Statistica Sinica **23** (2013), 1717-1741
doi:http://dx.doi.org/10.5705/ss.2012-082s

ALMOST OPTIMAL SEQUENTIAL TESTS OF DISCRETE COMPOSITE HYPOTHESES

Georgios Fellouris and Alexander G. Tartakovsky

The University of Southern California

Abstract: We consider the problem of sequentially testing a simple null hypothesis, H_0, versus a composite alternative hypothesis, H_1, that consists of a finite set of densities. We study sequential tests that are based on thresholding of mixture-based likelihood ratio statistics and weighted generalized likelihood ratio statistics. It is shown that both sequential tests have several asymptotic optimality properties as error probabilities go to zero. First, for any weights, they minimize the expected sample size within a constant term under every scenario in H_1 and at least to first order under H_0. Second, for appropriate weights that are specified up to a prior distribution, they minimize a weighted expected sample size in H_1 within an asymptotically negligible term. Third, for a particular prior distribution, they are almost minimax with respect to the expected Kullback–Leibler divergence until stopping. Furthermore, based on high-order asymptotic expansions for the operating characteristics, we propose prior distributions that lead to a robust behavior. Finally, based on asymptotic analysis as well as on simulation experiments, we argue that both tests have the same performance when they are designed with the same weights.

Key words and phrases: Asymptotic optimality, generalized likelihood ratio, minimax sequential tests, mixture-based tests.

1. Introduction

Let $\{X_t\}_{t\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random vectors with values in $\mathbb{R}^d$, $d \in \mathbb{N} = \{1, 2, \dots\}$, and common density f with respect to some non-degenerate, σ-finite measure $\nu(dx)$. We consider the problem of *sequentially* testing $\mathsf{H}_0 : f \in \mathcal{A}_0$ versus $\mathsf{H}_1 : f \in \mathcal{A}_1$, where $\mathcal{A}_0$ and $\mathcal{A}_1$ are two disjoint sets of densities with common support. That is, we assume that observations are acquired in a sequential manner and the goal is to select the correct hypothesis as soon as possible.

Let $\{\mathcal{F}_t\}$ be the observed filtration, $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$. A *sequential test* $\delta = (T, d_T)$ is a pair that consists of an $\{\mathcal{F}_t\}$-stopping time, T, and an $\mathcal{F}_T$-measurable (terminal) decision rule, $d_T = d_T(X_1, \dots, X_T) \in \{0, 1\}$, that specifies which hypothesis is to be accepted once observations have stopped. In particular, H_j is accepted if $d_T = j$, that is, $\{d_T = j\} = \{T < \infty,\ \delta \text{ accepts } \mathsf{H}_j\}$, $j = 0, 1$.

The ideal sequential test has the smallest possible expected sample size under both H_0 and H_1, while controlling error probabilities below given tolerance levels. Thus, if P_f is the underlying probability measure when X_1 has density f and E_f is the corresponding expectation, $\delta^o = (T^o, d^o_{T^o}) \in \mathcal{C}_{\alpha,\beta}$ is an *optimal* sequential test if

$$\mathsf{E}_f[T^o] = \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_f[T] \quad \forall\, f \in \mathcal{A}_0 \cup \mathcal{A}_1,$$

where $\mathcal{C}_{\alpha,\beta}$ is the class of sequential tests whose maximal type-I and type-II error probabilities are bounded above by α and β respectively,

$$\mathcal{C}_{\alpha,\beta} = \Big\{\delta : \sup_{f \in \mathcal{A}_0} \mathsf{P}_f(d_T = 1) \leq \alpha \quad \text{and} \quad \sup_{f \in \mathcal{A}_1} \mathsf{P}_f(d_T = 0) \leq \beta\Big\}.$$

Wald and Wolfowitz (1948) proved that an optimal sequential test exists when both hypotheses are simple, $\mathcal{A}_0 = \{f_0\}$ and $\mathcal{A}_1 = \{f_1\}$, and is given by the Sequential Probability Ratio Test (SPRT) that was proposed by Wald (1944) in his seminal work on Sequential Analysis:

$$S = \inf\{t \in \mathbb{N} : \Lambda^1_t \notin (A^{-1}, B)\}, \quad d_{\mathcal{S}} = \mathbb{1}_{\{\Lambda^1_S \geq B\}}, \tag{1.1}$$

where $A, B > 1$ are constant thresholds, selected so that $\mathsf{P}_0(d_S = 1) = \alpha$ and $\mathsf{P}_1(d_S = 0) = \beta$, and $\{\Lambda^1_t\}$ is the likelihood ratio statistic

$$\Lambda^1_t = \prod_{n=1}^{t} \frac{f_1(X_n)}{f_0(X_n)}, \quad t \in \mathbb{N}. \tag{1.2}$$

In the case of composite hypotheses, it has only been possible to find sequential tests that are optimal in an asymptotic sense. More specifically, we say that $\delta^0 \in \mathcal{C}_{\alpha,\beta}$ is *uniformly (first-order) asymptotically optimal* if

$$\mathsf{E}_f[T^0] = \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_f[T]\,(1 + o(1)) \quad \forall\, f \in \mathcal{A}_0 \cup \mathcal{A}_1,$$

as $\alpha, \beta \to 0$. When, in particular, $\mathcal{A}_0$ and $\mathcal{A}_1$ can be embedded in an exponential family $\{f_\theta, \theta \in \Theta\}$ and Θ_1 is a subset of the natural parameter space Θ such that $\theta_0 \notin \Theta_1$ and

$$\mathcal{A}_0 = \{f_{\theta_0}\} \quad \text{and} \quad \mathcal{A}_1 = \{f_\theta, \theta \in \Theta_1\}, \tag{1.3}$$

it is well known (see, for example, Lorden (1973); Pollak and Siegmund (1975)) that the sequential test (1.1) is uniformly asymptotically optimal if Λ^1_t is replaced either by the generalized likelihood-ratio (GLR) statistic, $\sup_{\theta \in \Theta} \Lambda^\theta_t$, or by a mixture-based likelihood ratio statistic, $\int_\Theta \Lambda^\theta_t\, w(\theta)\, d\theta$, where $w(\cdot)$ is some probability density function on Θ (weight function) and Λ^θ_t is defined as in (1.2) with f_1 replaced by f_θ. However, apart from certain tractable cases, these statistics are not in general recursive and, as a result, they cannot be easily implemented on-line. Moreover, their computation at each step may be approximate, since it often requires discretization of the parameter space. These problems can be overcome if one uses the adaptive likelihood-ratio statistic, $\Lambda_t = \Lambda_{t-1}(f_{\theta_t^*}(X_t)/f_0(X_t))$, where θ_t^* is an estimator of θ that depends on the first $t-1$ observations. However, this approach, initially developed by Robbins and Siegmund (1970, 1974) for power one tests and later extended by Pavlov (1990) and Dragalin and Novikov (1999) for multihypothesis sequential tests, generally leads to less efficient sequential tests, since one-stage delayed estimators use less information than the global MLE that is employed by the GLR statistic. Sequential testing of composite hypotheses in a Bayesian formulation with a small cost of observations was considered by Schwarz (1962); Kiefer and Sacks (1963); Chernoff (1972); Lorden (1967); Lai (1988), among others.

In the present paper, we consider the problem of sequential testing a simple null hypothesis against a *discrete* alternative consisting of a finite set of densities,

$$\mathcal{A}_0 = \{f_0\} \quad \text{and} \quad \mathcal{A}_1 = \{f_1, \ldots, f_K\}, \tag{1.4}$$

where K is a positive integer. This hypothesis testing problem has two main motivations. First, it serves as an approximation to the continuous-parameter testing problem (1.3), in which Θ_1 is replaced by a finite subset $\{\theta_1, \ldots, \theta_K\}$ of Θ_1 so that $f_j = f_{\theta_j}$, $j = 0, 1, \ldots, K$. This implies a loss of efficiency under P_θ when $\theta \notin \{\theta_1, \ldots, \theta_K\}$, but it leads to sequential tests that are easily implementable on-line.

Moreover, the hypothesis testing problem (1.4) applies to *multisample slippage* problems, which have a wide range of applications (see, e.g., Chernoff (1972); Tartakovsky et al. (2006, 2003)). As an example, consider the setup in which K sensors monitor different areas, a signal may be present in at most one of these areas and the goal is to detect signal presence without identifying its location. If additionally the sensors are statistically independent and sensor i takes i.i.d. observations $\{X_t^i\}_{t \in \mathbb{N}}$ with density g_1^i (resp. g_0^i) when signal is present (resp. absent), this problem turns out to be a special case of (1.4) with $X_t = (X_t^1, \ldots, X_t^K)$ and

$$f_0(X_t) = \prod_{j=1}^{K} g_0^j(X_t^j)\ , \quad f_i(X_t) = g_1^i(X_t^i) \prod_{\substack{j=1 \\ j \neq i}}^{K} g_0^j(X_t^j), \quad 1 \le i \le K. \tag{1.5}$$

In order to describe the main contributions of this paper, we need some additional terminology and notation. We say that $\boldsymbol{q} = (q^1, \ldots, q^K)$ is a *weight* if

$q_i > 0$ for every $1 \le i \le K$. For any weight $\boldsymbol{q}$, we set

$$\Lambda_t(\boldsymbol{q}) = \sum_{i=1}^{K} q^i \Lambda_t^i, \quad \hat{\Lambda}_t(\boldsymbol{q}) = \max_{1\le i\le K} \{q^i \Lambda_t^i\}, \tag{1.6}$$

where Λ_t^i is defined as in (1.2) with f_1 replaced by f_i. Introduce two sequential tests for problem (1.4):

$$M = \inf\{t : \Lambda_t(\boldsymbol{q}_1) \ge B \text{ or } \Lambda_t(\boldsymbol{q}_0) \le A^{-1}\},\ d_M = \mathbb{1}_{\{\Lambda_M(\boldsymbol{q}_1)\ge B\}},$$
$$N = \inf\{t : \hat{\Lambda}_t(\boldsymbol{q}_1) \ge B \text{ or } \hat{\Lambda}_t(\boldsymbol{q}_0) \le A^{-1}\},\ d_N = \mathbb{1}_{\{\hat{\Lambda}_N(\boldsymbol{q}_1)\ge B\}},$$

where $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$ are arbitrary weights. We call $\delta_{\text{mi}} = (M, d_M)$ the *Mixture Likelihood Ratio Test* (MiLRT) and $\delta_{\text{gl}} = (N, d_N)$ the *Weighted Generalized Likelihood Ratio Test* (WGLRT).

