this problem has been obtained by researchers previously (see, for example, [9]–[11]). In the following, we give a sketch of this recursive procedure using dynamic programming (DP) arguments.

A. Dynamic Programming Solution

It can easily be shown that the vector of posterior probabilities introduced in Section II is a sufficient statistic for a DP argument for the problem in (1).

Using Bayes' rule we obtain the following recursion for $\boldsymbol{p}_n$:

$$p_{n+1}^k = \mathrm{P}\{H = H_k \mid X_1, \cdots, X_{n+1}\} = \frac{p_n^k f_k(X_{n+1})}{\sum_{j=0}^{M-1} p_n^j f_j(X_{n+1})}$$

with initial condition $\boldsymbol{p}_0 = \boldsymbol{\pi} := (\pi_0, \cdots, \pi_{M-1})$. The conditional density of X_{n+1} given $\mathscr{X}_n$, which we denote by $f(\cdot; \boldsymbol{p}_n)$, is given by

$$f(x; \boldsymbol{p}_n) = \sum_{j=0}^{M-1} p_n^j f_j(x). \tag{2}$$

Furthermore, the vector of posterior probabilities is constrained to lie in the convex set $\mathscr{C}$ given by

$$\mathscr{C} = \{\boldsymbol{p} : 0 \leq p^j \leq 1,\ j = 0, \cdots, M-1 \text{ and } p^0 + \cdots + p^{M-1} = 1\}.$$

Finite-Horizon Optimization: First restrict the stopping time N to a finite interval, say $[0, T]$. The finite-horizon DP equations can then be derived. Toward this end, we note that the minimum expected cost-to-go at time n is a function of the sufficient statistic $\boldsymbol{p}_n$, and we denote it by $J_n^T(\boldsymbol{p}_n)$, defined for $0 \leq n \leq T$. It is easily seen that

$$J_T^T(\boldsymbol{p}_T) = g(\boldsymbol{p}_T)$$

where $g(\boldsymbol{p}) = \min\{(1 - p^0), \cdots, (1 - p^{M-1})\}$.

For $0 \leq n \leq T-1$, a standard DP argument gives the following backward recursion:

$$J_n^T(\boldsymbol{p}_n) = \min\{g(\boldsymbol{p}_n), c + A_n^T(\boldsymbol{p}_n)\} \tag{3}$$

where

$$A_n^T(\boldsymbol{p}_n) = \mathrm{E}\{J_{n+1}^T(\boldsymbol{p}_{n+1}) \mid \mathscr{X}_n\} = \int J_{n+1}^T\left(\frac{p_n^0 f_0(x)}{f(x; \boldsymbol{p}_n)}, \cdots, \frac{p_n^{M-1} f_{M-1}(x)}{f(x; \boldsymbol{p}_n)} \right) f(x; \boldsymbol{p}_n)\, dx.$$

Note that the functions J_n^T and A_n^T are defined on the set $\mathscr{C}$. Also note that the optimum cost for the finite-horizon sequential test is $J_0^T(\boldsymbol{\pi})$.

Infinite-Horizon Optimization: In order to solve the optimization problem in (1), the restriction that N belongs to a finite interval is removed by letting $T \to \infty$. As a first step, note that for all T, for all $n \le T$, and for all $\boldsymbol{p} \in \mathscr{C}, 0 \le J_n^T(\boldsymbol{p}) \le g(\boldsymbol{p})$. Furthermore, $J_n^{T+1}(\boldsymbol{p}) \le J_n^T(\boldsymbol{p})$, because the set of stopping times increases with T. This implies that, for each finite n, the following limit is well defined:

$$\lim_{T\to\infty} J_n^T(\boldsymbol{p}) = \inf_{T:T>n} J_n^T(\boldsymbol{p}).$$

Denote this limit by $J_n^\infty(\boldsymbol{p})$. Now, because the observations are i.i.d., a time-shift argument shows that $J_n^\infty(\boldsymbol{p})$ is independent of n. We hence denote the limit by $J(\boldsymbol{p})$, which we will refer to as the infinite-horizon cost-to-go function.

Now, by the dominated convergence theorem [19], the following limit is well defined for all n:

$$\begin{aligned}
&\lim_{T\to\infty} A_n^T(\boldsymbol{p}) \\
&= \lim_{T\to\infty} \int J_{n+1}^T\left(\frac{p_n^0 f_0(x)}{f(x;\boldsymbol{p}_n)},\cdots,\frac{p_n^{M-1} f_{M-1}(x)}{f(x;\boldsymbol{p}_n)}\right) f(x;\boldsymbol{p}_n)dx \\
&= \int J\left(\frac{p_n^0 f_0(x)}{f(x;\boldsymbol{p}_n)},\cdots,\frac{p_n^{M-1} f_{M-1}(x)}{f(x;\boldsymbol{p}_n)}\right) f(x;\boldsymbol{p}_n)\,dx.
\end{aligned}$$

The limit, which is independent of n, is denoted by $A_J(\boldsymbol{p})$.

It follows that the infinite-horizon cost-to-go function satisfies the following Bellman equation [20] whose solution can be shown to be *unique* (the proof is nearly identical to a parallel argument in [21]):

$$J(\boldsymbol{p}) = \min\{g(\boldsymbol{p}), c + A_J(\boldsymbol{p})\}. \tag{4}$$

Note that the optimum cost for the problem in (1) is $J(\boldsymbol{\pi})$. In addition, if $J(\boldsymbol{p})$ is computed for all $\boldsymbol{p} \in \mathscr{C}$, then the optimal sequential test can be obtained from (4). The uniqueness of the solution implies that a successive approximation technique can be used to compute $J(\boldsymbol{p})$ for all $\boldsymbol{p} \in \mathscr{C}$ (see [21] for a discussion of one such successive approximation procedure). However, the optimal test has a very complex structure [9] that makes implementation very impractical. Thus it is of interest to explore suboptimal or approximately optimal solutions which have a simpler structure.

B. Approximation to the Optimal Solution

We begin by investigating some properties of the functions $J(\boldsymbol{p})$ and $A_J(\boldsymbol{p})$ in the following lemma whose proof is quite straightforward (see [21] for the proof of a similar result).

Lemma 3.1: The functions $J(\boldsymbol{p})$ and $A_J(\boldsymbol{p})$ are nonnegative concave functions of $\boldsymbol{p}$ on the set $\mathscr{C}$. Furthermore,

$$\begin{aligned}
A_J(1,0,\cdots,0) &= A_J(0,1,0,\cdots,0) \\
&= \cdots = A_J(0,\cdots,0,1) = 0.
\end{aligned}$$

Now, on the convex set $\mathscr{C}$, $g(\boldsymbol{p})$ obtains its maximum value of $1 - 1/M$ at $\boldsymbol{p} = (1/M,\cdots,1/M)$. If $c + A_J(1/M,\cdots,1/M) > 1 - 1/M$, then the optimal sequential test ignores all the observations and bases its decision solely on the prior probability vector $\boldsymbol{\pi}$. If $c + A_J(1/M,\cdots,1/M) \le 1 - 1/M$, then the concave functions $g(\boldsymbol{p})$ and $c + A_J(\boldsymbol{p})$ intersect on some curve; the projection of this curve on $\mathscr{C}$ is a closed connected set, which we denote by $\mathscr{C}_{\text{in}}$. The optimal test γ_{opt} is then described by

$$\begin{aligned}
N_{\text{opt}} &= \text{first } n \ge 1 \text{ such that } \boldsymbol{p}_n \notin \mathscr{C}_{\text{in}} \\
\delta_{\text{opt}} &= H_m, \quad \text{where } m = \arg\min_j p_n^j.
\end{aligned}$$

If the set $\mathscr{C}_{\text{in}}$ has a simple characterization, then the structure of the optimal test can be much simplified. For $M = 2$, $\mathscr{C}$ is the line segment joining the points $(0,1)$ and $(1,0)$ in two-dimensional space, and the set $\mathscr{C}_{\text{in}}$ is a piece of this line segment. Hence for $M = 2$, $\mathscr{C}_{\text{in}} = \mathscr{C} \cap \{p^0 \le a_0\} \cap \{p^1 \le a_1\}$ for some thresholds a_0 and a_1, which leads to the well-known SPRT structure for the optimal test. Unfortunately, such a threshold characterization of the set $\mathscr{C}_{\text{in}}$ is not possible, in general, for $M > 2$. For $M > 2$, the set $\mathscr{C}$ is the $(M-1)$-dimensional set joining the points $(1,0,\cdots,0), (0,1,0,\cdots,0),\cdots,(0,\cdots,0,1)$ in M-dimensional space, and the set $\mathscr{C}_{\text{in}}$ is obtained from $\mathscr{C}$ by "pinching off" pieces at the corners. The piece that is pinched off from the jth corner, denoted by Δ_j, is the region of $\mathscr{C}$ in which H_j is chosen by the optimal test. The following property of the decision regions Δ_j is easily established.

