Multihypothesis Sequential Probability Ratio Tests—Part I: Asymptotic Optimality

Vladimir P. Dragalin, Alexander G. Tartakovsky, and Venugopal V. Veeravalli, *Senior Member, IEEE*

Abstract— The problem of sequential testing of multiple hypotheses is considered, and two candidate sequential test procedures are studied. Both tests are multihypothesis versions of the binary sequential probability ratio test (SPRT), and are referred to as MSPRT's. The first test is motivated by Bayesian optimality arguments, while the second corresponds to a generalized likelihood ratio test. It is shown that both MSPRT's are asymptotically optimal relative not only to the expected sample size but also to any positive moment of the stopping time distribution, when the error probabilities or, more generally, risks associated with incorrect decisions are small. The results are first derived for the discrete-time case of independent and identically distributed (i.i.d.) observations and simple hypotheses. They are then extended to general, possibly continuous-time, statistical models that may include correlated and nonhomogeneous observation processes. It also demonstrated that the results can be extended to hypothesis testing problems with nuisance parameters, where the composite hypotheses, due to nuisance parameters, can be reduced to simple ones by using the principle of invariance. These results provide a complete generalization of the results given in [36], where it was shown that the quasi-Bayesian MSPRT is asymptotically efficient with respect to the expected sample size for i.i.d. observations. In a companion paper [12], based on the nonlinear renewal theory we find higher order approximations, up to a vanishing term, for the expected sample size that take into account the overshoot over the boundaries of decision statistics.

Index Terms— Asymptotic optimality, invariant sequential tests, multialternative sequential tests, one-sided SPRT, renewal theory, slippage problems.

I. INTRODUCTION

THE goal of statistical hypothesis testing is to classify a sequence of observations into one of M $(M \geq 2)$ possible hypothesis based on some knowledge of statistical distributions of the observations under each of the hypotheses. For sequential testing problems, the number of observations used (sample size) is allowed to be variable, i.e., the sample size is a function of the observations. A sequential test picks a stopping time and a final decision rule to effect a tradeoff between sample size and decision accuracy.

The majority of research in sequential hypothesis testing has been restricted to two hypotheses. However, there are several situations, particularly in engineering applications, where it is natural to consider more than two hypotheses. Examples include, among a multitude of others, target detection in multiple-resolution radar [26] and infrared systems [13], signal acquisition in direct sequence code-division multiple access systems [37], and statistical pattern recognition [16]. It is thus of interest to study sequential testing of more than two hypotheses.

It is well known that for binary hypotheses testing $(M = 2)$, Wald's sequential probability ratio test (SPRT) is optimal, in the sense that it simultaneously minimizes both expectations of the sample size among all tests (sequential and nonsequential) for which the probabilities of error do not exceed predefined values (see, e.g., Wald and Wolfowitz [40], Chernoff [6], and Lehmann [22]). Unfortunately, if $M \geq 3$, it is not clear if there even exists a test that minimizes the expected sample size for all hypotheses. Moreover, existing research indicates that even if such a test exists, it would be very difficult to find its structure. For this reason a substantial part of the development of sequential multihypothesis testing has been directed toward the study of practical, suboptimal sequential tests and the evaluation of their asymptotic performance, see [3], [5]–[11], [21], [25], [30]–[38], and many others. A model for studying such asymptotics was first introduced by Chernoff [5], wherein he studied the independent and identically distributed (i.i.d.) case and assumed a cost per observation that was allowed to go to zero. Generalization to non-i.i.d. cases (where log-likelihood ratios fail to be random walks) have been made by Golubev and Khas'minskii [18], Lai [23], Tartakovsky [31], [32], [34], [35], as well as Verdenskaya and Tartakovskii [38].

For the binary SPRT, it is well known that renewal theory is useful in obtaining asymptotically exact expressions for expected sample sizes and error probabilities (see, e.g., [29], and [41]). An approach to applying renewal theory techniques to sequential multihypothesis tests was recently elucidated by Baum and Veeravalli [3], wherein they studied a generalization[1] of the SPRT to more than two hypotheses that is motivated by a Bayesian setting. This quasi-Bayesian

V. P. Dragalin is with the Department of Biostatistics, University of Rochester, Rochester, NY 14642 USA (e-mail: dragalin@bio1.bst.rochester.edu).

A. G. Tartakovsky is with the Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-1113 USA (e-mail: tartakov@cams.usc.edu).

V. V. Veeravalli is with the School of Electrical Engineering and the Department of Statistical Science, Cornell University, Ithaca, NY 14853 USA (e-mail: venu@ee.cornell.edu).

[1]This generalization also appears in other papers [18], [30]–[32] in various contexts.

multihypothesis SPRT (or MSPRT) was shown to be amenable to an asymptotic analysis using nonlinear renewal theory [41], and asymptotic expressions for the expected sample size and error probabilities were obtained in [3]. Furthermore, it was established in [36] that the quasi-Bayesian MSPRT is asymptotically efficient with respect to expected sample size, for the i.i.d. observation case.

In this paper, we continue to investigate the asymptotic behavior of the quasi-Bayesian MSPRT, as well as a related test that corresponds to a generalized likelihood ratio test. The latter test was suggested by Armitage [1], and studied in [8]–[10], [25], and [30]–[38] in various contexts. We provide a complete generalization of previous results, including that obtained in [36], in the following directions: i) we show that both MSPRT's are asymptotically optimal relative not only to the expected sample size but also to any positive moment of the stopping time distribution; and ii) we extend these results to general, possibly continuous-time, statistical models that may include correlated and nonhomogeneous observation processes. We also demonstrate that our results are applicable to hypothesis testing problems with nuisance parameters. This complete generalization justifies the use of the MSPRT's in a variety of applications, some of which are discussed in the paper.

The remainder of the paper is organized as follows. In Section II, we formulate the problem, present a motivation, and describe the structure of sequential test procedures. In Section III, we derive lower bounds for the finite moments of the stopping time distribution under very general conditions, which then are used in proving asymptotic optimality. In Section IV, first we consider the discrete-time i.i.d. case and prove that, under certain natural conditions, both tests minimize not only the expected sample size but also any positive moment of the stopping time distribution as error probabilities (or, more generally, decision risks) tend to zero. Then the asymptotic optimality results are shown to be valid for general statistical models that involve correlated and non-homogeneous observation processes. This generalization to non-i.i.d. cases is done in the spirit of the work in [23] and [35], and we borrow many auxiliary results from this work. Section V deals with a specific class of multipopulation problems—the so-called "slippage" problems. It is shown that slippage problems are related to problems that arise in target detection in multiresolution or multichannel systems. In Section VI, we illustrate the general results through examples related to detecting signals in multichannel systems, including composite testing situations with nuisance parameters. Finally, in Section VII we give some concluding remarks.

Higher order approximations for the expected sample size up to a vanishing term will be given in the companion paper [12]. These approximations take into account the overshoot over the boundaries of decision statistics, and their derivation is based on nonlinear renewal theory. In the companion paper, we also present simulation results which confirm that the obtained approximations are highly accurate not only for small but also for moderate error probabilities which are typical for many applications.

II. PROBLEM FORMULATION AND TEST PROCEDURES

A. Basic Notation

Throughout the paper $(\Omega, \mathcal{F}, \mathrm{P})$ is a probability space on which everything is defined, $\{X_t, \, t \in \mathcal{R}\}$ is a random process (defined on $(\Omega, \mathcal{F})$), which is observed either in discrete ($\mathcal{R} = \{1, 2, \cdots\}$) or continuous ($\mathcal{R} = [0, \infty)$) time. By $\mathcal{F}_t = \sigma(X^t)$ will be denoted a sub-σ-algebra of $\mathcal{F}$ generated by $X^t = \{X_u, 0 \leq u \leq t\}$ observed up to time t.

We first consider the problem of sequential testing of M simple hypotheses "$H_i : \mathrm{P} = \mathrm{P}_i$," $i = 0, 1, \cdots, M - 1$, where $M \geq 2$ and P_i are completely known distinct probability measures. An example where the measures P_i are not completely known (belong to some families of distributions) will be considered in Section VI.

A pair $\delta = (\tau, d)$ is said to be a sequential hypothesis test if τ is a Markov stopping time with respect to the family $\{\mathcal{F}_t\}$ (i.e., $\{\tau \leq t\} \in \mathcal{F}_t$), and $d = d(X^\tau)$ is an $\mathcal{F}_\tau$-measurable function (terminal decision function) with values in the set $\{0, 1, \cdots, M - 1\}$. Therefore, $d = i$ is identified with accepting the hypothesis H_i.

Next, let $(\pi_0, \cdots, \pi_{M-1})$ be the prior distribution of hypotheses, $\pi_i = \Pr(H_i) > 0$, and let the consequence of deciding $d = i$ when the hypothesis H_j is true be evaluated by the loss function $W(j, i) \in [0, \infty)$, $i, j = 0, 1, \cdots, M - 1$. Suppose, without loss of generality, that the losses due to correct decisions are zero (i.e., $W(i, i) = 0$). The risk associated with making the decision $d = i$ is then given by

$$R_i(\delta) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} \pi_j W(j, i) \alpha_{ji}(\delta)$$

where for $j \neq i$, $\alpha_{ji}(\delta) = \mathrm{P}_j(d = i)$ is the probability of accepting the hypothesis H_i when H_j is true (the conditional probability of error). Note that R_i is the risk associated with wrong decisions; their sum constitutes the average risk which does not include the observation or experiment cost. In the particular case of the *zero–one* loss function, where $W(j, i) = 1$ for $j \neq i$, the risk $R_i(\delta)$ is the same as *frequentist* error probability $\alpha_i(\delta)$, which is probability of accepting H_i incorrectly. That is, for the zero–one loss function

$$R_i(\delta) = \alpha_i(\delta) = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} \pi_j \alpha_{ji}(\delta).$$

We now introduce the following class of tests

$$\mathbf{\Delta}(\overline{\boldsymbol{R}}) = \{\delta \colon R_i(\delta) \leq \overline{R}_i, \qquad i = 0, 1, \cdots, M - 1\} \quad (2.1)$$

where $\overline{\boldsymbol{R}} = (\overline{R}_0, \overline{R}_1, \cdots, \overline{R}_{M-1})$ is a vector of positive finite numbers. In other words, the class $\mathbf{\Delta}(\overline{\boldsymbol{R}})$ consists of the tests for which the risks do not exceed the predefined numbers $\overline{R}_i$, $i = 0, 1, \cdots, M - 1$.

In what follows, E_i will denote the expectation with respect to the measure P_i. A reasonable formulation of the problem of optimizing sequential tests under given prior information is to find a test that minimizes the average observation time, $\overline{\mathrm{E}}\tau = \sum_{i=0}^{M-1} \pi_i \mathrm{E}_i \tau$, among all tests belonging to the class $\mathbf{\Delta}(\overline{\boldsymbol{R}})$.

