Annals of Mathematical Sciences and Applications
Volume 0, Number 0, 1, 2024, Special issue: Articles of optimization, wavelets, financial mathematics, and mathematical problems related to statistics, **in memory of Professor Tze Leung Lai**

Nearly Optimum Properties of Certain Multi-Decision Sequential Rules for General Non-i.i.d. Stochastic Models

Alexander G. Tartakovsky

Dedicated to the memory of Professor Tze Leung Lai, this paper introduces three multi-hypothesis sequential tests. These tests are derived from one-sided versions of the sequential probability ratio test and its modifications. They are proven to be first-order asymptotically optimal for testing simple and parametric composite hypotheses when error probabilities are small. These tests exhibit near optimality properties not only in classical i.i.d. observation models but also in general non-i.i.d. models, provided that the log-likelihood ratios between hypotheses converge r-completely to positive and finite numbers. These findings extend the seminal work of Lai (1981) on two hypotheses.

AMS 2000 subject classifications: Primary 62L10, 62L15; secondary 62G40.
Keywords and phrases: MULTI-DECISION SEQUENTIAL RULES, MULTI-HYPOTHESIS SEQUENTIAL PROBABILITY RATIO TEST, ASYMPTOTIC OPTIMALITY, COMPLETE CONVERGENCE.

1. Introduction

Let $X_1, X_2, \dots$ be independent and identically distributed (i.i.d.) random variables available for observation and $\{\mathsf{P}_\theta, \theta \in \Theta\}$ be a family of distributions with densities $p_\theta(x)$ with respect to some non-degenerate, sigma-finite measure, so the joint density of the vector $\mathbf{X}_1^n = (X_1, \dots, X_n)$ is $p_\theta(\mathbf{X}_1^n) = \prod_{t=1}^n p_\theta(X_t)$. For testing two simple hypotheses $\mathsf{H}_0 : \theta = \theta_0$ and $\mathsf{H}_1 : \theta = \theta_1$, Wald [33, 34] proposed a sequential probability ratio test (SPRT) that is based on comparing the likelihood ratio between these hypotheses, $\Lambda_{\theta_1,\theta_0}(n) = \prod_{t=1}^n [p_{\theta_1}(X_t)/p_{\theta_0}(X_t)]$, with two thresholds. Wald and Wolfowitz [35] proved that the SPRT has a remarkably strong optimality property – it minimizes the expected sample size under both hypotheses in the class of all tests with error probabilities of Type I and Type II upper-bounded by the given numbers. However, the SPRT may perform poorly for parameter values different from the putative values θ_0 and θ_1.

To address this issue, for testing composite hypotheses $\mathsf{H}_0 : \theta \in \Theta_0$ versus $\mathsf{H}_1 : \theta \in \Theta_1$, Wald [34] suggested the mixture-likelihood-ratio approach for modifying the SPRT. Let $\pi_0(\theta)$ and $\pi_1(\theta)$ be two "mixing" probability densities (more generally weights) and let

$$\Lambda^\pi(n) = \frac{\int_{\Theta_1} \prod_{k=1}^n p_\theta(X_k)\pi_1(\theta)\,\mathrm{d}\theta}{\int_{\Theta_0} \prod_{k=1}^n p_\theta(X_k)\pi_0(\theta)\,\mathrm{d}\theta}$$

be the mixture likelihood ratio. Then replacing the likelihood ratio $\Lambda_{\theta_1,\theta_0}(n)$ used in the SPRT by this mixture likelihood ratio $\Lambda^\pi(n)$ leads to the mixture SPRT. Applying Wald's likelihood ratio identity, one can easily obtain the upper bounds on the average error probabilities:

$$\int_{\Theta_0} \mathsf{P}_\theta(\text{accept } \mathsf{H}_1)\pi_0(\theta)\,\mathrm{d}\theta \quad \text{and} \quad \int_{\Theta_1} \mathsf{P}_\theta(\text{accept } \mathsf{H}_0)\pi_1(\theta)\,\mathrm{d}\theta.$$

However, for practical purposes, it is preferable to bound not only the average error probabilities but also the maximum error probabilities, represented by $\sup_{\theta\in\Theta_0} \mathsf{P}_\theta(\text{accept } \mathsf{H}_1)$ and $\sup_{\theta\in\Theta_1} \mathsf{P}_\theta(\text{accept } \mathsf{H}_0)$. Unfortunately, how to obtain the upper bounds on these maximal error probabilities for the mixture SPRT is generally unclear.