Lemma 3.2: The decision regions Δ_j are convex.

Fig. 1 illustrates the sets $\mathscr{C}$ and $\mathscr{C}_{\text{in}}$ for the case $M = 3$. In this figure, $P_0 = (1,0,0)$, $P_1 = (0,1,0)$, $P_2 = (0,0,1)$, and $\mathscr{C}$ is the set of all points in the triangle. If $\boldsymbol{p}_n$ lies within $\mathscr{C}_{\text{in}}$, another sample is taken, and if $\boldsymbol{p}_n$ lies within the region denoted by Δ_j, no further samples are taken and hypothesis H_j is chosen.

As mentioned in Section I, we are mainly interested in the asymptotic performance of the M-ary sequential test when the error probabilities are small and the expected stopping time is large. In the Bayesian formulation, this corresponds to asymptotics as $c \to 0$. We first investigate the behavior of $A_J(\boldsymbol{p})$ as $c \to 0$. The following lemma is proved in the Appendix.

Lemma 3.3: $A_J(\boldsymbol{p})$ is continuous in c and

$$A_J(\boldsymbol{p}) \downarrow 0 \quad \text{as } c \downarrow 0.$$

From Lemma 3.3 it follows that for $c \ll 1$, the decision regions Δ_j are also small. This fact along with the result of Lemma 3.2 leads us to propose to approximate the set $\mathscr{C}_{\text{in}}$ by a set of the form $\mathscr{C} \cap \bigcap_{i=0}^{M}\{p^i \le a_i\}$, for $c \ll 1$. That is, the approximation we propose to the optimal test, denoted by γ_{approx}, has the following form:

$$\begin{aligned}
N_{\text{approx}} &= \text{ first } n \ge 1 \ni p_n^k > a_k \text{ for at least one } k \\
\delta_{\text{approx}} &= H_m, \quad \text{where } m = \arg\min_j (1 - p_n^j).
\end{aligned}$$

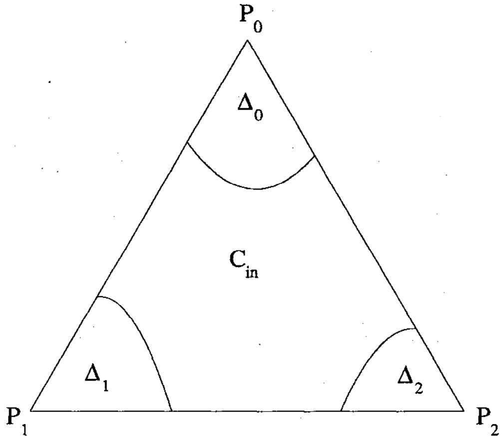

Fig. 1. Typical decision regions for an optimal test with $M = 3$.

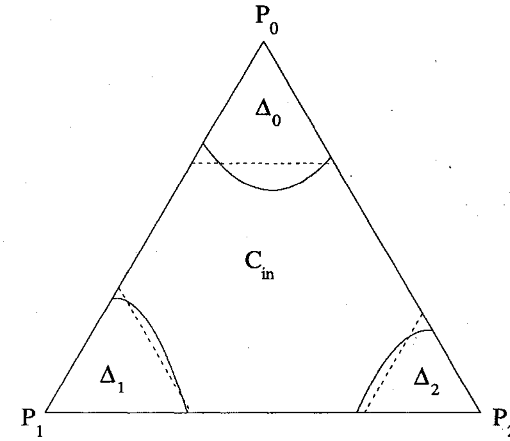

Fig. 2. Typical decision regions for an MSPRT with $M = 3$.

Setting $a_j = 1/1 + A_j$ in γ_{approx} gives us the MSPRT structure[1] of Section II.

An example of such an approximation for $M = 3$ is illustrated in Fig. 2. In this figure, the solid lines denote the boundary regions for an optimal test, and the dashed lines denote the boundary curves for an MSPRT. Note that if the optimal test were known, the parameters A_j could be chosen to provide a "best" approximation in some sense.

For $c \ll 1$, this test may perform quite well, as it captures the essential features of the optimal test. By way of example, consider a hypothesis testing problem for which $X_1, X_2, \cdots$, is a sequence of independent and identically distributed random variables such that X_i can take on the values 0, 1, and 2. Let H_0 be the hypothesis that $P(X_i = 1) = P(X_i = 2) = 1/2$, let H_1 be the hypothesis that $P(X_i = 0) = P(X_i = 2) = 1/2$, and let H_2 be the hypothesis that $P(X_i = 0) = P(X_i = 1) = 1/2$. In [11] it is shown that the optimal sequential test for this problem with a nonuniform cost function can be solved without the need of iterative techniques. (Note that only a handful of very specialized problems can be solved in this manner.) Using the approach of [11], we can solve this problem assuming a uniform cost function. As is done in [11], it is convenient to redefine the probability coordinates before proceeding with the solution. In the new coordinate system, a particular probability vector (p^0, p^1, p^2) is represented as a point in $\mathscr{C}$ at a distance p^0 from the line opposite P_0, a distance p^1 from the line opposite P_1, and a distance p^2 from the line opposite P_2. (Note that the third specification is redundant.)

The result is that the boundary between $\mathscr{C}_{\text{in}}$ and the decision region Δ_0 is given by a connected set of line segments successively joining $(1 - 2c, 2c, 0)$, $(1 - 2ac, 2ac - 4ac^2, 4ac^2)$, $(1 - 2bc + 4bc^2, 2bc - 8bc^2, 4bc^2)$, $(1 - 2bc - 4bc^2, 4bc^2, 2bc - 8bc^2)$, $(1 - 2ac, 4ac^2, 2ac - 4ac^2)$, and $(1 - 2c, 0, 2c)$, where

$$a = \frac{1}{1 - 2c + 2c^2}$$

and

$$b = \frac{1}{1 - 4c + 4c^2}.$$

The boundaries between $\mathscr{C}_{\text{in}}$ and the other decision regions can be obtained by successive cyclic permutation of the coordinates.

[1]Tartakovskii shows in [9] that the MSPRT structure is also obtained in the limiting case where the Kullback–Leibler distance between the conditional densities under each pair of hypotheses goes to infinity.

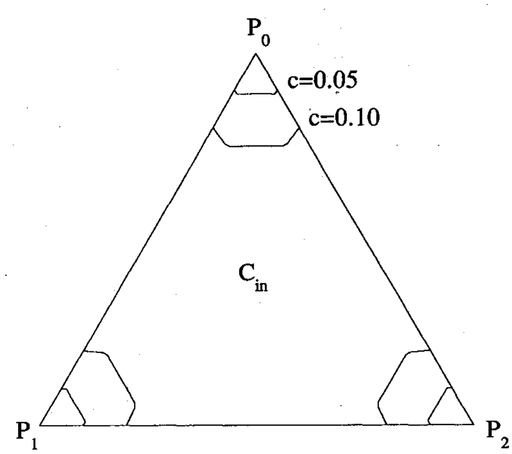

Fig. 3. Decision regions for an actual optimal test drawn in the redefined coordinated system.