Tartakovsky et al. (2003) studied the GLRT, that is, the WGLRT with *uniform* weights, $q_0^i = q_1^i = 1$, $1 \le i \le K$, in the multisample (multichannel) setup (1.5) and established its asymptotic optimality. More specifically, it was shown that the GLRT is *second-order* asymptotically optimal, in the sense that it attains $\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \mathsf{E}_i[T]$ within an $O(1)$ term for every $1 \le i \le K$, where $O(1)$ is asymptotically bounded as $\alpha, \beta \to 0$. Moreover, it was shown that, in the special case of completely asymmetric channels, the GLRT also attains $\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \mathsf{E}_0[T]$ within an $O(1)$ term. (Here and in what follows we denote by P_j the underlying probability measure when X_1 has density f_j and by E_j the corresponding expectation, $j = 0, 1, \dots, K$.)

In the present work, we establish this uniform, second-order asymptotic optimality property for both the MiLRT and the WGLRT with arbitrary weights $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$ in the more general setup of (1.4). However, the main question we want to answer is how to select the weights in order to obtain further "benefits". In this direction, we show that if $\mathbf{p} = (p_1, \dots, p_K)$ is an arbitrary probability mass function, which can be interpreted as a prior distribution on H_1, and $\boldsymbol{q}_0$, $\boldsymbol{q}_1$ are selected so that $q_0^i = p_i \mathcal{L}_i$ and $q_1^i = p_i/\mathcal{L}_i$, $1 \le i \le K$, then both tests attain $\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T]$ within an $o(1)$ term, where $\mathsf{E}^{\boldsymbol{p}}$ is expectation with respect to the weighted probability measure $\mathsf{P}^{\boldsymbol{p}} = \sum_{i=1}^{k} p_i \mathsf{P}_i$, whereas the $\mathcal{L}$-numbers $\{\mathcal{L}_i\}$, formally introduced in (2.1), provide overshoot corrections that allow us to achieve this refined asymptotic optimality property.

In addition, we find a prior distribution $\hat{\boldsymbol{p}}$ which makes both tests *almost minimax* with respect to the expected Kullback–Leibler (KL) divergence until stopping: they attain $\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \max_{1\le i\le K} (I_i \mathsf{E}_i[T])$ within an $o(1)$ term, where I_i is the KL information number (see (2.2)). In this way, we generalize the corresponding result in Fellouris and Tartakovsky (2012), where this minimax problem was considered in the context of open-ended, mixture-based sequential tests.

Furthermore, using high-order asymptotic expansions for the operating characteristics of both tests, we show that selecting p_i to be proportional to I_i or $\mathcal{L}_i$ leads to a much more robust behavior than the one induced by $\hat{\boldsymbol{p}}$. Finally, based on these asymptotic expansions as well as on Monte Carlo simulations, we argue that both the WGLRT and the MiLRT have essentially the same performance when they are designed with the same weights.

The remainder of the paper is organized as follows. In Section 2, we introduce notation, in Section 3 we obtain asymptotic approximations to the operating characteristics of the two tests, whereas in Section 4 we establish their asymptotic optimality properties. In Section 5, we compare different specifications for $\mathbf{p}$, and in Section 6 we compare the tests using Monte Carlo simulations. We conclude in Section 7.

2. Notation

For every $1 \leq i \leq K$, we set $Z_t^i = \log \Lambda_t^i$, where Λ_t^i is given by (1.2) with f_1 replaced by f_i. We denote by $\delta^i = (S^i, d_{S^i})$ the SPRT for testing f_0 against f_i, which is defined as in (1.1) with Λ_t^1 replaced by Λ_t^i.

We quantify the "distance" between f_i and f_0 using the $\mathcal{L}$-number

$$\mathcal{L}_i = \exp\bigg\{ -\sum_{n=1}^{\infty} n^{-1} \Big[\mathsf{P}_0(Z_n^i > 0) + \mathsf{P}_i(Z_n^i \leq 0) \Big] \bigg\}, \tag{2.1}$$

as well as the KL information numbers $I_i = \mathsf{E}_i[Z_1^i]$ and $I_0^i = \mathsf{E}_0[-Z_1^i]$,

$$I_i = \int \log\Big(\frac{f_i(x)}{f_0(x)}\Big) f_1(x)\, \nu(dx), \quad I_0^i = \int \log\Big(\frac{f_0(x)}{f_i(x)}\Big) f_0(x)\, \nu(dx). \tag{2.2}$$

Without loss of generality, we assume that $f_1, \ldots, f_K$ are ordered with respect to their KL divergence from f_0 so that

$$I_0 = \min_{1 \leq i \leq K} I_0^i = I_0^1 = \cdots = I_0^r < I_0^{r+1} \leq \cdots \leq I_0^K. \tag{2.3}$$

Note that $r = 1$ corresponds to the asymmetric situation in which I_0 is attained by a unique index $i = 1$. On the other hand, $r = K$ corresponds to the completely symmetric situation in which I_0^i is the same for every $1 \leq i \leq K$. The latter case occurs, for example, in the multisample slippage problem (1.5) when the densities do not depend on the population (or sensor in a multisensor context).

In order to avoid trivial cases, we assume that f_i and f_0 do not coincide almost everywhere, which implies that I_i, $I_0^i > 0$ for every $1 \leq i \leq K$. We also assume throughout the paper that Z_1^i is non-arithmetic under P_0 and P_i and that I_i, $I_0^i < \infty$ for every $1 \le i \le K$. Then, if we define the first hitting times

$$\tau_c^i = \inf\{t : Z_t^i \ge c\}, \quad \sigma_c^i = \inf\{t : Z_t^i \le -c\}, \quad c > 0,$$

it is well known that the overshoots $Z_{\tau_c^i}^i - c$ and $|Z_{\sigma_c^i}^i + c|$ have well defined asymptotic distributions under P_i and P_0 respectively,

$$\mathcal{H}_i(x) = \lim_{c\to\infty} \mathsf{P}_i(Z_{\tau_c^i}^i - c \le x), \quad \mathcal{H}_0^i(x) = \lim_{c\to\infty} \mathsf{P}_0(|Z_{\sigma_c^i}^i + c| \le x), \quad x > 0.$$

Consequently, we can define the Laplace transforms

$$\gamma_i = \int_0^\infty e^{-x}\, \mathcal{H}_i(dx), \quad \gamma_0^i = \int_0^\infty e^{-x}\, \mathcal{H}_0^i(dx)$$

that connect the KL-numbers with the $\mathcal{L}$-numbers as follows: $\mathcal{L}_i = \gamma_i\, I_i = \gamma_0^i\, I_0^i$ (see, e.g., Theorem 5 in Lorden (1977)). If, additionally, $\mathsf{E}_i[(Z_1^i)^2], \mathsf{E}_0[(Z_1^i)^2] < \infty$, then $\mathcal{H}_i$ and $\mathcal{H}_0^i$ have finite means (average limiting overshoots),

$$\kappa_i = \int_0^\infty x\, \mathcal{H}_i(dx), \quad \kappa_0^i = \int_0^\infty x\, \mathcal{H}_0^i(dx),$$

and we have the following asymptotic approximations to the expected sample sizes of the SPRT $\delta^i = (S^i, d_{S^i})$ under P_i and P_0:

$$\mathsf{E}_i[S^i] = \frac{1}{I_i}\left(|\log\alpha| + \kappa_i + \log\gamma_i\right) + o(1), \tag{2.4}$$

$$\mathsf{E}_0[S^i] = \frac{1}{I_0^i}\left(|\log\beta| + \kappa_0^i + \log\gamma_0^i\right) + o(1), \tag{2.5}$$

as $\alpha, \beta \to 0$ so that $\alpha|\log\beta| + \beta|\log\alpha| \to 0$.

3. Asymptotic Approximations for Operating Characteristics

In order to obtain asymptotic inequalities and approximations for the error probabilities and the expected sample sizes of the MiLRT and the WGLRT, we rely on the following decompositions for $Z_t(\boldsymbol{q}) = \log\Lambda_t(\boldsymbol{q})$ and $\hat{Z}_t(\boldsymbol{q}) = \log\hat{\Lambda}_t(\boldsymbol{q})$,

$$Z_t(\boldsymbol{q}) = Z_t^i + \log q^i + Y_t^i(\boldsymbol{q}), \quad t \in \mathbb{N}, \tag{3.1}$$

$$\hat{Z}_t(\boldsymbol{q}) = Z_t^i + \log q^i + \hat{Y}_t^i(\boldsymbol{q}), \quad t \in \mathbb{N}, \tag{3.2}$$

where

$$Y_t^i(\boldsymbol{q}) = \log\left(1 + \sum_{\substack{j=1\\ j\neq i}}^K \frac{q^j}{q^i}\,\frac{\Lambda_t^j}{\Lambda_t^i}\right), \quad t \in \mathbb{N}, \tag{3.3}$$

$$\hat{Y}_t^i(\boldsymbol{q}) = \log\Big(\max\Big\{1, \max_{1\le j\ne i\le K} \frac{q^j}{q^i}\frac{\Lambda_t^j}{\Lambda_t^i}\Big\}\Big), \quad t \in \mathbb{N}. \tag{3.4}$$

From the Strong Law of Large Numbers (SLLN) it follows that, for every $j \ne i$, $\mathsf{P}_i(\Lambda_t^j/\Lambda_t^i \to 0) = 1$. This implies that $Y^i(\boldsymbol{q})$ and $\hat{Y}^i(\boldsymbol{q})$ also converge to 0 P_i-a.s., and, consequently, they are *slowly changing* under P_i (for a precise definition of "slowly changing" we refer to Siegmund (1985), page 190). Since Z_t^i is a random walk under P_i, from this observation and decompositions (3.1)–(3.2) it follows that $Z(\boldsymbol{q})$ and $\hat{Z}(\boldsymbol{q})$ are *perturbed* random walks under P_i.

Similarly, the SLLN implies that, *in the special case where* $r = 1$, $\mathsf{P}_0(\Lambda_t^j/\Lambda_t^1 \to 0) = 1$ for every $j > 1$. Therefore, $Y^1(\boldsymbol{q})$ and $\hat{Y}^1(\boldsymbol{q})$ also converge to 0 P_0-a.s. and from (3.1)–(3.2) with $i = 1$ it follows that $Z(\boldsymbol{q})$ and $\hat{Z}(\boldsymbol{q})$ are perturbed random walks under P_0 when $r = 1$.

