Submitted to the IEEE Transactions on Information Theory, September, 1998.

MULTIHYPOTHESIS SEQUENTIAL PROBABILITY RATIO TESTS, PART II: ACCURATE ASYMPTOTIC EXPANSIONS FOR THE EXPECTED SAMPLE SIZE

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

Abstract – **In a companion paper [13], we proved that two specific constructions of multihypothesis sequential tests, which we refer to as Multihypothesis Sequential Probability Ratio Tests (MSPRT's), are asymptotically optimal as the decision risks (or error probabilities) go to zero. The MSPRT's asymptotically minimize not only the expected sample size but also any positive moment of the stopping time distribution, under very general statistical models for the observations. In this paper, based on nonlinear renewal theory we find accurate asymptotic approximations (up to a vanishing term) for the expected sample size that take into account the "overshoot" over the boundaries of decision statistics. The approximations are derived for the scenario where the hypotheses are simple, the observations are independent and identically distributed according to one of the underlying distributions, and the decision risks go to zero. Simulation results for practical examples show that these approximations are fairly accurate not only for large but also for moderate sample sizes. The asymptotic results given here complete the analysis initiated in [4], where first order asymptotics were obtained for the expected sample size under a specific restriction on the Kullback-Leibler distances between the hypotheses.**

Index Terms – **Multihypothesis sequential probability ratio tests, one-sided SPRT, expected sample size, nonlinear renewal theory.**

1. Introduction

The problem of sequential testing of more than two hypotheses is considerably more difficult than that of testing two hypotheses. Optimal solutions are, in general, intractable. Hence, research in sequential multihypothesis testing has been directed towards the study of practical, suboptimal sequential tests and the evaluation of their asymptotic performance – see, for example, [**1**], [4], [**7**]–[**12**], [**18**], [**22**], [**26**]–[**34**].

In a companion paper [**13**], we introduced two specific constructions of multihypothesis sequential tests, which we referred to as Multihypothesis Sequential Probability Ratio Tests (MSPRT's). The first test, which is motivated by a Bayesian framework, was considered by Baum and Veeravalli [**4, 32**], Golubev and Khas'minskii [**17**], Fishman [**16**], Tartakovsky [**27, 28**] in various contexts. The second

V.P. Dragalin is with the Department of Biostatistics, University of Rochester, Rochester, NY, 14642, 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, 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, e-mail: venu@ee.cornell.edu — **Correponding author**

This research was supported in part by the National Science Foundation under grant NCR-9523967, by the National Institutes of Health under grant RO1 HL58751, and by the Office of Naval Research under grant N00014-95-1-0229.

test which is formed by a specific combination of one-sided SPRT's was suggested by Armitage [**1**], and studied further in [**9**]–[**11**], [**22**], [**26**]–[**34**]. We showed in [**13**], that the MSPRT's are asymptotically optimal as the decision risks (or the probabilities of errors) go to zero. The MSPRT's asymptotically minimize not only the expected sample size but also any positive moment of the stopping time distribution, under very general statistical models for the observations. The asymptotic optimality of the MSPRT's makes them attractive candidates for various practical applications as indicated in [**13**]. It is hence of great interest to analyze the asymptotic performance of the MSPRT's.

A particularly effective way to analyze the asymptotic performance of sequential tests (for simple hypotheses and independent and identically distributed observations) is through the application of renewal theory. 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. [**25**]). An approach to applying renewal theory techniques to sequential multihypothesis tests was recently given by Baum and Veeravalli [**4**] in which they studied the quasi-Bayesian MSPRT. This MSPRT was shown to be amenable to an asymptotic analysis using *nonlinear* renewal theory [**35**], and asymptotic expressions for the expected sample size and error probabilities were obtained in [**4**]. While the work in [**4**] provided a starting point for the asymptotic performance analysis of MSPRT's, two important open problems remained. First, the asymptotic analysis in [**4**] was restricted to the "asymmetric" situation, where for each hypothesis, the hypothesis with minimum Kullback-Leibler distance among the remaining hypotheses is unique. Second, only first order asymptotic expressions were obtained for the expected sample size. These first order asymptotics were shown to be quite inaccurate for moderate sample sizes, particularly in the case of symmetric hypotheses.

In the present paper, we complete the asymptotic analysis initiated in [**4**] by addressing both of the open problems discussed above. The asymptotic expansions for the expected sample size are obtained using nonlinear renewal theory (see, e.g. Lai and Siegmund [**19, 20**], Siegmund [**25**], Woodroofe [**35**], Zhang [**36**]). We consider the asymmetric case first, and derive an asymptotically exact expression for the expected sample size under the assumption that the second moments of the log-likelihood ratios between the hypotheses are finite. Then, we go on to tackle the general case where the hypothesis with minimum Kullback-Leibler distance to the hypothesis under consideration is not necessarily unique. In the general case a much stronger Cramér-type condition is required on the log-likelihood functions of the observations. Finally, we present simulation results for practical examples to show that these approximations for the expected sample size are fairly accurate not only for large but also for moderate sample sizes.

2. Preliminaries

Let $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots\}$ be a sequence of independent and identically distributed (iid) random variables (generally vector-valued, $\boldsymbol{X}_n = (X_{1,n}, \dots, X_{l,n}) \in \boldsymbol{R}^l$, $l \geq 1$), and let P be their common probability distribution with the density $f(\boldsymbol{x})$ with respect to some sigma-finite measure. Our goal is to test sequentially the M hypotheses

$$H_i: \ f(\boldsymbol{x}) = f_i(\boldsymbol{x}) \ \forall \, \boldsymbol{x} \in \boldsymbol{R}^l, \ i = 0, 1, \dots, M-1,$$

where $f_i(\boldsymbol{x})$ are given probability densities.

A sequential test is a pair $\delta = (\tau, d)$, where τ is a Markov (stopping) time and $d = d(X_1, \dots, X_\tau)$ is a terminal decision function taking values in the set $\{0, 1, \dots, M-1\}$. That is, $d = j$ implies that the decision is in favor of hypothesis H_j. For each $i, j = 0, 1, \dots, M-1$, assume that the consequence of deciding $d = i$ when H_j is the true hypothesis is given by the loss function $W(j,i) \in [0, \infty)$. Also, without loss of generality, set the losses due to correct decisions to zero ($W(i,i) = 0$). Then the risk associated with making the decision $d = i$ is given by

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

where $\pi_j = \Pr(H_j)$ is the prior probability of the hypothesis H_j, and $\alpha_{ji} = \mathrm{P}_j(d = i)$ is the probability of deciding $d = i$ conditioned on H_j being the true hypothesis. The Bayes risk associated with the test δ is the sum $\sum_{i=0}^{M-1} R_i(\delta)$.

Note that for the special case of a *zero-one* loss function, where $W(j,i) = 1$ for $j \neq i$, the risk R_i is the same as the *frequentist* error probability α_i, which is defined to be the probability of deciding H_i incorrectly.

As in [**13**], we use ΔZ to denote the log-likelihood functions and ratios corresponding to individual observations, i.e.

$$\Delta Z_i(n) = \log f_i(\boldsymbol{X}_n), \text{ and } \Delta Z_{ij}(n) = \log \frac{f_i(\boldsymbol{X}_n)}{f_j(\boldsymbol{X}_n)}. \tag{2.1}$$

Furthermore, use Z to denote the log-likelihood functions and ratios corresponding to sequence of observations up to a given time n, i.e.

$$Z_i(n) = \sum_{t=1}^{n} \Delta Z_i(t), \text{ and } Z_{ij}(n) = \sum_{t=1}^{n} \Delta Z_{ij}(t) = Z_i(n) - Z_j(n). \tag{2.2}$$

Also, for convenience, define the parameters w_{ji} by

$$w_{ji} = \pi_j W(j,i)/\pi_i. \tag{2.3}$$

We now introduce the two constructions of multihypothesis sequential tests, which we refer to as MSPRT's, whose asymptotic performance we study in this paper.

Test δ_a. Introduce the Markov times

$$\tau_i = \inf\Big\{ n \geq 1 : Z_i(n) \geq a_i + \log\Big(\sum_{j \neq i} w_{ji} \exp[Z_j(n)] \Big) \Big\} \tag{2.4}$$

where a_i are positive thresholds ($\inf\{\varnothing\} = \infty$). Then the test procedure $\delta_a = (\tau_a, d_a)$ has stopping time τ_a and terminal decision d_a that are given by

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

This test is motivated by a Bayesian framework, and was considered earlier by Golubev and Khas'minskii [**17**], Fishman [**16**], Sosulin and Fishman [**26**], Tartakovsky [**27**], [**28**], Baum and Veeravalli [**4**], [**32**]. Indeed, for the zero-one loss function, the stopping times τ_i may be rewritten as

$$\tau_i = \inf\big\{ n \geq 1 : \Pi_i(n) \geq A_i \big\} \, , \quad \text{where} \quad A_i = \frac{\exp(a_i)}{1 + \exp(a_i)}$$

and where

$$\Pi_i(n) = \frac{\pi_i \exp[Z_i(n)]}{\sum_{j=0}^{M-1} \pi_j \exp[Z_j(n)]} = \mathrm{P}(H = H_i | X_1^n)$$

is the a posteriori probability of the hypothesis H_i.

Test δ_b. Let

$$\nu_i = \inf\Big\{ n \geq 1 : Z_i(n) \geq b_i + \max_{j \neq i} \Big[\log w_{ji} + Z_j(n) \Big] \Big\} \tag{2.5}$$

be the Markov "accepting" time for the hypothesis H_i, where b_i are positive thresholds ($\inf\{\varnothing\} = \infty$). The test $\delta_b = (\nu_b,\, d_b)$ is defined as follows

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

This test represents a modification of the matrix SPRT (the combination of one-sided SPRT's) that was suggested by Lorden [**22**]. It was considered earlier by Armitage [**1**], Lorden [**22**], Dragalin [**9**]–[**11**], Verdenskaya and Tartakovskii [**34**], Tartakovsky [**28**]–[**30**]. For the zero-one loss function, the stopping times ν_i may be rewritten as

$$\nu_i = \inf\big\{ n \geq 1 : L_i(n) \geq \exp(b_i) \big\}\,, \quad \text{where} \quad L_i(n) = \frac{\pi_i \exp[Z_i(n)]}{\max\limits_{\substack{0 \leq k \leq M-1 \\ k \neq i}} \pi_k \exp[Z_k(n)]}\,,$$

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

Note that δ_b has the obvious advantage that it is easier to implement than δ_a. However, as we shall see in the following section, δ_a has the advantage that it is easier to design the thresholds $\{a_i\}$ to precisely meet constraints on the risks $\{R_i\}$.

In [**13**], we established that if for some $m \geq 1$

$$\mathrm{E}_i\, |Z_{ij}(1)|^m < \infty\,, \tag{2.6}$$

then both δ_a and δ_b asymptotically minimize the k-th moment (for all $k \leq m$) of the stopping time distribution in the class of tests $\boldsymbol{\Delta}(\bar{\boldsymbol{R}})$ (sequential and non-sequential) for which $R_i(\delta) \leq \bar{R}_i$ as $\max_k \bar{R}_k \to 0$. Here $\bar{R}_i > 0$ is a predefined bound on the risk of choosing hypothesis H_i. Furthermore, it follows from Lemma 2.1 and Theorem 3.1 of [**13**] that if $a_i = \log(\pi_i/\bar{R}_i)$, $b_i = \log((M-1)\pi_i/\bar{R}_i)$, the ratios $\log \bar{R}_i / \log \bar{R}_j$ are bounded away from zero and infinity, and the condition (2.6) holds, then for $m \geq 1$

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

where $D_i = \min_{j \neq i} \mathrm{E}_i[Z_{ij}(1)]$ is the minimum Kullback-Leibler distance between H_i and other hypotheses.

Also regardless of the class $\boldsymbol{\Delta}(\bar{\boldsymbol{R}})$, if $\mathrm{E}_i|Z_{ij}(1)| < \infty$ and $\min_k a_k \to \infty$, $\min_k b_k \to \infty$, then the following first order approximations for the expected sample size hold [**13**]:

$$\mathrm{E}_i \tau_a \sim \frac{a_i}{D_i}\,, \qquad \mathrm{E}_i \nu_b \sim \frac{b_i}{D_i}\,. \tag{2.7}$$

These asymptotic formulas describe the behavior only of the first term of the expansion for the average sample size (ASS). The behavior of the second term remains to be determined. Simulation results given in Section 4 show that the first order approximations are usually inaccurate for moderate values of thresholds which are of main interest in practice. In this paper, we derive higher order approximations to the ASS up to a vanishing term. As we shall see, in a specific asymmetric case the second term is of the order of $O(1)$ (i.e. a constant), but in general it goes to infinity as the square root of the threshold.

3. Accurate Asymptotic Approximation for the ASS

In this section, we obtain asymptotic expansions for the ASS of the tests up to a vanishing term using nonlinear renewal theory . See, e.g., Lai and Siegmund [**19, 20**], Siegmund [**25**], Woodroofe [**35**], Zhang [**36**], for detailed discussions of the nonlinear renewal theory results used in this paper.

3.1. The Asymmetric Case. First we consider the asymmetric case studied in [**4, 32**], where the hypothesis $j^* = j^*(i)$ for which the Kullback-Leibler distance, $D_{ij} = \mathrm{E}_i Z_{ij}(1)$, attains its minimum (over $j \neq i$) is unique. (The reader is reminded of the definition of $\Delta Z_{ij}(t)$ and $Z_{ij}(n)$ given in (2.1) and (2.2).) Specifically, throughout Subsection 3.1 we assume that the following condition holds:

$$j^*(i) = \arg\min_{j \neq i} D_{ij} \qquad \text{is unique for any} \quad i \in \{0, 1, \dots, M-1\}\,. \tag{3.1}$$

In order to apply relevant results from nonlinear renewal theory, we rewrite the stopping times of (2.4) and (2.5) in the form of a random walk crossing a constant boundary plus a nonlinear term that is "slowly changing" in the sense defined below in Definition 3.1. By subtracting $Z_{j^*}(n)$ from both sides of the inequalities in (2.4) and (2.5), we see that

$$\tau_i = \inf\bigl\{n \geq 1 : Z_{ij^*}(n) \geq a_i + \xi_i(n)\bigr\}, \tag{3.2}$$

and

$$\nu_i = \inf\bigl\{n \geq 1 : Z_{ij^*}(n) \geq b_i + Y_i(n)\bigr\}, \tag{3.3}$$

where $\xi_i(n)$ and $Y_i(n)$ are defined by:

$$\xi_i(n) = \log\Bigl(w_{j^*i} + \sum_{j \neq i, j^*} w_{ji} \exp\bigl\{-Z_{j^*j}(n)\bigr\}\Bigr)\,; \tag{3.4}$$

$$Y_i(n) = \log\Bigl(\max\Bigl[w_{j^*i}\,,\, \max_{j \neq i, j^*} w_{ji} \exp\bigl\{-Z_{j^*j}(n)\bigr\}\Bigr]\Bigr)\,. \tag{3.5}$$

Definition 3.1. The process $\{\zeta_n\}_{n \geq 1}$ is said to be *slowly changing* if the following two conditions hold:

(i) as $n \to \infty$,

$$n^{-1} \max_{1 \leq t \leq n} |\zeta_t| \to 0 \qquad \text{in probability}\,; \tag{3.6}$$

(ii) for every $\varepsilon > 0$ and some $\lambda > 0$

$$\mathrm{P}\Bigl(\max_{0 \leq k \leq n\lambda} |\zeta_{n+k} - \zeta_n| > \varepsilon\Bigr) < \varepsilon \quad \forall\, n \geq 1\,, \tag{3.7}$$

i.e. $\{\zeta_n\}$ is uniformly continuous in probability [**35**].