These decision regions are shown in Fig. 3. From this figure, it would appear that the decision boundaries approach straight lines (parallel to the opposite side of the triangle) as $c \to 0$. In fact, it is straightforward to show that the boundary between $\mathscr{C}_{\text{in}}$ and Δ_0 is of the form

$$p^0 - (1 - 2c) = h(c, \boldsymbol{p})$$

where h is a function that is $O(c^2)$ for any fixed $\boldsymbol{p}$; that is, the boundary approaches the straight line $p^0 = 1 - 2c$ on $\mathscr{C}$, as $c \to 0$. This means that the optimal test for this example is asymptotically equivalent to the following symmetric MSPRT (as $c \to 0$):

$$N_c = \text{first } n \geq 1 \ni p_n^k > 1 - 2c \text{ for at least one } k$$
$$\delta_c = H_m, \quad \text{where } m = \arg\min_j (1 - p_n^j).$$

While we have not proved the asymptotic optimality of the MSPRT in general, we believe that the simplicity of the test makes it a good choice for implementation. Furthermore, the numerical results of Section VIII indicate that for specified levels of error probabilities, the MSPRT takes two to three times fewer samples on average than the corresponding fixed sample size test, a gain similar to that obtained by using the SPRT for binary hypothesis testing.

IV. Performance Bounds

We begin our analysis of the MSPRT by first deriving useful bounds on its stopping time and error probabilities. In the proofs of the results in this and subsequent sections, it is sometimes convenient to use the alternative description of N_A given below.

$$N_A = \text{ first } n \geq 1 \text{ such that } \sum_{\substack{j=0 \\ j\neq k}}^{M-1} \frac{\pi_j}{\pi_k} \prod_{i=1}^{n} \frac{f_j(X_i)}{f_k(X_i)} < A_k \text{ for at least one } k.$$

As with any other stopping time, it is important to determine the conditions or restrictions, if any, for which N_A is finite. The following theorem shows that the probability that N_A exceeds n decreases exponentially with n.

Theorem 4.1: The stopping time N_A is exponentially bounded, conditioned on each of the hypotheses H_k, $k = 0,\cdots, M-1$.

Proof: For any fixed k, it is easily verified that the following inequalities are satisfied:

$$\begin{aligned} N_A &\leq \text{ first } n \geq 1 \text{ such that } \sum_{\substack{j=0 \\ j\neq k}}^{M-1} \frac{\pi_j}{\pi_k} \prod_{i=1}^{n} \frac{f_j(X_i)}{f_k(X_i)} < \min_l A_l \\ &\leq \text{ first } n \geq 1 \text{ such that } \frac{\pi_j}{\pi_k} \prod_{i=1}^{n} \frac{f_j(X_i)}{f_k(X_i)} < \frac{\min_l A_l}{M-1} \quad \forall j \ni j \neq k. \end{aligned}$$

Thus, it is true that

$$\begin{aligned} P_{f_k}(N_A > n) &\leq P_{f_k}\left(\bigcup_{j:j\neq k} \frac{\pi_j}{\pi_k} \prod_{i=1}^{n} \frac{f_j(X_i)}{f_k(X_i)} \geq \frac{\min_l A_l}{M-1} \right) \\ &\leq \sum_{j:j\neq k} P_{f_k}\left(\frac{\pi_j}{\pi_k} \prod_{i=1}^{n} \frac{f_j(X_i)}{f_k(X_i)} \geq \frac{\min_l A_l}{M-1} \right) \\ &= \sum_{j:j\neq k} P_{f_k}\left(\prod_{i=1}^{n} \sqrt{\frac{f_k(X_i)}{f_j(X_i)}} \geq \sqrt{\frac{\pi_k}{\pi_j} \frac{\min_l A_l}{M-1}} \right) \\ &\leq \sum_{j:j\neq k} \sqrt{\frac{\pi_j}{\pi_k} \frac{M-1}{\min_l A_l}}\, E_{f_k}\left[\sqrt{\frac{f_j(X_1)}{f_k(X_1)}} \right]^n \\ &\leq \frac{(M-1)^{3/2}}{\sqrt{\pi_k \min_l A_l}} \max_{j:j\neq k} \sqrt{\pi_j}\,(\rho_j)^n \end{aligned}$$

where

$$\rho_j = \mathrm{E}_{f_k}\left[\sqrt{\frac{f_j(X_1)}{f_k(X_1)}} \right].$$

Note that the second-to last inequality is simply Markov's inequality applied to each term in the summation. All that remains is to show that, for $j \neq k$, $\rho_j < 1$. But this follows immediately from the Cauchy–Schwartz inequality since

$$\begin{aligned} \mathrm{E}_{f_k}\left[\sqrt{\frac{f_j(X_1)}{f_k(X_1)}} \right] &= \int \sqrt{f_k(x)}\sqrt{f_j(x)}\, dx \\ &< \sqrt{\left(\int f_k(x)\, dx\right)\left(\int f_j(x)\, dx\right)} = 1. \end{aligned}$$

(Note that the inequality is strict because, as we assumed in Section II, $f_k \neq f_j$ a.s.) □

A consequence of this theorem is that, conditioned on each hypothesis, all the moments of N_A are finite. As an immediate corollary, we have the following.

Corollary 4.1: Conditioned on each hypothesis, N_A is finite with probability one.

We now state and prove a result that gives upper bounds on all relevant error probabilities.

Theorem 4.2: Let $\alpha_{j,k} = P_{f_j}$ (accept H_k), let $\alpha_k =$ P(accept H_k incorrectly), and let α be the total probability of an incorrect decision. Then

a) $\alpha_k = \sum_{j:j\neq k} \pi_j \alpha_{j,k} \leq \pi_k A_k$ for all k
b) $\alpha = \sum_k \alpha_k \leq \sum_k \pi_k A_k$.

If, in addition, $A_0 = A_1 = \cdots = A_{M-1}$,

c) $\alpha \leq \dfrac{A}{1+A}$.

Proof: Because N_A is finite with probability one, we can write

$$\begin{aligned} \alpha_{k,k} &= \sum_{n=1}^{\infty} P_{f_k}(\text{accept } H_k,\, N_A = n) \\ &= \sum_{n=1}^{\infty} \int_{\{\text{accept } H_k,\, N_A = n\}} f_k(x_1) \\ &\quad \cdot f_k(x_2) \cdots f_k(x_n)\, dx_1\, dx_2 \cdots dx_n \\ &\geq \frac{1}{\pi_k(1+A_k)} \sum_{n=1}^{\infty} \sum_{j=0}^{M-1} \int_{\{\text{accept } H_k,\, N_A = n\}} \pi_j f_j(x_1) \\ &\quad \cdot f_j(x_2) \cdots f_j(x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= \frac{1}{\pi_k(1+A_k)} \sum_{j=0}^{M-1} \pi_j \alpha_{j,k} \\ &= \frac{1}{\pi_k(1+A_k)} (\pi_k \alpha_{k,k} + \alpha_k) \end{aligned}$$

where the inequality follows from the definition of the MSPRT. Subtracting $1/(1+A_k)\ \alpha_{k,k}$ from each side gives

$$\alpha_{k,k} \geq \frac{\alpha_k}{\pi_k A_k}$$

The fact that $\alpha_{k,k} \leq 1$ yields part a). Part b) follows trivially. Part c) is proven by starting with the fact [implied from the proof of part a)] that

$$\pi_k \alpha_{k,k} \geq \frac{1}{1+A} \sum_{j=0}^{M-1} \pi_j \alpha_{j,k}.$$

Summation over k and the fact that $\alpha = 1 - \sum_k \pi_k \alpha_{k,k}$ proves the desired result. □

The bounds in Theorem 4.2 are derived using techniques similar to the ones used by Wald for the SPRT [1]. It should be noted that the α_k's given above are *frequentist* error probabilities; they are not to be confused with the probabilities of error conditioned on particular hypotheses. The individual conditional error probabilities cannot be expressed in terms of frequentist error probabilities except for $M = 2$.

In addition to obtaining bounds for the error probabilities, Wald [1] obtained approximations for the expected stopping time of the SPRT conditioned on each hypothesis. These approximations were obtained by an application of Wald's lemma [1]. For the MPSRT, however, Wald's lemma is *not* applicable for $M > 2$ because the test statistic cannot be expressed as a sum of independent and identically distributed random variables. However, as we show in the following section, approximations for the expected stopping time of the MSPRT can be obtained using elementary renewal theory [3]. Furthermore, these approximations are shown to be asymptotically exact.