The structure of such a test has been determined in [31] and [32] (see also [3]). Specifically, if time is discrete, i.e., $t = n = 1, 2, \cdots$, and the observations are i.i.d., the optimal stopping time is given by $\tau_{\mathrm{opt}} = \min_i \tau_i^0$, where τ_i^0 is the first time n, $n \geq 1$, for which $\Pi_i(n) \geq A_i(\vec{\Pi}_i(n))$. Here $\Pi_i(n) = \Pr(H_i | X^n)$ is the posterior probability of the ith hypothesis based on the first n observations, $\vec{\Pi}_i(n)$ is the vector of posterior probabilities with ith component excluded, and $A_i(\cdot)$ is a nonlinear function depending on the distribution of observations. Thus the optimal test involves a comparison of posterior probabilities with random thresholds which must be determined for each model. In general, it is nearly impossible to find the form of the functions $A_i(\cdot)$, $i = 0, 1, \cdots, M - 1$ (for some special cases see [31], [32], and [3]). Furthermore, even if one is able to find the boundaries of the optimal test, it would be difficult to implement this procedure in practice. In addition, while in the case of two hypotheses Wald's SPRT minimizes not only the average observation time $\overline{\mathrm{E}}\tau$ but also both expectations $\mathrm{E}_0\tau$ and $\mathrm{E}_1\tau$, it is unclear if such a test exists for $M > 2$ (i.e., the test which minimizes $\mathrm{E}_i\tau$ for all $i = 0, 1, \cdots, M - 1$, $M \geq 3$). However, in an asymptotic setting when the risks (error probabilities) are sufficiently small, such tests may be found.

B. Sequential Test Procedures

In the following, we simplify the structure of the optimal test by replacing the nonlinear random thresholds with simple functions (constants in the case of the zero–one loss function). Specifically, we consider the two following multialternative sequential tests which will be called *multihypothesis sequential probability ratio tests* (MSPRT's).[2] The time parameter t may be either discrete or continuous. Let $\mathrm{P}_i^t(\cdot)$ be the restriction of the measure P_i to the σ-algebra $\mathcal{F}_t$ and let

$$Z_i(t) = \log \frac{d\mathrm{P}_i^t}{d\mathbf{Q}^t}(X^t), \qquad i = 0, 1, \cdots, M - 1$$

denote the log-likelihood ratio (LLR) processes with respect to a dominating measure $\mathbf{Q}^t$. If $\mathbf{Q}^t = \mathrm{P}_j^t$, the corresponding LLR process will be denoted $Z_{ij}(t)$.

1) Test δ_a: Introduce the Markov times

$$\tau_i = \inf\left\{ t \colon Z_i(t) \geq a_i + \log\left(\sum_{j \neq i} w_{ji} \exp[Z_j(t)] \right) \right\} \tag{2.2}$$

where $w_{ji} = \pi_j W(j, i)/\pi_i$ and a_i are positive thresholds. The test procedure $\delta_a = (\tau_a, d_a)$ is defined as follows:

$$\tau_a = \min_{0 \leq k \leq M - 1} \tau_k \qquad d_a = i \text{ if } \tau_a = \tau_i.$$

That is, we stop as soon as the threshold in (2.2) is exceeded for some i, and decide that this i is the true hypothesis. This test is motivated by a Bayesian framework and was considered earlier by Fishman [15], Golubev and Khas'minskii [18], Sosulin and Fishman [30], Tartakovsky [31], [32], as well

[2] This generalization also appears in other papers [18], [30], [31], and [32] in various contexts.

as Baum and Veeravalli [3], [36]. Indeed, in the special case of zero–one losses the stopping times τ_i may be rewritten as

$$\tau_i = \inf\{t \colon \Pi_i(t) \geq A_i\}, \qquad \text{where } A_i = \frac{\exp(a_i)}{1 + \exp(a_i)}$$

and where

$$\Pi_i(t) = \frac{\pi_i \exp[Z_i(t)]}{\sum_{j=0}^{M-1} \pi_j \exp[Z_j(t)]}$$

is the posterior probability of the hypothesis H_i. Note also that if $A_i = A$ (does not depend on i), the stopping time of the test is defined by

$$\tau_a = \inf\left\{ t \colon \max_i \Pi_i(t) \geq A \right\}, \qquad \text{where } A = \frac{\exp(a)}{1 + \exp(a)}$$

i.e., we stop as soon as the largest posterior probability exceeds a threshold.

2) Test δ_b: Let $b_{ij} > 0$ and

$$\nu_i = \inf\left\{ t \colon Z_i(t) \geq \max_{j \neq i}[b_{ij} + \log w_{ji} + Z_j(t)] \right\} \tag{2.3}$$

be the Markov "accepting" time for the hypothesis H_i. The test $\delta_b = (\nu_b, d_b)$ is defined as follows:

$$\nu_b = \min_{0 \leq i \leq M - 1} \nu_i \qquad d_b = i \text{ if } \nu_b = \nu_i.$$

This test represents a modification of the matrix SPRT (the combination of one-sided SPRT's) that was suggested by Lorden [25]. It was considered earlier by Armitage [1], Lorden [25], Dragalin [8]–[10], Tartakovsky [32], [34], [35], as well as Verdenskaya and Tartakovskii [38]. Note also that if $b_{ij} = b_i$ and the loss function is zero–one, then $w_{ji} = \pi_j/\pi_i$ and the Markov time ν_i may be represented in the form

$$\nu_i = \inf\{t \colon L_i(t) \geq \exp(b_i)\}$$

where

$$L_i(t) = \frac{\pi_i \exp[Z_i(t)]}{\max_{\substack{0 \leq k \leq M - 1 \\ k \neq i}} \pi_k \exp[Z_k(t)]}$$

i.e., $L_i(n)$ is the generalized likelihood ratio between H_i and the remaining hypotheses.

Suppose the distributions of the observations come from exponential families, which are good models for many applications. Then the test δ_b (as defined in (2.3)) has an advantage over the first test in that it does not require exponential transformations of the observations. This fact makes it more convenient for practical realization and simulation. Furthermore, δ_b may easily be modified to meet constraints on conditional risks (see [8]–[10], [32], [34], and [35]) that may be more relevant in some practical applications. However, as we shall see in [12], δ_a has the advantage that it is easier to design the thresholds $\{a_i\}$ to precisely meet constraints on the risks $\{R_i\}$.

We begin our analysis by deriving bounds on the risks in terms of the thresholds. This allows us to choose the thresholds in such a way as to guarantee that the tests belong to the class $\Delta(\overline{R})$ defined in (2.1). We emphasize that the bounds hold

under general assumptions, and require neither independence nor homogeneity of the observed data. The proof is given in the Appendix (see also [1], [3], [11], [32], [34], [35], and [38] for related results).

Lemma 2.1: Let $\{X_t, t \in \mathcal{R}\}$ be an arbitrary random process observed either in discrete or continuous time. For all $i = 0, 1, \cdots, M - 1$

$$R_i(\delta_a) \leq \pi_i \exp(-a_i) \qquad R_i(\delta_b) \leq \pi_i \sum_{\substack{j=0 \\ j \neq i}}^{M-1} \exp(-b_{ij}).$$

Corollary 2.1: Let

$$a_i = \log(\pi_i/\overline{R}_i) \qquad b_{ij} = b_i = \log[(M-1)\pi_i/\overline{R}_i]. \quad (2.4)$$

Then both tests belong to the class $\boldsymbol{\Delta}(\overline{R})$.

It should be pointed out that the following implication holds:

$$b_i = a_i + \log(M-1) \Longrightarrow \nu_{b-\log(M-1)} \leq \tau_a \leq \nu_b. \quad (2.5)$$

To show (2.5) we rewrite ν_i and τ_i as

$$\nu_i = \inf \left\{ t: \min_{j \neq i}[Z_{ij}(t) - \log w_{ji}] \geq a_i + \log(M-1) \right\}$$

and

$$\tau_i = \inf \left\{ t: \min_{j \neq i}[Z_{ij}(t) - \log w_{ji}] \right.$$
$$\geq a_i + \log \left(\sum_{j \neq i} w_{ji} e^{Z_j(t)} \right)$$
$$\left. - \log(\max_{j \neq i} w_{ji} e^{Z_j(t)}) \right\}.$$

Then, inequality (2.5) follows from the fact that

$$0 \leq \log \left(\sum_{j \neq i} w_{ji} e^{Z_j(t)} \right) - \log \left(\max_{j \neq i} w_{ji} e^{Z_j(t)} \right)$$
$$\leq \log(M-1).$$

A consequence of (2.5) is that if the thresholds are chosen according to (2.4), then $\tau_a \leq \nu_b$.

III. LOWER BOUNDS FOR MOMENTS OF THE STOPPING TIME DISTRIBUTION

The following theorem is of fundamental importance for proving asymptotic optimality. It establishes the potential asymptotic performance of any sequential or nonsequential multihypothesis test in the class $\boldsymbol{\Delta}(\overline{R})$ as the risks go to zero. For convenience, we write $\overline{R}_{\max} = \max_{0 \leq i \leq M-1} \overline{R}_i$. Also the standard abbreviation P-a.s. is used for the almost sure convergence under the measure P.

Theorem 3.1: Assume there exists an increasing nonnegative function $f(t)$ and positive finite constants q_{ij}, $i \neq j$, $i, j = 0, 1, \cdots, M - 1$, such that[3]

$$\frac{1}{f(t)} Z_{ij}(t) \xrightarrow{P_i\text{-a.s.}} q_{ij},$$
$$\text{as } t \to \infty, \ i \neq j, \ i, j = 0, 1, \cdots, M - 1. \quad (3.1)$$

In addition, assume that for all $L > 0$

$$P_i \left(\sup_{t \leq L} Z_{ij}^+(t) < \infty \right) = 1. \quad (3.2)$$

Then for all $m > 0$ and $i = 0, 1, \cdots, M - 1$

$$\inf_{\delta \in \boldsymbol{\Delta}(\overline{R})} E_i \tau^m \geq \left[F\left(\frac{|\log \overline{R}_i|}{\min_{j \neq i} q_{ij}} \right) \right]^m (1 + o(1)),$$
$$\text{as } \overline{R}_{\max} \to 0 \quad (3.3)$$

where $F(\cdot)$ is the inverse function of $f(t)$, and where $o(1) \to 0$ as $\overline{R}_{\max} \to 0$.

Proof: Let $\boldsymbol{\Delta}(\|\alpha_{ij}\|)$ denote the class of tests for which $P_i(d = j) \leq \alpha_{ij}$, $0 < \alpha_{ij} < 1$, and let $\alpha_{\max} = \max_{i,j} \alpha_{ij}$. It follows from Tartakovsky [35, Lemma 2.1] that under conditions (3.1) and (3.2)

$$\lim_{\alpha_{\max} \to 0} \inf_{\delta \in \boldsymbol{\Delta}(\|\alpha_{ij}\|)} P_i \left\{ \tau > \gamma F\left(\max_{j \neq i} \frac{|\log \alpha_{ji}|}{q_{ij}} \right) \right\} = 1,$$
$$\text{for every } 0 < \gamma < 1.$$

Since $\alpha_{ji}(\delta) \leq \overline{R}_i/\pi_j W(j, i)$ for any $\delta \in \boldsymbol{\Delta}(\overline{R})$, it follows that

$$\lim_{\overline{R}_{\max} \to 0} \inf_{\delta \in \boldsymbol{\Delta}(\overline{R})} P_i\{\tau > \gamma F(|\log \overline{R}_i|/q_i)\} = 1,$$
$$\text{for every } 0 < \gamma < 1 \quad (3.4)$$

where $q_i = \min_{j \neq i} q_{ij}$.