These properties allow us to apply nonlinear renewal theory for perturbed random walks (see Woodroofe (1976, 1982); Lai and Siegmund (1977, 1979); Siegmund (1985)) in order to obtain asymptotic approximations for the expected sample sizes of the tests δ_{mi} and δ_{gl} under P_i for every $1 \le i \le K$, as well as under P_0 when $r = 1$. An asymptotic approximation for $\mathsf{E}_0[N]$ when $r > 1$ can be obtained based on nonlinear renewal theory of Zhang (1988) using the following representation for N_A^0:

$$N_A^0 = \inf\Big\{t : \ell_t^0 \ge \log A + \max_{1\le i\le K}(\log q_0^i + \ell_t^i)\Big\}, \tag{3.5}$$

where

$$\ell_t^j = \sum_{n=1}^t \log f_j(X_n), \quad t \in \mathbb{N}. \tag{3.6}$$

For the latter approximation we also need some additional notation. For any $1 \le i \le K$, we set $\mu_i = \mathsf{E}_0[\log f_i(X_1)]$, so that $I_0^i = \mathsf{E}_0[\log f_0(X_1)] - \mu_i$. Moreover, we set $\mu = \max_{1\le i\le K} \mu_i$, so that $I_0 = \mathsf{E}_0[\log f_0(X_1)] - \mu$, we define the r-dimensional random vector

$$W = (\log f_1(X_1) - \mu, \dots, \log f_r(X_1) - \mu), \tag{3.7}$$

and we denote by Σ its covariance matrix under P_0. Finally, we set

$$d_r = \frac{h_r}{2\sqrt{I_0}}, \quad h_r = \int_{\mathbb{R}^r} (\max_{1\le i\le r} x_i)\, \phi_\Sigma(x)\, dx, \tag{3.8}$$

where ϕ_Σ is the density of an r-dimensional zero-mean Gaussian random vector with covariance matrix Σ.

3.1. Asymptotic bounds for the error probabilities

Define the overshoots associated with the MiLRT and the WGLRT,

$$\eta = [Z_M(\boldsymbol{q}_1) - \log B]\, \mathbb{1}_{\{d_M=1\}} - [Z_M(\boldsymbol{q}_0) + \log A]\, \mathbb{1}_{\{d_M=0\}}, \tag{3.9}$$

$$\hat{\eta} = [\hat{Z}_N(\boldsymbol{q}_1) - \log B]\, \mathbb{1}_{\{d_N=1\}} - [\hat{Z}_N(\boldsymbol{q}_0) + \log A]\, \mathbb{1}_{\{d_N=0\}}, \tag{3.10}$$

which play an important role in the asymptotic analysis of the operating characteristics. Observe also that the tests $\delta_{\rm mi}$ and $\delta_{\rm gl}$ can be equivalently defined as

$$M = \min\{M_A^0, M_B^1\}, \quad d_M = \mathbb{1}_{\{M_B^1 \le M_A^0\}}, \tag{3.11}$$

$$N = \min\{N_A^0, N_B^1\}, \quad d_N = \mathbb{1}_{\{N_B^1 \le N_A^0\}}, \tag{3.12}$$

where

$$M_B^1 = \inf\Big\{t : \Lambda_t(\boldsymbol{q}_1) \ge B\Big\}, \quad M_A^0 = \inf\Big\{t : \Lambda_t(\boldsymbol{q}_0) \le A^{-1}\Big\},$$
$$N_B^1 = \inf\Big\{t : \hat{\Lambda}_t(\boldsymbol{q}_1) \ge B\Big\}, \quad N_A^0 = \inf\Big\{t : \hat{\Lambda}_t(\boldsymbol{q}_0) \le A^{-1}\Big\}$$

are the corresponding one-sided stopping times.

Lemma 1. *For any $1 \le i \le K$,*

$$\mathsf{E}_i[e^{-\eta}\, \mathbb{1}_{\{d_M=1\}}] \to \gamma_i, \quad \mathsf{E}_i[e^{-\hat{\eta}}\, \mathbb{1}_{\{d_N=1\}}] \to \gamma_i \quad \text{as } A, B \to \infty. \tag{3.13}$$

If additionally $r = 1$, then

$$\mathsf{E}_0[e^{-\eta}\, \mathbb{1}_{\{d_M=0\}}] \to \gamma_0^1, \quad \mathsf{E}_0[e^{-\hat{\eta}}\, \mathbb{1}_{\{d_N=0\}}] \to \gamma_0^1 \quad \text{as } A, B \to \infty. \tag{3.14}$$

Proof. We only prove the first assertions in (3.13) and (3.14), since the others can be proven in an identical way.

Since $M = M_B^1 = \inf\{t : Z_t(\boldsymbol{q}_1) \ge \log B\}$ and $\eta = Z_{M_B^1}(\boldsymbol{q}_1) - \log B$ on $\{d_M = 1\} = \{M_B^1 \le M_A^0\}$, and $\{Z_t(\boldsymbol{q}_1) = Z_t^i + \log q_1^i + Y_t^i(\boldsymbol{q}_1)\}$ is a perturbed random walk under P_i, from nonlinear renewal theory (see, e.g., Theorem 9.12 in Siegmund (1985)) it follows that η converges in distribution to $\mathcal{H}_i$ under P_i on $\{d_M = 1\}$. Therefore, the Bounded Convergence Theorem yields $\mathsf{E}_i[e^{-\eta}\, \mathbb{1}_{\{d_M=1\}}] \to \gamma_i$.

Since $M = M_A^0 = \inf\{t : -Z_t(\boldsymbol{q}_0) \ge \log A\}$ and $\eta = |Z_{M_A^0}(\boldsymbol{q}_0) + \log A|$ on $\{d_M = 0\} = \{M_B^1 > M_A^0\}$, and $\{-Z_t(\boldsymbol{q}_0) = -Z_t^1 - \log q_0^1 - Y_t^1(\boldsymbol{q}_0)\}$ is a perturbed random walk under P_0 when $r = 1$, the same argument applies to show that $\mathsf{E}_0[e^{-\eta}\, \mathbb{1}_{\{d_M=0\}}] \to \gamma_0^1$.

In what follows, we set $|\boldsymbol{q}_1| = \sum_{j=1}^K q_1^j$.

Theorem 1. (a) *For any $A, B > 1$,*

$$\mathsf{P}_0(d_M = 1) \le \frac{|\boldsymbol{q}_1|}{B}, \quad \mathsf{P}_0(d_N = 1) \le \frac{|\boldsymbol{q}_1|}{B}, \tag{3.15}$$

$$\mathsf{P}_i(d_M = 0) \le \frac{1}{A\, q_0^i}\,, \quad \mathsf{P}_i(d_N = 0) \le \frac{1}{A\, q_0^i}, \quad 1 \le i \le K. \tag{3.16}$$

(b) *As $A, B \to \infty$,*

$$\mathsf{P}_0(d_M = 1) = \frac{1}{B}\Big(\sum_{j=1}^{K} q_1^j\, \gamma_j\Big)\,(1 + o(1)), \tag{3.17}$$

$$\mathsf{P}_0(d_N = 1) \le \frac{1}{B}\Big(\sum_{j=1}^{K} q_1^j\, \gamma_j\Big)\,(1 + o(1)). \tag{3.18}$$

If additionally $r = 1$, then for every $1 \le i \le K$

$$\mathsf{P}_i(d_M = 0) \le \frac{\gamma_0^1}{q_0^i A}(1 + o(1)), \quad \mathsf{P}_i(d_N = 0) \le \frac{\gamma_0^1}{q_0^i A}(1 + o(1)). \tag{3.19}$$

Proof. Let $\mathsf{P}^{\boldsymbol{q}_1} = \frac{1}{|\boldsymbol{q}_1|} \sum_{i=1}^{K} q_1^i\, \mathsf{P}_i$ and let $\mathsf{E}^{\boldsymbol{q}_1}$ denote expectation with respect to $\mathsf{P}^{\boldsymbol{q}_1}$. Since

$$\frac{d\mathsf{P}^{\boldsymbol{q}_1}}{d\mathsf{P}_0}\Big|_{\mathcal{F}_t} = \frac{1}{|\boldsymbol{q}_1|} \sum_{i=1}^{K} q_1^i \Lambda_t^i = \frac{1}{|\boldsymbol{q}_1|} e^{Z_t(\boldsymbol{q}_1)},$$

changing the measure $\mathsf{P}_0 \mapsto \mathsf{P}^{\boldsymbol{q}_1}$ we obtain

$$\begin{aligned}\mathsf{P}_0(d_M = 1) &= |\boldsymbol{q}_1|\; \mathsf{E}^{\boldsymbol{q}_1}[e^{-Z_M(\boldsymbol{q}_1)}\; \mathbb{1}_{\{d_M=1\}}] \\ &= \sum_{i=1}^{K} q_1^i\; \mathsf{E}_i[e^{-Z_M(\boldsymbol{q}_1)}\; \mathbb{1}_{\{d_M=1\}}] = \frac{1}{B} \sum_{i=1}^{K} q_1^i\; \mathsf{E}_i[e^{-\eta}\; \mathbb{1}_{\{d_M=1\}}],\end{aligned} \tag{3.20}$$

where the last equality follows from the fact that $Z_M(\boldsymbol{q}_1) = \log B + \eta$ on $\{d_M = 1\}$. Since η is positive, the first inequality in (3.15) follows from (3.20), whereas (3.17) follows from (3.13). A similar argument as the one that led to (3.20), along with the fact that $Z_t(\boldsymbol{q}_1) \ge \hat{Z}_t(\boldsymbol{q}_1)$, yields

$$\begin{aligned}\mathsf{P}_0(d_N = 1) &= \sum_{i=1}^{K} q_1^i\; \mathsf{E}_i[e^{-Z_N(\boldsymbol{q}_1)}\; \mathbb{1}_{\{d_N=1\}}] \\ &\le \sum_{i=1}^{K} q_1^i\; \mathsf{E}_i[e^{-\hat{Z}_N(\boldsymbol{q}_1)}\; \mathbb{1}_{\{d_N=1\}}] \le \frac{1}{B} \sum_{i=1}^{K} q_1^i\; \mathsf{E}_i[e^{-\hat{\eta}}\; \mathbb{1}_{\{d_N=1\}}].\end{aligned} \tag{3.21}$$

The last inequality and the fact that $\hat{\eta}$ is positive imply the second inequality in (3.15), whereas (3.18) follows from (3.13).