V. Asymptotics of Stopping Time

In order to proceed further with analysis of the MSPRT, it is helpful to denote the Kullback–Leibler distance between densities f and g by $D(f, g)$, so that

$$D(f,g) = \mathrm{E}_f\left[\log \frac{f(X)}{g(X)}\right] = \int f(x) \log \frac{f(x)}{g(x)}\, dx.$$

It is a well-known fact that $D(f, g) \geq 0$ with equality if and only if $f = g$ a.s.-f.

We are interested in the studying the behavior of the MSPRT when the parameters A_k, $k = 0, \cdots, M-1$, simultaneously approach zero, that is, when $\max_k A_k$ approaches zero. This corresponds to asymptotics where the error probabilities are small and the stopping time is large.

We consider the asymptotic behavior of N_A first. Towards this end, we state several relevant lemmas.

Lemma 5.1: Assume that, for fixed k, $\min_{j:j\neq k} D(f_k, f_j)$ is positive and finite. Then $N_A \to \infty$ a.s.-f_k as $\max_l A_l \to 0$.

Proof: Let $D(f_k, f_{j^*}) = \min_{j:j\neq k} D(f_k, f_j)$. We can write

$$\begin{aligned}
&\mathrm{P}_{f_k}(N_A \leq n)\\
&= \mathrm{P}_{f_k}\left(\text{for some } l,\ \min_{1\leq m\leq n} \sum_{j:j\neq l} \frac{\pi_j}{\pi_l} \prod_{i=1}^{m} \frac{f_j(X_i)}{f_l(X_i)} < A_l\right)\\
&\leq \mathrm{P}_{f_k}\left(\text{for some } l \text{ and any } j \neq l,\ \min_{1\leq m\leq n} \prod_{i=1}^{m} \frac{f_j(X_i)}{f_l(X_i)} < \frac{A_l \pi_l}{\pi_j}\right)\\
&= \mathrm{P}_{f_k}\left(\text{for some } l \text{ and any } j \neq l,\ \max_{1\leq m\leq n} \sum_{i=1}^{m} \log \frac{f_l(X_i)}{f_j(X_i)} > -\log \frac{\pi_l A_l}{\pi_j}\right)\\
&\leq \sum_{l\neq k} \mathrm{P}_{f_k}\left(\max_{1\leq m\leq n} \sum_{i=1}^{m} \log \frac{f_l(X_i)}{f_k(X_i)} > -\log \frac{\pi_l A_l}{\pi_k}\right)\\
&\quad + \mathrm{P}_{f_k}\left(\max_{1\leq m\leq n} \sum_{i=1}^{m} \log \frac{f_k(X_i)}{f_{j^*}(X_i)} > -\log \frac{\pi_k A_k}{\pi_{j^*}}\right).
\end{aligned}$$

The first term on the right-hand side converges to zero as $\max_l A_l \to 0$, since Kullback–Leibler distances are nonnegative. The second term converges to zero as $\max_l A_l \to 0$, since $D(f_k, f_{j^*})$ is finite. Thus, $N_A \to \infty$ in probability as $\max_l A_l \to 0$. This implies that a subsequence of the N_A's goes to infinity a.s.-f_k. However, it is easily seen that N_A is nondecreasing a.s.-f_k as each of the A_l's decreases to 0. This implies that N_A converges a.s.-f_k as $\max_l A_l \to 0$, proving the lemma. □

The next lemma is a technical result that can be proven from elementary real analysis on sequences.

Lemma 5.2: Suppose that, for $1 \leq j \leq m$, $Y_n^{(j)} \to \mu_j$ a.s., and let $\delta = \min_j \mu_j$. If $0 < \delta < \infty$, then

$$-\frac{1}{n} \log\left(\sum_{j=1}^{m} e^{-nY_n^{(j)}}\right) \to \delta \quad \text{a.s.}$$

as $n \to \infty$. □

Note that, in particular, there is no requirement of independence between the $Y_n^{(j)}$'s.

The following lemma establishes almost sure convergence of a stopping time closely related to N_A that is defined solely for use in the proof of Theorem 5.1. The proof is given in the Appendix.

Lemma 5.3: For any fixed k, let

$$\tilde{N}_A = \text{first } n \geq 1 \text{ such that } \frac{\pi_k \prod_{i=1}^{n} f_k(X_i)}{\sum_{j=0}^{M-1} \pi_j \left(\prod_{i=1}^{n} f_j(X_i)\right)} > \frac{1}{1+A_k}.$$

If $\min_{j:j\neq k} D(f_k, f_j)$ is positive and finite, then

$$\frac{\tilde{N}_A}{-\log A_k} \to \frac{1}{\min_{j:j\neq k} D(f_k, f_j)} \quad \text{a.s.-}f_k$$

as $\max_l A_l \to 0$.

We now state the main theorem on the asymptotics of the stopping time N_A. The proof is given in the Appendix.

Theorem 5.1: Assume that $\min_{j:j\neq k} D(f_k, f_j)$ is positive and finite. Then

$$\frac{N_A}{-\log A_k} \to \frac{1}{\min\limits_{j:j\neq k} D(f_k, f_j)} \quad \text{a.s.-}f_k$$

as $\max_l A_l \to 0$ and

$$\frac{\mathrm{E}_{f_k}[N_A]}{-\log A_k} \to \frac{1}{\min\limits_{j:j\neq k} D(f_k, f_j)}$$

as $\max_l A_l \to 0$.

VI. Asymptotics of Error Probabilities

We now turn to the problem of determining asymptotics for error probabilities. Results are obtained through the application of nonlinear renewal theory [3]. To proceed we need the following definition [3].

Definition 6.1: The sequence of random variables $\{\xi_n\}_{n=1}^{\infty}$ is said to be *slowly changing* if it is uniformly continuous in probability (u.c.i.p.) [3] and

$$\frac{\max\{|\xi_1|, |\xi_2|, \cdots, |\xi_n|\}}{n} \to 0$$

in probability as $n \to \infty$.

We now state two lemmas before proceeding to the main theorem on error probabilities.

Lemma 6.1: Let

$$\xi_n = \log\left(a + \sum_{j=1}^{m} b_j e^{-\sum_{i=1}^{n} V_i^{(j)}}\right)$$

where $V_1^{(j)}, V_2^{(j)}, \cdots$ are i.i.d. with mean $\mu_j > 0$, and a, b_1, $b_2, \cdots, b_m$ are all positive constants. Then $\{\xi_n\}_{n=1}^{\infty}$ is slowly changing.

Proof: It follows from the Strong Law of Large Numbers that $\xi_n \to 0$ a.s. The remainder of the proof is elementary. □

Lemma 6.2: Assume $X_1, X_2, \cdots,$ are i.i.d. and nonarithmetic [3], with $E[X_1] = \mu > 0$. Let t_c be the first $n \geq 1$ such that $\sum_{i=1}^{n} X_i - \xi_n > c$. If $\{\xi_n\}_{n=1}^{\infty}$ is slowly changing, then

$$\sum_{i=1}^{t_c} X_i - \xi_{t_c} - c \to W$$

in distribution as $c \to \infty$, where

$$\mathrm{P}(W \leq w) = \frac{\int_0^w \mathrm{P}\left(\sum_{i=1}^{\tau^+} X_i > s\right) ds}{\mathrm{E}\left[\sum_{i=1}^{\tau^+} X_i\right]}$$

and where τ^+ is the first $n \geq 1$ such that $\sum_{i=1}^{n} X_i > 0$.

Proof: Immediate from Theorem 4.1 of [3]. □

The following result, which is proved in the Appendix, gives the asymptotics of the error probabilities.

Theorem 6.1: Assume that $D(f_k, f_{j^*}) = \min_{j:j\neq k} D(f_k, f_j)$ is positive and finite, and that j^* is unique. Assume also that the log–likelihood ratios are nonarithmetic. Then

$$\frac{\alpha_k}{\pi_k A_k} \to \gamma_k$$

as $\max_l A_l \to 0$, where $\gamma_k = E_{f_k}[e^{-W_k}]$, and W_k has distribution

$$P_{f_k}(W_k \leq w) = \frac{\int_0^w P_{f_k}\left(\sum_{i=1}^{\tau_k^+} \log\frac{f_k(X_i)}{f_{j^*}(X_i)} > s\right) ds}{E_{f_k}\left[\sum_{i=1}^{\tau_k^+} \log\frac{f_k(X_i)}{f_{j^*}(X_i)}\right]},$$

and where τ_k^+ is the first $n \geq 1$ such that

$$\textstyle\sum_{i=1}^{n} \log f_k(X_i)/f_{j^*}(X_i) > 0.$$

It should be noted that the asymptotic expression for the error probability α_k differs from the bound given in Theorem 4.2 by the factor γ_k. It is easily shown that $0 < \gamma_k < 1$ for each k. Techniques for computing the γ_k's can be found in the SPRT literature [3].