Denoting

$$Y_i(\overline{R}) = \frac{\tau}{F(|\log \overline{R}_i|/q_i)}$$

and applying the Chebyshev inequality, we obtain

$$E_i[Y_i(\overline{R})]^m \geq \gamma^m P_i\{Y_i(\overline{R}) > \gamma\}, \qquad \text{for any } m > 0 \text{ and } \gamma > 0$$

where by (3.4)

$$\lim_{\overline{R}_{\max} \to 0} \inf_{\delta \in \boldsymbol{\Delta}(\overline{R})} P_i\{Y_i(\overline{R}) > \gamma\} = 1, \qquad \text{for every } 0 < \gamma < 1.$$

Thus

$$\varliminf_{\overline{R}_{\max} \to 0} \inf_{\delta \in \boldsymbol{\Delta}(\overline{R})} E_i[Y_i(\overline{R})]^m \geq 1, \qquad \text{for all } m > 0$$

and (3.3) follows. $\square$

IV. ASYMPTOTIC OPTIMALITY OF THE MSPRT'S

A. The Case of i.i.d. Observations

So far we considered quite a general case imposing only minor restrictions on the structure of the observed process. In this subsection we consider the discrete-time case ($t = n = 1, 2, \cdots$) and assume that, under hypothesis H_i, the random

---

[3]Note that (3.l) is nothing but the Strong Law of Large Numbers for the normalized LLR's $Z_{ij}(t)/f(t)$. The function $f(t)$ characterizes the degree of nonhomogeneity of the LLR's.

variables X_1, X_2, $\cdots$ are i.i.d. with a density $f_i(x)$ (relative to some sigma-finite measure) and that the densities f_k and f_j are distinct for all $k \neq j$. In other words, these densities do not coincide with probability one (w.p. 1) in the sense that $P_i(Z_{ij}(1) = 0) < 1$ for all i and all $j \neq i$.

In general, the X's will represent l-valued random vectors, i.e., $X_n = (X_{1,n}, \cdots, X_{l,n})$. Let us define $\Delta Z_{ij}(n)$ to be the log-likelihood ratio of the individual observation X_n, i.e.,

$$\Delta Z_{ij}(n) = \log \frac{f_i(X_n)}{f_j(X_n)} \quad \text{and} \quad Z_{ij}(n) = \sum_{k=1}^{n} \Delta Z_{ij}(k).$$

Further, we define

$$D_{ij} = E_i \Delta Z_{ij}(n) \qquad D_i = \min_{j \neq i} D_{ij}.$$

Note that the D_{ij} are the Kullback–Leibler (KL) information numbers ("distances"), and D_i is a minimal distance between the hypothesis H_i and other hypotheses. Due to the aforementioned condition of distinctness of measures, these KL distances are strictly positive. We shall also assume that $D_{ij} < \infty$.

In what follows, we will consider a general, asymmetric case (with respect to risk constraints) where it is assumed that the numbers $\overline{R}_k$ approach zero in such a way that for all i, j

$$\frac{\log \overline{R}_i}{\log \overline{R}_j} = c_{ij}, \qquad 0 < c_{ij} < \infty. \tag{4.1}$$

That is, we assume that the ratios $\log \overline{R}_i / \log \overline{R}_j$ (or more generally b_{ij}/b_{ks} and a_i/a_j) are bounded away from 0 and ∞. This guarantees that any $\overline{R}_i$ does not go to zero at an exponentially faster (slower) rate than any other $\overline{R}_j$. Note that we do not require that the $\overline{R}_i$'s go to zero at the same rate—if this were the case all the c_{ij}'s would equal 1. The reason for allowing c_{ij}'s other than 1 is that there are many interesting problems for which the risks for the various hypotheses may be orders of magnitude different.

The following lemma establishes the asymptotic lower bounds for the moments of the stopping time in the i.i.d. case. It is then used to prove the asymptotic optimality of the analyzed tests in the class (2.1) as $\overline{R}_{\max} \to 0$.

Lemma 4.1: If X_n, $n = 1, 2, \cdots$ are i.i.d. random vectors and $0 < D_{ij} < \infty$, then for any $m > 0$ and $i = 0, 1, \cdots, M-1$

$$\inf_{\delta \in \Delta(\overline{R})} E_i \tau^m \geq \left(\frac{|\log \overline{R}_i|}{D_i} \right)^m (1 + o(1)), \qquad \text{as } \overline{R}_{\max} \to 0. \tag{4.2}$$

Proof: In the i.i.d. case the condition $E_i|Z_{ij}(1)| < \infty$ is sufficient for conditions (3.1) and (3.2) to be satisfied with $f(n) = n$ and $q_{ij} = D_{ij}$. Thus the lemma follows directly from Theorem 3.1. $\square$

Now in order to prove first-order asymptotic optimality we have only to show that the lower bounds (4.2) are achieved for the considered tests. The details of the proof are given in the Appendix. The following theorem summarizes the main results on the asymptotic performance and the asymptotic optimality

of the MSPRT's with respect to any positive moment of the stopping time distribution. Everywhere below $x_\gamma \sim y_\gamma$ as $\gamma \to \gamma_0$ means that $\lim_{\gamma \to \gamma_0} (x_\gamma/y_\gamma) = 1$.

Theorem 4.1: Let $0 < D_{ij} < \infty$.

i) For all $m \geq 1$ and $i = 0, 1, \cdots M - 1$

$$E_i \tau_a^m \sim (a_i/D_i)^m, \qquad \text{as } a_{\min} \to \infty$$
$$E_i \nu_b^m \sim \max_{j \neq i} (b_{ij}/D_{ij})^m, \qquad \text{as } b_{\min} \to \infty \tag{4.3}$$

where $a_{\min} = \min_i a_i$ and $b_{\min} = \min_{i,j} b_{ij}$.

ii) If the thresholds are determined by (2.4), then

$$\inf_{\delta \in \Delta(\overline{R})} E_i \tau^m \sim E_i \tau_a^m \sim E_i \nu_b^m \sim \left(\frac{|\log \overline{R}_i|}{D_i} \right)^m,$$
$$\text{as } \overline{R}_{\max} \to 0 \text{ for all } m \geq 1. \tag{4.4}$$

For $m = 1$, the asymptotic formulas of (4.3) describe the behavior of the first term of expansion of the mean sample size. The behavior of the second term remains unclear—it may tend to infinity or may be of the order of a constant. In the companion paper [12], we derive higher order approximations to the expected sample size up to a vanishing term. These approximations require further conditions on the higher moments of the LLR's. As shown in [12], in a special asymmetric case, the second term is of the order of $O(1)$ (a constant) whenever the second moment of LLR's is finite. In general, however, the second term is shown to tend to infinity as a square root of the threshold.

Remark 4.1: Theorem (4.1) remains valid in the continuous-time case where the LLR's are random processes with independent and stationary increments with finite first absolute moment.

Remark 4.2: We emphasize that the tests considered asymptotically minimize all positive moments of the observation time, under only the condition of positiveness and finiteness of the KL distances—finiteness of higher order moments is not required. The required condition holds in most practical problems. In particular, in problems of signal detection, all that is required is that the signal-to-noise ratio (SNR) be finite and positive.

Remark 4.3: In the "symmetric" case (with respect to risks) where

$$\frac{\log \overline{R}_i}{\log \overline{R}_j} \sim 1, \qquad \text{for all } i, j; i \neq j, \text{ as } \overline{R}_{\max} \to 0 \tag{4.5}$$

(cf. (4.1)), a much stronger (near optimality) result is true for the expected observation time. More specifically, if thresholds are chosen so that $R_i(\delta_b) \sim \overline{R}_i$ and (4.5) is fulfilled, then

$$E_i \nu_b = \inf_{\delta \in \Delta(\overline{R})} E_i \tau + o(1), \qquad \text{as } \overline{R}_{\max} \to 0 \tag{4.6}$$

and the same is, of course, true for the test δ_a. This fact may be derived by using the results of Lorden [25]. We conjecture that (4.6) is also true for the more general case given in (4.1). We do not have a rigorous proof, but simulation results presented in [12] support our conjecture.

Remark 4.4: Since the two proposed MSPRT's have distinctly different structures, and are both asymptotically optimal, it is natural to ask if this asymptotic optimality is shared by a host of other tests. Clearly, since we only have first-order optimality, simple modifications of the proposed MSPRT's (such as adding a constant delay) preserve the optimality property. However, it is interesting to see if some other seemingly reasonable test with a very different structure will also have this property. As an example, consider the following. For simplicity, let $\overline{R}_i = \overline{R}$ for all $i = 0, 1, \cdots, M - 1$ and let $\delta_{\mathrm{ML}} = (\tau_{\mathrm{ML}}, d_{\mathrm{ML}})$ be the maximum-likelihood test procedure of the form

$$\tau_{\mathrm{ML}} = \inf\{t\colon \max_{0 \leq k \leq M-1} Z_k(t) \geq c_R\}$$

$$d_{\mathrm{ML}} = i, \qquad \text{if } \max_{0 \leq k \leq M-1} Z_k(\tau_{\mathrm{ML}}) = Z_i(\tau_{\mathrm{ML}})$$

where the threshold c_R is chosen so that $R_i(\delta_{\mathrm{ML}}) \leq \overline{R}$. Note that this test is a natural extension of the optimum fixed sample size test under a constraint on the Bayes risk $\sum_i R_i$. Now, using the results of Sosulin and Fishman [30, pp. 179–182] it can be proved that in a symmetric i.i.d. Gaussian case

$$\frac{\mathrm{E}_i \tau_{\mathrm{ML}}}{\inf_{\delta \in \boldsymbol{\Delta}(\overline{\boldsymbol{R}})} \mathrm{E}_i \tau} \sim 2, \qquad \text{as } \overline{R} \to 0.$$

That is, the seemingly reasonable maximum-likelihood test is not asymptotically optimum.

B. The Non-i.i.d. Case

Many results in sequential analysis, including the ones that we obtained above, are based on the random walk structure of LLR's. In this subsection we study the situation where the LLR's lose this nice property, and are general, possibly continuous-time, processes. The motivation for this study is as follows. The random walk structure is not valid in many practical applications where observations are nonhomogeneous and/or correlated. In addition, there is an important class of problems with nuisance parameters that admit invariant solutions. Invariant LLR's are not random walks even in the simplest cases where the models generate i.i.d. observations.

Note that the derivation of lower bounds for $\mathrm{E}_i \tau^m$ in Theorem 3.1 (see (3.3)) required only the Strong Law of Large Numbers

$$\frac{1}{f(t)} Z_{ij}(t) \to q_{ij}, \qquad \text{w.p. 1 as } t \to \infty$$

with some increasing function $f(t)$ where q_{ij} play the role of KL distances in this case. However, except in the i.i.d. case, this condition does not even guarantee finiteness of the moments of the stopping time. To obtain the asymptotics for moments of the stopping times τ_a and ν_b in the general non-i.i.d. case, we hence need to strengthen the convergence condition.

Let $\{\xi_t, t \in \mathcal{R}\}$ be a random process, and let $h > 0$. Further, let

$$T(h) = \sup\{t \in \mathcal{R}\colon |\xi_t - q| \geq h\} \tag{4.7}$$

denote the last entry time of ξ_t in the set $(-\infty, q - h] \cup [q + h, +\infty)$. The Strong Law "$\xi_t \to q$ P-a.s. as $t \to \infty$"

can be expressed in terms of $T(h)$ as $\mathrm{P}(T(h) < \infty) = 1$ for all positive h. In other words, $\xi_t \to q$ w.p. 1 if and only if $\mathrm{P}(T(h) < \infty) = 1$ for all positive h.

The following is the strengthening of the Strong Law to the so-called r-quick version (see, e.g., Lai [23] in the discrete-time case).

Definition: Let $\{\xi_t, t \in \mathcal{R}\}$ be a random process, and let $T(h)$ be as defined in (4.7). For $r > 0$, ξ_t is said to *converge r-quickly* under measure P to q if and only if

$$\mathrm{E}[T(h)]^r < \infty, \qquad \text{for all } h > 0$$

where E denotes expectation with respect to P. If the corresponding r-quick convergence condition holds for any positive r, we say that the process $\{\xi_t\}$ converges *strongly completely*[4] to q.