Finally, changing the measure $\mathsf{P}_i \mapsto \mathsf{P}_0$, we obtain

$$\mathsf{P}_i(d_M = 0) = \mathsf{E}_0[e^{Z_M^i} \mathbb{1}_{\{d_M=0\}}]. \tag{3.22}$$

Since $Z_M^i = Z_M(\boldsymbol{q}_0) - \log q_0^i - Y_M^i(\boldsymbol{q}_0)$ (recall (3.1)), $Z_M(\boldsymbol{q}_0) = -\log A - \eta$ on $\{d_M = 0\}$ (recall (3.9)), and $Y_M^i(\boldsymbol{q}_0) \geq 0$, it follows that $Z_M^i \leq -\log(A q_0^i) - \eta$ on $\{d_M = 0\}$ and, consequently, (3.22) becomes

$$\mathsf{P}_i(d_M = 0) \leq \frac{1}{A q_0^i} \, \mathsf{E}_0[e^{-\eta} \, \mathbb{1}_{\{d_M=0\}}].$$

Since η is positive, we obtain the first inequality in (3.16), whereas from (3.14) we obtain the first inequality in (3.19). The remaining inequalities in (3.16) and (3.19) can be shown in a similar way.

From Theorem 1(a) it is clear that when A, B are selected according to

$$A_\beta(\boldsymbol{q}_0) = \frac{1}{\beta \min_{1 \leq i \leq K} q_0^i}\,, \quad B_\alpha(\boldsymbol{q}_1) = \frac{|\boldsymbol{q}_1|}{\alpha}, \tag{3.23}$$

then δ_{mi}, $\delta_{\mathrm{gl}} \in \mathcal{C}_{\alpha,\beta}$. Moreover, from Theorem 1(b) it follows that we can obtain sharper inequalities if we correct for the overshoots selecting A and B as

$$A_\beta(\boldsymbol{q}_0) = \frac{\gamma_0^1}{\beta \min_{1 \leq i \leq K} q_0^i}, \quad B_\alpha(\boldsymbol{q}_1) = \frac{\sum_{j=1}^K q_1^j \gamma_j}{\alpha}. \tag{3.24}$$

Indeed, with this selection of the thresholds we have $\mathsf{P}_0(d_M = 1) = \alpha(1 + o(1))$, $\mathsf{P}_0(d_N = 1) \leq \alpha(1 + o(1))$ and, if additionally $r = 1$, $\max_{1 \leq i \leq K} \mathsf{P}_i(d_M = 0) \leq \beta(1 + o(1))$ and $\max_{1 \leq i \leq K} \mathsf{P}_i(d_N = 0) \leq \beta(1 + o(1))$.

3.2. Asymptotic approximations to expected sample sizes

In the rest of the paper, we need the following assumptions:

(A1) $\mathsf{E}_i[(Z_1^i)^2] < \infty \quad \text{and} \quad \mathsf{E}_0[(Z_1^i)^2] < \infty, \quad 1 \leq i \leq K.$

(A2) $\alpha, \beta \to 0$ so that $|\log \alpha| / |\log \beta| \to k$, where $k \in (0, \infty)$.

(A3) For $T = M$ or $T = N$, A and B are selected so that as $\alpha, \beta \to 0$

$$k_0 \, \alpha \, (1 + o(1)) \leq \mathsf{P}_0(d_T = 1) \leq \alpha \, (1 + o(1)), \tag{3.25}$$

$$k_1 \, \beta \, (1 + o(1)) \leq \max_{1 \leq i \leq K} \mathsf{P}_i(d_T = 0) \leq \beta \, (1 + o(1)) \tag{3.26}$$

or, equivalently,

$$|\log \alpha| + o(1) \leq |\log \mathsf{P}_0(d_T = 1)| \leq |\log \alpha| + |\log k_0| + o(1), \tag{3.27}$$

$$|\log\beta| + o(1) \le |\log \max_{1\le i\le K} \mathsf{P}_i(d_T = 0)| \le |\log\beta| + |\log k_1| + o(1), \qquad (3.28)$$

where k_0, $k_1 \in (0,1)$ are fixed constants, not necessarily the same for δ_{mi} and δ_{gl}.

The second moment condition (A1) on the log-likelihood ratio Z_1^i is required even for the asymptotic approximations (2.4)–(2.5) to the performance of the SPRT for testing f_0 against f_i. Assumption (A2) concerns the relative rates at which α and β go to 0 and requires that α should not go to 0 exponentially faster than β, and vice versa. Note, however, that α can still be much smaller than β (or vice versa), as may be natural in many applications. Assumption (A3) requires that the thresholds for both the MiLRT and the WGLRT are designed so that the probabilities of type-I and type-II error are *asymptotically* bounded by (and at the same time not much smaller than) α and β, respectively. As we show in the next lemma, (A3) connects the thresholds A and B with the desired error probabilities α and β.

Lemma 2. *If* (A3) *holds, then* $\log B = |\log\alpha| + O(1)$ *and* $\log A = |\log\beta| + O(1)$.

Proof. It follows from (3.15) that $\log B \le |\log \mathsf{P}_0(d_M = 1)| + |\boldsymbol{q}_1|$, whereas from (A3), and in particular (3.27), it follows that $|\log \mathsf{P}_0(d_M = 1)| \le |\log\alpha| + |\log k_0| + o(1)$, which proves $\log B = |\log\alpha| + O(1)$. The second relationship can be shown in a similar way.

Theorem 2. *If* (A1)–(A3) *hold, then*

(a) *for every* $1 \le i \le K$,

$$I_i\, \mathsf{E}_i[M] = \log B + \kappa_i - \log q_1^i + o(1), \qquad (3.29)$$

$$I_i\, \mathsf{E}_i[N] = \log B + \kappa_i - \log q_1^i + o(1); \qquad (3.30)$$

(b) *for* $r = 1$,

$$I_0\, \mathsf{E}_0[M] = \log A + \kappa_0^1 + \log q_0^1 + o(1), \qquad (3.31)$$

$$I_0\, \mathsf{E}_0[N] = \log A + \kappa_0^1 + \log q_0^1 + o(1); \qquad (3.32)$$

(c) *for* $r > 1$,

$$I_0\, \mathsf{E}_0[M] = \log A + 2\, d_r \sqrt{\log A} + O(1), \qquad (3.33)$$

$$I_0\, \mathsf{E}_0[N] = \log A + 2\, d_r \sqrt{\log A} + O(1), \qquad (3.34)$$

where d_r *is defined in* (3.8).

Proof. (a) Asymptotic approximations (3.29) and (3.30) can be relatively easily established using nonlinear renewal theory. Specifically, starting from (3.1) and applying the Nonlinear Renewal Theorem (see Theorem 9.28 in Siegmund (1985)), it can be shown (as in Theorem 2.1 of Fellouris and Tartakovsky (2012)) that $I_i\, \mathsf{E}_i[M_B^1]$ is equal to the right-hand side of (3.29) as $B \to \infty$. Therefore, to prove (3.29) it suffices to show that $\mathsf{E}_i[M_B^1 - M] = o(1)$ as $A, B \to \infty$ or, equivalently, as $\alpha, \beta \to 0$. To this end, note that

$$0 \le M_B^1 - M = [M_B^1 - M_A^0]\, 1\!\!1_{\{d_M=0\}} \le M_B^1\, 1\!\!1_{\{d_M=0\}}.$$

Applying the Cauchy–Schwartz inequality, we obtain

$$\mathsf{E}_i[M_B^1\, 1\!\!1_{\{d_M=0\}}] \le \sqrt{\mathsf{E}_i[(M_B^1)^2]\, \mathsf{P}_i(d_M = 0)}. \tag{3.35}$$

From (3.1) and (3.3) it is clear that $Z_t(\boldsymbol{q}_1) \ge Z_t^i + \log q_1^i$, $t \in \mathbb{N}$, thus

$$M_B^1 \le \inf\{t : Z_t^i \ge \log(B/q_1^i)\}.$$

Consequently, from Theorem 8.1 in Gut (2008) it follows that, since (A1) holds,

$$(I_i)^2\, \mathsf{E}_i[(M_B^1)^2] \le (\log(B/q_1^i))^2(1 + o(1)).$$

From the latter inequality and Lemma 2 we conclude that

$$\mathsf{E}_i[(M_B^1)^2] = O((\log B)^2) = O(|\log \alpha|^2).$$

Moreover, since (A3) implies $\mathsf{P}_i(d_M = 0) \le \beta(1 + o(1))$, (3.35) gives

$$\mathsf{E}_i[M_B^1\, 1\!\!1_{\{d_M=0\}}] = O(|\log \alpha|^2 \beta),$$

and from (A2) we conclude that the upper bound goes to 0. This completes the proof of (3.29). The proof of (3.30) is analogous.

(b) From (3.1) and the Nonlinear Renewal Theorem it follows that $I_0\, \mathsf{E}_0[M_A^0]$ is equal to the right-hand side of (3.31) as $A \to \infty$. Then, as in (a), we can show that $\mathsf{E}_0[M_A^0 - M] = o(1)$. The proof of (3.32) follows similar steps.

(c) In order to prove (3.34), we start from (3.5) and apply the nonlinear renewal theory of Zhang (1988). As a result, it can be shown (analogously to Lemma 2.1 of Dragalin (1999)) that $I_0\, \mathsf{E}_0[N_A^0]$ is equal to the right-hand side of (3.34). Thus, it suffices to show that $\mathsf{E}_0[N_A^0] = \mathsf{E}_0[N] + o(1)$, which can be done as in (a) and (b).

Remark 1. The asymptotic approximation (3.34) can be further improved (up to the negligible term $o(1)$), if stronger integrability conditions are postulated on the vector W defined in (3.7). Specifically, if in addition we assume the third moment condition $\mathsf{E}_0[||W||^3] < \infty$ as well as the Cramér-type condition $\limsup_{||t|| \to \infty} \mathsf{E}_0[e^{j<t,W>}] < 1$, where j is the imaginary unit, $t = (t_1, \dots, t_r)$ and $<t, W> = \sum_{l=1}^r t_l W_l$, then

$$I_0\, \mathsf{E}_0[N] = \log A + 2\, d_r \sqrt{\log A + d_r^2} + \frac{h_r^2}{2I_0} + \kappa_0^1 + \int_{\mathbb{R}^r} \left\{ \max_{1 \le i \le r} (x_i)\ \left[\mathcal{P}(x) + \lambda(\boldsymbol{q}_0)\, \Sigma^{-1} x'\right] \right\} \phi_\Sigma(x)\, dx + o(1),$$

where $\lambda(\boldsymbol{q}_0) = (\log q_0^1, \ldots, \log q_0^r)$ and $\mathcal{P}$ is a third-degree polynomial whose coefficients depend on the P_0-cumulants of W (see Bhattacharya and Rao (1986)). This approximation can be derived similarly to Theorem 3.3 of Dragalin et al. (2000) based on the nonlinear renewal theory of Zhang (1988).