VII. MSPRT Design

Employing the asymptotic results from the previous section as approximations, we have the following approximations:

$$\mathrm{E}_{f_k}[N_A] \approx \frac{-\log A_k}{\min\limits_{j:j\neq k} D(f_k, f_j)}$$

$$\alpha_k \approx \pi_k A_k \gamma_k$$

$$\alpha \approx \sum_{k=0}^{M-1} \pi_k A_k \gamma_k.$$

The design of the MSPRT for specific applications requires choosing values for threshold parameters A_k. If all the frequentist error probabilities $\alpha_0, \cdots, \alpha_{M-1}$ are specified, then we simply set

$$A_k = \frac{\alpha_k}{\pi_k \gamma_k}.$$

If the total error probability α or the expected stopping time $\mathrm{E}[N_A]$ is specified, then we can choose the parameters so as to minimize the Bayes risk defined in Section III, that is,

$$\mathrm{E}[cN_A + W(\delta, H)].$$

Now

$$\mathrm{E}[cN_A] = c\sum_{k=0}^{M-1} \pi_k \mathrm{E}_{f_k}[N_A] \approx c\sum_{k=0}^{M-1} \pi_k \frac{-\log A_k}{\min\limits_{j:j\neq k} D(f_k, f_j)}$$

and

$$\begin{aligned} \mathrm{E}[W(\delta, H)] &= 1 \cdot \mathrm{P(incorrect\ decision)} \\ &= \alpha \approx \sum_{k=0}^{M-1} \pi_k A_k \gamma_k. \end{aligned}$$

Putting these approximations together, we get

$$\mathrm{E}[cN_A + W(\delta, H)] = \sum_{k=0}^{M-1} \pi_k \left\{ \frac{-c \log A_k}{\min_{j:j \neq k} D(f_k, f_j)} + A_k \gamma_k \right\}.$$

It can be shown that the expression on the right-hand side is minimized over the parameters A_k by setting

$$A_k = \frac{c}{\delta_k \gamma_k} \tag{5}$$

for $k = 0, 1, \cdots, M-1$, where $\delta_k = \min_{j:j \neq k} D(f_k, f_j)$.

With this choice,

$$\begin{aligned} \mathrm{E}[N_A] &\approx \sum_{k=0}^{M-1} \pi_k \frac{-\log A_k}{\delta_k} \\ &= -\log c \sum_{k=0}^{M-1} \frac{\pi_k}{\delta_k} + \sum_{k=0}^{M-1} \frac{\pi_k \log(\delta_k \gamma_k)}{\delta_k} \end{aligned}$$

and

$$\alpha \approx \sum_{k=0}^{M-1} \pi_k A_k \gamma_k = \sum_{k=0}^{M-1} \frac{\pi_k c}{\delta_k}.$$

If it is desired that $\mathrm{E}[N_A]$ is fixed to a desired level, then setting

$$c = \exp\left(\frac{-\mathrm{E}[N_A]}{\sum_{k=0}^{M-1} \frac{\pi_k}{\delta_k}} + \frac{\sum_{k=0}^{M-1} \frac{\pi_k \log(\delta_k \gamma_k)}{\delta_k}}{\sum_{k=0}^{M-1} \frac{\pi_k}{\delta_k}} \right) \tag{6}$$

is appropriate. Alternatively, if α is fixed,

$$c = \frac{\alpha}{\sum_{k=0}^{M-1} \frac{\pi_k}{\delta_k}} \tag{7}$$

should be chosen. Equation (5), in combination with (6) or (7), as desired, completely specifies the MSPRT.

VIII. Examples

In this section we consider two examples employing the MSPRT. The first is for sequential detection of a signal with one of M amplitudes, and the second is for sequential detection of one of M orthogonal signals. In both examples, the signals are corrupted by additive white Gaussian noise.

Example 1: Consider the problem of determining the amplitude of a signal in noise. For $k = 0, 1, \cdots, M-1$, H_k is the hypothesis that $X_1, X_2, \cdots,$ is a sequence of independent Gaussian random variables with mean θ_k and variance σ^2, where the θ_ks and σ^2 are known quantities.

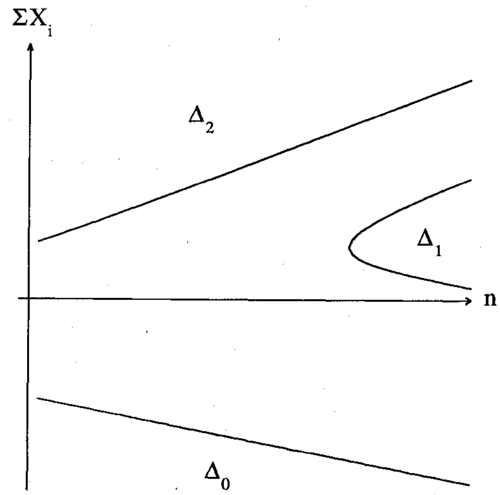

Fig. 4. Decision regions for Example 1.

With no loss of generality, it is assumed that the θ_ks are strictly increasing with k. With the assumptions of equal priors and $\sigma^2 = 1$, it is easily verified that the MSPRT takes the following form:

$$N_A = \text{first } n \geq 1 \text{ such that } \sum_{\substack{j=0 \\ j \neq k}}^{M-1} e^{\sum_{i=1}^{n} X_i(\theta_j - \theta_k) + \frac{n}{2}(\theta_k^2 - \theta_j^2)} < A_k \text{ for some } k$$

$$\delta = H_k.$$

The decision regions are plotted in Fig. 4 for $M = 3$. It can be seen that the curves become straight lines asymptotically as $n \to \infty$. For general M, expressions for the straight lines that these curves approach are given by

$$\sum_{i=1}^{n} X_i = \frac{n(\theta_k + \theta_{k+1})}{2} + \frac{\log A_k}{\theta_{k+1} - \theta_k}$$

and

$$\sum_{i=1}^{n} X_i = \frac{n(\theta_k + \theta_{k+1})}{2} - \frac{\log A_{k+1}}{\theta_{k+1} - \theta_k}$$

where k ranges from zero to $M-2$. It is interesting to note that these asymptotic expressions correspond to the decision regions for $M-1$ SPRT's, one for H_0 versus H_1, one for H_1 versus H_2, and so forth. The test which uses these $M-1$ SPRT's is precisely the *ad-hoc* test that is discussed in much of the previous work on M-ary sequential hypothesis testing (see, for example, [12]–[14], [22], [23]). The decision regions shown in Fig. 4 are very similar to those in [12] and [14].

The performance of this test for $M = 3$ and for specific values of the θ_ks and the A_ks is given in Table I. The test has been designed for specific values of α using the techniques of the previous section. Here, equal priors have been assumed, and $\mathrm{E}[N_A]$ denotes the average (unconditional) expected stopping time. For the simulations, the sample size is taken to be such that the standard

TABLE I
ERROR PROBABILITIES AND STOPPING TIMES FOR EXAMPLE 1

		Sequential (MSPRT)				Fixed Sample Size	
		Asymptotics		Simulation			
$A_0 = A_1$	A_2	α	$\mathrm{E}[N_A]$	α	$\mathrm{E}[N_A]$	α	N
$\theta_0 = -0.3,\ \theta_1 = 0.0,\ \theta_2 = 0.6$							
1.59×10^{-1}	4.72×10^{-2}	0.10	32.9	8.99×10^{-2}	32.7	9.02×10^{-2}	59
4.76×10^{-2}	1.42×10^{-2}	0.03	53.0	2.90×10^{-2}	55.6	2.89×10^{-3}	131
1.59×10^{-2}	4.72×10^{-3}	0.01	71.3	9.90×10^{-3}	75.5	9.91×10^{-3}	210
$\theta_0 = -0.4,\ \theta_1 = 0.0,\ \theta_2 = 0.5$							
1.43×10^{-1}	9.73×10^{-2}	0.10	22.4	9.27×10^{-2}	26.0	9.40×10^{-2}	44
4.30×10^{-2}	2.92×10^{-2}	0.03	35.6	2.94×10^{-2}	42.0	2.88×10^{-2}	85
1.43×10^{-2}	9.73×10^{-3}	0.01	47.7	9.92×10^{-3}	55.6	9.93×10^{-3}	126

deviations are less than 1 percent of the simulated quantities. In the calculation of the asymptotic expression for α, the γ values are obtained from [3, Table 3.1]. From the results, we see that the asymptotic expressions for the error probabilities and stopping times are reasonably accurate even for moderate values. The expressions for the error probabilities are especially accurate.