Let $f(t)$ be an increasing nonnegative function with $F(t)$ being the inverse function. In many practical applications $f(t) = t^\lambda$, $\lambda > 0$ (see, e.g., examples in Section VI). We now strengthen the a.s. convergence condition (3.1) in Theorem 3.1 into the following r-quick convergence for some positive r:

$$\frac{1}{f(t)} Z_{ij}(t) \xrightarrow{\mathrm{P}_i\text{-}r\text{-quickly}} q_{ij},$$

$$i \neq j, \ i, j = 0, 1, \cdots, M - 1. \tag{4.8}$$

The following theorem establishes the asymptotic optimality result for general statistical models (without the i.i.d. assumption) when the r-quick convergence condition (4.8) holds.

Theorem 4.2: Let the condition (4.8) hold for some $r > 0$. Then the following three assertions are true.

i) $\mathrm{E}_i \tau_a^r < \infty$, $\mathrm{E}_i \nu_b^r < \infty$, $i = 0, 1, \cdots, M - 1$, for any finite $\{a_i\}$ and $\{b_{ij}\}$.

ii) In addition, if the condition (3.2) holds, then for all $i = 0, 1, \cdots, M - 1$ and $m \leq r$

$$\mathrm{E}_i \tau_a^m \sim \left[F\left(\frac{a_i}{q_i}\right) \right]^m, \qquad \text{as } a_{\min} \to \infty \tag{4.9}$$

$$\mathrm{E}_i \nu_b^m \sim \left[F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right) \right]^m, \qquad \text{as } b_{\min} \to \infty \tag{4.10}$$

where $q_i = \min_{j \neq i} q_{ij}$, $a_{\min} = \min_i a_i$, and $b_{\min} = \min_{i,j} b_{ij}$.

iii) If $a_i = \log(\pi_i / \overline{R}_i)$, $b_{ij} = \log[(M - 1)\pi_i / \overline{R}_i]$, and the condition (4.11) holds, then both tests belong to the class $\boldsymbol{\Delta}(\overline{\boldsymbol{R}})$, and for all $m \leq r$ and $i = 0, 1, \cdots, M - 1$

$$\inf_{\delta \in \boldsymbol{\Delta}(\overline{\boldsymbol{R}})} \mathrm{E}_i \tau^m \sim \mathrm{E}_i \tau_a^m \sim \mathrm{E}_i \nu_b^m \sim \left[F\left(\frac{|\log \overline{R}_i|}{q_i}\right) \right]^m,$$

$$\text{as } \overline{R}_{\max} \to 0. \tag{4.11}$$

In particular, if $f(t) = t^\lambda$ with $\lambda > 0$, then

$$\inf_{\delta \in \boldsymbol{\Delta}(\overline{\boldsymbol{R}})} \mathrm{E}_i \tau^m \sim \mathrm{E}_i \tau_a^m \sim \mathrm{E}_i \nu_b^m \sim \left(\frac{|\log \overline{R}_i|}{q_i}\right)^{m/\lambda},$$

$$i = 0, 1, \cdots, M - 1. \tag{4.12}$$

---

[4]This definition of strong complete convergence should not be confused with that of complete convergence (which is equivalent to 1-quick convergence) given in Hsu and Robbins [20].

Note that in [35], Tartakovsky established a result similar to (4.11) for the MSPRT δ_b but for the situation where constraints are imposed on the conditional error probabilities $P_i(d \neq i)$. The asymptotics (4.9) and (4.10) are new and perhaps the most useful, since they allow us to obtain asymptotics under a variety of constraints.

The rigorous proof of Theorem 4.2 is given in the Appendix. Here we present a heuristic argument that leads to this theorem. We focus our attention on the test δ_b, since, as is clear from (2.5), the corresponding asymptotics for δ_a follow from those for δ_b.

Introduce the following notation:

$$\tilde{b}_{ij} = q_{ij}^{-1}[b_{ij} + \log w_{ji}]; \quad Y_{ij}(t) = q_{ij}^{-1} Z_{ij}(t)$$

$$\tilde{\nu}_i = \inf \left\{ t \in \mathcal{R} \colon \min_{j \neq i} Y_{ij}(t) \geq \max_{j \neq i} \tilde{b}_{ij} \right\}.$$

Note that the condition (4.8) implies in particular that the mean value of the LLR $Z_{ij}(t)$ is asymptotically equal to $q_{ij} f(t)$, as $t \to \infty$. Thus we may expect that for the large b_{ij}

$$\min_{j \neq i}[f(\tilde{\nu}_i) + L_{ij}(\tilde{\nu}_i)] \approx \max_{j \neq i} \tilde{b}_{ij}$$

where $L_{ij}(t) = Y_{ij}(t) - f(t)$ is the stochastic part of $Y_{ij}(t)$. Now, if fluctuations of $Z_{ij}(t)$ are not too large, which is expressed in terms of P_i-r-quick convergence $L_{ij}(t)/f(t) \to 0$, then it may be expected that the stochastic part $L_{ij}(\tilde{\nu}_i)$ is substantially less than $f(\tilde{\nu}_i)$ (in average). Therefore,

$$f(\tilde{\nu}_i) \approx \max_{j \neq i} \tilde{b}_{ij}$$

which for large b_{ij} and $r > 0$ gives

$$E_i \tilde{\nu}_i^r \approx \left[F\left(\max_{j \neq i} \tilde{b}_{ij} \right) \right]^r.$$

Since $\nu_b \leq \tilde{\nu}_i$, we also expect

$$E_i \nu_b^r \lesssim \left[F\left(\max_{j \neq i} \tilde{b}_{ij} \right) \right]^r.$$

On the other hand, if we choose the numbers b_{ij} as in (2.4), the right-hand side of the latter inequality is the first term of expansion in the lower bound for $E_i \nu_b^r$ in the corresponding class (see Theorem 3.1). Thus we expect that the tests δ_a and δ_b are asymptotically optimal when the error probabilities approach zero.

Corollary 4.1: If the normalized LLR's $Z_{ij}(t)/f(t)$ converge strongly completely to q_{ij} under P_i, then the asymptotic relationships (4.9)–(4.11) hold for all $m > 0$, and hence the test procedures δ_a and δ_b minimize any positive moment of the stopping time distribution.

It should be pointed out that since $T_{ij}^{(f)}(h)$ is not a Markov time, it is usually not easy to check the r-quick convergence condition (4.8). The following condition implies (4.8) and is sometimes useful. If, for any $\varepsilon > 0$

$$\int_0^\infty t^{r-1} P_i \left\{ \left| \sup_{0 \leq u \leq t} [Z_{ij}(u) - q_{ij}f(u)] \right| > \varepsilon f(t) \right\} dt < \infty \tag{4.13}$$

then (4.8) is true. The integral in (4.13) is replaced by a sum in the discrete-time case. (See Chow and Lai [7] for some related inequalities and conditions in the i.i.d. case, and Tartakovsky [35] for the general case.) This condition will be checked in a number of examples in Section VI.

Remark 4.5: In the i.i.d. case, the condition $E_i|Z_{ij}(1)|^{r+1} < \infty$ is necessary and sufficient for the r-quick convergence of $t^{-1} Z_{ij}(t) \to D_{ij}$ (see (4.13) and [7]). Therefore, Theorem 4.2 implies the asymptotic formulas (4.4) in Theorem 4.1 for $m \leq r$ under conditions $E_i|Z_{ij}(1)|^{r+1} < \infty$. Since Theorem 4.1 holds for any $m \geq 1$ under the weaker condition $E_i|Z_{ij}(1)| < \infty$, this shows that Theorem 4.1 does not follow from Theorem 4.2.

V. SLIPPAGE PROBLEMS

In this section we apply the general results obtained above to a specific multiple decision problem called the *slippage problem* (see Ferguson [14], Mosteller [27]). Suppose there are N populations whose distribution functions $F(x - \theta_1), \cdots, F(x - \theta_N)$ are identical except possibly for different shifts $\theta_1, \cdots, \theta_N$. On the basis of samples from the N populations we wish to decide whether or not the populations are equal or one of these populations has slipped to the right of the rest and, if so, which one (e.g., which is the "odd" one). In other words, we may ask whether or not for some i, we have $\theta_i > \theta_k = \theta, \forall k \neq i$. In the language of hypothesis testing we want to test the null hypothesis "$H_0: \theta_1 = \theta_2 = \cdots = \theta_N = \theta$" against N alternatives "$H_i: \theta_i = \theta + \Delta$," $i = 1, \cdots, N$, where $\Delta \neq 0$. Thus in this case $M = N + 1$. The slippage problem is of considerable practical importance and closely related to the so-called ranking and selection problem in which the goal of an experimenter is to select the best population [4], [17].

A problem of this kind was first discussed by Mosteller [27] in a nonsequential setting for the nonparametric case when both the form of distribution function $F(\cdot)$, and values of θ and Δ are unknown. Ferguson [14] generalized this result for the case of arbitrary (but known) distributions $F_0(x)$ and $F_1(x)$ (not necessarily just with different means), i.e., when, under hypothesis H_0, all N populations have the same distribution $F_0(\cdot)$ and, under H_i, the ith representative has some different distribution $F_1(\cdot)$. Tartakovsky [33] considered the case of possibly different (and unknown) distribution functions $F_i(\cdot)$, $i = 1, \cdots, N$, in a minimax (again nonsequential) setting. As a result, the minimax-invariant solution to this problem has been obtained.

Another interesting application of this model is in signal (target) detection in a multichannel (multiresolution) system. There may be no useful signal at all (hypothesis H_0) or a signal may be present in one of the N channels, in the ith, say (hypothesis H_i). It is necessary to detect a signal as soon as possible and to indicate the number of the channel where the signal is located. This important practical problem will be emphasized in the subsequent examples. Moreover, we will consider a more general case of possibly correlated and nonidentically distributed observations in each population

while populations will be assumed statistically independent of each other.

To be specific, let $\boldsymbol{X}_t = (X_{1,t}, \cdots, X_{N,t})$, $N \geq 2$, be an N-component process observed either in discrete or continuous time. The component $X_{j,t}$ corresponds to the observation in the jth channel. It is assumed that all components may be observed simultaneously and may have a fairly general structure. We also suppose that they are mutually independent. For the problem under consideration, the following three-valued loss function is appropriate:

$$W(j,i) = \begin{cases} W_0, & \text{for } j = 0, \ i = 1, \cdots, N \\ W_1, & \text{for } j = 1, \cdots, N, \ i = 0 \\ W_2, & \text{for } j = 1, \cdots, N, \ i = 1, \cdots, N, \ j \neq i \\ 0, & \text{otherwise.} \end{cases}$$

$$(5.1)$$

That is, we assume that the losses associated with false alarms, missing the signal, and choosing the wrong signal are, respectively, given by W_0, W_1, and W_2. The decision risks are then given by

$$R_0(\delta) = W_1 \cdot \sum_{j=1}^{N} \pi_j \alpha_{j0}(\delta) \qquad (5.2)$$

$$R_i(\delta) = W_2 \cdot \sum_{\substack{j=1 \\ j \neq i}}^{N} \pi_j \alpha_{ji}(\delta) + \pi_0 W_0 \alpha_{0i}(\delta), \qquad i = 1, \cdots, N$$

$$(5.3)$$

where $\alpha_{ji}(\delta) = \mathrm{P}_j(d = i)$, $j \neq i$, are the corresponding error probabilities.