Lemma 3.1. *Let the condition* (3.1) *be fulfilled and* $\mathrm{E}_i|Z_{ij}(1)| < \infty$ *for all* i *and* j, $i \neq j$. *Then the processes* $\{\xi_i(n),\, n \geq 1\}$ *and* $\{Y_i(n),\, n \geq 1\}$ *are slowly changing under* P_i.

PROOF. It is easy to see that $\mathrm{E}_i\Delta Z_{j^*j}(t) = \mathrm{E}_i\Delta Z_{ij}(t) - \mathrm{E}_i\Delta Z_{j^*i}(t) = D_{ij} - D_{ij^*}$. We denote this difference by Δ_{ij}. We can then rewrite $\xi_i(n)$ of (3.4) as

$$\xi_i(n) = \log\Big(w_{j^*i} + \sum_{j\neq i,j^*} w_{ji}\exp\big\{-n\big[\Delta_{ij} + n^{-1}M_{ij}(n)\big]\big\}\Big)$$

where $M_{ij}(n) = \sum_{t=1}^{n}[\Delta Z_{j^*j}(t) - \Delta_{ij}]$ is a zero-mean random walk with respect to P_i. By the strong law of large numbers $n^{-1}M_{ij}(n) \to 0$ P_i–a.s. and

$$\xi_i(n) \to \log w_{j^*i} \qquad \text{w.p. } 1 \quad \text{as} \quad n\to\infty\,, \tag{3.8}$$

which implies (3.6) and (3.7). Evidently, $\log w_{j^*i} \le Y_i(n) \le \xi_i(n)$ and hence, by a standard sandwich argument, the process $\{Y_i(n)\}$ is also slowly changing. ■

An important consequence of the slowly changing property is that limiting distributions of the overshoot of a random walk (with positive mean) over a fixed threshold are unchanged by the addition of a slowly changing nonlinear term (see Theorem 4.1 of Woodroofe [**35**]). In particular, suppose we define the stopping time

$$\tau_{ij^*}(c) = \inf\{n \ge 1 : Z_{ij^*}(n) \ge c\}\,, \quad c > 0\,,$$

which corresponds to the one-sided SPRT which tests the hypothesis H_i against the closest one H_{j^*}. Now let

$$\rho_i(c) = Z_{ij^*}(\tau_{ij^*}) - c$$

denote the overshoot at the stopping time. Furthermore, let

$$G_i(y) = \lim_{c\to\infty} \mathrm{P}_i\big\{\rho_i(c) \le y\big\}$$

denote the limiting distribution of the overshoot.

Now, let $\{\zeta_n\}$ be a slowly changing process and consider the stopping time

$$t_{ij^*}(c) = \inf\{n \ge 1 : Z_{ij^*}(n) \ge c + \zeta_n\}\,, \quad c > 0\,.$$

Then, by Theorem 4.1 of [**35**], the overshoot at the stopping time

$$r_i(c) = Z_{ij^*}(t_{ij^*}) - c - \zeta_{t_{ij^*}}$$

has the same limiting distribution as $\rho_i(c)$, i.e.

$$\lim_{c\to\infty} \mathrm{P}_i\big\{r_i(c) \le y\big\} = G_i(y)\,.$$

Now note that from (3.2),

$$Z_{ij^*}(\tau_i) = a_i + \xi_i(\tau_i) + \chi_i \qquad \text{on} \quad \{\tau_i < \infty\}, \tag{3.9}$$

where χ_i is the overshoot of the process $Z_{ij^*}(n) - \xi_i(n)$ over the level a_i at time τ_i. Taking the expectations in both sides of this last equality and applying the Wald identity, we obtain

$$D_i\,\mathrm{E}_i\tau_i = a_i + \mathrm{E}_i\xi_i(\tau_i) + \mathrm{E}_i\chi_i\,. \tag{3.10}$$

A similar equality is true for the Markov time ν_i. Indeed, from (3.3),

$$Z_{ij^*}(\nu_i) = b_i + Y_i(\nu_i) + \tilde\chi_i \qquad \text{on} \quad \{\nu_i < \infty\} \tag{3.11}$$

where $\tilde{\chi}_i$ is the overshoot of the process $Z_{ij^*}(n) - Y_i(n)$ over the level b_i at time ν_i. Applying Wald's identity, we get

$$D_i \, \mathrm{E}_i \nu_i = b_i + \mathrm{E}_i Y_i(\nu_i) + \mathrm{E}_i \tilde{\chi}_i \,. \tag{3.12}$$

We are now in a position to understand the role of Lemma 3.1. The key point is that since the sequences $\{\xi_i(n)\}$ and $\{Y_i(n)\}$ are slowly changing, the overshoots χ_i and $\tilde{\chi}_i$ have the same limiting distributions as $\rho_i(c)$ (when c, a_i and b_i tend to infinity). Furthermore, both $\xi_i(n)$ and $Y_i(n)$ converge to the constant $\log w_{j^*i}$ as $n \to \infty$. Thus, for large a_i and b_i, the ASS's $\mathrm{E}_i \tau_a$ and $\mathrm{E}_i \nu_b$ should be approximately equal to $D_i^{-1}(a_i + \log w_{j^*i} + \varkappa_i)$ and $D_i^{-1}(b_i + \log w_{j^*i} + \varkappa_i)$, respectively, where

$$\varkappa_i = \int_0^\infty y \, dG_i(y)$$

is the expected limiting overshoot. The following theorem is a formal statement of this result and is proved in the Appendix.

Theorem 3.1. *Suppose that the condition* (3.1) *holds and a_i/a_j, b_i/b_j are bounded away from zero and infinity, i.e.*

$$a_i/a_j \sim b_i/b_j \sim c_{ij} \qquad as \quad \min_k a_k \to \infty, \ \min_k b_k \to \infty \quad where \quad 0 < c_{ij} < \infty \,. \tag{3.13}$$

In addition, assume that $\mathrm{E}_i |Z_{ij}(1)|^2 < \infty$ *and the* $Z_{ij}(1)$ *are* P_i*−non-arithmetic*[1]. *Then*

$$\mathrm{E}_i \tau_a = \frac{1}{D_i}\big(a_i + \log w_{j^*i} + \varkappa_i\big) + o(1) \qquad as \quad \min_k a_k \to \infty \,; \tag{3.14}$$

$$\mathrm{E}_i \nu_b = \frac{1}{D_i}\big(b_i + \log w_{j^*i} + \varkappa_i\big) + o(1) \qquad as \quad \min_k b_k \to \infty \,. \tag{3.15}$$

Remark 3.1. The condition that $Z_{ij}(1)$ are P_i−non-arithmetic is imposed due to the necessity of considering certain discrete cases separately in the renewal theorem (see, e.g., Woodroofe [**35**], Section 2.1). If $Z_{ij}(1)$ are P_i-arithmetic with span $d > 0$, the results of Theorem 3.1 (as well as of Theorem 3.3 below) hold true as $\min_k a_k \to \infty$ and $\min_k b_k \to \infty$ through multiples of d (i.e. $a_i = kd$, $k \to \infty$) and with respective modification of definition of $\varkappa_i$.

We note that in the definitions of τ_i and ν_i of (2.4) and (2.5), and in Theorem 3.1 above, the numbers w_{ij} may be assumed to be arbitrary positive constants not necessarily equal to $\pi_j W(j,i)/\pi_i$ as we defined them before in (2.3). However, by setting w_{ij} as specified in (2.3), we can precisely control the risks R_i corresponding to the tests δ_a and δ_b.

Now let

$$\gamma_i = \int_0^\infty \exp(-y) \, dG_i(y) \,.$$

Recall that $G_i(y)$ is the limiting distribution of the overshoot $\rho_i(c)$ in the one-sided SPRT as $c \to \infty$ (under the measure P_i). The following theorem provides asymptotic expressions for the risks R_i in terms of γ_i.

[1]A random variable X is called arithmetic (has arithmetic distribution with span d) if $d \cdot X$ is integer-valued for some non-zero constant d. A typical example is the Bernoulli sequence. Otherwise X is called non-arithmetic (or non-lattice), i.e. $\mathrm{P}(X = dm$ for some $m) < 1$.

Theorem 3.2. *Let* $R_i(\delta) = \sum_{j\neq i} \pi_j W(j,i)\alpha_{ji}(\delta)$ *be the risk of deciding* $d = i$. *If we set* $w_{ji} = \pi_j W(j,i)/\pi_i$, *then under the same conditions as in Theorem* 3.1

$$R_i(\delta_a) = \pi_i \gamma_i e^{-a_i}\big(1 + o(1)\big) \quad \textit{as} \quad \min_k a_k \to \infty\,; \tag{3.16}$$

$$\pi_i \gamma_i e^{-b_i}\big(1 + o(1)\big) \le R_i(\delta_b) \le (M-1)\pi_i \gamma_i e^{-b_i}\big(1 + o(1)\big) \quad \textit{as} \quad \min_k b_k \to \infty\,. \tag{3.17}$$

Proof. For the zero-one loss function, the asymptotic equality (3.16) follows from Theorem 6.1 of Baum and Veeravalli [**4**]. To prove this equality for an arbitrary loss function we observe first that the risk $R_i(\delta_a)$ can be written in the form

$$R_i(\delta_a) = \pi_i \operatorname{E}_i\Big\{\exp\big[-Z_{ij^*}(\tau_i) + \xi_i(\tau_i)\big]\boldsymbol{I}(\tau_a = \tau_i)\Big\}\,. \tag{3.18}$$

Indeed,

$$\begin{aligned} R_i(\delta_a) &= \sum_{j\neq i} \pi_j W(j,i) \mathrm{P}_j(\tau_a = \tau_i) = \sum_{j\neq i} \pi_j W(j,i) \int_{\tau_a=\tau_i} \exp[Z_{ji}(\tau_i)]\, d\mathrm{P}_i \\ &= \mathrm{E}_i\Big\{\boldsymbol{I}(\tau_a = \tau_i) \sum_{j\neq i} \pi_j W(j,i) \exp[Z_{ji}(\tau_i)]\Big\} \\ &= \pi_i \operatorname{E}_i\Big\{\exp[-Z_{ij^*}(\tau_i)]\boldsymbol{I}(\tau_a = \tau_i) \sum_{j\neq i} w_{ji} \exp[-Z_{j^*j}(\tau_i)]\Big\} \\ &= \pi_i \operatorname{E}_i\Big\{\exp[-Z_{ij^*}(\tau_i)]\boldsymbol{I}(\tau_a = \tau_i)\Big[w_{j^*i} + \sum_{j\neq i,j^*} w_{ji} \exp[-Z_{j^*j}(\tau_i)]\Big]\Big\}, \end{aligned} \tag{3.19}$$

from which (3.18) follows in an obvious manner. Now, since by (3.9)

$$Z_{ij^*}(\tau_i) - \xi_i(\tau_i) = a_i + \chi_i \qquad \text{on} \quad \tau_i < \infty\,,$$

we obtain

$$R_i(\delta_a) = \pi_i\, \exp\big(-a_i\big)\mathrm{E}_i\Big\{\exp\big(-\chi_i\big)\boldsymbol{I}(\tau_a = \tau_i)\Big\}\,.$$

Due to the fact that $\xi_i(n)$ is slowly changing, $\mathrm{E}_i \exp(-\chi_i) \to \gamma_i$ (see Theorem 4.1 in [**35**]). Furthermore, $\mathrm{P}_i(\tau_a = \tau_i) \to 1$ as $a_i \to \infty$. Hence the value of $\mathrm{E}_i\{\exp(-\chi_i)\boldsymbol{I}(\tau_a = \tau_i)\}$ converges to γ_i which along with the previous equality implies (3.16).

Consider the second test δ_b. Obviously, the equality (3.19) is true for $R_i(\delta_b)$ if τ_a and τ_i are replaced with ν_b and ν_i, respectively. In turn, this equality implies

$$R_i(\delta_b) = \pi_i \operatorname{E}_i\Big\{\exp\big[-\big(Z_{ij^*}(\nu_i) - \mathrm{Y}_i(\nu_i) + \beta_i(\nu_i)\big)\big]\boldsymbol{I}(\nu_b = \nu_i)\Big\}$$

where

$$\beta_i(\nu_i) = \mathrm{Y}_i(\nu_i) - \xi_i(\nu_i). \tag{3.20}$$

Next, by (3.11),

$$Z_{ij^*}(\nu_i) - \mathrm{Y}_i(\nu_i) = b_i + \tilde{\chi}_i \qquad \text{on} \quad \nu_i < \infty\,,$$

where $\tilde{\chi}_i$ is the overshoot of the process $Z_{ij^*}(n) - \mathrm{Y}_i(n)$ over the level b_i at time instant ν_i, and hence

$$R_i(\delta_b) = \pi_i\, \exp(-b_i)\, \mathrm{E}_i\Big\{\exp\big[-\tilde{\chi}_i - \beta_i(\nu_i)\big]\boldsymbol{I}(\nu_b = \nu_i)\Big\}\,.$$

Note that by considering the difference of (3.4) and (3.5), $\beta_i(\nu_i)$ of (3.20) can be shown to satisfy

$$-\log(M-1) \le \beta_i(\nu_i) \le 0 \quad \text{on } \{\nu_i < \infty\}\,,$$

which implies the inequalities

$$\pi_i\, e^{-b_i}\, \mathrm{E}_i\Big\{e^{-\tilde{\chi}_i}\boldsymbol{I}(\nu_b=\nu_i)\Big\} \le R_i(\delta_b) \le (M-1)\,\pi_i\, e^{-b_i}\, \mathrm{E}_i\Big\{e^{-\tilde{\chi}_i}\boldsymbol{I}(\nu_b=\nu_i)\Big\}.$$

Now, $\mathrm{P}_i(\nu_b=\nu_i) \to 1$ as $b_i \to \infty$, and by Theorem 4.1 of Woodroofe [**35**] $\mathrm{E}_i \exp(-\tilde{\chi}_i) \to \gamma_i$, due to the fact that $Y_i(n)$ is slowly changing. Thus $\mathrm{E}_i\{\exp(-\tilde{\chi}_i)\boldsymbol{I}(\nu_b=\nu_i)\}$ converges to γ_i. This fact together with the previous inequalities yields (3.17) and the theorem follows. ■

As a consequence of Theorem 3.1 and Theorem 3.2 we have the following important result.