Corollary 1. *Suppose that* (A1)–(A3) *hold and that A and B are selected so that $\mathsf{P}_0(d_M = 1) \sim \alpha$ and $\mathsf{P}_0(d_N = 1) \sim \alpha$. Then,*

$$I_i \, \mathsf{E}_i[M] = |\log \alpha| + \log\Big(\sum_{j=1}^{K} q_1^j \gamma_j\Big) + \kappa_i - \log q_1^i + o(1), \tag{3.36}$$

$$I_i \, \mathsf{E}_i[N] \leq |\log \alpha| + \log\Big(\sum_{j=1}^{K} q_1^j \gamma_j\Big) + \kappa_i - \log q_1^i + o(1). \tag{3.37}$$

Proof. From (3.17)–(3.18) it follows that

$$\log B = |\log \mathsf{P}_0(d_M = 1)| + \log\Big(\sum_{j=1}^{K} q_1^j \gamma_j\Big) + o(1),$$

$$\log B \leq |\log \mathsf{P}_0(d_N = 1)| + \log\Big(\sum_{j=1}^{K} q_1^j \gamma_j\Big) + o(1).$$

Moreover, setting $k_0 = 1$ in (3.27), we obtain

$$|\log \mathsf{P}_0(d_M = 1)| = |\log \alpha| + o(1) \quad \text{and} \quad |\log \mathsf{P}_0(d_N = 1)| = |\log \alpha| + o(1).$$

From these two relationships and Theorem 2(a) we obtain the desired result.

4. Asymptotic Optimality Properties

In this section, we establish the asymptotic optimality properties of the MiLRT and the WGLRT.

4.1. Uniform asymptotic optimality

We start by showing that both tests minimize the expected sample size within an $O(1)$ term under P_i, for every $1 \leq i \leq K$, and at least to first order under P_0.

Theorem 3. *If* (A1)–(A3) *hold and A, B are selected so that $\delta_{\rm mi} \in \mathcal{C}_{\alpha,\beta}$, then*

$$\mathsf{E}_i[M] = \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_i[T] + O(1), \quad 1 \leq i \leq K, \tag{4.1}$$

$$\mathsf{E}_0[M] = \begin{cases} \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_0[T] + O(1), & \text{if } r = 1, \\ \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_0[T] \, (1 + o(1)), & \text{if } r > 1, \end{cases} \tag{4.2}$$

and similar results hold for $\delta_{\rm gl}$.

Proof. From (2.4) it is clear that

$$I_i \inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_i[T] \geq |\log \alpha| + O(1), \tag{4.3}$$

whereas from Theorem 2(a) and Lemma 2 it follows that

$$I_i \, \mathsf{E}_i[M] = \log B + O(1) = |\log \alpha| + O(1),$$

which proves (4.1). From (2.5) and the fact that $I_0 = \min_{1 \leq i \leq K} I_0^i$ it is clear that

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_0[T] \geq \frac{|\log \beta|}{I_0} + O(1). \tag{4.4}$$

Then, from Theorem 2(b) and Lemma 2 it follows that

$$I_0 \, \mathsf{E}_0[M] = \begin{cases} \log A + O(1) = |\log \beta| + O(1), & \text{if } r = 1, \\ \log A \, (1 + o(1)) = |\log \beta| \, (1 + o(1)), & \text{if } r > 1, \end{cases} \tag{4.5}$$

which implies (4.2).

4.2. Almost optimality

In what follows, we denote by $\delta^*_{\text{mi}}(\boldsymbol{p}) = (M^*(\boldsymbol{p}), d_{M^*(\boldsymbol{p})})$ and $\delta^*_{\text{gl}}(\boldsymbol{p}) = (N^*(\boldsymbol{p}), d_{N*(\boldsymbol{p})})$ the MiLRT and the WGLRT with weights

$$q_1^i = \frac{p_i}{\mathcal{L}_i} \quad \text{and} \quad q_0^i = p_i \, \mathcal{L}_i, \quad 1 \leq i \leq K, \tag{4.6}$$

where $\boldsymbol{p} = (p_1, \ldots, p_K)$, $p_i > 0$ for every $1 \leq i \leq K$, and $\sum_{i=1}^K p_i = 1$. Then, from Corollary 1 and the fact that $\mathcal{L}_i = \gamma_i I_i$, $1 \leq i \leq K$, it follows that if B is selected so that $\mathsf{P}_0(d_{M^*(\boldsymbol{p})} = 1) \sim \alpha$ and $\mathsf{P}_0(d_{N^*(\boldsymbol{p})} = 1) \sim \alpha$, then

$$\mathsf{E}_i[M^*(\mathbf{p})] = \frac{1}{I_i} \Big[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \Big] + o(1), \tag{4.7}$$

$$\mathsf{E}_i[N^*(\mathbf{p})] \leq \frac{1}{I_i} \Big[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \Big] + o(1), \tag{4.8}$$

where

$$C_i(\mathbf{p}) = \log \Big(\sum_{j=1}^K \frac{p_j}{I_j} \Big) - \log \frac{p_i}{I_i}, \quad 1 \leq i \leq K. \tag{4.9}$$

The next theorem states that $\delta^*_{\text{mi}}(\boldsymbol{p})$ and $\delta^*_{\text{gl}}(\boldsymbol{p})$ attain $\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T]$ asymptotically within an $o(1)$ term, where $\mathsf{E}^{\boldsymbol{p}}$ is expectation with respect to the weighted probability measure $\mathsf{P}^{\boldsymbol{p}} = \sum_{i=1}^K p_i \, \mathsf{P}_i$.

Theorem 4. *Suppose that* (A1)–(A3) *hold and* $\alpha, \beta \to 0$ *so that* $|\log \alpha| \sim |\log \beta|$. *Then*

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] = \sum_{i=1}^{K} \frac{p_i}{I_i} \Big[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \Big] + o(1). \tag{4.10}$$

If, additionally, A and B are selected so that $\delta^*_{\mathrm{mi}}(\boldsymbol{p})$ and $\delta^*_{\mathrm{gl}}(\boldsymbol{p})$ belong to $\mathcal{C}_{\alpha,\beta}$ and $\mathsf{P}_0(d_{M^*(\boldsymbol{p})} = 1) \sim \alpha$ and $\mathsf{P}_0(d_{N^*(\boldsymbol{p})} = 1) \sim \alpha$, then

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] = \mathsf{E}^{\boldsymbol{p}}[M^*(\boldsymbol{p})] + o(1),$$
$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] = \mathsf{E}^{\boldsymbol{p}}[N^*(\boldsymbol{p})] + o(1).$$

To prove this theorem, we formulate our sequential testing problem as a Bayesian sequential decision problem with $K+1$ states, $\mathsf{H}_0 : f = f_0$ and $\mathsf{H}_1^i : f = f_i$, $1 \le i \le K$, and two possible actions upon stopping, selecting either H_0 or $\mathsf{H}_1 = \cup_i H_1^i$. Let c denote the sampling cost per observation and let w_1 (resp. w_0) be the loss associated with accepting H_0 (resp. H_1) when the correct hypothesis is H_1 (resp. H_0). Define $\mathbf{P}^\pi = \pi\, \mathsf{P}_0 + (1-\pi)\, \mathsf{P}^{\boldsymbol{p}}$, which means that $\pi = \mathbf{P}^\pi(\mathsf{H}_0)$ is the prior probability of H_0 and $p_i = \mathbf{P}^\pi(\mathsf{H}_1^i | \mathsf{H}_1)$ is the prior probability of $f = f_i$ given that H_1 is correct.

The integrated risk of a sequential test $\delta = (T, d_T)$ is the sum $\mathcal{R}(\delta) = \mathcal{R}_c(T) + \mathcal{R}_s(d_T)$, where $\mathcal{R}_c(T)$ is the integrated risk due to sampling and $\mathcal{R}_s(d_T)$ is the integrated risk due to a wrong decision upon stopping,

$$\begin{aligned} \mathcal{R}_c(T) &= c\, \mathbf{E}^\pi[T] = c\Big[\pi\, \mathsf{E}_0[T] +\ (1-\pi)\, \mathsf{E}^{\boldsymbol{p}}[T]\Big], \\ \mathcal{R}_s(d_T) &= \mathbf{E}^\pi[w_0\, \mathbb{1}_{\{d_T=1\}} | \mathsf{H}_0] + \mathbf{E}^\pi[w_1\, \mathbb{1}_{\{d_T=0\}} | \mathsf{H}_1] \\ &= \pi\, w_0\, \mathsf{P}_0(d_T = 1) + (1-\pi)\, w_1\, \mathsf{P}^{\boldsymbol{p}}(d_T = 0). \end{aligned}$$

The Bayesian sequential decision problem is to find an optimal (*Bayes*) sequential test that attains the *Bayes risk*, $\mathcal{R}^* = \inf_\delta \mathcal{R}(\delta)$. It is well known that the solution to this problem does not have a simple structure (see, e.g., Chow et al. (1971)). However, from the seminal work of Lorden (1977) on finite-state sequential decision making it follows that $\delta^*_{\mathrm{mi}}(\mathbf{p})$ and $\delta^*_{\mathrm{gl}}(\mathbf{p})$ are *almost Bayes* when their thresholds A and B are chosen as

$$A_c = \frac{1-\pi}{\pi} \frac{w_1}{c} \quad \text{and} \quad B_c = \frac{\pi}{1-\pi} \frac{w_0}{c}. \tag{4.11}$$

More specifically, denote by $\delta^*_{\mathrm{mi,c}}(\boldsymbol{p}) = (M^*_c(\boldsymbol{p}),\, d_{M^*_c(\boldsymbol{p})})$ and $\delta^*_{\mathrm{gl,c}}(\boldsymbol{p}) = (N^*_c(\boldsymbol{p}),\, d_{N^*_c(\boldsymbol{p})})$ the sequential tests $\delta^*_{\mathrm{mi}}(\boldsymbol{p})$ and $\delta^*_{\mathrm{gl}}(\boldsymbol{p})$ whose thresholds are given by A_c

and B_c. Under the integrability condition (A1), it follows from Lorden (1977) that

$$\mathcal{R}(\delta^*_{\mathrm{mi,c}}(\boldsymbol{p})) - \mathcal{R}^* = o(c) \quad \text{and} \quad \mathcal{R}(\delta^*_{\mathrm{gl,c}}(\boldsymbol{p})) - \mathcal{R}^* = o(c). \tag{4.12}$$

The proof of Theorem 4 relies on this third-order Bayesian asymptotic optimality property.

Proof. In order to lighten the notation, we omit the dependence on the prior distribution $\boldsymbol{p}$ and write simply $\delta^*_{\mathrm{mi}} = (M^*, d_{M^*})$ and $\delta^*_{\mathrm{mi,c}} = (M^*_c, d_{M^*_c})$, and similarly for the WGLRT.