The performance of this test can also be compared with the performance of an appropriately designed fixed sample size test. For $M = 3$, $\sigma^2 = 1$, and equal priors, the minimum-probability-of-error fixed sample size test chooses H_0 if $\overline{X}_N \le (\theta_0 + \theta_1)/2$, chooses H_2 if $\overline{X}_N \ge (\theta_1 + \theta_2)/2$, and chooses H_1 otherwise, where

$$\overline{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$$

and N is the (fixed) number of observations. It can be shown that the probability of error for this test is given by

$$\alpha = \frac{2}{3} Q\big(\sqrt{N}(\theta_1 - \theta_0)/2\big) + \frac{2}{3} Q\big(\sqrt{N}(\theta_2 - \theta_1)/2\big)$$

where $Q(\cdot)$ is defined to be

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt. \tag{8}$$

Evaluation of α shows that significantly fewer samples are needed to achieve a desired error probability by using the MSPRT. For the θ values in Table I, the MSPRT requires, on the average, roughly one-third to one-half the number of samples needed for the fixed sample size test. The table also shows that the speedup factor increases as α decreases.

Two additional observations merit further discussion. The first is that the absolute error between the asymptotic and simulated values of $\mathrm{E}[N_A]$ appears to increase as A_0, A_1, and A_2 decrease. This should not cause concern, because the asymptotic expressions for $\mathrm{E}[N_A]$ can only guarantee that the *relative* error decreases toward zero.

The second observation is that the asymptotic values for $\mathrm{E}[N_A]$ appear to be more accurate for the $\theta_0 = -0.3$, $\theta_1 = 0.0$, $\theta_2 = 0.6$ case than for $\theta_0 = -0.4$, $\theta_1 = 0.0$, $\theta_2 = 0.5$. This should match intuition, because the asymptotic expression neglects all but the nearest (in terms of the informational divergence) hypothesis. One would expect, then, that the further away the other hypotheses, the more accurate the asymptotic expression.

Example 2: Now consider the problem of detecting one of M orthogonal signals in Gaussian noise. For simplicity it is assumed that the M signals contain equal energy, and that exactly one signal is present. The data consists of an infinite stream of independent and identically distributed random vectors, denoted $X_1, X_2, \cdots$. Furthermore, X_i, is written as $(X_{i,0}, X_{i,1}, \cdots, X_{i,M-1})$. For $k = 0, 1, \cdots, M-1$, H_k is the hypothesis that, for each i, $X_{i,k}$ is Gaussian with mean θ and variance σ^2, and $X_{i,j}$ is Gaussian with mean zero and variance σ^2 for all $j \ne k$. It is also assumed that, conditioned on each hypothesis, $X_{i,0}, X_{i,1}, \cdots, X_{i,M-1}$ are mutually independent.

With the assumption of equal priors, it can be shown by a symmetry argument that the MSPRT should be designed with $A_0 = A_1 = \cdots = A_{M-1}$, and the common value is denoted by A. If in addition $\sigma^2 = 1$, the MSPRT can then be written as follows:

$$N_A = \text{first } n \ge 1 \text{ such that } \sum_{\substack{j=0 \\ j \ne k}}^{M-1} e^{\theta \sum_{i=1}^{n} (X_{i,j} - X_{i,k})} < A \text{ for some } k.$$

At time N_A, hypothesis H_m is chosen if m is the value of k that satisfies the above inequality. Equivalently, m is the value of k for which

$$\sum_{i=1}^{N_A} X_{i,k}$$

is maximized. Unfortunately, the decision regions are difficult to plot, because M axes are required.

The performance of this test for $M = 4$ and $\theta^2 = 0.125$ is given in Table II. The test has been designed for specific values of α using the techniques of the previous section assuming equal priors. Note that multiple hypotheses are equidistant from the true hypothesis. Thus, Theorem 6.1 has not been proven in this case. However, the simulation results in Table II clearly seem to indicate that the theorem does indeed hold. For the simulations, the sample size is taken to be such that the standard deviations are less than 1 percent of the simulated quantities. From the results, we see that the asymptotic expressions are quite accurate in estimating α (even though Theorem 6.1, as stated, does not apply), but they are not at all accurate in estimating $\mathrm{E}[N_A]$. The reason for this is that multiple hypotheses are equidistant from the true hypothesis, and, as explained in Example 1, the nearer the other hypotheses, the less accurate the asymptotic expression. However, it can be seen that the percentage error is decreasing as A decreases.

The performance of this test can also be compared with the performance of an appropriately designed fixed sample size test. With equal priors, the minimum-probability-

TABLE II
ERROR PROBABILITIES AND STOPPING TIMES FOR EXAMPLE 2

	Sequential (MSPRT)				Fixed Sample Size	
	Asymptotics		Simulation			
A	α	$E[N_A]$	α	$E[N_A]$	α	N
$M = 4, \theta = 1/\sqrt{8}$						
1.34×10^{-1}	0.100	16.1	9.37×10^{-2}	29.9	9.35×10^{-2}	50
1.34×10^{-2}	0.010	34.5	1.03×10^{-2}	55.0	1.01×10^{-2}	114
1.34×10^{-3}	0.001	52.9	1.00×10^{-3}	76.9	1.01×10^{-3}	184

of-error fixed sample size test chooses H_k if $\overline{X}_N^{(k)} = \max_j \overline{X}_N^{(j)}$, where

$$\overline{X}_N^{(k)} = \frac{1}{N} \sum_{i=1}^{N} X_{i,k}$$

and N is the number of observations. It can be shown that the probability of error for this test, assuming $\sigma^2 = 1$, is given by

$$\alpha = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t-\sqrt{N}\theta)^2/2} [1 - Q(t)]^{M-1} \, dt$$

where $Q(\cdot)$ is given by (8).

From Table II it can be seen that many fewer samples are needed to achieve a desired probability by using the MSPRT. As with the first example, the average number of required samples for the MSPRT is roughly one-third to one-half the number needed for the fixed sample size test.

IX. CONCLUSIONS

In this paper we have presented a sequential test for multihypothesis testing that generalizes the two-hypothesis SPRT. The test has been motivated by examining the optimality of sequential tests under a Bayesian framework when the cost per observation is small. We have also provided one example for which the optimal test can be computed straightforwardly, and we have shown that the MSPRT is asymptotically optimal in this case. In addition, we have given bounds on error probabilities, and we have found asymptotic expressions for both the error probabilities and stopping times. A design procedure has been provided to determine the parameters A_k. We have also presented two examples and compared asymptotic results with simulations, as well as with appropriately designed fixed sample size tests.

There are a number of questions regarding the MSPRT that merit further research. First, although the arguments presented in Section III certainly provide a justification for the using the MSPRT when the cost per observation is small compared to the cost of an incorrect decision, they do not constitute a proof of the asymptotic optimality of the MSPRT. It would be of interest to investigate if the MSPRT is indeed asymptotically optimal, and to determine the conditions, if any, under which it is so.

Second, it should be noted that Theorem 6.1 assumes that $\min_{j:j \neq k} D(f_k, f_j)$ is achieved for a unique j. An important question is whether this theorem can be extended to the situation in which ties occur. The proof as given cannot be extended easily, because the nonlinear term ξ_n, in general, will no longer be slowly changing. Because problems in M-ary hypothesis testing may often contain symmetry, the extension of the theorem merits further research. Furthermore, Example 2 clearly indicates that there are at least some cases for which the theorem *can* be extended.