Corollary 3.1. *If* $a_i = \log(\pi_i\gamma_i/\bar{R}_i)$ *and* $b_i = \log((M-1)\pi_i\gamma_i/\bar{R}_i)$, *then as* $\max_k \bar{R}_k \to 0$

$$R_i(\delta_a) = \bar{R}_i + o\big(\bar{R}_i\big)\,; \qquad \bar{R}_i/(M-1) + o\big(\bar{R}_i\big) \le R_i(\delta_b) \le \bar{R}_i + o\big(\bar{R}_i\big)\,; \tag{3.21}$$

$$\mathrm{E}_i\tau_a = \frac{1}{D_i}\left[\log\Big(\frac{\gamma_i\,\pi_{j^*}\,W(j^*,i)}{\bar{R}_i}\Big) + \varkappa_i\right] + o(1)\,; \tag{3.22}$$

$$\mathrm{E}_i\nu_b = \frac{1}{D_i}\left[\log\Big(\frac{\gamma_i\,\pi_{j^*}\,W(j^*,i)}{\bar{R}_i}\Big) + \log(M-1) + \varkappa_i\right] + o(1)\,. \tag{3.23}$$

It is important to emphasize that the test δ_a performs better than the test δ_b. Indeed, according to (3.22) and (3.23) the difference between $\mathrm{E}_i\nu_b$ and $\mathrm{E}_i\tau_a$ is equal to $D_i^{-1}\log(M-1) + o(1)$ if we take $b_i = \log((M-1)\pi_i\gamma_i/\bar{R}_i)$. For large M the difference may appear to be substantial. However, for the procedure δ_b we guarantee only the inequality $R_i(\delta_b) \le \bar{R}_i$ while for the procedure δ_a the corresponding equality holds. Thus, the real difference between the two tests would be less than $D_i^{-1}\log(M-1)$ if the thresholds b_i can be chosen so that $R_i(\delta_b) \sim \bar{R}_i$. In that case, the difference between $\mathrm{E}_i\nu_b$ and $\mathrm{E}_i\tau_a$ can be made negligible. For the symmetric case with respect to decision risks, where for all i, j, $i \ne j$

$$c_{ij} = \frac{\log \bar{R}_i}{\log \bar{R}_j} \sim 1 \qquad \text{as} \quad \max_k \bar{R}_k \to 0\,, \tag{3.24}$$

the fact that $\mathrm{E}_i\nu_b = \mathrm{E}_i\tau_a + o(1)$ may be implicitly derived from the results of Lorden [**22**]. More specifically, if $R_i(\delta_b) \sim \bar{R}_i$ and (3.24) is fulfilled, then

$$\mathrm{E}_i\nu_b - \inf_{\delta \in \boldsymbol{\Delta}(\bar{\boldsymbol{R}})} \mathrm{E}_i\tau = o(1) \qquad \text{as} \quad \max_k \bar{R}_k \to 0\,. \tag{3.25}$$

We suspect that the same result is true as long as c_{ij} are arbitrary numbers bounded away from 0 and ∞. We do not have a rigorous proof for this case; however, simulation results given in Section 4 agree with this conjecture.

3.2. The General Case. Thus far we have assumed that the minimum distance $D_i = \min_{j\ne i} D_{ij}$ is achieved uniquely (see the condition (3.1)). We now relax this condition. For fixed i, let

$$D_{i[1]} \le D_{i[2]} \le \cdots \le D_{i[M-1]}$$

be the ordered values of D_{ij}, $j \ne i$ ($D_{i[M]} = +\infty$). Throughout the rest of this section we assume that

$$D_{i[1]} = \cdots = D_{i[r]} < D_{i[r+1]} \qquad \text{for some} \quad r \in \{1,\dots,M-1\}\,. \tag{3.26}$$

Note that condition (3.26) includes the fully symmetric situation

$$D_{ij} = D_i \qquad \text{for all} \quad j \in \{0,1,\dots,M-1\} \setminus i \tag{3.27}$$

when $r = M - 1$. Note also that the previous asymmetric case is covered by setting $r = 1$.

The derivation of the asymptotics for the general case is more complicated than in the asymmetric case. The reason is that if we write the Markov times τ_i and ν_i in the form given in (3.2) and (3.3), the sequences corresponding to $\{\xi_i(n)\}$ and $\{Y_i(n)\}$ are not necessarily slowly changing in the general case. A different approach and stronger conditions (see the Cramér-type condition (3.34) below) are required to tackle this general case.

Recall that $Z_i(n) = \sum_{t=1}^n \log f_i(\boldsymbol{X}_t)$ is the log-likelihood function of the observations up to time n. Now, let

$$\mu_{ij} = \mathrm{E}_i Z_j(1)\,, \quad j \in \{0, 1, \dots, M-1\} \setminus i$$

denote the corresponding mean value of the increment of the log-likelihood function under H_i.

Let $\mu_{i[1]} \le \mu_{i[2]} \le \cdots \le \mu_{i[M-1]}$ be the ordered values of μ_{ij}, $j \ne i$, with $\langle j \rangle$ denoting the index of the ordered values, i.e.

$$\langle j \rangle = k, \quad \text{if} \;\; \mu_{i[j]} = \mu_{ik}, \tag{3.28}$$

with an arbitrary index assignment in case of ties.

Now, under assumption (3.26), there are r values that achieve the minimum distance D_i. Since $D_{ij} = \mathrm{E}_i Z_i(1) - \mu_{ij}$, we must have r values that achieve the maximum of $\{\mu_{ij}, j \ne i\}$, i.e.

$$\mu_{i[M-1-r]} < \mu_{i[M-r]} = \cdots = \mu_{i[M-1]} \quad (\mu_{i[0]} = -\infty)\,.$$

Next, define an r-dimensional vector $\boldsymbol{Y}_i = (Y_{1,i}, Y_{2,i}, \dots, Y_{r,i})$ with components

$$Y_{k,i} = Z_{\langle M-1-r+k\rangle}(1) - \mu_{i[M-1-r+k]} = Z_{\langle M-1-r+k\rangle}(1) - \mu_{i[M-1]}\,, \quad k = 1, \dots, r.$$

Obviously, $\boldsymbol{Y}_i$ is zero-mean. Let $\boldsymbol{V}_i = \mathrm{Cov}_i(\boldsymbol{Y}_i)$ denote its covariance matrix with respect to P_i.

Now let

$$\phi_{0,V_i}(\boldsymbol{x}) = \left[(2\pi)^r \, |\boldsymbol{V}_i|\right]^{-1/2} \exp\left\{-\tfrac{1}{2}\boldsymbol{x}\boldsymbol{V}_i^{-1}\boldsymbol{x}^\top\right\}$$

be the density of a multivariate normal distribution function with covariance matrix $\boldsymbol{V}_i$.

The asymptotic expansions in the general case are derived using normal approximations (see Bhattacharya and Rao [**5**]). In this context, we introduce the variables $h_{r,i}$ and $C_{r,i}$. The variable $h_{r,i}$ is the expected value of the maximum of r zero-mean normal random variables with the density function $\phi_{0,V_i}(\boldsymbol{x})$, i.e.

$$h_{r,i} = \int_{\boldsymbol{R}^r} \Big(\max_{1\le k\le r} x_k\Big) \phi_{0,V_i}(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{3.29}$$

and $C_{r,i}$ is given by

$$C_{r,i} = \int_{\boldsymbol{R}^r} \Big(\max_{1\le k\le r} x_k\Big) \left(\mathcal{P}_i(\boldsymbol{x}) + \boldsymbol{\lambda}_i \boldsymbol{V}_i^{-1}\boldsymbol{x}^\top\right) \phi_{0,V_i}(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{3.30}$$

where

$$\boldsymbol{\lambda}_i = (\lambda_{1,i}, \dots, \lambda_{r,i})\,; \quad \lambda_{k,i} = \log w_{\langle M-1-r+k\rangle i}\,, \tag{3.31}$$

and where $\mathcal{P}_i(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x} \in \boldsymbol{R}^r$ of degree 3 whose coefficients involve $\boldsymbol{V}_i$ and the P_i-cumulants of $\boldsymbol{Y}_i$ up to order 3 and is given explicitly by formula (7.19) in Bhattacharya and Rao [**5**]. Due to the lengthiness of this formula we have not included it here. Computing the constant $h_{r,i}$ for a given application is relatively straighforward, since the integral in (3.29) involves only the covariance matrix of the vector $\boldsymbol{Y}_i$; further simplification results in the case where $\boldsymbol{V}_i$ is diagonal. Computing the constant $C_{r,i}$ is, in general, quite difficult due to the fact that the polynomial $\mathcal{P}_i(\boldsymbol{x})$ is a complicated function of the statistics of $\boldsymbol{Y}_i$. However, in the symmetric situation (which is usually of interest), where $\boldsymbol{V}_i$ is of the form $v_i^2\mathbb{I}+\varepsilon$ (with $\mathbb{I}$ being the identity matrix), and $\lambda_{k,i}=\lambda_i$, for $k=1,\dots,r$, a considerable simplication is possible (see Section 3.3.2 below). Note that the application of the normal approximations requires a Cramér-type condition on the joint characteristic function of the vector $\boldsymbol{Y}_i$ given in (3.34) below.

We now present a "heuristic" outline of our approach to finding asymptotics in the general case. We focus on the second test δ_b, since our approach works more naturally for this test. As we shall see in the proof of Theorem 3.3, the asymptotic ASS results for the δ_a follow from those for δ_b.

The Markov time ν_i of (2.5) may be written in the form:

$$\nu_i=\inf\Big\{n: S_i(n)\ge b_i+\xi(n)+h_{r,i}\sqrt{n}\Big\}\,, \tag{3.32}$$

where

$$S_i(n)=Z_i(n)-n\mu_{i[M-1]}$$

is a random walk with increments having positive mean $\mathrm{E}_i Z_i(1)-\mu_{i[M-1]} \;=\; D_i$, and

$$\xi(n)=\max_{k\ne i}\left[\log w_{ki}+Z_k(n)-n\mu_{i[M-1]}\right]-h_{r,i}\sqrt{n}\,.$$

It is established in the proof of Theorem 3.3 that $\{\xi(n)\}$ is a slowly changing sequence that converges in distribution to a random variable ξ, and that

$$\mathrm{E}_i\xi(n)\to\mathrm{E}_i\xi=C_{r,i}\qquad\text{as}\quad n\to\infty\,,$$

where $C_{r,i}$ is defined in (3.30). In comparing with the corresponding equation for the asymmetric case given in (3.3), we see that $S_i(n)$ is written in terms of the log-likelihood function of the observations under H_i as opposed to the log-likelihood ratios. Defining $S_i(n)$ in this fashion is required for the corresponding $\{\xi(n)\}$ sequence to be slowly changing in the general case. Also note that in addition to the slowly changing term there is a nonlinear deterministic term that is added to the threshold in (3.32).

Now note that from (3.32),

$$S_i(\nu_i)=b_i+\xi(\nu_i)+h_{r,i}\sqrt{\nu_i}+\chi_i\qquad\text{on}\quad\{\nu_i<\infty\}\,, \tag{3.33}$$

where χ_i is the overshoot of the process $S_i(n)-\xi(n)-h_{r,i}\sqrt{n}$ over the level b_i at time ν_i. In the limit as $b_i\to\infty$, we expect ν_i to approximately satisfy the equation

$$\nu_i D_i=b_i+\xi+h_{r,i}\sqrt{\nu_i}+\chi$$

where χ is the limiting overshoot. Solving this equation for ν_i gives

$$\nu_i\approx\frac{1}{D_i}\left[b_i+\frac{h_{r,i}^2}{2D_i}+h_{r,i}\sqrt{\frac{b_i}{D_i}+\frac{h_{r,i}^2}{4D_i^2}+\chi+\xi}\,\right]\,.$$

Finally, using uniform integrability arguments, we expect that

$$\mathrm{E}_i\nu_i\approx\frac{1}{D_i}\left[b_i+\frac{h_{r,i}^2}{2D_i}+h_{r,i}\sqrt{\frac{b_i}{D_i}+\frac{h_{r,i}^2}{4D_i^2}}+\varkappa_i+C_{r,i}\right]\,,$$

where $\varkappa_i$ is the expected limiting overshoot, and $C_{r,i}$ is expectation of the limit of the slowly changing sequence $\{\xi(n)\}$. The following theorem formalizes this result, and the detailed proof based on the renewal theory of Zhang [**36**] is given in the Appendix. We write $a = \min_k a_k$ and $b = \min_k b_k$ for brevity.

THEOREM 3.3. *Suppose that $Z_i(1)$ is P_i−non-arithmetic, the covariance matrix $\boldsymbol{V}_i$ of the vector $\boldsymbol{Y}_i$ is positive definite, $\mathrm{E}_i\|\boldsymbol{Y}_i\|^3 < \infty$, the condition* (3.13) *holds, and the Cramér condition*

$$\overline{\lim}_{\|\boldsymbol{t}\|\to\infty} \mathrm{E}_i \exp\{\jmath \cdot (\boldsymbol{t}, \boldsymbol{Y}_i)\} < 1 \tag{3.34}$$

on the joint characteristic function of $\boldsymbol{Y}_i$ is satisfied. Then

$$\mathrm{E}_i\nu_b = \frac{1}{D_i}\left[b_i + h_{r,i}\sqrt{\frac{b_i}{D_i} + \frac{{h_{r,i}}^2}{4D_i^2}} + \frac{{h_{r,i}}^2}{2D_i} + \varkappa_i + C_{r,i}\right] + o(1) \quad \text{as} \quad b \to \infty, \tag{3.35}$$

$$\mathrm{E}_i\tau_a = \frac{1}{D_i}\left[a_i + h_{r,i}\sqrt{\frac{a_i}{D_i} + \frac{{h_{r,i}}^2}{4D_i^2}} + \frac{{h_{r,i}}^2}{2D_i} + \varkappa_i + C_{r,i} + K\right] + o(1) \quad \text{as} \quad a \to \infty, \tag{3.36}$$

where $0 \le K \le \log(M-1)$ is a constant which does not depend on a_i, and $\varkappa_i$ is the expectation of the limiting overshoot in the one-sided SPRT based on the log-likelihood ratio $Z_{i\langle M-1\rangle}(n) = Z_i(n) - Z_{\langle M-1\rangle}(n)$.

REMARK 3.2. Numerical results given in Section 4 indicate that setting $K = 0$ in (3.36) consistently produces the best match with simulated ASS values for test δ_a.

REMARK 3.3. We note that in the general case we have not been able to obtain the counterpart of Theorem 3.2. Thus, we need to rely on weaker Wald type inequalities for the risks in order set the thresholds to meet risk constraints. In particular, as we established in Lemma 2.1 of [**13**],

$$R_i(\delta_a) \le \pi_i \exp(-a_i), \qquad R_i(\delta_b) \le (M-1)\pi_i \exp(-b_i)\,. \tag{3.37}$$

3.3. Some Special Cases. In the following we address the issue of computing the constants $h_{r,i}$ and $C_{r,i}$ appearing in asymptotic expansions for the ASS given in Theorem 3.3.

3.3.1. *Case* 1. In the asymmetric case (3.1), $r = 1$, and it is easily shown that $h_{1,i} = 0$. Also, as shown in (7.21) of [**5**], $\mathcal{P}_i(x) = \mathrm{E}_i Y_{1,i}^3\,(x^3 - 3x)/6$ in this case. Now, since for the standard Gaussian random variable $\mathrm{E}X^4 = 3\mathrm{E}X^2$, we see from (3.30) that $C_{1,i} = \log w_{j^*i}$. Thus the resulting expression for the ASS is consistent with the result of Theorem 3.1.

3.3.2. *Case* 2. Consider the symmetric case where $\{Y_{k,i}, k = 1, \dots, r\}$ are identically distributed (but not necessarily independent), and where $\boldsymbol{V}_i$ is of the form $v_i^2\mathbb{I}+\varepsilon$ (with $\mathbb{I}$ being the identity matrix), and $\lambda_{k,i} = \lambda_i$, for $k = 1, \dots, r$. This case often arises in practice (see examples in Section 4).