From Corollary 1 it is clear that the right-hand side in (4.10) is attained by δ^*_{mi} and δ^*_{gl} when their thresholds are selected so that $\mathsf{P}_0(d_{M^*} = 1) \sim \alpha$ and $\mathsf{P}_0(d_{N^*} = 1) \sim \alpha$. If additionally $\delta^*_{\mathrm{mi}}, \delta^*_{\mathrm{gl}} \in \mathcal{C}_{\alpha,\beta}$, then $\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T]$ is attained by these two tests to within an $o(1)$ term. Thus it suffices to establish (4.10).

Consider the class of sequential tests

$$\mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta} = \{\delta = (T, d_T) : \mathsf{P}_0(d_T = 1) \le \alpha \quad \text{and} \quad \mathsf{P}^{\boldsymbol{p}}(d_T = 0) \le \beta\}.$$

Since $\mathcal{C}_{\alpha,\beta} \subset \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}$, we have $\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] \ge \inf_{\delta \in \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T]$. Thus, it suffices to show that

$$\inf_{\delta \in \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] = \sum_{i=1}^{K} \frac{p_i}{I_i} \Big[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \Big] + o(1). \tag{4.13}$$

Consider now the sequential test $\delta^*_{\mathrm{mi,c}} = (M^*_c, d_{M^*_c})$ with thresholds A_c and B_c selected so that $\mathsf{P}_0(d_{M^*_c} = 1) = \alpha$ and $\mathsf{P}^{\boldsymbol{p}}(d_{M^*_c} = 0) = \beta$. From Corollary 1 it is clear that $\mathsf{E}^{\boldsymbol{p}}[M^*_c]$ is equal to the right-hand side in (4.13) as $c \to 0$, which means that it suffices to show that

$$\inf_{\delta \in \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}} \mathsf{E}^{\boldsymbol{p}}[T] = \mathsf{E}^{\boldsymbol{p}}[M^*_c] + o(1),$$

where $o(1)$ is an asymptotically negligible term as $c \to 0$. More specifically, if δ is an arbitrary sequential test in $\mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}$, we need to show that, for sufficiently small c, $|\mathsf{E}^{\boldsymbol{p}}[T] - \mathsf{E}^{\boldsymbol{p}}[M^*_c]|$ is bounded above by an arbitrarily small, but fixed number.

We observe that

$$\begin{aligned} \mathcal{R}_s(d_T) &= \pi\, w_0\, \mathsf{P}_0(d_T = 1) + (1-\pi)\, w_1\, \mathsf{P}^{\boldsymbol{p}}(d_T = 0) \\ &\le \pi\, w_0\, \alpha + (1-\pi)\, w_1\, \beta = \mathcal{R}_s(d_{M^*_c}), \end{aligned} \tag{4.14}$$

where the inequality is due to $\delta \in \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}$ and the second equality follows from the assumption that $\mathsf{P}_0(d_{M^*_c} = 1) = \alpha$ and $\mathsf{P}^{\boldsymbol{p}}(M^*_c = 0) = \beta$.

From (3.15)–(3.16) and the definition of A_c and B_c in (4.11) we have

$$\begin{aligned}\mathcal{R}_s(d_{M_c^*}) &= \pi\, w_0\, \mathsf{P}_0(d_{M_c^*}=1) + (1-\pi)\, w_1\, \mathsf{P}^{\boldsymbol{p}}(d_{M_c^*}=0)\\ &\le \pi\, w_0\, \frac{|\boldsymbol{q}_1|}{B_c} + (1-\pi)\, w_1 \sum_{i=1}^K p_i \frac{1}{A_c q_0^i}\\ &\le |\boldsymbol{q}_1|(1-\pi)c + \sum_{i=1}^K p_i \frac{\pi c}{q_0^i} \le (Q-1)c, \end{aligned} \tag{4.15}$$

where $Q > 1$ is some constant that does not depend on c or π.

Fix $\epsilon > 0$ and introduce the sequential test

$$T_{\epsilon c} = \min\{M_{\epsilon c}^*, T\}\,, \quad d_{T_{\epsilon c}} = d_T \mathbb{1}_{\{T \le M_{\epsilon c}^*\}} + d_{M_{\epsilon c}^*} \mathbb{1}_{\{T > M_{\epsilon c}^*\}}.$$

Obviously,

$$\begin{aligned}\mathcal{R}_s(d_{T_{\epsilon c}}) \le \mathcal{R}_s(d_T) + \mathcal{R}_s(d_{M_{\epsilon c}^*}) &\le \mathcal{R}_s(d_{M_c^*}) + \mathcal{R}_s(d_{M_{\epsilon c}^*})\\ &\le \mathcal{R}_s(d_{M_c^*}) + (Q-1)c\,\epsilon, \end{aligned} \tag{4.16}$$

where the second inequality is due to (4.14) and the third is due to (4.15).

Since M_c^* is almost Bayes (recall (4.12)), for sufficiently small c,

$$\mathcal{R}_c(M_c^*) + \mathcal{R}_s(d_{M_c^*}) \le \mathcal{R}_c(T_{\epsilon c}) + \mathcal{R}_s(d_{T_{\epsilon c}}) + c\,\epsilon. \tag{4.17}$$

From (4.16) we obtain $\mathcal{R}_c(M_c^*) \le \mathcal{R}_c(T_{\epsilon c}) + Q\,c\,\epsilon$ and, consequently,

$$\begin{aligned}\pi\, \mathsf{E}_0[M_c^*] + (1-\pi)\, \mathsf{E}^{\boldsymbol{p}}[M_c^*] &\le \pi\, \mathsf{E}_0[T_{\epsilon c}] + (1-\pi)\, \mathsf{E}^{\boldsymbol{p}}[T_{\epsilon c}] + Q\,\epsilon\\ &\le \pi\, \mathsf{E}_0[M_{\epsilon c}^*] + (1-\pi)\, \mathsf{E}^{\boldsymbol{p}}[T] + Q\,\epsilon, \end{aligned} \tag{4.18}$$

where the second inequality follows from the definition of $T_{\epsilon c}$. Rearranging terms, we obtain from (4.18) that

$$\mathsf{E}^{\boldsymbol{p}}[M_c^*] - \mathsf{E}^{\boldsymbol{p}}[T] \le \frac{\pi}{1-\pi}\Big(\mathsf{E}_0[M_{\epsilon c}^*] - \mathsf{E}_0[M_c^*]\Big) + \frac{Q\,\epsilon}{1-\pi}. \tag{4.19}$$

Since the last inequality holds for any $\pi \in (0,1)$, we can set $\pi = \epsilon/(1+\epsilon)$, which implies $B_c = \epsilon w_0/c$ and $A_c = w_1/(\epsilon\, c)$, whereas (4.19) becomes

$$\mathsf{E}^{\boldsymbol{p}}[M_c^*] - \mathsf{E}^{\boldsymbol{p}}[T] \le \epsilon\,(\mathsf{E}_0[M_{\epsilon c}^*] - \mathsf{E}_0[M_c^*]) + Q\,\epsilon(1+\epsilon). \tag{4.20}$$

But from (3.31) and (3.34) it follows that, as $c \to 0$,

$$I_0\,(\mathsf{E}_0[M_{\epsilon c}^*] - \mathsf{E}_0[M_c^*]) = O(\log A_{\epsilon c} - \log A_c)$$

and, from (4.11), we have $\log A_{\epsilon c} - \log A_c = |\log \epsilon| + O(1)$ as $c \to 0$, which completes the proof.

Remark 2. With a similar argument as the one used in the proof of Theorem 4 it can be shown that if $\mathsf{P}_0(d_{M^*(\boldsymbol{p})} = 1) = \alpha$ and $\mathsf{P}^{\boldsymbol{p}}(d_{M^*(\boldsymbol{p})} = 0) = \beta$, then

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathsf{E}_0[T] \geq \inf_{\delta \in \mathcal{C}^{\boldsymbol{p}}_{\alpha,\beta}} \mathsf{E}_0[T] = \mathsf{E}_0[M^*(\boldsymbol{p})] + o(1),$$

and similarly for δ_{gl}. However, the right-hand side in this asymptotic lower bound is generally not attained by $\delta^*_{\mathrm{mi}}(\boldsymbol{p})$ or $\delta^*_{\mathrm{gl}}(\boldsymbol{p})$ when their thresholds are selected so that $\delta_{\mathrm{mi}}, \delta_{\mathrm{gl}} \in \mathcal{C}_{\alpha,\beta}$.

Remark 3. While we have no rigorous proof, we believe that the assertions of Theorem 4 (as well as of Theorem 5 below) hold true in the more general case where α and β approach zero in such a way that the ratio $\log \alpha / \log \beta$ is bounded away from zero and infinity, which allows one to cover the asymptotically asymmetric case as well.

4.3. Almost minimaxity

For any stopping time T and $1 \leq i \leq K$, we set $\mathcal{I}_i(T) = I_i \mathsf{E}_i[T]$. Without loss of generality, we restrict ourselves to P_i-integrable stopping times. Thus, from Wald's identity it follows that $\mathcal{I}_i(T) = \mathsf{E}_i[Z^i_T]$, that is, $\mathcal{I}_i(T)$ is the expected KL divergence between P_i and P_0 that has been accumulated up to time T.