An alternative approach is the generalized likelihood ratio (GLR) method from the classical Neyman–Pearson fixed-sample-size theory. This approach replaces the likelihood ratio $\Lambda_{\theta_1,\theta_0}(n)$ used in the SPRT by the GLR statistics

$$(1) \qquad \widehat{\Lambda}_i(n) = \frac{\sup_{\theta\in\Theta} \prod_{t=1}^n p_\theta(X_t)}{\sup_{\theta\in\Theta_i} \prod_{t=1}^n p_\theta(X_t)} = \frac{\prod_{t=1}^n p_{\hat\theta_n}(X_t)}{\sup_{\theta\in\Theta_i} \prod_{t=1}^n p_\theta(X_t)}, \quad i = 0, 1,$$

where $\hat\theta_n = \arg\sup_{\theta\in\Theta} \prod_{t=1}^n p_\theta(X_t)$ is the maximum likelihood estimator of θ. Obtaining upper bounds for maximal error probabilities is also a challenge since the GLR statistics are not viable likelihood ratios anymore and as a result, Wald's likelihood ratio identity cannot be used for this purpose. The GLR method has been developed in numerous publications, encompassing both Bayesian and frequentist settings, particularly when the cost of observations or the error probabilities are small. See, e.g, Schwarz [23], Wong [37], Lorden [11, 12, 15], Lai [7], Lai and Zhang [8], Chan and Lai [1].

The other way is to employ a combination of mixture-based and GLR approaches exploiting the statistics

$$(2) \qquad \widehat{\Lambda}_i^\pi(n) = \frac{\int_\Theta \prod_{t=1}^n p_\theta(X_t)\pi(\theta)\,\mathrm{d}\theta}{\sup_{\theta\in\Theta_i} \prod_{t=1}^n p_\theta(X_t)}, \quad i = 0, 1,$$

where $\Theta = \Theta_0 \cup \Theta_1$. This modification of the SPRT has the advantage that upper bounds on the maximal error probabilities $\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(\text{accept } \mathsf{H}_1)$ and $\sup_{\theta \in \Theta_1} \mathsf{P}_\theta(\text{accept } \mathsf{H}_0)$ can be obtained in the same way as for the SPRT (see Lemma 5.1 in Section 5.2 for the more general multi-hypothesis and non-i.i.d. case).

Note that the GLR method is adaptive. Yet another adaptive approach is to replace the "global" maximum likelihood estimator in the GLR statistic (1) by one-step delayed estimators (at each step), that is, instead of $\widehat{\Lambda}_i(n)$ to use the adaptive likelihood ratio statistics

$$\widehat{\Lambda}_i^*(n) = \frac{\prod_{t=1}^n p_{\hat{\theta}_{t-1}}(X_t)}{\sup_{\theta \in \Theta_i} \prod_{t=1}^n p_\theta(X_t)}, \quad i = 0, 1.$$

In this case, the Wald likelihood ratio identity can still be applied to upper-bound the error probabilities of the corresponding adaptive SPRT. The adaptive SPRT is therefore a very attractive alternative to the GLR-SPRT and the mixture-based SPRT. The idea of this test goes back to the works by Robbins and Siegmund [19, 20] who were the first who suggested a one-sided adaptive test in the context of power 1 tests.

In the present paper, we develop the asymptotic theory of multi-hypothesis sequential testing for general stochastic models not restricted to the i.i.d. assumption. Specifically, we consider models with dependent and non-identically distributed data of a very general structure with minimal assumptions related to the r-complete convergence of normalized log-likelihood ratio processes to certain positive numbers. The approach modifies and extends the ideas in Lai's seminal work [6] for two simple hypotheses, and Tartakovsky's contributions [25] for multiple simple hypotheses.

2. Problem formulation

We are interested in the following general discrete-time multi-hypothesis model with parametric composite hypotheses. Let $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathsf{P}_\theta)$, $n \geqslant 1$ be a filtered probability space with standard assumptions about monotonicity of the σ-algebras $\mathscr{F}_n$. The parameter $\theta = (\theta_1, \dots, \theta_l)$ belongs to a subset Θ of the l-dimensional Euclidean space $\mathbb{R}^l$. Random variables $X_1, X_2, \dots$ are observed sequentially taking values in a measurable space $(\Omega, \mathscr{F})$. The sub-σ-algebra $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$ of $\mathscr{F}$ is generated by the sequence $\{X_t\}_{t \geqslant 1}$ observed up to time n. The hypotheses to be tested are $\mathsf{H}_i : \ \theta \in \Theta_i$, $i = 0, 1, \dots, N$ $(N \geqslant 1)$, where Θ_i are disjoint subsets of Θ. We also suppose that there is an indifference zone $\Theta_{\text{in}} \subset \Theta$ in which there are no restrictions on the error probabilities. It is assumed that the subsets Θ_{in} and Θ_i