First suppose that $\varepsilon = \lambda_i = 0$. In this special case, it is easy to see from (3.29) that

$$h_{r,i} = v_i h_r^* \tag{3.38}$$

where h_r^* is the expected value of the standard normal order statistic. (See Table 1 for computed values of h_r^*.) Furthermore, we may use the results in [**11**] to get the following relatively simple expression for $C_{r,i}$. (Note that the second term inside the integral in (3.30) is zero since $\boldsymbol{\lambda} = 0$.)

$$C_{r,i} = \left(\mathrm{E}_i Y_{1,i}^3\right) \frac{C_r^*}{6v_i^2} \tag{3.39}$$

where

$$C_r^* = r \int_{-\infty}^{\infty} x\varphi(x)\Phi(x)^{r-2} \Big[(r-1)\varphi(x)(1-x^2) + (x^3 - 3x)\Phi(x)\Big]\, dx\,, \tag{3.40}$$

and where $\varphi(x)$ and $\Phi(x)$ are standard normal density and distribution functions, respectively. See Table 2 for computed values of C_r^*.

We now remove the restriction that $\varepsilon = 0$ (still assuming $\lambda_i = 0$). Note that for $\boldsymbol{V}_i$ to be positive definite, we require that $\varepsilon > -v_i^2/r$. Since $\boldsymbol{Y}_i$ is zero mean with covariance $\boldsymbol{V}_i = v_i^2\mathbb{I} + \varepsilon$ under H_i, we may write

$$Y_{k,i} = \tilde{Y}_{k,i} + \eta\bar{Y}\,,$$

where

$$\tilde{Y}_{k,i} = Y_{k,i} - \frac{\eta}{1+\eta}\left(\tfrac{1}{r}\sum_{k=1}^{r} Y_{k,i}\right) \tag{3.41}$$

is zero mean with covariance $\boldsymbol{V}_i = v_i^2\mathbb{I}$, $\bar{Y} = \frac{1}{r}\sum_{k=1}^{r}\tilde{Y}_{k,i}$, and

$$\eta = -1 + \sqrt{1 + \frac{\varepsilon r}{v_i^2}}\,.$$

Thus, for $\varepsilon \neq 0$, the slowly changing term $\zeta(n)$ in equation (A.12) in the proof of Theorem 3.3 simply gets modified, relative to the case $\varepsilon = 0$, by the addition of the term $\eta\bar{Y}$. It is easy to see that the addition of $\eta\bar{Y}$ does not affect the conditions needed on $\xi(n)$ given in (A.13)–(A.18). Furthermore, $\mathrm{E}_i\bar{Y} = 0$ implies that $\mathrm{E}_i\xi$ is also unaffected. Thus, using (3.39), we can see that $C_{r,i}$ is given by

$$C_{r,i} = \left(\mathrm{E}_i \tilde{Y}_{1,i}^3\right) \frac{C_r^*}{6v_i^2} \tag{3.42}$$

for all allowable ε, i.e. $\varepsilon > -v_i^2/r$. It is also easily shown that $h_{r,i}$ is given by (3.38) for all $\varepsilon > -v_i^2/r$. Indeed, if $(X_1, \ldots, X_r) \sim \mathcal{N}(0, \boldsymbol{V}_0)$ with such $\boldsymbol{V}_0$, then

$$\mathrm{E}\max\{X_1, \ldots, X_r\} = \mathrm{E}\max\{\tilde{X}_1 + \eta\bar{X}, \ldots, \tilde{X}_r + \eta\bar{X}\} = \mathrm{E}\max\{\tilde{X}_1, \ldots, \tilde{X}_r\} + \eta\mathrm{E}\bar{X}\,,$$

where $(\tilde{X}_1, \ldots, \tilde{X}_r) \sim \mathcal{N}(0, v_i^2\mathbb{I})$, and $\bar{X} = \frac{1}{r}\sum_{k=1}^{r}\tilde{X}_k$ is zero mean.

TABLE 1. **Expected values of standard normal order statistics.** The computations were done on a HP computer using Maple V. The first four values coincide with known "exact" values.

r	h_r^*	r	h_r^*	r	h_r^*
2	0.56418 95835	14	1.70338 1555	80	2.42677 4421
3	0.84628 43753	15	1.73591 3445	90	2.46970 0479
4	1.02937 5373	16	1.76599 1393	100	2.50759 3639
5	1.16296 4473	17	1.79394 1981	200	2.74604 2451
6	1.26720 6361	18	1.82003 1880	300	2.87776 6853
7	1.35217 8376	19	1.84448 1512	400	2.96817 8187
8	1.42360 0306	20	1.86747 5060	500	3.03669 9351
9	1.48501 3162	30	2.04276 0846	600	3.09170 2266
10	1.53875 2731	40	2.16077 7180	700	3.13754 7901
11	1.58643 6352	50	2.24907 3631	800	3.17679 1412
12	1.62922 7640	60	2.31927 8210	900	3.21105 5997
13	1.66799 0177	70	2.37735 9241	1000	3.24143 5777

TABLE 2. **Values of the absolute constants C_r^* for the case $\boldsymbol{\lambda} = \mathbf{0}$, $\boldsymbol{V} = \mathbb{I}$.** The computations were done on an HP computer using Maple V.

r	C_r^*	r	C_r^*	r	C_r^*
2	0.0	14	2.20924	80	5.08274
3	0.27566	15	2.31444	90	5.28802
4	0.55133	16	2.41374	100	5.47243
5	0.80002	17	2.50776	200	6.70147
6	1.02174	18	2.59705	300	7.43096
7	1.22030	19	2.68205	400	7.95237
8	1.39953	20	2.76316	500	8.35874
9	1.56262	30	3.41871	600	8.69193
10	1.71210	40	3.89695	700	8.97438
11	1.85003	50	4.27404	800	9.21958
12	1.97802	60	4.58561	900	9.43625
13	2.09740	70	4.85120	1000	9.63036

Consider the case $\lambda_i \neq 0$. In this case, we simply subtract λ_i from $\xi(n)$ and add it to b_i in equation (3.32). Then the modified $\xi(n)$ is identical to that obtained in the case $\lambda_i = 0$, and hence its limiting value is given by $C_{r,i}$ of (3.39).

To summarize, for $\varepsilon > -v_i^2/r$ and $\lambda_i \in I\!R$,

$$\mathrm{E}_i \nu_b = \frac{1}{D_i} \left[F_r\big(b_i + \lambda_i, D_i, v_i, \mathrm{E}_i \tilde{Y}_{1,i}^3\big) + \varkappa_i \right] , \tag{3.43}$$

where

$$F_r(a,q,u,g) = a + uh_r^* \sqrt{\frac{a}{q} + \frac{u^2(h_r^*)^2}{4q^2}} + \frac{u^2(h_r^*)^2}{2q} + \frac{gC_r^*}{6u^2} \,. \tag{3.44}$$

A similar expression holds for $\mathrm{E}_i\tau_a$ with b_i replaced by a_i and K added to $\varkappa_i$ in (3.43)

4. Applications and Results of Simulation

In this section, we consider examples which are meaningful in many applications such as target detection by multiple-element-resolution radar systems [**6**], [**23**], acquisition in code division multiple access (CDMA) communication systems [**33**] and others [**15**], [**14**], [**26**]. Simulation results are provided for these examples that verify the accuracy of the approximations obtained in the previous section. In addition, the performance of the MSPRT's is compared with that of non-sequential (fixed sample size) tests.

4.1. Example 1: Testing the Mean of an iid Gaussian Sequence. Suppose the observations are given by

$$X_n = \theta + \varsigma_n \,, \qquad n = 1, 2, \dots$$

where, under H_i, $\theta = \theta_i$, and $\varsigma_n \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise. The log-likelihood ratios of the observations are easily computed as

$$\Delta Z_{ij}(n) = \frac{\theta_i - \theta_j}{\sigma^2} X_n - \frac{\theta_i^2 - \theta_j^2}{2\sigma^2}$$

and the Kullback-Leibler distances are given by

$$D_{ij} = (\theta_i - \theta_j)^2 / 2\sigma^2 \,.$$

Consider the case of three hypotheses ($M = 3$) where $0 < \theta_1 - \theta_0 = \Delta_0 < \theta_2 - \theta_1 = \Delta_1$. Then $\min_{j \neq i} D_{ij}$ is achieved for only one j, and

$$D_0 = D_{01} = D_1 = D_{10} = \rho_0 \,; \quad D_2 = \rho_1$$

where we denoted $\rho_0 = \Delta_0^2/2\sigma^2$, $\rho_1 = \Delta_1^2/2\sigma^2$.

Also, suppose that the prior distribution is uniform, $\pi_i = 1/3$, and the loss function is zero-one (i.e. $W(j,i) = 1$ for all $j \neq i$). As noted earlier, for the zero-one loss function, $\{R_i(\delta)\}$ represent frequentist error probabilities. Suppose we are given risk constraints $\{\bar{R}_i\}$. Then, if we set $a_i = \log(\gamma_i/3\bar{R}_i)$ and $b_i = \log(2\gamma_i/3\bar{R}_i)$, we can apply the results of Corollary 3.1 to get

$$R_i(\delta_a) \approx \bar{R}_i \,, \quad \bar{R}_i/2 \leq R_i(\delta_b) \leq \bar{R}_i \,;$$

$$\mathrm{E}_i\tau_a \approx \frac{1}{D_i}\left[\log\left(\frac{\gamma_i}{3\bar{R}_i}\right) + \varkappa_i\right], \quad \mathrm{E}_i\nu_b \approx \frac{1}{D_i}\left[\log\left(\frac{2\gamma_i}{3\bar{R}_i}\right) + \varkappa_i\right], \quad i = 0, 1, 2$$

where γ_0, γ_1, $\varkappa_0$ and $\varkappa_1$ are calculated using techniques described in [**35**] and are given below:

$$\gamma_0 = \gamma_1 = \frac{1}{\rho_0}\exp\left\{-2\sum_{k=1}^{\infty}\frac{1}{k}\Phi\left(-\sqrt{\tfrac{\rho_0}{2}k}\right)\right\}; \tag{4.1}$$

$$\gamma_2 = \frac{1}{\rho_1}\exp\left\{-2\sum_{k=1}^{\infty}\frac{1}{k}\Phi\left(-\sqrt{\tfrac{\rho_1}{2}k}\right)\right\}; \tag{4.2}$$

$$\varkappa_0 = \varkappa_1 = 1 + \frac{\rho_0}{2} - \sqrt{2\rho_0}\sum_{k=1}^{\infty}\left[\frac{1}{\sqrt{k}}\varphi\left(\sqrt{\tfrac{\rho_0}{2}k}\right) - \sqrt{\tfrac{\rho_0}{2}k}\,\Phi\left(-\sqrt{\tfrac{\rho_0}{2}k}\right)\right]; \tag{4.3}$$

$$\varkappa_2 = 1 + \frac{\rho_1}{2} - \sqrt{2\rho_1}\sum_{k=1}^{\infty}\left[\frac{1}{\sqrt{k}}\varphi\left(\sqrt{\tfrac{\rho_1}{2}k}\right) - \sqrt{\tfrac{\rho_1}{2}k}\,\Phi\left(-\sqrt{\tfrac{\rho_1}{2}k}\right)\right]. \tag{4.4}$$

The performance of the tests δ_a and δ_b for two different test cases is given in Table 3. In the table, $\hat{R}_i$ and $\hat{\mathrm{E}}_i\tau$ are the estimates of the risks and ASS obtained by Monte Carlo techniques. Note that the second order asymptotics are considerably more accurate than the first order asymptotics. Note that for the designed values of the thresholds, the expected sample size for δ_b is slightly larger than that for δ_a. This is consistent with the result of Corollary 3.1. As we noted in the discussion following Corollary 3.1, the difference between the two tests will be negligible if δ_b can be designed to meet the risk constraints $\bar{R}_i$ more tightly. Experimentation indicates that $\bar{R}_i$ is better approximated by setting $\{b_i\}$ using the lower bound in (3.17), i.e. $b_i = a_i$. These results are also shown in Table 3.

Note that when $\Delta_0 = \Delta_1 = \Delta$, we have $D_{10} = D_{12}$, and, therefore, the above approximations for $\mathrm{E}_1\tau_a$ and $\mathrm{E}_1\nu_b$ are no longer valid. In that case, Theorem 3.3 needs to be used for H_1. It is interesting to note that $\boldsymbol{V}_1$ has the form $v_1^2\mathbb{I} + \varepsilon$ considered in Section 3.3.2 with $v_1^2 = 4\rho$, $\varepsilon = 1/2 - 2\rho$, and $\lambda_{k,1} = \lambda_1 = 0$. Thus (3.43) may be applied to calculate accurate approximations for the ASS in this case. Note that a further simplification results in this case since $C_2^* = 0$ (see Table 2). Typical results for this symmetric case are given in Table 4.

The performance of the sequential tests may also be compared with fixed sample size (FSS) tests. It is not easy to design an FSS test that meets the individual risk constraints $\{\bar{R}_i\}$. However, the FSS test that meets the corresponding constraint $\bar{R}_{\mathrm{B}}$ on the total Bayes risk $\sum_i R_i$ is easily shown to be of the form:

$$\begin{array}{ll} \text{choose } H_0 & \text{if } \bar{X}_T \le (\theta_0+\theta_1)/2 \\ \text{choose } H_2 & \text{if } \bar{X}_T \ge (\theta_1+\theta_2)/2 \\ \text{choose } H_1 & \text{otherwise} \end{array}$$

where $\bar{X}_T = \sum_{n=1}^{T} X_n/T$, and T is the (fixed) number of observations. It can be shown that the Bayes risk for this test is given by

$$R_{\mathrm{B}} = \frac{2}{3}\left[2 - \Phi\left(\sqrt{T}(\theta_1-\theta_0)/2\sigma\right) - \Phi\left(\sqrt{T}(\theta_2-\theta_1)/2\sigma\right)\right].$$

Using this equation, the value of T that meets the constraint $\bar{R}_{\mathrm{B}}$ can be found. Comparing T, $\mathrm{E}\tau_a = \sum_i \pi_i \mathrm{E}_i\tau_a$, and $\mathrm{E}\nu_b = \sum_i \pi_i \mathrm{E}_i\nu_b$ in Tables 3 and 4, we see that the sequential tests are two to three times faster than the corresponding FSS tests.

4.2. Example 2: The Slippage Problem. As a second application of the above results, we consider the problem of detecting a single target in a multi-channel (multi-resolution) system which is essentially a multi-sample slippage problem [**13**]. Suppose there are N channels. In the i-th channel one observes the process $X_{i,n}$ and all components may be observed simultaneously, i.e. $\boldsymbol{X}_n = (X_{1,n}, \dots, X_{N,n})$, $n \ge 1$. 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 i-th, say (hypothesis H_i). Thus, the number of hypotheses $M = N+1$. The goal is to detect a signal as soon as possible and to indicate the number of the channel where the signal is located.

Under hypothesis H_0, $X_{1,n}, \dots, X_{N,n}$ are mutually independent and distributed with common density $g_0(x)$ which describes the distribution of noise, and, under H_i, all $X_{k,n}$ are mutually independent, $X_{1,n}, \dots, X_{i-1,n},\ X_{i+1,n}, \dots, X_{N,n}$ are distributed with common density $g_0(x)$ and $X_{i,n}$ has density $g_i(x)$. The latter describes the distribution of a mixture of signal and noise.