Let $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_K)$ denote the prior distribution for which

$$\hat{p}_i = \frac{\mathcal{L}_i e^{\kappa_i}}{\sum_{j=1}^K \mathcal{L}_j \, e^{\kappa_j}}, \quad 1 \leq i \leq K. \tag{4.21}$$

Then, from (3.29)–(3.30) it follows that $\hat{\boldsymbol{p}}$ (almost) equalizes the KL divergence that is accumulated by both the MiLRT and the WGLRT until stopping, in the sense that $\mathcal{I}_i(M^*(\hat{\boldsymbol{p}}))$ and $\mathcal{I}_i(N^*(\hat{\boldsymbol{p}}))$ are independent of i up to an $o(1)$ term. Indeed,

$$\mathcal{I}_i(M^*(\hat{\boldsymbol{p}})) = \log B + \log\Big(\sum_{j=1}^K e^{\mathcal{L}_j \, \kappa_j}\Big) + o(1), \tag{4.22}$$

$$\mathcal{I}_i(N^*(\hat{\boldsymbol{p}})) = \log B + \log\Big(\sum_{j=1}^K e^{\mathcal{L}_j \, \kappa_j}\Big) + o(1), \tag{4.23}$$

where only negligible terms $o(1)$ may depend on i. If additionally B is selected so that $\mathsf{P}(d_{M^*(\hat{\boldsymbol{p}})} = 1) \sim \alpha$ and $\mathsf{P}(d_{N^*(\hat{\boldsymbol{p}})} = 1) \sim \alpha$, then (3.36)–(3.37) imply that for every $1 \leq i \leq K$,

$$\mathcal{I}_i(M^*(\hat{\mathbf{p}})) = |\log \alpha| + \log\Big(\sum_{j=1}^K \gamma_j e^{\kappa_j}\Big) + o(1), \tag{4.24}$$

$$\mathcal{I}_i(N^*(\hat{\mathbf{p}})) \leq |\log \alpha| + \log\Big(\sum_{j=1}^{K} \gamma_j e^{\kappa_j}\Big) + o(1). \tag{4.25}$$

Consequently, if we denote by $\hat{\mathcal{I}}(T) = \max_{1\leq i\leq K} \mathcal{I}_i(T)$ the maximal expected KL divergence until stopping, we have

$$\hat{\mathcal{I}}(M^*(\hat{\mathbf{p}})) = |\log \alpha| + \log\Big(\sum_{j=1}^{K} \gamma_j e^{\kappa_j}\Big) + o(1), \tag{4.26}$$

$$\hat{\mathcal{I}}(N^*(\hat{\mathbf{p}})) \leq |\log \alpha| + \log\Big(\sum_{j=1}^{K} \gamma_j e^{\kappa_j}\Big) + o(1). \tag{4.27}$$

In the following theorem we show that $\delta_{\rm mi}(\hat{\boldsymbol{p}})$ and $\delta_{\rm gl}(\hat{\boldsymbol{p}})$ are almost minimax in this KL sense.

Theorem 5. *Suppose that* (A1)–(A3) *hold and* $\alpha, \beta \to 0$ *so that* $|\log \alpha| \sim |\log \beta|$. *Then,*

$$\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \hat{\mathcal{I}}[T] = |\log \alpha| + \log\Big(\sum_{j=1}^{K} \gamma_j e^{\kappa_j}\Big) + o(1). \tag{4.28}$$

If, additionally, the thresholds A and B are selected so that $\delta_{\rm mi}(\hat{\boldsymbol{p}})$, $\delta_{\rm gl}(\hat{\boldsymbol{p}}) \in \mathcal{C}_{\alpha,\beta}$ *and* $\mathsf{P}(d_{M^*(\hat{\boldsymbol{p}})} = 1) \sim \alpha$ *and* $\mathsf{P}(d_{N^*(\hat{\boldsymbol{p}})} = 1) \sim \alpha$, *then*

$$\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \hat{\mathcal{I}}(T) = \hat{\mathcal{I}}[M^*(\hat{\boldsymbol{p}})] + o(1), \tag{4.29}$$

$$\inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \hat{\mathcal{I}}(T) = \hat{\mathcal{I}}[N^*(\hat{\boldsymbol{p}})] + o(1). \tag{4.30}$$

Proof. If A and B are selected so that $\delta_{\rm mi}(\hat{\boldsymbol{p}}) \in \mathcal{C}_{\alpha,\beta}$ and $\mathsf{P}(d_{M^*(\hat{\boldsymbol{p}})} = 1) \sim \alpha$, then it follows from Theorem 4 that

$$\begin{aligned}\sum_{i=1}^{K} \hat{p}_i\, \mathsf{E}_i[M^*(\hat{\mathbf{p}})] + o(1) &\leq \inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \sum_{i=1}^{K} \hat{p}_i\, \mathsf{E}_i[T] = \inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \sum_{i=1}^{K} \frac{\hat{p}_i}{I_i}\, \mathcal{I}_i(T) \\ &\leq \Big(\sum_{i=1}^{K} \frac{\hat{p}_i}{I_i}\Big) \inf_{\delta\in\mathcal{C}_{\alpha,\beta}} \hat{\mathcal{I}}(T),\end{aligned} \tag{4.31}$$

whereas from (4.24) and (4.26) we have

$$\sum_{i=1}^{K} \hat{p}_i \mathsf{E}_i[M^*(\hat{\mathbf{p}})] = \Big(\sum_{i=1}^{K} \frac{\hat{p}_i}{I_i}\Big) \Big[|\log \alpha| + \log\Big(\sum_{j=1}^{K} \gamma_j e^{\kappa_j}\Big) + o(1)\Big] \tag{4.32}$$

$$= \Big(\sum_{i=1}^{K} \frac{\hat{p}_i}{I_i}\Big)\, \hat{\mathcal{I}}(M^*(\hat{\mathbf{p}})) + o(1). \tag{4.33}$$

From (4.31) and (4.32) we obtain (4.28), and from (4.31) and (4.33) we obtain (4.29). Finally, from (4.27) and (4.28) we obtain (4.30).

5. How to Select $\boldsymbol{p}$

In this section, we consider the specification of the prior distribution $\boldsymbol{p}$ that determines the weights $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$ of the MiLRT and the WGLRT, when these are selected according to (4.6). In order to do so, we quantify the resulting performance loss for the MiLRT (and similarly for the WGLRT) under P_i using the measure

$$\mathcal{J}_i(\boldsymbol{p}) = \frac{\mathsf{E}_i[M^*(\boldsymbol{p})] - \mathsf{E}_i[S^i]}{\mathsf{E}_i[S^i]}, \quad 1 \le i \le K,$$

where S^i is the SPRT for testing f_0 against f_i with type-I and type-II error probabilities α and β, respectively. Thus, $\mathcal{J}_i(\boldsymbol{p})$ represents the *additional* expected sample size due to the uncertainty in the alternative hypothesis divided by the smallest possible expected sample size that is required for testing f_0 against f_i.

Then, if (A1)–(A3) hold and $k_0 = 1$, from (2.4) and (4.7) it follows that

$$\mathcal{J}_i(\mathbf{p}) \approx \frac{C_i(\boldsymbol{p})}{|\log\alpha| + \kappa_i + \log\gamma_i} = \frac{\log\Big[\sum_{j=1}^{K}(p_j/I_j)\Big] + \log I_i - \log p_i}{|\log\alpha| + \kappa_i + \log\gamma_i}, \tag{5.1}$$

where by $\approx$ we mean that the two sides differ by an $o(1)$ term. From this expression we can see that the magnitude of $\mathcal{J}_i(\boldsymbol{p})$ is mainly determined by K, the cardinality of $\mathcal{A}_1$, and the probability of type-I error, α. In particular, for every $1 \le i \le K$ and $\mathbf{p}$, $\mathcal{J}_i(\mathbf{p})$ will be "small" when $|\log\alpha|$ is much larger than $\log K$, which implies that the choice of $\boldsymbol{p}$ may make a difference only when $|\log\alpha|$ is not much larger than $\log K$.

In Figure 1, we compare different priors with respect to this criterion, in particular $\hat{\boldsymbol{p}}$, defined in (4.21), as well as $\boldsymbol{p}^I$, $\mathbf{p}^{\mathcal{L}}$, and $\mathbf{p}^{\mathbf{u}}$ defined so that $p_i^I \propto I_i$, $p_i^{\mathcal{L}} \propto \mathcal{L}_i$, and $p_i^u \propto 1$. We note that $\mathbf{p}^{\mathcal{L}}$, $\mathbf{p}^{\mathcal{I}}$, and $\hat{\mathbf{p}}$ are ranked in the sense that $\mathcal{L}_i \le I_i \le e^{\kappa_i}\mathcal{L}_i$, since $\mathcal{L}_i = \gamma_i I_i$ and $\gamma_i \le 1 \le e^{\kappa_i}\gamma_i$. Thus, $\boldsymbol{p}^{\mathcal{L}}$ (resp. $\hat{\boldsymbol{p}}$) assigns relatively less (resp. more) weight than $\boldsymbol{p}^{\mathcal{I}}$ to a hypothesis as its "signal-to-noise ratio" increases. We also note that $\mathbf{p}^{\mathcal{L}}$ and $\hat{\mathbf{p}}$ reduce to $\mathbf{p}^{\mathcal{I}}$ when there is no overshoot effect, in which case $\kappa_i = 0$ and $\gamma_i = 1$, whereas all these priors reduce to $\mathbf{p}^{\mathbf{u}}$ in the symmetric case where I_i and $\mathcal{H}_i$ do not depend on i.

In order to make some concrete comparisons, we focus on the multichannel setup (1.5), assuming that $g_0^i(x) = g(x;0)$ and $g_1^i(x) = g(x;\theta_i)$ for every $1 \le i \le$

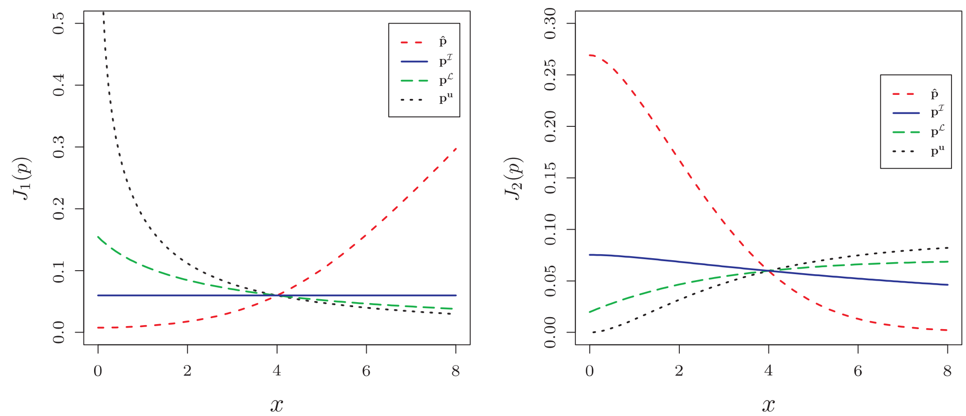

Figure 1. Performance loss for different prior distributions in a multichannel problem with exponential data.

$K = 2$, where $g(x;\theta) = (1+\theta)^{-1}\exp\{-x/(1+\theta)\}$, $x > 0$, is a density of the exponential distribution with mean $1+\theta$, in which case $I_i = \theta_i - \log(1+\theta_i)$, $\kappa_i = \theta_i$, $\gamma_i = (1+\theta_i)^{-1}$. We also assume that $\theta_1 = 4$ and we let $\theta_2 = x$ vary. Thus, the signal-to-noise ratio in the first (resp. second) channel is stronger (resp. weaker) than that in the second (resp. first) channel when $x < 4$ (resp. $x > 4$). In Figure 1, we plot $\mathcal{J}_1(\boldsymbol{p})$ and $\mathcal{J}_2(\boldsymbol{p})$, the inflicted performance loss when signal is present in the first and second channel, respectively, as a function of x, for the above priors. We do so using asymptotic approximation (5.1), in which we set $\alpha = 10^{-4}$.