($i = 0, 1, \dots, N$) are disjoint. The indifference zone, where any decision is acceptable, is usually introduced because the correct action is not critical and often not even possible when the hypotheses are too close, which is perhaps the case in most practical applications. However, if needed $\Theta_{\rm in}$ may be an empty set. The probability measures P_θ and P_ϑ are assumed to be locally mutually absolutely continuous, i.e., the restrictions P_θ^n and P_ϑ^n of these measures to the sub-σ-algebras $\mathscr{F}_n$ are equivalent for all $1 \leqslant n < \infty$ and all distinct values $\theta, \vartheta \in \Theta$.

Write $\mathbf{X}_1^n = (X_1, \dots, X_n)$ for the sample of size n. It is convenient to represent the general probabilistic model in terms of densities as

$$p_{\theta,n}(\mathbf{X}_1^n) = \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1}), \quad \theta \in \Theta, \tag{3}$$

where $p_{\theta,n}(\mathbf{X}_1^n)$ is the joint density of the sample $\mathbf{X}_1^n$ (i.e., density of P_θ^n with respect to some sigma-finite measure) and $f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})$ is the corresponding conditional density.

For the sake of brevity in what follows we will often write $\mathscr{N}_0$ for the set $\{0, 1, \dots, N\}$.

A multi-hypothesis sequential test $\mathsf{D} = (T, d)$, where T is a stopping time with respect to the filtration $\{\mathscr{F}_n\}_{n \geqslant 1}$, and $d = d_T(\mathbf{X}_1^T) \in \{0, 1, \dots, N\}$ is an $\mathscr{F}_T$-measurable terminal decision rule specifying which hypothesis is to be accepted once observations have stopped. The hypothesis H_i is accepted if $d = i$ and rejected if $d \neq i$, i.e., $\{d = i\} = \{T < \infty, \ \text{accepts } H_i\}$, $i \in \mathscr{N}_0$.

The quality of a sequential test is judged based on its error probabilities and expected sample sizes or more generally on the moments of the sample size. Let E_θ denote expectation under P_θ, $\theta \in \Theta$ and let

$$\alpha_{ij}(\mathsf{D}, \theta) = \mathsf{E}_\theta\left[\mathbb{1}_{\{d=j\}}\right] = \mathsf{P}_\theta(d = j), \quad \theta \in \Theta_i, \quad i \neq j \quad (i, j = 0, 1, \dots, N)$$

denote the probability that the test D accepts the hypothesis H_j when the true value of the parameter θ is fixed within the subset Θ_i. Hereafter $\mathbb{1}_{\{\mathcal{A}\}}$ denotes an indicator of a set $\mathcal{A}$. By

$$\mathbb{C}(\boldsymbol{\alpha}) = \left\{\mathsf{D} : \sup_{\theta \in \Theta_i} \alpha_{ij}(\mathsf{D}, \theta) \leqslant \alpha_{ij} \text{ for } i, j = 0, 1, \dots, N, \, i \neq j\right\} \tag{4}$$

denote the class of tests with the maximal error probabilities that do not exceed the predefined values $\alpha_{ij} \in (0, 1)$, where $\boldsymbol{\alpha} = (\alpha_{ij})_{i,j \in \mathscr{N}_0, i \neq j}$ is the matrix of given constraints.

Note that the class $\mathbb{C}(\boldsymbol{\alpha})$ confines the error probabilities in the regions Θ_i but not in the indifference zone $\Theta_{\rm in}$ where the hypotheses are too close to be distinguished with the given and relatively low error probabilities. However, ideally, we would like to minimize the expected sample size $\mathsf{E}_\theta[T]$ for all possible parameter values, including those in the indifference zone. Unfortunately, there is no such test, since the structure of the test that minimizes the expected sample size $\mathsf{E}_\theta[T]$ at a specific parameter value $\theta = \tilde{\theta}$ depends on $\tilde{\theta}$. However, this problem may be solved asymptotically when the error probabilities are small. That is, we are interested in finding multi-hypothesis tests $\mathsf{D} = (T, d)$ that minimize the expected sample size $\mathsf{E}_\theta[T]$ uniformly for all $\theta \in \Theta$ in the class of tests $\mathbb{C}(\boldsymbol{\alpha})$ as α_{ij} approach zero. More generally, we are interested in finding the tests that minimize asymptotically the moments of the stopping time distribution up to some order $r \geqslant 1$:

$$(5) \qquad \inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \quad \text{as } \alpha_{\max} \to 0 \quad \text{uniformly in} \quad \theta \in \Theta,$$

where $\alpha_{\max} = \max_{i,j\in\mathscr{N}_0, i\neq j} \alpha_{ij}$ and $\Theta = \sum_{i=0}^N \Theta_i + \Theta_{\rm in}$. Specifically, we need to construct a sequential test $\mathsf{D}_* = (T_*, d_*)$ that is first-order asymptotically optimal under certain quite general conditions, i.e.,

$$(6) \qquad \frac{\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r]}{\mathsf{E}_\theta[T_*^r]} = 1 + o(1) \quad \text{as } \alpha_{\max} \to 0 \quad \text{for all } \theta \in \Theta.$$

Hereafter $o(1) \to 0$.

We will also consider the case of simple hypotheses where the parameter θ takes $N+1$ values $\theta_0, \theta_1, \dots, \theta_N$ or, more generally, the distributions under hypotheses H_i have joint distinct densities

$$(7) \qquad p(\mathbf{X}_1^n|\mathsf{H}_i) = p_{i,n}(\mathbf{X}_1^n) = \prod_{t=1}^n f_{i,t}(X_t|\mathbf{X}_1^{t-1}), \quad i = 0, 1, \dots, N,$$

where $f_{i,t}(X_t|\mathbf{X}_1^{t-1})$ are the corresponding conditional densities. We need to construct a sequential test $\mathsf{D}_* = (T_*, d_*)$ such that

$$(8) \qquad \frac{\inf_{\mathsf{D}\in\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})} \mathsf{E}_i[T^r]}{\mathsf{E}_i[T_*^r]} = 1 + o(1) \quad \text{as } \alpha_{\max} \to 0 \quad \text{for all } i = 0, 1, \dots, N,$$

where E_i is expectation under measure P_i and

$$(9) \qquad \mathbb{C}_{\rm sim}(\boldsymbol{\alpha}) = \left\{\mathsf{D} : \mathsf{P}_i(d = j) \leqslant \alpha_{ij} \text{ for } i, j = 0, 1, \dots, N,\, i \neq j\right\}.$$

Remark 2.1. In what follows, for practical purposes, we will suppose that the ratios of logarithms of error probabilities $\log \alpha_{ij}/\log \alpha_{ks}$ are bounded away from zero and infinity, i.e., the error probabilities approach zero in such a way that for all i, j

$$\frac{|\log \alpha_{ij}|}{|\log \alpha_{\max}|} \sim c_{ij} \quad \text{as } \alpha_{\max} \to 0, \quad 0 < c_{ij} \leqslant 1. \tag{10}$$

This assumption guarantees that any error probability α_{ij} does not go to zero at an exponentially faster (or slower) rate than any other α_{ks}. However, we do not require (as it often happens in the literature) that the α_{ij}'s go to zero at the same rate, in which case all c_{ij} are equal to 1. In other words, we consider an asymptotically asymmetric case rather than an asymptotically symmetric one when $c_{ij} = 1$. The reason is that there are many important problems for which the error probabilities may be orders of magnitude different. A typical example is the problem of target detection when the probabilities of a false alarm are required to be substantially smaller than the probabilities of target missing (miss detection).

3. Asymptotic lower bounds for performance metrics

For $n \geqslant 1$ and $\theta, \vartheta \in \Theta$ define the likelihood ratio (LR) and the log-likelihood ratio (LLR) for the sample $\mathbf{X}_1^n$ between the distinct points θ and ϑ

$$\begin{aligned} \Lambda_{\theta,\vartheta}(n) &= \frac{p_{\theta,n}(\mathbf{X}_1^n)}{p_{\vartheta,n}(\mathbf{X}_1^n)} = \prod_{t=1}^{n} \frac{f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}{f_{\vartheta,t}(X_t|\mathbf{X}_1^{t-1})}, \\ \lambda_{\theta,\vartheta}(n) &= \log \Lambda_{\theta,\vartheta}(n) = \sum_{t=1}^{n} \log \left[\frac{f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}{f_{\vartheta,t}(X_t|\mathbf{X}_1^{t-1})} \right]. \end{aligned} \tag{11}$$

Below we show that the lower bounds in class $\mathbb{C}(\boldsymbol{\alpha})$ for the performance metrics – the moments of the stopping time $\mathsf{E}_\theta[T^r]$ – can be established as long as the LLR $\lambda_{\theta,\vartheta}(n)$ obeys the strong law of large numbers (SLLN) with a certain rate $\psi(n)$.