Consider a commonly used additive model [2], [30], [32] where an observed process in the jth channel represents either the additive mixture of a useful signal $S_{j,t}$ with a noise $\xi_{j,t}$ or only noise

$$X_{j,t}^{(i)} = \begin{cases} S_{j,t} + \xi_{j,t}, & \text{if } i = j, \ j = 1, \cdots, N \\ \xi_{j,t}, & \text{if } i \neq j, \ i = 0, 1, \cdots, N \end{cases} \qquad (5.4)$$

where the superscript i means that the process $X_{j,t}$ is regarded under H_i, i.e., when a signal is present in the ith channel. In general, the signal $S_{j,t}$ could be random and its structure may be different for the various channels.

Since $Z_{ij}(t) = Z_i(t) - Z_j(t)$ and

$$Z_\ell(t) := \log \frac{d\mathrm{P}_\ell^t}{d\mathrm{P}_0^t}(\boldsymbol{X}^t) = \log \frac{d\mathrm{P}_\ell^t}{d\mathrm{P}_0^t}(X_\ell^t)$$

the LLR $Z_{ij}(t)$ depends on the observation process $\boldsymbol{X}_t$ through only the components $X_{i,t}$ and $X_{j,t}$.

In order to apply general results of Section IV, we have to check the validity of the required conditions. This will be done in the next section for several specific examples.

VI. EXAMPLES

In this section we consider examples which are meaningful for many applications including target detection in multiple-resolution radar [26] and infrared systems [13], signal acquisition in direct sequence code-division multiple-access

systems [37], and statistical pattern recognition [16]. The first is a discrete-time but non-i.i.d. example, the second is a continuous-time example, and the third includes nonparametric models and invariant sequential tests. The main goal in these examples is to show that the conditions required for our results, particularly in the non-i.i.d. cases, are reasonable and are usually met in practice. Detailed numerical and simulation results for the i.i.d. case are presented in the companion paper [12].

Example 1: Discrete-time detection/identification of a deterministic signal in the presence of correlated noise in a multichannel system. Assume that the functions $S_{1,n}, \cdots, S_{N,n}$ in (5.4) are deterministic and the noise processes $\xi_{i,n}$, $n = 0, 1, \cdots$, are stable first-order autoregressive Gaussian processes, i.e.,

$$\xi_{i,n} = \beta \xi_{i,n-1} + \zeta_{i,n}, \qquad n \geq 1, \ \xi_{i,0} = 0$$

where $\zeta_{i,1}, \zeta_{i,2}, \cdots$ are i.i.d. Gaussian variables with zero mean and variance σ^2 ($\zeta_{i,n}$ and $\zeta_{j,n}$ are independent), and $|\beta| < 1$. It is easy to show that the LLR's are of the form

$$Z_{ij}(n) = \frac{1}{\sigma^2} \sum_{k=1}^{n} \tilde{S}_{i,k} \tilde{X}_{i,k} - \frac{1}{\sigma^2} \sum_{k=1}^{n} \tilde{S}_{j,k} \tilde{X}_{j,k} - \frac{1}{2\sigma^2} \sum_{k=1}^{n} [\tilde{S}_{i,k}^2 - \tilde{S}_{j,k}^2],$$

and hence under H_i

$$Z_{ij}^{(i)}(n) = \frac{1}{2\sigma^2} \sum_{k=1}^{n} [\tilde{S}_{i,k}^2 + \tilde{S}_{j,k}^2] + \frac{1}{\sigma^2} \sum_{k=1}^{n} \tilde{S}_{i,k} \zeta_{i,k} - \frac{1}{\sigma^2} \sum_{k=1}^{n} \tilde{S}_{j,k} \zeta_{j,k}.$$

Here the "tilde values" are $\tilde{X}_{i,n} = X_{i,n} - \beta X_{i,n-1}$ for $t \geq 2$ and $\tilde{X}_{i,1} = X_{i,1}$, and similarly for $\tilde{S}_{i,k}$.

We denote the accumulated SNR for channel ℓ up to time n by

$$\mu_\ell(n) = \frac{1}{\sigma^2} \sum_{k=1}^{n} \tilde{S}_{\ell,k}^2.$$

Assume that

$$\lim_{n \to \infty} n^{-\lambda} \mu_\ell(n) = q_\ell, \qquad \text{for some } \lambda > 0 \qquad (6.1)$$

where q_ℓ, $\ell = 1, \cdots, N$, are finite positive numbers. Now we will establish that

$$n^{-\lambda} Z_{ij}(n) \to (q_i + q_j)/2 \qquad \mathrm{P}_i\text{-strongly completely.}$$

To establish this it is sufficient to show that (see (4.13))

$$\sum_{n=1}^{\infty} n^{r-1} \mathrm{P}_i(|W_\ell(n)| > \varepsilon n^\lambda) < \infty,$$

$$\text{for some } \varepsilon > 0 \text{ and all } r > 0 \qquad (6.2)$$

where $W_\ell(n) = \sum_{k=1}^{n} \tilde{S}_{\ell,k} \zeta_{i,k}$. Since $W_\ell(n)$ is a weighted Gaussian sample sum with mean zero and variance $\sim q_\ell n^\lambda$

(for large n), it is easy to show that there is a number $\gamma < 1$ such that

$$\mathbf{P}_i(|W_\ell(n)| > \varepsilon n^\lambda) \le O(\gamma^n)$$

and hence (6.2) is fulfilled. Thus by Theorem 4.2, the asymptotic equalities (4.11) hold with $q_{ij} = (q_i + q_j)/2$, $i, j \ne 0$; $q_{i0} = q_i/2$, $q_{0j} = q_j/2$, $i, j = 1, \cdots, N$, and the tests δ_a and δ_b are asymptotically optimal.

It should be noted that in the application to signal detection, the quantity $\overline{R}_0$ is a constraint on the risk associated with missing the signal when it is actually present on one of the channels (see (5.2)). Similarly, $\overline{R}_i$, $i \ne 0$ is a constraint on the risk associated with deciding that the signal is present on channel i, which is a weighted sum of false alarm probability and the probability of incorrect classification (see (5.3)). Symmetry assumptions generally yield $\overline{R}_i = \overline{R}$ for $i = 1, \cdots, N$. Now, since typically the accumulated SNR $\mu_i(n)$ in the ith channel does not depend on the number of the channel, i.e., q_i is same for all i and equal to ρ (say), the asymptotic formulas become

$$\mathbf{E}_i \nu_b^m \sim \mathbf{E}_i \tau_a^m \sim \left(\frac{2}{\rho} |\log \overline{R}|\right)^{m/\lambda},$$
$$i = 1, \cdots, N, \quad m > 0 \qquad (6.3)$$

$$\mathbf{E}_0 \nu_b^m \sim \mathbf{E}_0 \tau_a^m \sim \left(\frac{2}{\rho} |\log \overline{R}_0|\right)^{m/\lambda}, \qquad m > 0. \qquad (6.4)$$

In particular, if $S_{i,t} = \theta$, then it is easy to see that $\lambda = 1$ works in (6.1), as well as in (6.3) and (6.4). It is also easy to see that the corresponding SNR is $q_i = \rho = \theta^2(1 - \beta)^2/\sigma^2$. In this case, the expected stopping times are proportional to the risk constraints and inversely proportional to the SNR ρ.

Now, from the analysis of Section IV-B, it is not clear if condition (4.8) is necessary to guarantee the corresponding optimality of the tests. In the following, we show that if (4.8) is not fulfilled, the sequential tests considered may have infinite moments. At the same time the best fixed sample size test has a finite sample size.

Instead of (6.1), which postulates the growth of the SNR as a power of n for sufficiently large n, suppose the following condition holds:

$$\log(\log n)^{-1} \mu_\ell(n) \sim q_\ell, \qquad \text{as } n \to \infty \qquad (6.5)$$

i.e., the SNR grows as $\log n$ for sufficiently large n. Since this "law" is too slow, one may expect that r-quick convergence does not hold. Indeed, it is easy to see that

$$\frac{1}{\log(1+n)} Z_{ij}(n) \xrightarrow{P_i\text{-a.s.}} q_{ij} = (q_i + q_j)/2$$

but not r-quickly (for any $r \ge 1$), since for sufficiently small ε

$$\mathbf{E}_i T_{ij}^{(f)}(\varepsilon) \ge \int_0^\infty \mathbf{P}_i\{|W_\ell(t)| > \varepsilon \ln(1+t)\}\, dt$$
$$= 4 \int_0^\infty v(e^{v^2} - 1)\Phi(-\varepsilon v/\sqrt{q_m})\, dv = \infty$$

where, as above in (6.2)

$$W_\ell(n) = \sum_{k=1}^n \tilde{S}_{\ell,k} \zeta_{i,k}$$

and $W_\ell(t) = W_\ell(n)$ for $n \le t < n+1$, and where $\Phi(y)$ is the standard normal distribution function. Thus $\mathbf{E}_i T_{ij}^{(f)}(h) = \infty$ and Theorem 4.2 cannot be applied. In fact, in this case $\mathbf{E}_i \nu = \infty$ for sufficiently small $\overline{R}$ (actually for $\overline{R} < 0.305$, see, e.g., Golubev and Khas'minskii [18]).

Example 2: Slippage problem for nonhomogeneous Poisson process. In applications involving infrared and optical warning systems, an appropriate model for noise and clutter is a point random process. Below we consider a multichannel system in which the noise process is a nonhomogeneous Poisson process and target appearance leads to a change in the intensity of this process. Specifically, under hypothesis H_0 (target absent), let the observed process $\boldsymbol{X}_t = (X_{1,t}, \cdots, X_{N,t})$, $t \ge 0$, be a vector nonstationary Poisson random process with independent components each of which has intensity $\gamma_0(t)$ while, under H_i (target is located in the ith channel), the ith component has the intensity $\gamma_i(t) \ne \gamma_0(t)$. Then

$$Z_{ij}(t) = \int_0^t \log\left(\frac{\gamma_i(u)}{\gamma_0(u)}\right) dX_{i,u}$$
$$- \int_0^t \log\left(\frac{\gamma_j(u)}{\gamma_0(u)}\right) dX_{j,u} - \Lambda_{ij}(t), \qquad t \ge 0$$

$$\mathbf{E}_i Z_{ij}(t) = \int_0^t \gamma_i(u) \log\left(\frac{\gamma_i(u)}{\gamma_0(u)}\right) du$$
$$- \int_0^t \gamma_0(u) \log\left(\frac{\gamma_j(u)}{\gamma_0(u)}\right) du - \Lambda_{ij}(t)$$

$$\mathrm{Var}_i[Z_{ij}(t)] = \int_0^t \gamma_i(u)\left(\log \frac{\gamma_i(u)}{\gamma_0(u)}\right)^2 du$$
$$+ \int_0^t \gamma_0(u)\left(\log \frac{\gamma_j(u)}{\gamma_0(u)}\right)^2 du$$

where

$$\Lambda_{ij}(t) = \int_0^t [\gamma_i(u) - \gamma_j(u)]\, du.$$

Note that the LLR's $Z_{ij}(t)$ are processes with independent but (generally) nonstationary increments

$$\Delta Z_{ij}(t) = Z_{ij}(t) - \lim_{s \uparrow t} Z_{ij}(s).$$

Now assume that $\gamma_i(t)$ is a power function, $\gamma_i(t) = Q_i t^{\lambda-1}$, where $\lambda > 0$ and $Q_i > 0$. Then

$$\mathbf{E}_i Z_{ij}(t) = \frac{1}{\lambda}\left[Q_i \log \frac{Q_i}{Q_0} - Q_0 \log \frac{Q_j}{Q_0} - (Q_i - Q_j)\right] t^\lambda$$

$$\mathrm{Var}_i[Z_{ij}(t)] = \frac{1}{\lambda}\left[Q_i\left(\log \frac{Q_i}{Q_0}\right)^2 + Q_0\left(\log \frac{Q_j}{Q_0}\right)^2\right] t^\lambda$$

and the increments $\Delta Z_{ij}(t)$ are bounded as

$$|\Delta Z_{ij}(t)| \le \log \frac{Q_i}{Q_0} + \log \frac{Q_j}{Q_0}.$$

Thus $t^{-\lambda} Z_{ij}(t)$ converges P_i-r-quickly for all positive r (i.e., P_i-strongly completely) to the numbers

$$q_{ij} = \frac{1}{\lambda}\left[Q_i \log \frac{Q_i}{Q_0} - Q_0 \log \frac{Q_j}{Q_0} - (Q_i - Q_j)\right]. \qquad (6.6)$$

By Theorem 4.2, the MSPRT's are asymptotically optimal relative to any positive moment of the stopping time distribution. Note that $f(t) = t^\lambda$ in this case and hence (4.12) applies.