Assume, for simplicity, that $g_i(x) = g_1(x)$, $i = 1, \dots, N$. In other words, the statistical properties of the observed data do not depend on the number of the channel where the signal is located. Then

$$f_0(\boldsymbol{x}_n) = \prod_{k=1}^{N} g_0(x_{k,n}); \quad f_i(\boldsymbol{x}_n) = g_1(x_{i,n}) \prod_{\substack{k=1 \\ k \ne i}}^{N} g_0(x_{k,n}), \quad i = 1, \dots, N\,. \tag{4.5}$$

Therefore, the log-likelihood ratios are given by

$$\Delta Z_{i0}(n) = -\Delta Z_{0i}(n) = \log \frac{g_1(X_{i,n})}{g_0(X_{i,n})}\,, \quad i = 1, \dots, N;$$

$$\Delta Z_{ij}(n) = \Delta Z_{i0}(n) + \Delta Z_{0j}(n)\,, \quad i \ne j;\ i, j \ne 0\,.$$

By the symmetry of the problem, the distances $D_{i0} = q_1$ (say) are the same for $i = 1, \dots, N$, and so are $D_{0i} = q_0$ (say) where

$$q_1 = \int \log\left[\frac{g_1(x)}{g_0(x)}\right] g_1(x)dx\,, \quad \text{and } q_0 = \int \log\left[\frac{g_0(x)}{g_1(x)}\right] g_0(x)dx \tag{4.6}$$

Also the distances between the non-null hypotheses are given by $D_{ij} = q_1 + q_0$, $\forall\ i, j \ne 0$. Hence $D_0 = q_0$, $D_i = q_1$, $i = 1, \dots, N$. This means that for hypothesis H_0, we have the fully symmetric case with $D_0 = \min_{j \ne 0} D_{0j} = D_{0i} = q_0$, $i = 1, \dots, N$; while for any other hypothesis H_i, $i \ne 0$, the asymmetric condition (3.1) holds with $j^*(i) = 0$.

Further, we assume that the conditional prior distribution of the signal location is uniform, i.e. $\Pr(H_i \mid H_0 \text{ is incorrect}) = 1/N$. In other words, if $\pi_0 = \Pr(H_0)$ is the prior probability of signal absence, then $\pi_i = \Pr(H_i) = (1-\pi_0)/N$, $i = 1, \dots, N$.

Finally, given the symmetry of the problem, the following three-valued loss function is appropriate.

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

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

$$R_i(\delta) = \frac{(1-\pi_0)W_2}{N} \sum_{\substack{j=1,\\ j\neq i}}^{N} \alpha_{ji} + \pi_0 W_0 \alpha_{0i} \,, \quad i = 1, \dots, N \,,$$

$$R_0(\delta) = \frac{(1-\pi_0)W_1}{N} \sum_{j=1}^{N} \alpha_{j0} \,.$$

Under these assumptions, it is clear that we require to specify only two thresholds for each sequential test, a_0, $a_i = a_1$ and b_0, $b_i = b_1$ $(i = 1, \dots, N)$. Then, by symmetry, the conditional error probabilities for both tests satisfy the following properties:

$$\alpha_{j0} = \alpha_{10} \;\forall j \neq 0 \,, \quad \alpha_{ji} = \alpha_{12} \;\forall i \neq 0, j \neq 0 \text{ s.t. } i \neq j \,, \quad \text{and } \alpha_{0j} = \alpha_{01} \;\forall j \neq 0.$$

Thus, we have

$$\begin{aligned} R_i(\delta) &= R_1(\delta) = \frac{(1-\pi_0)W_2(N-1)}{N}\alpha_{12} + \pi_0 W_0 \alpha_{01} \,, \quad i = 1, \dots, N \,; \\ R_0(\delta) &= (1-\pi_0)W_1 \alpha_{10} \,. \end{aligned}$$

To meet constraints $\{\bar{R}_i\}$, we set $a_i = \log[\gamma_1(1-\pi_0)/N\bar{R}_i]$ and $b_i = \log[\gamma_1(1-\pi_0)/\bar{R}_i]$, for $i = 1, \dots, N$. Then, by Corollary 3.1, we get

$$R_i(\delta_a) \approx \bar{R}_i \,, \quad \frac{\bar{R}_i}{N} \leq R_i(\delta_b) \leq \bar{R}_i \,, \quad i = 1, \dots, N \,. \tag{4.7}$$

For hypothesis H_0, we may use the bounds given in (3.37) to set $a_0 = \log(\pi_0/\bar{R}_0)$ and $b_0 = \log(N\pi_0/\bar{R}_0)$ and thus guarantee that the risk constraint $\bar{R}_0$ is met. However, as we see in numerical results, setting $b_i = a_i, i = 1, \dots, N$, $a_0 = \log(\gamma_0\pi_0/\bar{R}_0)$, and $b_0 = \log(2\gamma_0\pi_0/\bar{R}_0)$ result in tests that approximate $\bar{R}_0$ more accurately (of course, without the guarantee of being below $\bar{R}_0$).

To compute the expected sample sizes, for $i \neq 0$, we apply Theorem 3.1 to get

$$\mathrm{E}_i \tau_a \approx \frac{1}{q_1}\big(a_1 + \lambda_1 + \varkappa_1\big) \,, \quad \mathrm{E}_i \nu_b \approx \frac{1}{q_1}\big(b_1 + \lambda_1 + \varkappa_1\big) \,, \tag{4.8}$$

where

$$\lambda_1 = \log w_{0i} = \log\big[W_0 \pi_0 N/(1-\pi_0)\big]$$

To compute the ASS under H_0, we need to use Theorem 3.3. By the symmetry of the problem, $r = N$ in this case. In order to compute the constants $h_{r,i}$ and $C_{r,i}$, it is convenient to use the measure $\mu(\boldsymbol{x}_n) = \prod_{k=1}^{n} g_1(x_{k,n})$ as the dominating measure for defining densities. Then the likelihood functions of (4.5) get modified to

$$f_0(\boldsymbol{x}_n) = \prod_{k=1}^{N} \frac{g_0(x_{k,n})}{g_1(x_{k,n})}; \quad f_i(\boldsymbol{x}_n) = g_1(x_{i,n}) \prod_{\substack{k=1\\ k\neq i}}^{N} \frac{g_0(x_{k,n})}{g_1(x_{k,n})}, \quad i = 1, \dots, N \,.$$

With these likelihood functions, the vector $\boldsymbol{Y}_0$ has components given by:

$$Y_{k,0} = \sum_{\substack{m=1\\ m\neq k}}^{N} \frac{g_0(x_{m,n})}{g_1(x_{m,n})} - \mathrm{E}_0 \left[\sum_{\substack{m=1\\ m\neq k}}^{N} \frac{g_0(x_{m,n})}{g_1(x_{m,n})} \right] .$$

It is easy to show that the covariance matrix $\boldsymbol{V}_0$ has the form $v_0^2\mathbb{I}+\varepsilon$, considered in Section 3.3.2, where

$$v_0^2 = \mathrm{Var}_0\left[\log\frac{g_1(X)}{g_0(X)}\right]$$

with $\mathrm{Var}_0[\cdot]$ being the variance relative to the density function $g_0(x)$, and where

$$\varepsilon = (N-2)v_0^2\ .$$

Furthermore

$$\lambda_{k,0} = \lambda_0 = \log w_{k,0} = \log[(1-\pi_0)W_1/N\pi_0]\ .$$

Thus (3.43) may be applied to get

$$\mathrm{E}_0\nu_b \approx \frac{1}{q_0}\left[F_N\big(b_0+\lambda_0, q_0, v_0, \mathrm{E}_0\tilde{Y}_{1,0}^3\big) + \varkappa_0\right]\ , \tag{4.9}$$

$$\mathrm{E}_0\tau_a \approx \frac{1}{q_0}\left[F_N\big(a_0+\lambda_0, q_0, v_0, \mathrm{E}_0\tilde{Y}_{1,0}^3\big) + \varkappa_0 + K\right]\ , \tag{4.10}$$

where $0 \le K \le \log N$, and $F_r(\cdot,\cdot,\cdot,\cdot)$ is as defined in (3.44).

Special case: signal always present

We now consider the situation where the null hypothesis H_0 is excluded from consideration, i.e. we are certain that the signal is present in the system and only its location is to be determined. In this case we have a fully symmetric set of N hypotheses.

It is easy to see that the distances between the hypotheses are $D_{ij} = q_1+q_0$ for all $i,j = 1,\dots,N, j \ne i$, where q_1 and q_0 are as defined in (4.6).

Due to the symmetry, one may set $a_i = a$, $b_i = b$, $\pi_i = 1/N$ for $i = 1,\dots,N$, and assume a zero-one loss function. The risks $\{R_i(\delta)\}$ are then all equal:

$$R_i(\delta) = R(\delta) = (N-1)\alpha_{12}/N$$

where $\alpha_{12} = \mathrm{P}_j(d=i), \forall j \ne i$ is the probability of signal mixing. In order to meet a risk constraint $\bar{R}$, we use the bounds given in (3.37) to set $a = \log[1/(N\bar{R})]$ and $b = \log[(N-1)/(N\bar{R})]$. This will guarantee that the constraint $\bar{R}$ is met. However, as seen in numerical results, setting $a = \log[\gamma/(N\bar{R})]$ and $b = \log[2\gamma/(N\bar{R})]$ (where $\gamma = \gamma_i$) results in tests that approximate $\bar{R}$ more accurately.

To compute the ASS under any of the hypotheses, say H_1, we first note that we have a situation covered under Theorem 3.3 with $r = N-1$. In order to compute the required constants, it is convenient to use the measure $\mu(\boldsymbol{x}_n) = \prod_{k=1}^n g_0(x_{k,n})$ as the dominating measure for defining densities. Then the likelihood functions are given by

$$f_i(\boldsymbol{x}_n) = \frac{g_1(x_{i,n})}{g_0(x_{i,n})}, \quad i = 1,\dots,N\ .$$

The components $Y_{k,1}$ of the vector $\boldsymbol{Y}_1$ are obviously i.i.d. in this case, with variance given by

$$v^2 = \mathrm{Var}_0\left[\log\frac{g_1(X)}{g_0(X)}\right]$$

where $\mathrm{Var}_0[\cdot]$ denotes the variance relative to the density function $g_0(x)$. Thus $\boldsymbol{V}_1 = v^2\mathbb{I}$. Furthermore, by symmetry, $\lambda_{k,1} = \lambda_1 = 0$. Applying (3.43), we get, for all $i = 1, \dots, N$,

$$\mathrm{E}_i\nu_b \approx \frac{1}{q_0+q_1}\Big[F_{N-1}\big(b, q_0+q_1, v, \mathrm{E}_1 Y_{1,1}^3\big) + \varkappa\Big] , \tag{4.11}$$

$$\mathrm{E}_i\tau_a \approx \frac{1}{q_0+q_1}\Big[F_{N-1}\big(a, q_0+q_1, v, \mathrm{E}_1 Y_{1,1}^3\big) + \varkappa + K\Big] . \tag{4.12}$$

Here $0 \le K \le \log(N-1)$, $\varkappa = \varkappa_i$, and $F_r(\cdot,\cdot,\cdot,\cdot)$ is as defined in (3.44).

4.2.1. *Detection of deterministic signals in white Gaussian noise.* Consider the problem of detection of a deterministic pulse signal in an N-channel radar in the presence of additive white Gaussian noise. The pre-processing scheme consists of a matched filter, matched to the pulse. Then the hypotheses are

$$H_0: \;\; X_{k,n} = \varsigma_{k,n} \quad \text{for} \quad k = 1, \dots, N ,$$

$$H_i: \;\; X_{k,n} = \varsigma_{k,n} \quad \text{for} \quad k \ne i; \qquad X_{i,n} = \theta + \varsigma_{i,n}$$

where $\varsigma_{k,n} \sim \mathcal{N}(0,\sigma^2)$ are iid Gaussian variables (both θ and σ^2 are assumed to be known). Note that

$$g_0(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\Big\{-\frac{x^2}{2\sigma^2}\Big\} , \quad g_1(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\Big\{-\frac{(x-\theta)^2}{2\sigma^2}\Big\} .$$

Let $\rho = \theta^2/2\sigma^2$ denote the signal-to-noise ratio (SNR). Then it is easy to show that q_0 and q_1 of (4.6) are both equal to ρ. The constants $\gamma_1 = \gamma_0$ and $\varkappa_1 = \varkappa_0$ are obtained by substituting $\rho_0 = \rho$ in (4.1) and (4.3), respectively. The vector $\boldsymbol{Y}_0$ is Gaussian and zero mean. From (3.41), it is clear that $\tilde{\boldsymbol{Y}}_0$ is Gaussian and zero mean as well. Hence $\mathrm{E}_0 Y_{1,0}^3 = 0$. The constant v_0^2 can be shown to equal 2ρ. Using these constants, we can compute the ASS for both tests for any given costs, priors and risk constraints. Sample results are given in Table 5. Note that the second order asymptotics are considerably more accurate than the first order asymptotics, particularly for hypothesis H_0.

In the completely symmetric case (signal is always present but its location is unknown), the required constants are given by $q_0 + q_1 = 2\rho$, $v_1^2 = 2\rho$, $\mathrm{E}_0 Y_{1,1}^3 = 0$, and γ and $\varkappa$ are obtained by substituting $\rho_0 = 2\rho$ in (4.1) and (4.3), respectively.

For the completely symmetric case, the best FSS test chooses H_i if $\bar{X}_{i,T} = \max_j \bar{X}_{j,T}$, where $\bar{X}_{i,T} = \sum_{n=1}^{T} X_{i,n}/T$, and T is the (fixed) number of observations. It can be shown that the risk R for this test is given by

$$R = \frac{1}{N} - \frac{1}{N\sqrt{2\pi}}\int_{-\infty}^{\infty} \big[\Phi(x)\big]^{N-1} \exp\left[-\frac{\big(x-\sqrt{T}\theta\big)^2}{2\sigma^2}\right] dx.$$

Using this equation, the value of T that meets the constraint $\bar{R}$ can be found. Comparing T, $\mathrm{E}\tau_a$ and $\mathrm{E}\nu_b$ in Table 6, we see that the sequential tests are two to three times faster than the FSS test.