Third, it has been noted through the examples that the asymptotic expressions for $E[N_A]$ are not especially accurate; this is particularly true when multiple hypotheses are equally distant from each other (in the sense of the informational divergence). Through the use of nonlinear renewal theory, it may be possible to obtain second order results for asymptotics on $E[N_A]$ in a manner similar to the second order results for the stopping time of the SPRT (see [3, chap. 3]). It is suspected that such a result will be considerably more difficult to obtain when multiple hypotheses are equidistant than when a unique j achieves $\min_{j:j \neq k} D(f_k, f_j)$.

Finally, it would be helpful to analyze the MSPRT when the data fits *none* of the hypotheses, and to develop techniques for the design of the MSPRT for composite hypotheses. Some work in this area has been done for the SPRT [3]. Such work should increase the number of potential applications for the MSPRT.

APPENDIX

Proof of Lemma 3.3: For the purposes of this proof, we abuse notation slightly and write the cost functions explicitly as functions of c. For example, $A_n^T(\boldsymbol{p})$ is written as $A_n^T(\boldsymbol{p}; c)$.

Claim: For any T and any $n < T$, $A_n^T(\boldsymbol{p}; c)$ is a concave, continuous and monotonically increasing (c.c.m.i.) function of c on the interval $[0, \infty]$.

Proof of Claim: The claim is true for $n = T - 1$ since the function $A_{T-1}^T(\boldsymbol{p}; c)$ does not depend on c. Now, suppose the claim is true for $n = m + 1 < T$. Then it is easily checked that

$$J_{m+1}^T(\boldsymbol{p}; c) = \min\{g(\boldsymbol{p}), c + A_{m+1}^T(\boldsymbol{p}; c)\}$$

is a c.c.m.i. function of c on $[0, \infty]$. This implies that the claim is true for $n = m$. By induction, the claim follows for all $n < T$. □

Now, for each $c \in [0, \infty]$,

$$A_0^T(\boldsymbol{p}; c) \downarrow A_J(\boldsymbol{p}; c) \text{ as } T \uparrow \infty. \tag{9}$$

The infimum of concave and monotonically increasing functions is also concave and monotonically increasing. Hence $A_J(\boldsymbol{p}; c)$ is a concave, monotonically increasing function of c on $[0, \infty]$. This implies that $A_J(\boldsymbol{p}; c)$ is a c.c.m.i. function of c on the interval $(0, \infty)$.

For $c = 0$, the optimal test simply takes an infinite number of observations to determine the true hypothesis precisely. (This is possible since we have assumed that $f_k \neq f_j$ almost surely for all $j \neq k$.) Thus, $J(\boldsymbol{p}; 0) = 0$, and consequently, $A_J(\boldsymbol{p}, 0) = 0$.

To finish the proof of the lemma, the continuity of $A_J(\boldsymbol{p}; c)$ at $c = 0$ needs to be established. Given $\epsilon > 0$, pick τ such that

$$A_0^T(\boldsymbol{p}; 0) - A_J(\boldsymbol{p}; 0) \le \frac{\epsilon}{2}, \quad \text{for all } T \ge \tau. \tag{10}$$

(This is possible by (9).) Also pick $\delta > 0$ such that

$$A_0^\tau(\boldsymbol{p}; c) - A_0^\tau(\boldsymbol{p}; 0) \le \frac{\epsilon}{2} \quad \text{for all } c \le \delta. \tag{11}$$

(This is possible by the continuity of $A_0^\tau(\boldsymbol{p}; c)$ at $c = 0$.)

From (9), (10), and (11), it is easy to show that the following is true:

$$A_0^T(\boldsymbol{p}; c) - A_J(\boldsymbol{p}; 0) \le \epsilon \text{ for all } c \le \delta \text{ and for all } T \ge \tau.$$

Taking limits as $T \to \infty$ in the above equation gives

$$A_J(\boldsymbol{p}; c) - A_J(\boldsymbol{p}; 0) \le \epsilon \text{ for all } c \le \delta$$

which establishes the continuity of $A_J(\boldsymbol{p}; c)$ at $c = 0$. □

Proof of Lemma 5.3: It is clear that $\tilde{N}_A \ge N_A$. Thus, by Lemma 5.1, $\tilde{N}_A \to \infty$ a.s.-f_k as $\max_l A_l \to 0$. Now, note that $\tilde{N}_A$ may be written as

$$\tilde{N}_A = \text{first } n \ge 1 \ni$$
$$-\frac{1}{n} \log \left(\sum_{j:j \ne k} \exp\left(-n \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} + \frac{1}{n} \log \frac{\pi_k}{\pi_j} \right] \right) \right) > \frac{-\log A_k}{n}.$$

Also note that

$$\frac{1}{n} \log \frac{\pi_k}{\pi_j} + \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} \to D(f_k, f_j) \quad \text{a.s.-}f_k$$

as $n \to \infty$. Thus, by Lemma 5.2,

$$-\frac{1}{n} \log \left(\sum_{j:j \ne k} \exp\left(-n \left[\frac{1}{n} \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} + \frac{1}{n} \log \frac{\pi_k}{\pi_j} \right] \right) \right) \to \min_{j:j \ne k} D(f_k, f_j) \quad \text{a.s.-}f_k$$

as $n \to \infty$. Therefore,

$$-\frac{1}{\tilde{N}_A} \log \left(\sum_{j:j \ne k} \exp\left[-\sum_{i=1}^{\tilde{N}_A} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k} \right] \right) \to \min_{j:j \ne k} D(f_k, f_j) \quad \text{a.s.-}f_k$$

and

$$-\frac{1}{\tilde{N}_A - 1} \log \left(\sum_{j:j \ne k} \exp\left[-\sum_{i=1}^{\tilde{N}_A - 1} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k} \right] \right) \to \min_{j:j \ne k} D(f_k, f_j) \quad \text{a.s.-}f_k$$

both as $\max_l A_l \to 0$. Furthermore, by the definition of $\tilde{N}_A$, we have

$$-\frac{1}{\tilde{N}_A} \log \left(\sum_{j:j \ne k} \exp\left[-\sum_{i=1}^{\tilde{N}_A} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k} \right] \right) > \frac{-\log A_k}{\tilde{N}_A}$$

and

$$-\frac{1}{\tilde{N}_A - 1} \log \left(\sum_{j:j \ne k} \exp\left[-\sum_{i=1}^{\tilde{N}_A - 1} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k} \right] \right) \le \frac{-\log A_k}{\tilde{N}_A - 1}.$$

Taking appropriate infimums and supremums gives the desired result. □

Proof of Theorem 5.1: To show almost sure convergence, we can write, for all $\epsilon > 0$,

$$\begin{aligned}
&\mathrm{P}_{f_k}\left(\left| \frac{N_A}{-\log A_k} - \frac{1}{\min_{j:j \ne k} D(f_k, f_j)} \right| > \epsilon \right) \\
&= \sum_{l=0}^{M-1} \mathrm{P}_{f_k}\left(\left| \frac{N_A}{-\log A_k} - \frac{1}{\min_{j:j \ne k} D(f_k, f_j)} \right| > \epsilon \text{ and accept } H_l \right) \\
&= \mathrm{P}_{f_k}\left(\left| \frac{\tilde{N}_A}{-\log A_k} - \frac{1}{\min_{j:j \ne k} D(f_k, f_j)} \right| > \epsilon \right) \\
&\quad + \sum_{l:l \ne k} \mathrm{P}_{f_k}\left(\left| \frac{N_A}{-\log A_k} - \frac{1}{\min_{j:j \ne k} D(f_k, f_j)} \right| > \epsilon \text{ and accept } H_l \right) \\
&\le \mathrm{P}_{f_k}\left(\left| \frac{\tilde{N}_A}{-\log A_k} - \frac{1}{\min_{j:j \ne k} D(f_k, f_j)} \right| > \epsilon \right) + \sum_{l:l \ne k} \mathrm{P}_{f_k}(\text{accept } H_l).
\end{aligned}$$

Now, the term on the left-hand side converges to zero as $\max_l A_l \to 0$ by Lemma 5.3, and the terms on the right-hand side converge to zero as $\max_l A_l \to 0$ by application of Theorem 4.2. This proves the first part of the theorem.