If $Q_i = Q_1$ and $\overline{R}_i = \overline{R}$ for all $i = 1, \cdots, N$, then using (6.6), we obtain

$$q_1 = q_{10} = \frac{1}{\lambda} \left[Q_1 \log \frac{Q_1}{Q_0} - Q_1 + Q_0 \right]$$

and

$$q_0 = q_{01} = \frac{1}{\lambda} \left[Q_1 - Q_0 - Q_0 \log \frac{Q_1}{Q_0} \right].$$

Thus from (4.12)

$$\mathrm{E}_i \nu_b^m \sim \mathrm{E}_i \tau_a^m \sim \left(\frac{\lambda |\log \overline{R}|}{Q_1 \log(Q_1/Q_0) - Q_1 + Q_0} \right)^{m/\lambda}$$
$$i = 1, \cdots, N,$$
$$\mathrm{E}_0 \nu_b^m \sim \mathrm{E}_0 \tau_a^m \sim \left(\frac{\lambda |\log \overline{R}_0|}{Q_1 - Q_0 - Q_0 \log(Q_1/Q_0)} \right)^{m/\lambda}.$$

Example 3: Nonparametric detection of a target in a multichannel system when training clutter data is available. So far we considered the case of simple hypotheses assuming that the measures $\mathrm{P}_0, \mathrm{P}_1, \cdots, \mathrm{P}_{M-1}$ were completely known. However, in many practical applications, the models are known only partially (parametric uncertainty) or they may even be unknown (nonparametric uncertainty). The above ideas, particularly Theorem 4.2, may be used to prove asymptotic optimality of invariant multihypothesis sequential tests for composite hypotheses "$H_i: \mathrm{P} \in \mathcal{P}_i$," $i = 0, 1, \cdots, M-1$. Here $\mathcal{P}_i$ are families of distributions, either parametric or nonparametric, and it is assumed that they can be reduced to simple ones by using the principle of invariance (for further details related to the invariance principle in testing hypotheses, see, e.g., Ferguson [14] and Lehmann [22]).[5]

In fact, Theorem 4.2 and Corollary 4.1 remain true if we use the LLR's $Z_{ij}(t)$ constructed for a maximal invariant and if the class $\Delta(\overline{R})$ includes only invariant tests with the constraints (2.1). In particular, if $Z_{ij}(t)/f(t)$ converges strongly completely to q_{ij}, then the corresponding invariant MSPRT's minimize all positive moments of the stopping time distribution in the class of invariant sequential tests with the corresponding constraints imposed on risks.

To illustrate this point, we apply Theorem 4.2 to an $(N+1)$-sample $(N \geq 1)$ nonparametric problem with Lehmann hypotheses (or proportional hazards, if we replace P by $1 - \mathrm{P}$ below) and show that the extended multialternative version of the Savage sequential test is asymptotically optimal with respect to any positive moment of the stopping time distribution.

Let $\boldsymbol{X}_n = (Y_n, X_{1,n}, \cdots, X_{N,n})$, where $\{Y_n\}_{n \geq 1}$ are i.i.d. with a continuous distribution function P_0, and independent of $\{X_{i,n}\}_{n \geq 1}$, which are i.i.d. either with P_0 or one of them (say, the kth one) has the distribution $\mathrm{P}_0^{A_k}$, where A_k are specified positive constants, $A_k \neq 1$, and P_0 is completely unknown. In other words, the hypotheses are

"$H_0: \mathrm{P}_i = \mathrm{P}_0, \qquad$ for all $i = 1, \cdots, N$"

---

[5] Another possible approach is to use adaptive tests that employ estimates of the unknown nuisance parameters.

and

"$H_i: X_{j,n} \sim \mathrm{P}_0, \qquad$ for $j \neq i$ and $X_{i,n} \sim \mathrm{P}_i = \mathrm{P}_0^{A_i}$."

An interesting application of this problem is in target detection in the N-channel system in the presence of clutter with unknown continuous distribution P_0, based on unclassified data $(X_{n,1}, \cdots X_{N,n})$ when a classified training sequence Y_n is also available. That is, it is known in advance that the clutter only generates the data Y_n.

A maximal invariant with respect to the group

$$\boldsymbol{X}_n \to (\psi(Y_n), \psi(X_{1,n}), \cdots, \psi(X_{N,n}))$$

where ψ is any continuous increasing function, is the vector of ranks of Y among $(Y, X_1, \cdots, X_N)$. Let

$$\hat{\mathrm{P}}_{n,i}(x) = n^{-1} \sum_{k=1}^n \mathbf{1}_{\{X_{i,k} \leq x\}}, \qquad i = 0, 1, \cdots, N$$

be the empirical distribution functions where $X_{0,k} = Y_k$. For the sake of simplicity consider $N = 1$. Then the LLR of the maximal invariant is [28]

$$Z(n) = \log \frac{A_1^n (2n)!}{n^{2n}}$$
$$- \sum_{k=1}^n \log \left([\hat{\mathrm{P}}_{n,0}(Y_k) + A_1 \hat{\mathrm{P}}_{n,1}(Y_k)] \right.$$
$$\left. \cdot [\hat{\mathrm{P}}_{n,0}(X_{1,k}) + A_1 \hat{\mathrm{P}}_{n,1}(X_{1,k})] \right).$$

Define

$$q_0(A_1) = S(A_1, \mathrm{P}_0, \mathrm{P}_0) < 0$$
$$q_1(A_1) = S(A_1, \mathrm{P}_0, \mathrm{P}_0^{A_1}) > 0$$
$$S(A_1, \mathrm{P}_0, \mathrm{P}_1) = \log(4A_1) - 2 - \int \log[\mathrm{P}_0(x) + A_1 \mathrm{P}_1(x)]$$
$$\cdot (d\mathrm{P}_0(x) + d\mathrm{P}_1(x)).$$

By [28, Lemma 2], for any $\varepsilon > 0$ there exists a number $\rho < 1$ such that

$$\mathrm{P}_i(|n^{-1} Z(n) - q_i(A_1)| > \varepsilon) \leq O(\rho^n), \qquad i = 0, 1$$

which obviously implies the strong complete convergence of $n^{-1} Z(n)$ to $q_i(A_1)$ under P_i.

Thus based on Theorem 4.2, one may conclude that Savage's nonparametric sequential rank-order test asymptotically minimizes all the moments of the stopping time distribution in the class of invariant tests when the risks $\overline{R}_0$ and $\overline{R}_1$ approach zero. A similar result is valid for the more general case $N \geq 2$ when $\overline{R}_i \to 0$.

VII. CONCLUSIONS

Most of the research on sequential hypothesis testing over the last fifty years, starting with the classical works of Wald [39] and Wald and Wolfowitz [40], has dealt with the case of i.i.d. observations and the optimality (or asymptotic optimality) of sequential tests relative to the expected sample size. However, in many practical applications the i.i.d. assumption does not hold. Furthermore, the behavior of higher order moments

may be of interest. The results presented in this paper show that the proposed MSPRT's are asymptotically optimal under fairly general conditions that cover nonhomogeneous and correlated stochastic processes observed either in discrete or continuous time, stochastic models with nuisance parameters, and even nonparametric models. In addition, the tests are shown to be asymptotically optimal not only with respect to the expected sample size but also with respect to any moment of the stopping time distribution. This justifies the use of the MSPRT's in a variety of applications, some of which have been described in this paper.

Throughout the analysis of the paper we assumed that the hypotheses were simple with respect to the informative parameters. If the hypotheses are composite, then an adaptive approach may be applied (see, e.g., Dragalin and Novikov [11]) to find asymptotically optimal solutions. However, we have reasons to believe that in spite of their asymptotic optimality, the corresponding adaptive tests will not perform well in practice. (The asymptotic convergence is generally too slow for such tests.) Thus a reasonable partitioning of composite hypotheses into a number of simple ones and subsequent application of the MSPRT's may be beneficial in practice.

The results presented above do not cover an important case for many applications, namely, one where there is an "indifference" zone. This important extension will be considered elsewhere.

APPENDIX

Proof of Lemma 2.1: Obviously,

$$R_i(\delta_a) = \sum_{j \neq i} \pi_j W(j, i) \int_{\{\tau_a = \tau_i, \, \tau_i < \infty\}} e^{Z_{ji}(\tau_i)} \, d\mathbf{P}_i$$

$$= \pi_i \mathrm{E}_i \left\{ \mathbf{1}_{\{\tau_a = \tau_i, \, \tau_i < \infty\}} \left[\sum_{j \neq i} w_{ji} e^{Z_{ji}(\tau_i)} \right] \right\}$$

$$\leq \pi_i e^{-a_i} \mathrm{E}_i \mathbf{1}_{\{\tau_a = \tau_i, \, \tau_i < \infty\}} \leq \pi_i e^{-a_i}$$

where $\mathbf{1}_{\{\omega\}}$ is the indicator of the event ω. Here we used the Fubini theorem and the definition of the stopping time τ_i according to which

$$\sum_{j \neq i} w_{ji} e^{Z_{ji}(\tau_i)} \leq e^{-a_i}, \qquad \text{on } \{\tau_i < \infty\}.$$

Consider the second test. For the Markov times ν_i, we have the equivalent representation

$$\nu_i = \inf \left\{ t: \min_{j \neq i} [Z_{ij}(t) - b_{ij} - \log w_{ji}] \geq 0 \right\}$$

which implies the inequality

$$e^{Z_{ij}(\nu_i)} \geq w_{ji} e^{b_{ij}}, \qquad \text{for all } j \neq i \text{ on } \{\nu_i < \infty\}.$$

Next, by Wald's likelihood ratio identity [29], [39], [41]

$$\mathrm{E}_j \mathbf{1}_{\{\nu_i < \infty\}} e^{Z_{ij}(\nu_i)} = 1.$$

Combining the last two relationships, we get

$$\pi_j W(j, i) \mathbf{P}_j(\nu_i < \infty) \leq \pi_i e^{-b_{ij}}, \qquad \text{for all } j \neq i.$$

Since the event $\{d_b = i\} = \{\nu_b = \nu_i, \, \nu_i < \infty\}$ implies the event $\{\nu_i < \infty\}$, it follows that

$$\sum_{j \neq i} \pi_j W(j, i) \mathbf{P}_j(d_b = i) \leq \pi_i \sum_{j \neq i} e^{-b_{ij}}$$

and the proof is complete. $\square$

Proof of Theorem 4.1: To prove the theorem we shall need the following result, which may be of independent interest.