4.2.2. *Detection of fluctuating signals.* Now suppose that one wants to detect a fluctuating signal in additive white Gaussian noise from data at the output of a pre-processing scheme which consists of a match filter and square-law detector [**2**]. Under the assumption that the signal has slow Gaussian fluctuations within pulses and fast fluctuations between pulses (the Swerling II model), the observed data is exponentially distributed and independent. After appropriate normalization,

$$g_0(x) = \exp(-x)\, \boldsymbol{I}\{x \ge 0\}\,, \quad g_1(x) = \frac{1}{1+\rho}\, \exp\left[-\frac{x}{1+\rho}\right] \boldsymbol{I}\{x \ge 0\}\,.$$

It is easy to show that q_0 and q_1 of (4.6) are given by:

$$q_0 = \log(1+\rho) - \rho/(1+\rho)\,, \quad \text{and} \quad q_1 = \rho - \log(1+\rho)$$

and that

$$\Delta Z_{i0}(n) = \tfrac{\rho}{1+\rho} X_{i,n} - \log(1+\rho), \quad \Delta Z_{ij}(n) = \tfrac{\rho}{1+\rho}[X_{i,n} - X_{j,n}]\,.$$

Since under H_i the distribution of $\Delta Z_{i0}(n)$ has an exponential right tail,

$$\mathrm{P}_i\big\{\Delta Z_{i0}(n) > z\big\} = \frac{1}{(1+\rho)^{1/\rho}} \exp\left\{-\tfrac{1}{\rho} z\right\} \boldsymbol{I}\{z \ge -\log(1+\rho)\}, \tag{4.13}$$

the distribution of the overshoot $r_i(c) = Z_{i0}(\tau_{i0}) - c$ is exponential for all $c > 0$ [**31**],[**35**]:

$$\mathrm{P}_i\big\{r_i(c) > r\big\} = \exp\left\{-\tfrac{1}{\rho} r\right\} \boldsymbol{I}\{r \ge 0\}.$$

Thus,

$$\varkappa_1 = \rho, \quad \gamma_1 = 1/(1+\rho)$$

and it remains to compute the constants γ_0 and $\varkappa_0$. By Port [**24**], the density of the overshoot $r_0(c) = Z_{0i}(\tau_{0i}) - c$ under H_0 may be written in the form

$$p_0(y) = \tfrac{1}{q_0} \mathrm{P}_0\Big\{\min_{n \ge 1} Z_{0i}(n) > y\Big\} = \tfrac{1}{q_0} \mathrm{P}_0\big\{Z_{0i}(n) > y, n \ge 1\big\}\,.$$

Introduce the stopping time

$$\tau_-(y) = \inf\{n \ge 1 : Z_{0i}(n) \le y\} \quad \text{for } y \le \ln(1+\rho)\,.$$

Obviously,

$$\begin{aligned} \mathrm{P}_0\{Z_{0i}(n) > y, n \ge 1\} &= \mathrm{P}_0\{\tau_-(y) = \infty\} = 1 - \mathrm{P}_0\{\tau_-(y) < \infty\} \\ &= 1 - \int_{\{\tau_- < \infty\}} \exp\{Z_{0i}(\tau_-)\}\, d\mathrm{P}_i = 1 - e^y \mathrm{E}_i e^{-\chi_-(y)} \end{aligned}$$

where $\chi_-(y) = y + Z_{i0}(\tau_-)$. It follows from (4.13) that for any $y \le \ln(1+\rho)$

$$\mathrm{P}_i\{\chi_-(y) \ge t\} = \exp\left\{-\tfrac{1}{\rho} t\right\} \boldsymbol{I}\{t \ge 0\}, \quad \mathrm{E}_i e^{-\chi_-(y)} = (1+\rho)^{-1}$$

and hence

$$p_0(y) = \frac{1}{q_0}\left[1 - \tfrac{1}{1+\rho} e^y\right] \boldsymbol{I}\{0 \le y \le \ln(1+\rho)\}\,.$$

Using this last expression we finally obtain

$$\gamma_0 = \frac{q_1}{q_0}\gamma_1, \quad \varkappa_0 = \frac{1}{2q_0}\left[\log(1+\rho)\right]^2 - 1\,.$$

The constant v_0^2 is given by

$$v_0^2 = \frac{\rho^2}{(1+\rho)^2}\,.$$

It is easy to show that

$$Y_{k,0} = \frac{\rho}{1+\rho} \sum_{\substack{m=1 \\ m\neq k}}^{N} (1 - X_{m,n}).$$

Now, starting with the definition of $\tilde{Y}_{1,0}$ given in (3.41), a series of straightforward calculations leads to

$$\mathrm{E}_0 \tilde{Y}_{1,0}^3 = \frac{2\rho^3}{(1+\rho)^3} \left[1 + \frac{4}{N^2} - \frac{6}{N} \right] .$$

Using these constants, we can compute the ASS for both tests for any given costs, priors and risk constraints. Sample results are given in Table 7.

For the completely symmetric case (we don't test the hypothesis H_0),

$$\begin{aligned} \mathrm{P}_i\big\{\Delta Z_{ij}(n) > z\big\} &= \mathrm{P}\left\{X_i > \tfrac{1+\rho}{\rho} z + X_j\right\} \\ &= \int_0^\infty \mathrm{P}_i\left\{X_i > \tfrac{1+\rho}{\rho} z + x\right\} e^{-x} dx \\ &= \int_{-\infty}^\infty \boldsymbol{I}\{x \ge 0\} \boldsymbol{I}\left\{x \ge \tfrac{1+\rho}{\rho} z\right\} \exp\left\{-\left(\tfrac{1}{\rho} z + \tfrac{1}{1+\rho} x\right)\right\} e^{-x} dx \\ &\quad + \int_{-\infty}^\infty \boldsymbol{I}\{x \ge 0\} \boldsymbol{I}\left\{x \le \tfrac{1+\rho}{\rho} z\right\} e^{-x} dx \end{aligned}$$

After some simple algebra, we obtain

$$\begin{aligned} \mathrm{P}_i\big\{\Delta Z_{ij}(n) > z\big\} = &\left[1 + \frac{1+\rho}{2+\rho} \exp\left\{\tfrac{1+\rho}{2+\rho} z\right\} - \exp\left\{\tfrac{1+\rho}{\rho} z\right\} \right] \boldsymbol{I}\{z < 0\} \\ &+ \frac{1+\rho}{2+\rho} \exp\left\{-\tfrac{1}{\rho} z\right\} \boldsymbol{I}\{z \ge 0\}. \end{aligned}$$

Hence again the distribution has an exponential right tail and the overshoot $r_{ij}(c)$ has the exponential distribution with the parameter $1/\rho$. Thus

$$\varkappa_i = \rho, \quad \gamma_i = 1/(1+\rho).$$

The constant v^2 obviously equals v_0^2, and $\mathrm{E}_1 Y_{1,1}^3$ is easily seen to be given by

$$\mathrm{E}_i Y_{j,i}^3 = \frac{2\rho^3}{(1+\rho)^3} .$$

Sample results for this case are given in Table 8.

5. Conclusions

We studied two constructions of sequential tests for multiple hypotheses. The MSPRT δ_b has the advantage that it is easier to implement, while, as shown in Sections 3 and 4, δ_a is easier to design to meet given risk requirements. We established in [**13**] that both MSPRT's asymptotically minimize any positive moment of the stopping time distribution, under general statistical models for the observations. This makes the MSPRT's attractive candidates for practical applications, and it is hence of interest to obtain analytical approximations for the expected sample sizes of these tests.

Simulation results for several examples and various conditions show that while the first order approximations to the expected sample size are fairly inaccurate in most cases, the derived higher order approximations (up to a vanishing term) are accurate not only for large but also for moderate sample sizes, which are typical for many applications. This is especially true in cases where the set $\Gamma_i = \{j : D_{ij} = D_i\}$ consists of more than a single point $j^*(i)$. But even in the asymmetric situation when $j^*(i)$ is unique, the higher order approximations are substantially more accurate compared to the first order ones.

For the test δ_a, the risk constraints are sharply met by setting the thresholds as specified in Corollary 3.1. This is true not just in the asymmetric case (as may be expected from Corollary 3.1) but also in the general case. For δ_b, on the other hand, it is more difficult to meet the risk constraints. But if the thresholds for δ_b are chosen (by trial and error) to meet the constraints tightly, then the average sample sizes are the same as those for δ_a.

The results presented in this paper complete the asymptotic analysis initiated in [**4**] for the case of i.i.d. observations. Since the MSPRT's studied in this paper are asymptotically optimal under more general statistical models for observations, it would be of interest to analyze the performance of the MSPRT's under these models. We leave this as an open problem for future research.

Appendix

Proof of Theorem 3.1. Consider the test δ_a. We first show that

$$\mathrm{E}_i\tau_i = \frac{1}{D_i}\bigl(a_i + \log w_{j^*i} + \varkappa_i\bigr) + o(1) \qquad \text{as} \quad \min_k a_k \to \infty\,. \tag{A.1}$$

The above result follows from (3.10), (3.8) and an application of Theorem 4.5 of [**35**]. However, before we can apply Theorem 4.5 of [**35**], the validity of the following four conditions has to be checked:

$$\xi_i(n) \quad \text{converges in distribution to a random variable} \quad \xi\,; \tag{A.2}$$

$$\sum_{n=1}^{\infty} \mathrm{P}_i\bigl\{\xi_i(n) \le -\varepsilon n\bigr\} < \infty \quad \text{for some} \quad 0 < \varepsilon < D_i\,; \tag{A.3}$$

$$\mathrm{P}_i\bigl(\tau_i \le \varepsilon a_i/D_i\bigr) = o\bigl(1/a_i\bigr) \quad \text{for some} \quad 0 < \varepsilon < 1 \qquad \text{as} \quad a_i \to \infty\,; \tag{A.4}$$

$$\max_{0\le k\le n} |\xi_i(n+k)|\,,\ n \ge 1, \quad \text{are uniformly integrable.} \tag{A.5}$$

By (3.8), the condition (A.2) is true with $\xi = \log w_{j^*i}$. Condition (A.3) holds since $\xi_i(n) \ge \log w_{j^*i}$. Obviously,

$$\tau_i \ge t_i = \inf\{n \ge 1 : Z_{ij^*}(n) \ge \tilde a_i\}\,, \quad \tilde a_i = a_i + \log w_{j^*i}$$

and hence

$$\begin{aligned}\mathrm{P}_i(\tau_i \le m) \le \mathrm{P}_i(t_i \le m) &= \mathrm{P}_i\left\{\max_{n\le m} Z_{ij^*}(n) \ge \tilde a_i\right\}\\ &= \mathrm{P}_i\left\{\max_{n\le m} \exp[\lambda Z_{ij^*}(n)] \ge \exp(\lambda \tilde a_i)\right\}, \quad \lambda > 0\,.\end{aligned} \tag{A.6}$$

It is easily checked that the process $\exp[\lambda Z_{ij^*}(n)]$, $n \ge 1$, is a P_i-submartingale for any $\lambda > 0$. Applying Doob's inequality for submartingales [**21**] to the last term in (A.6), we obtain

$$\mathrm{P}_i(\tau_i \le m) \le \exp(-\lambda\tilde a_i)\mathrm{E}_i \exp[\lambda Z_{ij^*}(m)] = \exp(-\lambda\tilde a_i)\exp[m\rho_i(\lambda)]$$

where

$$\rho_i(\lambda) = \log\left(\mathrm{E}_{j^*}\left[\frac{f_i(\boldsymbol{X}_1)}{f_{j^*}(\boldsymbol{X}_1)}\right]^{1+\lambda}\right) > 0 \qquad \text{for} \quad \lambda > 0\,.$$

Taking $m = \varepsilon a_i/D_i$, we finally obtain

$$\mathrm{P}_i(\tau_i \le \varepsilon a_i/D_i) \le \exp[-\tilde{a}_i(\lambda - \varepsilon\rho_i(\lambda)/D_i)]\,. \tag{A.7}$$

Choosing ε sufficiently small, $\varepsilon < \lambda D_i/\rho_i(\lambda)$, one may see that $\mathrm{P}_i(\tau_i \le \varepsilon a_i/D_i)$ is bounded by $\exp(-Ca_i)$, $C > 0$, and hence the condition (A.4) holds true.

Thus, it remains to check the condition (A.5) (uniform integrability) which is a straightforward but tedious task (see, e.g., [**4**]).

Therefore, all conditions of Theorem 4.5 in [**35**] are satisfied. The use of this theorem yields (A.1) for large a_i. To prove the assertion of the theorem for the test procedure δ_a, it remains to prove that

$$\mathrm{E}_i(\tau_i - \tau_a) = o(1) \qquad \text{as} \quad \min_k a_k \to \infty\,. \tag{A.8}$$

To this end, we first observe that

$$\mathrm{E}_i(\tau_i - \tau_a) = \mathrm{E}_i\big\{(\tau_i - \tau_a)\boldsymbol{I}(\tau_a \ne \tau_i)\big\} \le 2\mathrm{E}_i\big\{\tau_i\,\boldsymbol{I}(\tau_a \ne \tau_i)\big\}\,.$$

Using Schwarz's inequality, we get

$$\mathrm{E}_i(\tau_i - \tau_a) \le 2\sqrt{\mathrm{E}_i\tau_i^2}\sqrt{\mathrm{P}_i(\tau_a \ne \tau_i)}\,.$$

By Lemma 2.1 in [**13**]

$$\sum_{i\ne k} \pi_i W(i,k)\mathrm{P}_i(d_a = k) \le \pi_k\, e^{-a_k}$$

which implies

$$\mathrm{P}_i(d_a = k) \le w_{ik}^{-1}\, e^{-a_k} \quad \text{for all} \quad k \ne i\,.$$

In turn, the latter inequality yields

$$\begin{aligned}\mathrm{P}_i(\tau_a \ne \tau_i) = \mathrm{P}_i(d_a \ne i) &= \sum_{k\ne i} \mathrm{P}_i(d_a = k) \le \sum_{k\ne i} w_{ik}^{-1}\, \exp(-a_k)\\ &\le \frac{N}{\min_{k\ne i} w_{ik}} \exp(-\min_k a_k)\end{aligned}$$

Now $\mathrm{E}_i\tau_i^2 \sim a_i^2/D_i^2$ by Corollary 3.1 in [**13**]. Thus,

$$\mathrm{E}_i(\tau_i - \tau_a) \le Ca_i \exp\Big(-\tfrac{1}{2}\min_k a_k\Big) \to 0 \qquad \text{as} \quad \min_k a_k \to \infty$$

and the theorem follows for the MSPRT δ_a.

For the test δ_b the argument is quite similar. In just the same way as above one can prove that the conditions (A.2)–(A.5) hold for the process $Y_i(n)$, $n \ge 1$. Then using (3.3)–(3.12) and Theorem 4.5 in [**35**], we obtain

$$\mathrm{E}_i\nu_i = \frac{1}{D_i}\big(b_i + \log w_{j^*i} + \varkappa_i\big) + o(1) \qquad \text{as} \quad \min_k b_k \to \infty\,.$$

The rest of the argument is essentially the same. ■

Proof of Theorem 3.3. Consider the test procedure δ_b. Arguments identical to those used in the proof of Theorem 3.1 may be used to establish that it is sufficient to prove (3.35) only for ν_i.

Now recall from (3.32) that

$$\nu_i = \inf\big\{n : S_i(n) \ge \xi(n) + A(n, b_i)\big\}\,, \tag{A.9}$$

where $S_i(n) = Z_i(n) - n\mu_{i[M-1]}$ is a random walk with increments having positive mean $\mathrm{E}_i Z_i(1) - \mu_{i[M-1]} = D_i$, $A(n, b_i) = b_i + h_{r,i}\sqrt{n}$ and

$$\xi(n) = \max_{k \neq i}\bigl[\log w_{ki} + Z_k(n) - n\mu_{i[M-1]}\bigr] - h_{r,i}\sqrt{n}\,. \tag{A.10}$$

It is desirable to replace the maximization over $k \neq i$ in (A.10) by a maximization over only those k corresponding to the r nearest hypotheses. To this end, let

$$L = \sup\Bigl\{n : \max_{j=M-r,\dots,M-1}\bigl(\log w_{\langle j\rangle i} + Z_{\langle j\rangle}(n)\bigr) < \max_{j=1,\dots,M-r-1}\bigl(\log w_{\langle j\rangle i} + Z_{\langle j\rangle}(n)\bigr)\Bigr\}\,,$$

where $\langle j\rangle$ is as defined in (3.28).