To show convergence in mean, it suffices to establish uniform integrability of $\{N_A/(-\log A_k), \max_l A_l < 1\}$ (see [19] for explanation and definition). First note that

$$N_A \le \text{first } n \ge 1 \text{ such that } \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} > \left\lceil \log \frac{(M-1)\pi_j}{A_k \pi_k} \right\rceil \quad \forall j \ni j \ne k.$$

Now, let

$$\tau_1 = \text{first } n \geq 1 \text{ such that } \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} > 1 \quad \forall j \ni j \neq k$$

$$\tau_2 = \text{first } n \geq 1 \text{ such that } \sum_{i=\tau_1+1}^{n+\tau_1} \log \frac{f_k(X_i)}{f_j(X_i)} > 1$$

$$\forall j \ni j \neq k$$

and define τ_i for $i > 2$ in the obvious way. Note that $\tau_1, \tau_2, \cdots,$ are independent and identically distributed. Also note that $E_{f_k}[\tau_1] < \infty$. To see this, define

$$\tau_1^j = \text{first } n \geq 1 \text{ such that } \sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} > \tilde{n}_j$$

for all $j \neq k$, where

$$\tilde{n}_j = \left\lceil \log \frac{(M-1)\pi_j}{A_k \pi_k} \right\rceil$$

and note that

$$\tau_1 = \max_{j:j\neq k} \tau_1^j \leq \sum_{j:j\neq k} \tau_1^j.$$

Furthermore, the fact that

$$E_{f_k}\left[\log \frac{f_k(X_1)}{f_j(X_1)}\right] = D(f_k, f_j) > 0 \quad \forall j \ni j \neq k$$

implies that $E_{f_k}[\tau_1^j] < \infty$ for all $j \neq k$ (see [24] for proof). Thus, τ_1 also has finite expected value.

Now, note that

$$\sum_{i=1}^{\tau_1 + \cdots + \tau_{\tilde{n}_j}} \log \frac{f_k(X_i)}{f_j(X_i)} > \tilde{n}_j$$

for all $j \neq k$, so that $\tau_1 + \cdots + \tau_{\tilde{n}_j} \geq N_A$. Since, by the Strong Law of Large Numbers,

$$\frac{1}{\tilde{n}_j} \sum_{i=1}^{\tilde{n}_j} \tau_i \to E_{f_k}[\tau_1] \text{ a.s.-}f_k \quad \text{as } \max_l A_l \to 0$$

and since

$$E_{f_k}\left[\frac{1}{\tilde{n}_j} \sum_{i=1}^{\tilde{n}_j} \tau_i\right] = E_{f_k}[\tau_1]$$

it follows that (see [19])

$$\left\{\frac{1}{\tilde{n}_j} \sum_{i=1}^{\tilde{n}_j} \tau_i, \quad \max_l A_l < 1\right\}$$

is uniformly integrable, which implies that $\{N_A/(-\log A_k), \max_l A_l < 1\}$ is also uniformly integrable. □

Proof of Theorem 6.1: We can write

$$\begin{aligned}
\alpha_k &= \sum_{j:j\neq k} \sum_{n=1}^{\infty} \cdot \int_{\{\text{accept } H_k,\ N_A = n\}} \pi_j f_j(x_1) \cdots f_j(x_n)\, dx_1 \cdots dx_n \\
&= \sum_{n=1}^{\infty} \int_{\{\text{accept } H_k,\ N_A = n\}} \cdot \sum_{j:j\neq k} \exp\left(-\sum_{i=1}^{n} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k}\right) \cdot \pi_k f_k(x_1) \cdots f_k(x_n)\, dx_1 \cdots dx_n \\
&= \pi_k E_{f_k}\left[\sum_{j:j\neq k} \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)} + \log \frac{\pi_j}{\pi_k}\right) \cdot I_{\left\{\sum_{j:j\neq k} \pi_j \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)}\right) < \pi_k A_k\right\}}\right] \\
&= \pi_k A_k E_{f_k}\left[\sum_{j:j\neq k} \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)} - \log \frac{A_k \pi_k}{\pi_j}\right) \cdot I_{\left\{\sum_{j:j\neq k} \pi_j \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)}\right) < \pi_k A_k\right\}}\right].
\end{aligned}$$

Now, it can be shown that

$$\begin{aligned}
-\log&\left[\sum_{j:j\neq k} \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)} - \log \frac{A_k \pi_k}{\pi_j}\right)\right] \\
&= \sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_{j^*}(X_i)} - \log\left(A_k + \sum_{j:j\neq k, j^*} A_k \frac{\pi_k}{\pi_j} \cdot \exp\left(\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_{j^*}(X_i)} - \log \frac{f_k(X_i)}{f_j(X_i)}\right)\right).
\end{aligned}$$

If the second term is denoted ξ_{N_A}, then by Lemmas 6.1 and 6.2, it follows that

$$\sum_{j:j\neq k} \exp\left(-\sum_{i=1}^{N_A} \log \frac{f_k(X_i)}{f_j(X_i)} - \log \frac{A_k \pi_k}{\pi_j}\right) \to e^{-W_k}$$

in distribution as $\max_l A_l \to 0$. Furthermore, since the indicator function in the above equation for α_k converges to 1 in probability, the entire expression inside the expectation converges to e^{-W_k} in distribution. Since it is bounded between 0 and 1, the expectation converges to γ_k, which gives the desired result. □

Acknowledgment

This research was presented in part as a final project in a course on sequential analysis taught by A. T. Martinsek. The authors wish to thank Dr. Martinsek, as well as M. B. Pursley, H. V. Poor, and T. Başar, for their support and advice. The authors also wish to thank the reviewers for their helpful comments.

References

[1] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
[2] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

[3] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis, SIAM*, 1982.

[4] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Wiley, 1959.

[5] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*, vol. III. Rockville, MD: Comput. Sci., 1985.

[6] M. B. Marcus and P. Swerling, "Sequential detection in radar with multiple resolution elements," *IRE Trans. Inform. Theory*, pp. 237–245, 1962.

[7] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[8] K. S. Fu, *Sequential Methods in Pattern Recognition and Learning*. New York: Academic, 1968.

[9] A. G. Tartakovskii, "Sequential testing of many simple hypotheses with independent observations," *Probl. Inform. Transm.*, vol. 24, no. 4, pp. 299–309, 1989 (translated from the Russian version of Oct.–Dec. 1988).

[10] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.

[11] D. Blackwell and M. A. Girschik, *Theory of Games and Statistical Decisions*. New York: Wiley, 1970.

[12] M. Sobel and A. Wald, "A sequential design procedure for choosing one of three hypotheses concerning the unknown mean of normal distribution," *Ann. Mathemat. Statist.*, vol. 20, pp. 502–522, 1949.

[13] P. Armitage, "Sequential analysis with more than two alternative hypotheses and its relation to discriminant function analysis," *J. Roy. Statist. Soc. Suppl.*, vol. 9, pp. 250–263, 1947.

[14] G. Simons, "A sequential three hypothesis test for determining the mean of a normal population with known variance," *Ann. Mathemat. Statist.*, vol. 38, pp. 1365–1375, 1967.

[15] G. Lorden, "Likelihood ratio tests for sequential k-decision problems," *Ann. Mathemat. Statist.*, vol. 43, pp. 1412–1427, 1972.

[16] ——, "2-SPRT's and the modified Kiefer–Weiss problem of minimizing an expected sample size," *Ann. Statist.*, vol. 4, pp. 281–291, 1976.

[17] B. Eisenberg, "Multihypothesis problems," in *Handbook of Sequential Analysis*, B. K. Ghosh and P. K. Sen, Eds. New York: Marcel Dekker, 1991.

[18] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[19] P. Billingsley, *Probability and Measure*. New York: Wiley, 1986.

[20] D. Bertsekas, *Dynamic Programming*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[21] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433–442, Mar. 1993.

[22] E. Paulson, "A sequential decision procedure for choosing one of k hypotheses concerning the unknown mean of a normal distribution," *Ann. Mathemat. Statist.*, vol. 34, pp. 549–554, 1963.

[23] G. Simons, "A class of sequential procedures for choosing one of k hypotheses concerning the unknown drift parameter of the Wiener process," *Ann. Mathemat. Statist.*, vol. 38, pp. 1376–1383, 1967.

[24] S. M. Ross, *Stochastic Processes*. New York: Wiley, 1983.