Lemma A.1: Let $\boldsymbol{X}_n$, $n = 1, 2, \cdots$ be i.i.d., and suppose $0 < D_{ij} < \infty$ for all $i \neq j$. Then
   i) the stopping times τ_a and ν_b are exponentially bounded and hence $\mathrm{E}_i \tau_a^m < \infty$, $\mathrm{E}_i \nu_b^m < \infty$ for any positive finite m, a_i, and b_{ij};
  ii) the families

$$\left\{ \left(\frac{\tau_a}{a_i} \right)^m, \, a_i > 0 \right\} \quad \text{and} \quad \left\{ \left(\frac{\nu_b}{b_{ij}} \right)^m, \, b_{ij} > 0 \right\}$$

are uniformly integrable with respect to $\mathbf{P}_i$ for all $m \geq 1$.

Proof: Note first that both stopping times τ_a and ν_b do not exceed the stopping time

$$T_c = \inf \left\{ n \geq 1: \min_{j \neq i} Z_{ij}(n) > c \right\}$$

if

$$c = a_i + \log[(M-1) \max_{j \neq i} w_{ji}]$$

and

$$c = \max_{j \neq i} b_{ij} + \log(\max_{j \neq i} w_{ji})$$

for tests δ_a and δ_b, respectively. Thus to prove i) it is sufficient to prove that T_c is exponentially bounded for $0 < c < \infty$. To prove ii) we have only to prove the uniform integrability of $\{(T_c/c)^m, \, c \geq 1\}$.
   i) Obviously,

$$\mathbf{P}_i(T_c > n) \leq \mathbf{P}_i \left(\max_{j \neq i} e^{Z_{ji}(n)} \geq e^{-c} \right)$$

$$\leq \sum_{j \neq i} \mathbf{P}_i \left(\prod_{t=1}^{n} e^{p \Delta Z_{ji}(t)} \geq e^{-pc} \right).$$

Applying Markov's inequality

$$\mathbf{P}(|X| > y) \leq y^{-1} \mathrm{E}|X|$$

we obtain that for any $p > 0$

$$\mathbf{P}_i(T_c > n) \leq e^{pc} \sum_{j \neq i} \mathrm{E}_i \left[\prod_{t=1}^{n} e^{p \Delta Z_{ji}(t)} \right]$$

$$= e^{pc} \sum_{j \neq i} (\mathrm{E}_i e^{p \Delta Z_{ji}(1)})^n$$

$$\leq (M-1) e^{pc} \max_{j \neq i} \rho_j^n$$

where

$$\rho_j = \rho_j(p) = \mathrm{E}_i \left[\frac{f_j(\boldsymbol{X}_1)}{f_i(\boldsymbol{X}_1)} \right]^p.$$

Evidently, $0 < \rho_j(p) < 1$ for any $0 < p < 1$ when $0 < D_{ij} < \infty$ (in fact, $P_i(Z_{ij}(1) \neq 0) > 0$ as $D_{ij} > 0$). Thus there is a finite positive constant C and a number ρ, $0 < \rho < 1$, such that $P_i(T_c > n) \leq C\rho^n$, $n = 1, 2, \cdots$, and hence the Markov time T_c is exponentially bounded for all finite positive c, which implies the assertion i).

ii) For $k = 1, 2, \cdots$, define the random variables

$$M_k = \inf\left\{ n > M_{k-1}: \min_{j \neq i}(Z_{ij}(n) - Z_{ij}(M_{k-1})) > 0 \right\},$$
$$M_0 = 0$$
$$N_k = M_k - M_{k-1} \qquad Y_{ij}(k) = Z_{ij}(M_k) - Z_{ij}(M_{k-1}).$$

Since $Z_{ij}(0) = 0$

$$N_1 = M_1 = \inf\{n \geq 1: \min_{j \neq i} Z_{ij}(n) > 0\} = T_0$$
$$\text{and } Y_{ij}(1) = Z_{ij}(M_1).$$

It follows from the construction that $\{N_k\}_{k \geq 1}$ are i.i.d. and exponentially bounded and that $\{Y_{ij}(k)\}_{k \geq 1}$ are i.i.d. positive random variables (under P_i). Further, define the stopping time

$$\tau_c^* = \inf\left\{ n \geq 1: \min_{j \neq i} \sum_{k=1}^{n} Y_{ij}(k) \geq c \right\}.$$

It is easy to see that T_c is a sum of τ_c^* i.i.d. and exponentially bounded random variables N_k

$$T_c = M_{\tau_c^*} = \sum_{k=1}^{\tau_c^*} N_k.$$

By Theorem I.6.1 of Gut [19], the family $\{(T_c/c)^m, c \geq 1\}$ is uniformly integrable whenever $\{(\tau_c^*/c)^m, c \geq 1\}$ is uniformly integrable. Thus to prove the desired result we have only to show the uniform integrability of the latter family for all $m \geq 1$.

Since the increments of the random walks $\sum_{k=1}^{n} Y_{ij}(k)$ are positive

$$\tau_c^* = \max_{j \neq i} T_c^{(j)}$$

where

$$T_c^{(j)} = \inf\left\{ n \geq 1: \sum_{k=1}^{n} Y_{jk} \geq c \right\}.$$

We apply [19, Theorem III.7.1] to show that $\{(T_c^{(j)}/c)^m, c \geq 1\}$ is uniformly integrable under P_i for all $m \geq 1$ and hence the family $\{(\tau_c^*/c)^m, c \geq 1\}$ is uniformly integrable too (for all $m \geq 1$). This completes the proof of ii). $\square$

Proof of i) in Theorem 4.1: Consider the MSPRT δ_a, and let $a_{\min} = \min_i a_i$. By Gut [19, Theorem A.1.1], the following convergence of moments

$$E_i\left(\frac{\tau_a}{a_i}\right)^m \to \frac{1}{D_i^m}, \qquad \text{as } a_{\min} \to \infty$$

holds whenever

(C1) $E_i \tau_a^m < \infty$ for all $0 < a_i < \infty$,

(C2) the family $\{(\tau_a/a_i)^m, a_i > 0\}$ is P_i-uniformly integrable,

(C3) $(\tau_a/a_i) \to D_i^{-1}$ P_i-a.s. as $a_{\min} \to \infty$.

By Lemma A.1, Conditions (C1) and (C2) are satisfied. The a.s. convergence (C3) follows from Baum and Veeravalli [3, Theorem 5.1].

For the MSPRT δ_b Conditions (C1) and (C2) follow from Lemma A.1. The convergence

$$\left(\frac{\nu_b}{\max_{j \neq i}[b_{ij}/D_{ij}]} \right) \xrightarrow{P_i\text{-a.s.}} 1, \qquad \text{as } b_{\min} \to \infty$$

where $b_{\min} = \min_{i,j} b_{ij}$, is proved in essentially the same way as (C3).

Proof of ii) in Theorem 4.1: To prove (4.4) it suffices to show that the lower bound (4.2) is achieved for the tests δ_a and δ_b. This follows immediately by substitution of

$$a_i = \log\left(\pi_i/\overline{R}_i\right)$$

and

$$b_{ij} = \log\left[(M-1)\pi_i/\overline{R}_i\right]$$

in (4.3). This completes the proof. $\square$

Proof of Theorem 4.2: Theorem 4.2 will be proved only for the MSPRT δ_b. For the MSPRT δ_a the proof is essentially the same and is omitted. In fact, most of the results, including the asymptotic optimality of δ_a, follow immediately from (2.5).

Proof of i) and ii). Write

$$\tilde{b}_i = \max_{j \neq i}[q_{ij}^{-1}(b_{ij} + \log w_{ji})] \qquad Y_{ij}(t) = q_{ij}^{-1} Z_{ij}(t)$$
$$\tilde{\nu}_i = \inf\left\{ t \in \mathcal{R}: \min_{j \neq i} Y_{ij}(t) \geq \tilde{b}_i \right\}.$$

Let $h \in (0, 1)$ and $T_i^{(f)}(h) = \max_{j \neq i} T_{i,j}^{(f)}(h)$. By the definition of the stopping time $\tilde{\nu}_i$

$$\min_{j \neq i} Y_{ij}(\tilde{\nu}_i - 1) < \tilde{b}_i$$

and by condition (4.8)

$$\min_{j \neq i} Y_{ij}(\tilde{\nu}_i - 1) > f(\tilde{\nu}_i - 1)(1 - h)$$

on the set

$$\{\tilde{\nu}_i > T_i^{(f)}(h) + 1\}.$$

These two inequalities show that

$$\tilde{\nu}_i < 1 + F\left(\frac{\tilde{b}_i}{1-h}\right), \qquad \text{on } \{\tilde{\nu}_i > T_i^{(f)}(h) + 1\}.$$

Hence for any $h \in (0, 1)$

$$\tilde{\nu}_i < \left[1 + F\left(\frac{\tilde{b}_i}{1-h}\right)\right] \mathbf{1}_{\{\tilde{\nu}_i > 1 + T_i^{(f)}(h)\}} + [1 + T_i^{(f)}(h)]$$
$$\cdot \mathbf{1}_{\{\tilde{\nu}_i \leq 1 + T_i^{(f)}(h)\}}$$
$$\leq 1 + T_i^{(f)}(h) + F\left(\frac{\tilde{b}_i}{1-h}\right) \tag{A.1}$$

where $\mathbf{1}_{\{\omega\}}$ is the indicator of the set ω.

Since $\nu_b \leq \tilde{\nu}_i$, and since by the assumptions of the theorem

$$\mathrm{E}_i[T_i^{(f)}(h)]^r \leq (M-1) \max_{j \neq i} \mathrm{E}_i[T_{i,j}^{(f)}(h)]^r < \infty$$

it follows from (A.1) that $\mathrm{E}_i \nu_b^r < \infty$ for any positive set of thresholds b_{ij}. Thus assertion i) holds.

Furthermore, (A.1) implies

$$\mathrm{E}_i \nu_b^r \leq \left[F\left(\frac{\tilde{b}_i}{1-h}\right)\right]^r (1 + o(1)),$$

$$\text{for all } 0 < h < 1 \text{ as } b_{\min} \to \infty$$

where $b_{\min} = \min_{i,j} b_{ij}$. Letting $h \to 0$, we obtain the following upper estimate:

$$\mathrm{E}_i \nu_b^r \leq \left[F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right)\right]^r (1 + o(1)), \qquad \text{as } b_{\min} \to \infty. \tag{A.2}$$

To prove ii) for the MSPRT δ_b (the relation 4.10)) it suffices to show that the right-hand side of (A.2) is also the lower estimate for $\mathrm{E}_i \nu_b^r$. To this end, we first show that (similarly to (3.4))

$$\mathrm{P}_i\left\{\nu_b > \gamma F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right)\right\} \to 1$$

$$\text{as } b_{\min} \to \infty \text{ for every } 0 < \gamma < 1. \tag{A.3}$$

Write $\Omega_{i,s} = \{d_b = i\} \cap \{\nu_b \leq s\}$. Obviously, for any $s > 0$ and $C > 0$

$$\begin{aligned}
\mathrm{P}_j(d_b = i) &= \mathrm{E}_i\{\mathbf{1}_{\{d_b = i\}} \exp[Z_{ji}(\nu_b)]\} \\
&\geq \mathrm{E}_i\{\mathbf{1}_{\{\Omega_{i,s}, Z_{ij}(\nu_b) < C\}} \exp[-Z_{ij}(\nu_b)]\} \\
&\geq e^{-C} \mathrm{P}_i\left(\Omega_{i,s}, \sup_{t \leq s} Z_{ij}(t) < C\right) \\
&\geq e^{-C}\left\{\mathrm{P}_i\left(\Omega_{i,s}\right) - \mathrm{P}_i(\sup_{t \leq s} Z_{ij}(t) \geq C)\right\}.
\end{aligned}$$