Due to assumption (3.26), we have

$$\begin{aligned}
\mathrm{P}_i(L \geq n) &= \mathrm{P}_i\left\{\max_{j \leq M-r-1}\Bigl(\log w_{\langle j\rangle i} + Z_{\langle j\rangle}(k)\Bigr) > \max_{j > M-r-1}\Bigl(\log w_{\langle j\rangle i} + Z_{\langle j\rangle}(k)\Bigr) \text{ for some } k \geq n\right\} \\
&\leq \mathrm{P}_i\left\{\max_{j \leq M-r-1}\left(\frac{\log w_{\langle j\rangle i}}{k} + \frac{\tilde{Z}_j(k)}{k}\right) > \varepsilon \text{ for some } k \geq n\right\} \\
&\leq \mathrm{P}_i\left\{\sup_{k \geq n}\left(\max_{j \leq M-r-1}\frac{\tilde{Z}_j(k)}{k}\right) > \varepsilon - \frac{\max_{j \leq M-r-1}\log w_{\langle j\rangle i}}{n}\right\} \\
&\leq \sum_{j \leq M-r-1}\mathrm{P}_i\left\{\sup_{k \geq n}\left(\tfrac{1}{k}\bigl|\tilde{Z}_j(k)\bigr|\right) > \varepsilon_1\right\}\,,
\end{aligned}$$

where

$$\varepsilon = \mu_{i[M-1]} - \mu_{i[M-r-1]} > 0\,, \quad \varepsilon_1 = \varepsilon - \frac{1}{n}\max_{j \leq M-r-1}\log w_{\langle j\rangle i}\,,$$
$$\tilde{Z}_j(k) = Z_{\langle j\rangle}(k) - k\mu_{i[j]} - Z_{\langle M-1\rangle}(k) + k\mu_{i[M-1]}\,.$$

Since $\tilde{Z}_j(k)$ is a random walk with mean zero and finite second moment, $\mathrm{E}_i|\tilde{Z}_j(1)|^2 < \infty$, by the Baum-Katz rate of convergence in the law of large numbers [**3**][1],

$$\sum_{n \geq 1}\mathrm{P}_i\left(\sup_{k \geq n}\frac{|\tilde{Z}_j(k)|}{k} \geq \varepsilon_1\right) < \infty \quad \text{for all } \ \varepsilon_1 > 0\,,$$

which along with the above inequality yields

$$\mathrm{E}_i L = \sum_{n \geq 1}\mathrm{P}_i(L \geq n) < \infty\,.$$

Hence, for $n > L$, we can write $\xi(n)$ of (A.10) as

$$\xi(n) = \sqrt{n}[\zeta(n) - h_{r,i}] \tag{A.11}$$

with

$$\zeta(n) = \frac{1}{\sqrt{n}}\max_{1 \leq k \leq r}\bigl[\lambda_{k,i} + Z_{\langle M-r-1+k\rangle}(n) - n\mu_{i[M-1]}\bigr]\,, \tag{A.12}$$

[1]See also Chow and Lai [**8**] for a one-sided versions that may be applied in the case considered.

where $\lambda_{k,i}$ is defined in (3.31). The proof of (3.35) for ν_i runs almost parallel to that of Lemma 1 in Dragalin [**11**] and is based on Theorem 3 of Zhang [**36**]. In order to apply this theorem the following conditions need to be checked.

$$\mathrm{P}_i\big\{\nu_i \le (1-\varepsilon)b_i/D_i\big\} = o\big(b_i^{-1}\big) \text{ as } b_i \to \infty \text{ for some } 0<\varepsilon<1; \tag{A.13}$$

$$n\mathrm{P}_i\Big\{\max\nolimits_{0\le i\le n} \xi(n+i) \ge \varepsilon n\Big\} \to 0 \text{ as } n\to\infty,\ \forall \varepsilon>0; \tag{A.14}$$

$$\textstyle\sum_{n=1}^{\infty} \mathrm{P}_i\big\{\xi(n) \le -\omega n\big\} < \infty \text{ for some } 0<\omega<D_i; \tag{A.15}$$

$$\big\{\max\nolimits_{1\le i\le n} |\xi(n+i)|\,, n\ge 1\big\} \text{ is uniformly integrable;} \tag{A.16}$$

$$\lim\nolimits_{n\to\infty} \mathrm{P}_i\Big\{\max\nolimits_{i\le\sqrt{n}} |\xi(n+i)-\xi(n)| \ge \varepsilon\Big\} = 0 \text{ for any } \varepsilon>0; \tag{A.17}$$

$$\xi(n) \text{ converges in distribution to an integrable random variable } \xi. \tag{A.18}$$

Then, by Theorem 3 of Zhang [**36**],

$$\mathrm{E}_i\nu_i = b + \frac{1}{D_i}\big(\varkappa_i + \mathrm{E}_i\xi\big) + o(1) \qquad \text{as} \quad b_i\to\infty \tag{A.19}$$

where

$$b = b_{b_i} = \sup\{t\ge 1 : A(t,b_i) \ge D_i t\} = \frac{b_i}{D_i} + \frac{h_r}{D_i}\sqrt{\frac{b_i}{D_i} + \frac{h_r^2}{4D_i^2}} + \frac{h_r^2}{2D_i^2}. \tag{A.20}$$

To prove (A.13) we rewrite ν_i in the following form

$$\nu_i = \inf\{n : S_i(n) + \gamma(n) \ge b_i\}\,,$$

where $\gamma(n) = -\max_{1\le j\le r}(S_n^j + \lambda_j)$ with $S_n^j = Z_{\langle M-r-1+j\rangle}(n) - n\mu_{i[M-r-1+j]}$ which is a zero-mean random walk. Let $K = [(1-\varepsilon)b_i/D_i]$ and $\Lambda = \max(S_i(1)-D_i, S_i(2)-2D_i,\dots,S_i(K)-KD_i)$. (In what follows we omit the index i in $h_{r,i}$, λ_{ki}, $S_i(n)$, etc. for brevity.) Then

$$\begin{aligned}\mathrm{P}_i\big(\nu_i\le K\big) &= \mathrm{P}_i\Big\{\max_{n\le K}(S(n)+\gamma(n)) \ge b_i\Big\}\\ &\le \mathrm{P}_i\Big\{\max_{n\le K} S(n) \ge (1-\varepsilon/2)b_i\Big\} + \mathrm{P}_i\Big\{\max_{n\le K}\gamma(n) \ge \varepsilon b_i/2\Big\}\,.\end{aligned} \tag{A.21}$$

By the submartingale inequality, the first probability in (A.21) can be estimated above

$$\begin{aligned}\mathrm{P}_i\Big\{\max_{n\le K} S(n) \ge (1-\varepsilon/2)b_i\Big\} &\le \mathrm{P}_i\big(\Lambda \ge \varepsilon D_i K/2\big)\\ &\le \left(\frac{2}{\varepsilon D_i K}\right)^2 \int_{\{\Lambda\ge\varepsilon D_i K/2\}} (S(K)-KD_i)^2\, d\mathrm{P}_i\,.\end{aligned}$$

Since $\mathrm{Var}(S(K)) = K\sigma_0^2$, we have that $\mathrm{P}_i(\Lambda \ge \varepsilon D_i K/2) \to 0$ as $b_i\to\infty$. Then, by the uniform integrability of $(S(K)-KD_i)/\sigma_0\sqrt{K}$,

$$\int_{\{\Lambda\ge\varepsilon D_i K/2\}} \left(\frac{S(K)-KD_i}{\sigma_0\sqrt{K}}\right)^2 d\mathrm{P}_i \to 0\,,$$

and hence $\mathrm{P}_i\{\max_{n\le K} S(n) \ge (1-\varepsilon/2)b_i\} = o(1/K)$.

Using the same arguments, it may be shown that the second probability in (A.21)

$$\mathrm{P}_i\Big\{\max_{n\le K}\gamma(n) \ge \varepsilon b_i/2\Big\} \le \mathrm{P}_i\Big\{\max_{n\le K}(-S_n^1-\lambda_1) \ge \varepsilon D_i K/2(1-\varepsilon)\Big\} = o\big(1/K\big)\,.$$

The proof of (A.13) is complete.

Let $N = \varepsilon n - h_r\sqrt{2n}$. The proof of (A.14) is a direct application of a submartingale inequality and uniform integrability of $S^j_{2n}/\sqrt{2n}$:

$$\mathrm{P}_i\Big\{\max_{k\le n}\xi(n+k)\ge\varepsilon n\Big\}\le \mathrm{P}_i\Big\{\max_{k\le 2n}\xi(k)\ge\varepsilon n\Big\}\le \mathrm{P}_i\Big\{\max_{k\le 2n}\min_j(-S^j_k-\lambda_j)\ge N\Big\}$$
$$\le \mathrm{P}_i\Big\{\max_{k\le 2n}(-S^1_k-\lambda_1)\ge N\Big\}\le \frac{1}{(N+\lambda_1)^2}\int_{\{\max_{k\le 2n}(-S^1_k-\lambda_1)\ge N\}}(S^1_{2n})^2\,d\mathrm{P}_i\,.$$

Similar to the proof of (A.4) we have that $\mathrm{P}_i\{\max_{k\le 2n}(-S^1_k-\lambda_1)\ge N\}\to 0$ as $b_i\to\infty$. Now, the uniform integrability of $S^1_{2n}/\sqrt{2n}$ yields $\mathrm{P}_i\{\max_{k\le n}\xi(n+k)\ge\varepsilon n\}=o(1/n)$ which proves (A.14).

Verification of (A.15) is similar to, but simpler than, that presented in the proof of (A.14) and is omitted.

Conditions (3.6)–(3.7) (slowly changing) and conditions (A.16)–(A.18) are used in the proof of Theorem 3 in [**36**] to obtain the uniform integrability of the overshoot $\rho_{b_i}=S_i(\nu_i)+\gamma(\nu_i)-b_i$ and to prove that $\mathrm{E}_i\xi(n)\to\mathrm{E}_i\xi$ as $n\to\infty$. The former follows from Lemma 4.2 in [**10**] which states that ρ_{b_i} is bounded above by a uniformly integrable random variable. The latter convergence is proved in the next paragraph.

Using Theorem 20.1 of Bhattacharya and Rao [**5**] with $f(\boldsymbol{x})=\max_{1\le k\le r}\{x_k\}$ and $s=3$, we have (under the assumption of our theorem)

$$\begin{aligned}\mathrm{E}_i\zeta(n)&=\mathrm{E}\max_{1\le k\le r}\left\{\varsigma_k+\frac{\lambda_k}{\sqrt{n}}\right\}\\&\quad+\frac{1}{\sqrt{n}}\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\left\{x_k+\frac{\lambda_k}{\sqrt{n}}\right\}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}+o\big(n^{-1/2}\big)\end{aligned}\tag{A.22}$$

where $\boldsymbol{\varsigma}=(\varsigma_1,\dots,\varsigma_r)\sim\mathcal{N}_r(0,\boldsymbol{V})$. On the other hand, transformation of variables and first-order Taylor expansion for $\phi_{0,V}(\boldsymbol{x})$ in the first integral yield (as $n\to\infty$)

$$\begin{aligned}\mathrm{E}\max_{1\le k\le r}\left\{\varsigma_k+\frac{\lambda_k}{\sqrt{n}}\right\}&=\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\left\{y_k+\frac{\lambda_k}{\sqrt{n}}\right\}\phi_{0,V}(\boldsymbol{y})\,d\boldsymbol{y}\\&=\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\phi_{0,V}(\boldsymbol{x}-\boldsymbol{\lambda}/\sqrt{n})\,d\boldsymbol{x}\\&=\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}\\&\quad+\frac{1}{\sqrt{n}}\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\boldsymbol{\lambda}\boldsymbol{V}^{-1}\boldsymbol{x}^\top\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}+o\big(n^{-1/2}\big)\,.\end{aligned}\tag{A.23}$$

By J_2 denote the second integral in (A.22). Obviously we have the following estimates for J_2

$$\frac{1}{n}\min_{1\le k\le r}\{\lambda_k\}\int_{\boldsymbol{R}^r}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}+\frac{1}{\sqrt{n}}\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}\le J_2$$
$$\le\frac{1}{\sqrt{n}}\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}+\frac{1}{n}\max_{1\le k\le r}\{\lambda_k\}\int_{\boldsymbol{R}^r}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}\,,$$

which show that

$$J_2=\frac{1}{\sqrt{n}}\int_{\boldsymbol{R}^r}\max_{1\le k\le r}\{x_k\}\mathcal{P}(\boldsymbol{x})\phi_{0,V}(\boldsymbol{x})\,d\boldsymbol{x}+o\big(n^{-1/2}\big)\qquad\text{as}\quad n\to\infty\,.\tag{A.24}$$

The relations (A.11) and (A.22)–(A.24) give

$$\mathrm{E}_i\xi(n)=C_{r,i}+o(1)\qquad\text{as}\quad n\to\infty\,.$$

Substituting the above limiting value for $\mathrm{E}_i\xi$ in (A.19), the required asymptotic result (3.35) follows.

Now, consider the MSPRT δ_a. It follows from (2.2) in [**13**] that if $b_i = a_i, i = 0, 1, \dots, M-1$, then

$$\mathrm{E}_i \nu_b \le \mathrm{E}_i \tau_a \le \mathrm{E}_i \nu_{b+\log(M-1)}$$

Thus, $\mathrm{E}_i \tau_a$ may be obtained to within a constant factor K, $0 \le K \le M-1$, by replacing b_i with a_i in (3.35). ■

References

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

[2] P.A. Bakut, I.A. Bolshakov *et al*, *Statistical Radar Theory*, vol. 1, G.P. Tartakovsky Ed. Moscow: Sovetskoe Radio, 1963 (in Russian).

[3] L.E. Baum and M. Katz, " Convergence rates in the law of large numbers," *Trans. Amer. Math. Soc.*, vol. 120, pp. 108-123, 1965.

[4] C.W. Baum and V.V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1994-2007, November 1994.

[5] R.N. Bhattacharya and R.R. Rao, *Normal Approximations and Asymptotic Expansions*. New York: Wiley, 1986.

[6] J.J. Bussgang, "Sequential methods in radar detection," *Proc. IEEE*, vol. 58, pp. 731-743, 1970.

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

[8] 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.

[9] V.P. Dragalin, "Asymptotic solution of a problem of detecting a signal from k channels," *Russian Math. Surveys*, vol. 42, no. 3, pp. 213-214, 1987.

[10] V.P. Dragalin, "Asymptotic solution of a problem of signal searching in multi-channel system," *Mat. Issled (Academy of Sciences of Moldova)*, vol. 109, pp. 15-35, 1989 (in Russian).

[11] V.P. Dragalin, "Third-order asymptotics for a sequential selection procedure," Technical Report, Institute of Applications of Mathematics and Informatics, Milan, Italy, 1994. To appear in *Statistics* & *Decisions*, Special Issue in Honor of Professor Bechhofer.

[12] V.P. Dragalin and A.A. Novikov, "Adaptive sequential tests for composite hypotheses," *Technical Report* 94.4, Institute for Applications of Mathematics and Informatics, Milan, 1994. To appear in *Proc. Steklov Math. Inst.: Statistics and Control of Stochastic Processes, Frontier in Pure and Applied Math.*, TVP Science Publ., Moscow.

[13] V.P. Dragalin, A.G. Tartakovsky, and V. Veeravalli, "Multihypothesis sequential probability ratio tests, I: asymptotic optimality," *IEEE Trans. Inform. Theory* (submitted 2/98).

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

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

[16] M.M. Fishman, "Average duration of asymptotically optimal multi-alternative sequential procedure for recognition of processes," *Soviet Journ. Communic. Technol. Electron*, vol. 30, pp. 2541-2548, 1987.

[17] 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.

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

[19] T.L. Lai and D. Siegmund, "A non-linear renewal theory with applications to sequential analysis, I," *Ann. Statist.*, vol. 5, pp. 946-954, 1977.

[20] T.L. Lai and D. Siegmund, "A non-linear renewal theory with applications to sequential analysis, II," *Ann. Statist.*, vol. 7, pp. 60-76, 1979.

[21] R.Sh. Liptser and A.N. Shiryaev, *Theory of Martingales*. Dordrecht: Kluwer, 1989.

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

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

[24] S.C. Port, "Escape probability for a half-line," *Ann. Math. Statist.*, vol. 35, pp. 1351-1355, 1964.