These plots show that $\boldsymbol{p} = \hat{\mathbf{p}}$ (resp. $\boldsymbol{p} = \mathbf{p}^{\mathbf{u}}$) leads to a better performance when signal is present in the channel with stronger (resp. weaker) signal-to-noise ratio. However, the inflicted performance loss when the signal is present in the other channel can be very high. On the other hand, $\boldsymbol{p} = \mathbf{p}^{\mathcal{I}}$ or $\boldsymbol{p} = \mathbf{p}^{\mathcal{L}}$ lead to a more robust behavior, since the resulting performance loss is relatively low and stable for various signal strengths.

Note that this exponential example has practical meaning in radar applications when detecting a signal (from a target), representing a sequence of slowly fluctuating pulses in white Gaussian noise after square-law pre-processing (see, e.g., Tartakovsky (1991)), in which case θ is the signal-to-noise ratio at the output of the square-law detector.

6. Monte Carlo Simulations

In this section, we present a simulation study to verify the accuracy of the asymptotic approximations established in Section 4 and compare the MiLRT with the WGLRT for realistic probabilities of errors. We considered the multichannel

Table 1. Parameter values in a multichannel problem with exponential data.

θ_i	I_i	κ_i	γ_i	q_1^i	q_0^i
0.5	0.095	0.5	0.67	0.308	0.013
1	0.584	1	0.4	0.837	0.078
2	0.901	2	0.33	1.380	0.138

Table 2. Type-I error probabilities and the expected sample sizes under P_i, $i = 1, 2, 3$ for different values of the target probability α when $\beta = 10^{-2}$.

α	$\frac{\mathsf{P}_0(d_{M^*}=1)}{\alpha}$	$\frac{\mathsf{P}_0(d_{N^*}=1)}{\alpha}$	$\mathsf{E}_1[M^*]$	$\mathsf{E}_1[N^*]$	$\mathsf{E}_2[M^*]$	$\mathsf{E}_2[N^*]$	$\mathsf{E}_3[M^*]$	$\mathsf{E}_3[N^*]$
10^{-2}	1.051	0.994	59.9	59.4	17.8	19.4	6.2	7.3
10^{-3}	1.033	0.995	84.1	84.1	25.7	27.1	9.0	9.9
10^{-4}	1.025	0.996	108.5	108.3	33.7	34.6	11.7	12.4
10^{-5}	1.017	0.996	132.5	132.3	41.4	42.0	14.3	15.0

setup (1.5) with $K = 3$ channels, $g_0^i(x) = g(x; 0)$ and $g_1^i(x) = g(x; \theta_i)$ for every $1 \le i \le K$, where $g(x;\theta) = (1+\theta)^{-1} \exp\{-x/(1+\theta)\}$, $x > 0$, and selected the parameter values according to Table 1. Since our main emphasis is on the fast detection of signal, we set $\beta = 10^{-2}$ (miss detection probability) and considered different values of α (false alarm probability). We chose the thresholds A and B according to (3.24) and selected the weights according to (4.6) with prior $\boldsymbol{p} = \boldsymbol{p}^{\mathcal{I}}$.

In the first three columns of Table 2 we compare the type-I error probabilities for the two tests, computed based on simulation experiments, against the target level α. More specifically, these error probabilities were computed using (3.20), (3.21) and *importance sampling*, a simulation technique whose application in Sequential Analysis goes back to Siegmund (1976). These results indicate that selecting B according to (3.23) leads to type-I error probabilities very close to α for both tests, even for not too small α. In particular, we see that $\mathsf{P}_0(d_{M^*} = 1)$ is slightly larger than α, which is expected, since (3.23) implies $\mathsf{P}_0(d_{M^*} = 1) \sim \alpha$, whereas α is a sharp upper bound for $\mathsf{P}_0(d_{N^*} = 1)$, the type-I error probability of the WGLRT.

In the remaining columns of Table 2, we present the (simulated) expected sample size under P_i, $i = 1, 2, 3$, and in Figure 2 we plot these values against the corresponding (simulated) type-I error probabilities. In these graphs, we also superimpose asymptotic approximation (3.36) (dashed lines), as well as the asymptotic performance of the corresponding SPRT (solid lines), given by (2.4). Triangles correspond to the WGLRT and circles to the MiLRT. From these results we can see that (3.36) is very accurate for both tests. Also, the two tests have almost the same performance; in particular, their performance is practically identical when the signal is present in the channel with the smallest signal strength. In the other two cases, the MiLRT seems to perform slightly better.

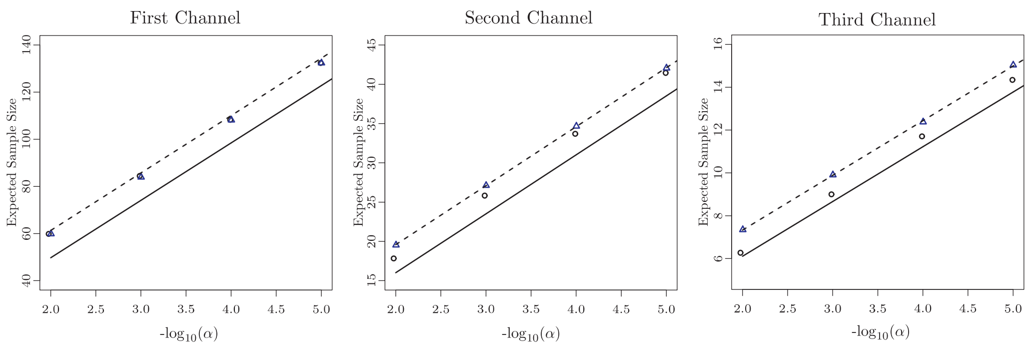

Figure 2. Simulated expected sample size of MiLRT (circles) and WGLRT (triangles) under P_i against type-I error probability in logarithmic scale, $i = 1, 2, 3$. The dashed line represents asymptotic approximation (3.36) and the solid line refers to (2.4), the asymptotic performance of the corresponding SPRT.

7. Conclusion

We performed a detailed analysis and optimization of weighted GLR and mixture-based sequential tests when the null hypothesis is simple and the alternative hypothesis is composite but discrete. Independently of the choice of weights, both tests minimize asymptotically, at least to first order and often to second order, the expected sample size under each possible scenario as error probabilities go to 0. With an appropriate selection of weights, both tests achieve higher-order asymptotic optimality properties. Specifically, they minimize a weighted expected sample size, as well as the expected Kullback–Leibler divergence in the least favorable scenario, to within asymptotically negligible terms as error probabilities go to zero. Based on simulation experiments, we found that the two tests perform similarly even for not too small error probabilities. The proposed approach can be extended to sequential testing of multiple hypotheses, a substantially more complex problem that we plan to consider elsewhere.

Acknowledgements

We are grateful to a referee and an associate editor whose comments improved the presentation. This work was supported by the U.S. Air Force Office of Scientific Research under MURI grant FA9550-10-1-0569, by the U.S. Defense Threat Reduction Agency under grant HDTRA1-10-1-0086, by the U.S. Defense Advanced Research Projects Agency under grant W911NF-12-1-0034 and by the U.S. National Science Foundation under grants CCF-0830419, EFRI-1025043, and DMS-1221888 at the University of Southern California, Department of Mathematics.

References

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Chernoff, H. (1972). *Sequential Analysis and Optimal Design.* SIAM, Philadelphia.

Chow, Y. S. , Robbins, H., and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston.

Dragalin, V. (1999). Asymptotics for a sequential selection procedure. *Statist. Decisions* **4**, 123-137.

Dragalin, V. P., Tartakovsky, A. G., and Veeravalli, V. V. (2000). Multihypothesis sequential probability ratio tests - Part II: Accurate asymptotic expansions for the expected sample size. *IEEE Trans. Inform. Theory* **46**, 1366-1343.

Dragalin, V. and Novikov, A. (1999). Adaptive sequential tests for composite hypotheses. *Surveys in Applied and Industrial Mathematics, TVP press* **6**, 387-398.

Fellouris, G. and Tartakovsky, A. G. (2012). Nearly minimax mixture-based open-ended sequential tests. *Sequential Analysis* **31**, 297-325.

Gut, A. (2008). *Stopped Random Walks.* Second Edition. Springer.

Kiefer, J. and Sacks, J. (1963). Asymptotically optimal sequential inference and design. *Ann. Math. Statist.* **34**, 705-750.

Lai, T. L. (1988). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16**, 856-886.

Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5**, 628-643.

Lai, T. L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7**, 60-76.

Lorden, G. (1967). Integrated risk of asymptotically Bayes sequential tests. *Ann. Math. Statist.* **38**, 1399-1422.

Lorden, G. (1973). Open-ended tests for Koopman–Darmois families. *Ann. Statist.* **1**, 633-643.

Lorden, G. (1977). Nearly optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5**, 1-21.

Pavlov, I. V. (1990). A sequential procedure for testing composite hypotheses with application to the Kiefer–Weiss problem. *Theory Probab. Appl.* **35**, 280-292.

Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3**, 1267-1282.

Robbins, H. and Siegmund, D. (1970). A class of stopping rules for testing parameter hypotheses. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **4** (Edited by K. M. Le Cam, J. Neyman, and E. L. Scott), 37-41. University of California Press, Berkeley.

Robbins, H. and Siegmund, D. (1974). The expected sample size of some tests of power one. *Ann. Math. Statist.* **2**, 415-436.

Schwarz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33**, 224-236.

Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4**, 673-684.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer-Verlag, New York.

Tartakovsky, A. G. (1991). *Sequential Methods in the Theory of Information Systems.* Radio & Communications, Moscow (in Russian).

Tartakovsky, A. G., Li, X. R., and Yaralov, G. (2003). Sequential detection of targets in multichannel systems. *IEEE Trans. Inform. Theory* **49**, 425-445.

Tartakovsky, A. G., Rozovskii, B. L., Blaźek, R. B., and Kim, H. (2006). Detection of intrusions in information systems by sequential change-point methods. *Statist. Methodol.* **3**, 252-293.

Wald, A. (1944). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16**, 117-186.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19**, 326-339.

Woodroofe, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probab.* **4**, 67-80.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* SIAM, Philadelphia.

Zhang, C. H. (1988). A nonlinear renewal theory. *Ann. Probab.* **16**, 793-825.

University of Southern California, Department of Mathematics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA.

E-mail: fellouri@usc.edu

University of Southern California, Department of Mathematics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA.

E-mail: tartakov@usc.edu

(Received March 2012; accepted March 2013)