Since

$$\mathrm{P}_i(\Omega_{i,s}) \geq \mathrm{P}_i(d_b = i) - \mathrm{P}_i(\nu_b > s),$$

$$\mathrm{P}_i(d_b = i) = 1 - \sum_{k \neq i} \mathrm{P}_i(d_b = k)$$

and since by Theorem 2.1 of Tartakovsky [35]

$$\mathrm{P}_i(d_b = j) \leq \exp(-b_{ij} - \log w_{ji})$$

we obtain

$$\begin{aligned}
\mathrm{P}_i(\nu_b > s) &\geq 1 - \sum_{k \neq i} w_{ki}^{-1} e^{-b_{ik}} - w_{ij}^{-1} e^{-b_{ji}+C} \\
&\quad - \mathrm{P}_i\left(\sup_{t \leq s} Z_{ij}(t) \geq C\right). \tag{A.4}
\end{aligned}$$

Now, set $C = c q_{ij} f(s)$ with $c > 1$. Then

$$\mathrm{P}_i\left(\sup_{t \leq s} Z_{ij}(t) \geq C\right)$$

$$= \mathrm{P}_i\left(\sup_{t \leq s} Z_{ij}(t) \geq c q_{ij} f(s)\right)$$

$$\leq \mathrm{P}_i\left(\sup_{t \leq K} Z_{ij}^+(t) + \sup_{K \leq t \leq s} Z_{ij}^+(t) \geq c q_{ij} f(s)\right)$$

$$\leq \mathrm{P}_i\left\{\sup_{t \leq K} Z_{ij}^+(t) + f(s) \sup_{K \leq t \leq s} \left(\frac{1}{f(t)} Z_{ij}^+(t) - q_{ij}\right)\right.$$

$$\left. \geq (c-1) q_{ij} f(s)\right\}$$

$$\leq \mathrm{P}_i\left(\frac{1}{f(s)} \sup_{t \leq K} Z_{ij}^+(t) + \sup_{t \geq K}\left|\frac{1}{f(t)} Z_{ij}^+(t) - q_{ij}\right|\right.$$

$$\left. \geq (c-1) q_{ij}\right).$$

By the condition (3.2), for any $\varepsilon > 0$ there is a finite w.p. 1 random variable $K_{ij}(\varepsilon)$ such that

$$\left|\frac{Z_{ij}^+(K_{ij}(\varepsilon))}{f(K_{ij}(\varepsilon))} - q_{ij}\right| \leq \varepsilon$$

(in fact, one may take $K_{ij}(\varepsilon) = T_{ij}^{(f)}(\varepsilon) + 1$) and hence

$$\mathrm{P}_i\left(\sup_{t \leq s} Z_{ij}(t) \geq c q_{ij} f(s)\right)$$

$$\leq \mathrm{P}_i\left(\frac{1}{f(s)} \sup_{t \leq K_{ij}(\varepsilon)} Z_{ij}^+(t) \geq (c-1) q_{ij} - \varepsilon\right).$$

By the conditions (4.8) and (3.2), the right-hand side of the latter inequality approaches zero when $s \to \infty$ and $c > 1 + \varepsilon/q_{ij}$. Thus

$$\lim_{s \to \infty} \mathrm{P}_i\left(\sup_{t \leq s} Z_{ij}(t) \geq c q_{ij} f(s)\right) = 0, \qquad \text{for every } c > 1.$$

Now, setting $s = s_b = F(\gamma q_{ij}^{-1} b_{ji})$ with $0 < \gamma < 1/c$ and using (A.4) and (A.5), we obtain that for all $j \neq i$

$$\begin{aligned}
\mathrm{P}_i\{\nu_b > F(\gamma q_{ij}^{-1} b_{ji})\} &\geq 1 - \sum_{k \neq i} w_{ki}^{-1} e^{-b_{ik}} - e^{-(1-\gamma c) b_{ji}} \\
&\quad - \mathrm{P}_i\left(\sup_{t \leq s_b} Z_{ij}(t) \geq c q_{ij} f(s_b)\right) \to 1, \\
&\qquad\qquad\qquad \text{as } b_{\min} \to \infty
\end{aligned}$$

which proves (A.3).

Finally, by Chebyshev's inequality

$$\mathrm{E}_i\left[\frac{\nu_b}{F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right)}\right]^r \geq \gamma^r \mathrm{P}_i\left\{\frac{\nu_b}{F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right)} > \gamma\right\}$$

$$\text{for any } r > 0, \gamma > 0$$

where by (A.3) the probability in the right-hand side tends to 1 as $b_{\min} \to \infty$ for every $0 < \gamma < 1$. This shows that

$$\varliminf_{b_{\min} \to \infty} \mathrm{E}_i\left\{\frac{\nu_b}{F\left(\max_{j \neq i} \frac{b_{ij}}{q_{ij}}\right)}\right\}^r \geq 1, \qquad \text{for any } r > 0$$

which along with (A.2) proves (4.10).

Proof of iii): To prove (4.11) it remains to set $b_i = \log[(M-1)\pi_i/\overline{R}_i]$, and to use Corollary 2.1, Theorem 3.1, and (4.10). This completes the proof. $\square$

ACKNOWLEDGMENT

The authors are grateful to the reviewers whose comments have led to improvements in the paper.

REFERENCES

[1] P. Armitage, "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," *J. Roy. Statist. Soc. B*, vol. 12, pp. 137–144, 1950.

[2] P. A. Bakut, I. A. Bolshakov, B. M. Gerasimov, A. A. Kuriksha, V. G. Repin, G. P. Tartakovsky, and V. V. Shirokov, *Statistical Radar Theory*, vol. 1, G. P. Tartakovsky, Ed. Moscow, USSR: Sov. Radio, 1963 (in Russian).

[3] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1994–2007, Nov. 1994.

[4] R. E. Bechhofer, J. Kiefer, and M. Sobel, *Sequential Identification and Ranking Procedures (with special reference to Koopman–Darmois families)*. Chicago, IL: Univ. Chicago Press, 1968.

[5] H. Chernoff, "Sequential design of experiments," *Ann. Math. Statist.*, vol. 30, pp. 755–770, 1959.

[6] ——, *Sequential Analysis and Optimal Design*. Philadelphia, PA: SIAM, 1972.

[7] Y.S. Chow and T. L. Lai, "Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossing," *Trans. Amer. Math. Soc.*, vol. 208, pp. 51–72, 1975.

[8] V. P. Dragalin, "Asymptotic solution of a problem of detecting a signal from k channels," *Russ. Math. Surv.*, vol. 42, no. 3, pp. 213–214, 1987.

[9] ——, "Asymptotic solution of a problem of signal searching in multichannel system," *Mat. Issled (Acad, Sci. Moldova)*, vol. 109, pp. 15–35, 1989 (in Russian).

[10] ——, "Third-order asymptotics for a sequential selection procedure," Inst. Applications of Math. and Informatics, Milan, Italy, Tech. Rep., 1994; to appear in *Statist. Decisions* (Special Issue in Honor of Professor Bechhofer).

[11] V. P. Dragalin and A. A. Novikov, "Adaptive sequential tests for composite hypotheses," Inst. Applications of Math. and Informatics, Milan, Italy, Tech. Rep. 94.4, 1994.

[12] V. P. Dragalin, A. G. Tartakovsky, and V. Veeravalli, "Multihypothesis sequential probability ratio tests, part II: Accurate asymptotic expansions for the expected sample size," *IEEE Trans. Inform. Theory*, submitted for publication..

[13] B. Emlsee, S. K. Rogers, M. P. DeSimio, and R. A. Raines, "Complete automatic target recognition system for tactical forward-looking infrared images," *Opt. Eng.*, vol. 36, pp. 2593–2603, Sept. 1997.

[14] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. New York/London: Academic, 1967.

[15] M. M. Fishman, "Average duration of asymptotically optimal multialternative sequential procedure for recognition of processes," *Sov. J. Commun. Technol. Electron.*, vol. 30, pp. 2541–2548, 1987.

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

[17] J. D. Gibbons, I. Olkin, and M. Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*. New York: Wiley, 1977.

[18] G. K. Golubev and R. Z. Khas'minskii, "Sequential testing for several signals in Gaussian white noise," *Theory Prob. Appl.*, vol. 28, pp. 573–584, 1983.

[19] A. Gut, *Stopped Random Walks: Limit Theorems and Applications*. New York: Springer-Verlag, 1988.

[20] P. L. Hsu and H. Robbins, "Complete convergence and the law of large numbers," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 33, pp. 25–31, 1947.

[21] J. Kiefer and J. Sacks, "Asymptotically optimal sequential inference and design," *Ann. Math. Statist.*, vol. 34, pp. 705–750, 1963.

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

[23] T. L. Lai, "Asymptotic optimality of invariant sequential probability ratio test," *Ann. Statist.*, vol. 9, pp. 318–333, 1981.

[24] R. Sh. Liptser and A. N. Shiryaev, *Statistics of Random Processes*, vol. 1. New York: Springer-Verlag, 1977.

[25] G. Lorden, "Nearly-optimal sequential tests for finitely many parameter values," *Ann. Statist.*, vol. 5, pp. 1–21, 1977.

[26] M. B. Marcus and P. Swerling, "Sequential detection in radar with multiple resolution elements," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 237–245, 1962.

[27] F. Mosteller, "A k-sample slippage test for an extreme population," *Ann. Math. Statist.*, vol. 19, pp. 58–65, 1948.

[28] I. R. Savage and J. Sethuraman, "Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives," *Ann. Math. Statist.*, vol. 37, pp. 1154–1160, 1966.

[29] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

[30] Y. G. Sosulin and M. M. Fishman, *Theory of Sequential Decisions and Its Applications*. Moscow, USSR: Radio i Svyaz', 1985 (in Russian).

[31] A. G. Tartakovskii, "Sequential testing of many simple hypotheses with dependent observations," *Probl. Inform. Transm.*, vol. 24, pp. 299–309, 1988.

[32] ——, *Sequential Methods in the Theory of Information Systems*. Moscow, Russia: Radio i Svyaz', 1991 (in Russian).

[33] ——, "Minimax-invariant regret solution to the N-sample slippage problem," *Math. Meth. Statist.*, vol. 6, no. 4, pp. 491–508, 1997.

[34] ——, "Asymptotically optimal sequential tests for nonhomogeneous processes," *Sequential Anal.*, vol. 17, no. 1, pp. 33–61, 1998.

[35] ——, "Asymptotic optimality of certain multihypothesis sequential tests: non-i.i.d. case," *Statist. Inference for Stochastic Processes*, 1999.

[36] V. V. Veeravalli and C. W. Baum, "Asymptotic efficiency of a sequential multihypothesis test," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1994–1997, Nov. 1995.

[37] ——, "Hybrid acquisition of direct sequence CDMA signals," *Int. J. Wireless Inform. Networks,* vol. 3, pp. 55–65, Jan. 1996.

[38] N. V. Verdenskaya and A. G. Tartakovskii, "Asymptotically optimal sequential testing of multiple hypotheses for nonhomogeneous Gaussian processes in an asymmetric situation," *Theory Probab. Appl.*, vol. 36, pp. 536–547, 1991.

[39] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[40] A. Wald and J. Wolfowitz, "Optimal character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, pp. 326–339, 1948.

[41] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia, PA: SIAM, 1982.