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

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

[27] A.G. Tartakovskii, "Sequential testing of many simple hypotheses with dependent observations," *Probl. Information Transmis.*, vol. 24, pp. 299-309, 1988.

[28] A.G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*. Moscow: Radio i Svyaz', 1991 (in Russian).

[29] A.G. Tartakovsky, "Asymptotically optimal sequential tests for nonhomogeneous processes," *Sequential Analysis*, vol. 17, No. 1, pp. 33-61, 1998.

[30] A.G. Tartakovsky, "Asymptotic optimality of certain multi-alternative sequential tests: non-i.i.d. case," *Statistical Inference for Stochastic Processes* (submitted).

[31] A.G. Tartakovsky and I.A. Ivanova, "Approximations in sequential rules for testing composite hypotheses and their accuracy in the problem of signal detection from post-detector data," *Probl. Information Transmis.*, vol. 28, pp. 63-74, 1992.

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

[33] V.V. Veeravalli and C.W. Baum, "Hybrid acquisition of direct sequence CDMA signals," *Int. J. Wireless Inform. Networks*, vol. 3, pp. 55-65, January 1996.

[34] N.V. Verdenskaya and A.G. Tartakovskii, "Asymptotically optimal sequential testing of multiple hypotheses for non-homogeneous Gaussian processes in an asymmetric situation," *Theory Prob. Appl.*, vol. 36, pp. 536-547, 1991.

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

[36] C.H. Zhang, "A nonlinear renewal theory," *Ann. Probab.*, vol. 16, pp. 793-825, 1988.

TABLE 3. **Results for Example 1 with distinct distances between hypotheses.** The number of trials used in the simulations was 10^5 and 10^6 for cases (i) and (ii), respectively. The best fixed sample size test that meets the constraint on the Bayes risk $\sum_i R_i$ takes 34 and 111 samples for cases (i) and (ii), respectively.

(i) $\theta_0 = -0.6$, $\theta_1 = 0.0$, $\theta_2 = 1.0$; $\sigma^2 = 1.0$

Results for Test δ_a									
	Risks & Thresholds			Expected Sample Size					
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	3.16	0.0097	18.83	17.54	6.82	19.75	4.88
H_1	0.01	3.16	0.0091	21.46	17.54	18.2	19.75	7.99
H_2	0.01	2.93	0.0098	7.24	5.85	19.1	7.29	0.66
Results for Test δ_b (using $b_i = \log(2\gamma_i/3\bar{R}_i)$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	3.85	0.0049	22.98	21.39	6.91	23.59	2.68
H_1	0.01	3.85	0.0053	25.27	21.39	15.34	23.59	6.62
H_2	0.01	3.62	0.0051	8.61	7.24	15.90	8.67	0.66
Results for Test δ_b (using $b_i = a_i$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	3.16	0.0097	18.75	17.54	6.46	19.74	5.30
H_1	0.01	3.16	0.0106	20.85	17.54	15.89	19.74	5.33
H_2	0.01	2.93	0.0100	7.17	5.85	18.39	7.29	1.67

(ii) $\theta_0 = -0.5$, $\theta_1 = 0.0$, $\theta_2 = 1.0$; $\sigma^2 = 1.0$.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.001	5.52	1.0×10^{-3}	46.48	44.15	5.02	46.73	0.55
H_1	0.001	5.52	1.0×10^{-3}	48.39	44.15	8.77	46.73	3.42
H_2	0.001	5.23	1.0×10^{-3}	11.90	10.46	12.10	11.89	0.03
Results for Test δ_b (using $b_i = \log(2\gamma_i/3\bar{R}_i)$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.001	6.21	4.9×10^{-4}	52.12	49.69	4.66	52.28	0.31
H_1	0.001	6.21	5.6×10^{-4}	53.60	49.69	7.27	52.28	2.44
H_2	0.001	5.92	5.4×10^{-4}	13.28	11.84	10.79	13.28	0.01
Results for Test δ_b (using $b_i = a_i$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.001	5.52	1.0×10^{-3}	46.47	44.15	5.01	46.73	0.57
H_1	0.001	5.52	1.1×10^{-3}	48.03	44.15	8.10	46.73	2.70
H_2	0.001	5.23	1.0×10^{-3}	11.89	10.46	12.01	11.89	0.07

TABLE 4. **Results for Example 1 with $\theta_0 = -0.5$, $\theta_1 = 0.0$, $\theta_2 = 0.5$; $\sigma^2 = 1.0$.** The number of trials used in the simulations was 10^6. The fixed sample size test that meets the constraint on the Bayes risk $\sum_i R_i$ takes 131 samples.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.001	5.52	1.0×10^{-3}	46.59	44.15	5.24	46.74	0.32
H_1	0.001	5.52	1.1×10^{-3}	69.43	44.15	36.41	73.64	6.07
H_2	0.001	5.52	1.0×10^{-3}	46.60	44.15	5.27	46.74	0.29
Results for Test δ_b (using $b_0 = a_0$, $b_2 = a_2$, $b_1 = \log(\gamma_1/2\bar{R}_1)$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.001	5.52	1.0×10^{-3}	46.61	44.15	5.29	46.74	0.26
H_1	0.001	5.92	8.0×10^{-4}	71.67	47.39	33.88	77.63	8.31
H_2	0.001	5.23	1.0×10^{-3}	46.64	44.15	5.33	46.73	0.22

Table 5. **Results for Example 2 with Gaussian observations and H_0 present.** The number of trials used in the simulations was 10^5 and 10^6 for $\bar{R}_0$ values of 0.01 and 0.001, respectively.

(i) $N = 4$, $\pi_0 = 0.5$, $\theta = 0.5$, $\sigma^2 = 1$.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	3.62	1.0×10^{-2}	47.56	28.97	39.09	48.31	1.57
H_1	0.001	4.54	1.0×10^{-3}	49.29	36.30	26.36	49.98	1.39
H_0	0.001	5.92	1.0×10^{-3}	72.37	47.39	34.52	73.58	1.67
H_1	0.0001	6.84	1.0×10^{-4}	68.81	54.72	20.48	68.40	0.60
Results for Test δ_b (using $b_0 = \log(2\gamma_0\pi_0/\bar{R}_0)$, and $b_i = a_i$, $i \neq 0$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	4.31	7.8×10^{-3}	50.50	34.51	31.66	56.14	11.18
H_1	0.001	4.54	1.1×10^{-3}	48.87	36.30	25.73	49.98	2.27
H_0	0.001	6.61	7.3×10^{-4}	75.99	52.93	30.34	80.86	6.41
H_1	0.0001	6.84	1.0×10^{-4}	68.45	54.72	20.05	68.40	0.06

(ii) $N = 10$, $\pi_0 = 0.5$, $\theta = 0.5$, $\sigma^2 = 1$.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	3.62	1.1×10^{-2}	58.65	28.97	50.60	59.62	1.65
H_1	0.001	3.62	1.0×10^{-3}	49.98	28.97	42.04	49.98	0.00
H_0	0.001	5.92	1.1×10^{-3}	87.10	47.39	45.59	88.66	1.79
H_1	0.0001	5.92	1.0×10^{-4}	69.30	47.39	31.61	68.40	1.29
Results for Test δ_b (using $b_0 = \log(2\gamma_0\pi_0/\bar{R}_0)$, and $b_i = a_i$, $i \neq 0$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_{b_{\mathrm{so}}}$	$\epsilon_{\mathrm{so}}\ \%$
H_0	0.01	4.31	1.0×10^{-2}	59.15	34.51	41.65	68.74	16.21
H_1	0.001	3.62	1.2×10^{-3}	48.70	28.96	40.51	49.98	2.64
H_0	0.001	6.61	9.0×10^{-4}	89.17	52.93	40.64	96.87	8.63
H_1	0.0001	5.92	1.1×10^{-4}	68.59	47.39	30.91	68.40	0.28

TABLE 6. **Results for Example 2 with Gaussian observations and H_0 absent.** The number of trials used in the simulations was 10^5 and 10^6 for $\bar{R}$ values of 0.01 and 0.001, respectively.

(i) $N = 4$, $\theta = 0.25$, $\sigma^2 = 1$.

Results for Test δ_a								
Risks & Thresholds			Expected Sample Size					
$\bar{R}$	a	$\hat{R}$	$\hat{\mathrm{E}}\tau_a$	$(\mathrm{E}\tau_a)_{\mathrm{fo}}$	ϵ_{fo}%	$(\mathrm{E}\tau_a)_{\mathrm{so}}$	ϵ_{so} %	T (FSS)
0.01	3.01	9.8×10^{-3}	80.64	48.20	40.22	81.68	1.28	151
0.001	5.32	1.0×10^{-3}	127.67	85.05	33.38	126.07	1.25	287
Results for Test δ_b (using $b = \log(2\gamma/(N\bar{R}))$)								
$\bar{R}$	b	$\hat{R}$	$\hat{\mathrm{E}}\nu_b$	$(\mathrm{E}\nu_b)_{\mathrm{fo}}$	ϵ_{fo}%	$(\mathrm{E}\nu_b)_{\mathrm{so}}$	ϵ_{so} %	T (FSS)
0.01	3.71	7.5×10^{-3}	87.07	59.30	31.89	95.27	9.42	151
0.001	6.01	7.3×10^{-4}	134.54	96.14	28.54	139.11	3.39	287

(i) $N = 10$, $\theta = 0.25$, $\sigma^2 = 1$.

Results for Test δ_a								
Risks & Thresholds			Expected Sample Size					
$\bar{R}$	a	$\hat{R}$	$\hat{\mathrm{E}}\tau_a$	$(\mathrm{E}\tau_a)_{\mathrm{fo}}$	ϵ_{fo}%	$(\mathrm{E}\tau_a)_{\mathrm{so}}$	ϵ_{so} %	T (FSS)
0.01	2.10	9.4×10^{-3}	92.74	33.55	63.83	93.41	0.72	144
0.001	4.40	1.0×10^{-3}	148.51	70.39	52.60	144.45	2.73	290
Results for Test δ_b (using $b = \log(2\gamma/(N\bar{R}))$)								
$\bar{R}$	b	$\hat{R}$	$\hat{\mathrm{E}}\nu_b$	$(\mathrm{E}\nu_b)_{\mathrm{fo}}$	ϵ_{fo}%	$(\mathrm{E}\nu_b)_{\mathrm{so}}$	ϵ_{so} %	T (FSS)
0.01	2.79	1.1×10^{-2}	89.26	44.64	49.99	109.26	22.41	144
0.001	5.09	9.7×10^{-4}	150.15	81.48	45.73	159.12	5.97	290

TABLE 7. **Results for Example 2 with exponential observations and H_0 present.** The number of trials used in the simulations was 10^5 and 10^6 for $\bar{R}_0$ values of 0.01 and 0.001, respectively.

(i) $N = 4$, $\pi_0 = 0.5$, $\rho = 0.5$.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	ϵ_{so} %
H_0	0.01	3.78	9.1×10^{-3}	77.48	52.36	32.41	75.81	2.15
H_1	0.001	4.42	9.0×10^{-4}	66.78	46.78	29.93	66.74	0.06
H_0	0.001	6.08	7.8×10^{-4}	118.92	84.28	29.13	118.10	0.68
H_1	0.0001	6.72	9.4×10^{-5}	92.01	71.14	22.68	92.02	1.00
Results for Test δ_b (using $b_0 = \log(2\gamma_0\pi_0/\bar{R}_0)$, and $b_i = a_i$, $i \neq 0$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	ϵ_{so} %
H_0	0.01	4.47	7.0×10^{-3}	81.64	61.97	24.09	88.89	8.88
H_1	0.001	4.42	1.1×10^{-3}	65.83	46.78	28.93	66.74	1.38
H_0	0.001	6.77	5.7×10^{-4}	124.41	93.89	24.53	130.35	4.76
H_1	0.0001	6.72	1.0×10^{-4}	91.28	71.14	22.06	91.10	0.21

(ii) $N = 10$, $\pi_0 = 0.5$, $\rho = 0.75$.

Results for Test δ_a								
	Risks & Thresholds			Expected Sample Size				
	$\bar{R}_i$	a_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\tau_a$	$(\mathrm{E}_i\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\tau_a)_{\mathrm{so}}$	ϵ_{so} %
H_0	0.01	3.72	9.8×10^{-3}	49.72	28.43	42.81	46.70	6.07
H_1	0.001	3.35	1.0×10^{-3}	33.78	17.60	47.88	33.64	0.42
H_0	0.001	6.02	8.1×10^{-4}	75.28	46.00	38.89	73.07	2.93
H_1	0.0001	5.65	1.0×10^{-4}	46.38	29.70	35.96	45.74	1.39
Results for Test δ_b (using $b_0 = \log(2\gamma_0\pi_0/\bar{R}_0)$, and $b_i = a_i$, $i \neq 0$)								
	$\bar{R}_i$	b_i	$\hat{R}_i$	$\hat{\mathrm{E}}_i\nu_b$	$(\mathrm{E}_i\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}_i\nu_b)_{\mathrm{so}}$	ϵ_{so} %
H_0	0.01	4.42	9.9×10^{-3}	49.56	33.72	31.96	54.98	10.93
H_1	0.001	3.35	1.0×10^{-3}	32.64	17.61	46.05	33.64	3.07
H_0	0.001	6.72	7.0×10^{-4}	76.74	51.30	33.16	80.54	4.95
H_1	0.0001	5.65	1.2×10^{-4}	45.70	29.70	35.00	45.74	0.09

TABLE 8. **Results for Example 2 with exponential observations and H_0 absent.** The number of trials used in the simulations was 10^5 and 10^6 for $\bar{R}_i$ values of 0.01 and 0.001, respectively.

(i) $N = 4$, $\rho = 0.3$.

Results for Test δ_a							
Risks & Thresholds			Expected Sample Size				
$\bar{R}$	a	$\hat{R}$	$\hat{\mathrm{E}}\tau_a$	$(\mathrm{E}\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
0.01	2.96	9.4×10^{-3}	71.20	42.70	40.02	70.18	1.43
0.001	5.26	9.4×10^{-3}	112.40	75.96	32.42	109.49	2.60
Results for Test δ_b (using $b = \log(2\gamma/(N\bar{R}))$)							
$\bar{R}$	b	$\hat{R}$	$\hat{\mathrm{E}}\nu_b$	$(\mathrm{E}\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
0.01	3.65	7.6×10^{-3}	75.95	52.71	30.58	82.20	8.23
0.001	5.95	6.8×10^{-4}	117.95	85.98	27.10	121.05	2.63

(i) $N = 10$, $\rho = 0.5$

Results for Test δ_a							
Risks & Thresholds			Expected Sample Size				
$\bar{R}$	a	$\hat{R}$	$\hat{\mathrm{E}}\tau_a$	$(\mathrm{E}\tau_a)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}\tau_a)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
0.001	4.20	9.8×10^{-4}	52.37	25.20	51.88	49.19	6.06
0.0001	6.50	1.0×10^{-4}	69.49	39.01	43.86	66.53	4.25
Results for Test δ_b (using $b = \log(2\gamma/(N\bar{R}))$)							
$\bar{R}$	b	$\hat{R}$	$\hat{\mathrm{E}}\nu_b$	$(\mathrm{E}\nu_b)_{\mathrm{fo}}$	$\epsilon_{\mathrm{fo}}\%$	$(\mathrm{E}\nu_b)_{\mathrm{so}}$	$\epsilon_{\mathrm{so}}\ \%$
0.001	4.89	1.0×10^{-3}	52.35	29.35	43.93	54.50	4.08
0.0001	7.19	9.0×10^{-5}	70.18	43.17	38.49	71.63	2.05