Throughout the paper, we assume that $\psi(t)$ is an increasing one-to-one $\mathbb{R}_+ \to \mathbb{R}_+$ function. By Ψ we denote its inverse ψ^{-1} and we assume that both ψ and Ψ go to infinity as $t \to \infty$, $\lim_{t\to\infty} \psi(t) = \lim_{t\to\infty} \Psi(t) = \infty$. In addition, we assume the following conditions on the inverse function

$$\lim_{\delta\to 1} \lim_{t\to\infty} \frac{\Psi(\delta\, t)}{\Psi(t)} = 1 \tag{12}$$

and

$$\lim_{t\to\infty} t^{-1} \log \Psi(t) = 0. \tag{13}$$

Note that conditions (12)-(13) are satisfied for the power function $\psi(t) = t^\beta$ with $\beta > 0$, but not for the logarithmic function $\psi(t) = \log t$, i.e., the function $\psi(t)$ has to increase not too slowly – faster than the logarithmic function, $\psi(t)/\log t \to \infty$ as $t \to \infty$.

To obtain asymptotic lower bounds we impose the following condition.

C1. *Right-tail Condition.* Assume that there are a positive increasing function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and a positive continuous function $I(\theta, \vartheta)$ with

$$\begin{aligned} &\min_{j\in\mathscr{N}_0\setminus i} \inf_{\vartheta\in\Theta_j} I(\theta,\vartheta) > 0 \quad \text{for all } \theta\in\Theta_i \text{ and } i\in\mathscr{N}_0, \\ &\min_{i\in\mathscr{N}_0} \inf_{\vartheta\in\Theta_i} I(\theta,\vartheta) > 0 \quad \text{for all } \theta\in\Theta_{\mathrm{in}}, \end{aligned} \tag{14}$$

such that for any $\varepsilon > 0$ and all $\theta, \vartheta \in \Theta$, $\theta \neq \vartheta$

$$\lim_{L\to\infty} \mathsf{P}_\theta \left\{ \frac{1}{\psi(L)} \max_{1\leqslant n\leqslant L} \lambda_{\theta,\vartheta}(n) > (1+\varepsilon) I(\theta,\vartheta) \right\} = 0. \tag{15}$$

Remark 3.1 below shows that $I(\theta, \vartheta)$ can be interpreted as an information "distance" between the parameter points θ and ϑ and $\inf_{\vartheta\in\Theta_i} I(\theta,\vartheta) = I(\theta,\Theta_i)$ as a "distance" between the parameter value θ and the subset Θ_i. Thus, conditions (14) represent separability restrictions between subsets Θ_i, Θ_j and Θ_{in} ($i \neq j$). It is intuitively obvious that if the corresponding distances are zero, then the hypotheses become indistinguishable.

Additional notation: $I_i(\theta) = I(\theta, \Theta_i) = \inf_{\vartheta\in\Theta_i} I(\theta,\vartheta)$,

$$F_{i,\theta}(\varepsilon, \boldsymbol{\alpha}) = \Psi\left((1-\varepsilon) \max_{j\in\mathscr{N}_0\setminus i} \frac{|\log \alpha_{ji}|}{I_j(\theta)} \right), \tag{16}$$

so that

$$F_{i,\theta}(0, \boldsymbol{\alpha}) = \Psi\left(\max_{j\in\mathscr{N}_0\setminus i} \frac{|\log \alpha_{ji}|}{I_j(\theta)} \right). \tag{17}$$

In the following, we will omit 0 in (17) and will write $F_{i,\theta}(\boldsymbol{\alpha})$ for $F_{i,\theta}(0, \boldsymbol{\alpha})$.

The following theorem establishes lower bounds for arbitrary moments of the stopping time distribution in the class of sequential or non-sequential multi-hypothesis tests $\mathbb{C}(\boldsymbol{\alpha})$ defined in (4). In the sequel, we will often write θ_i for θ when θ belongs to the subset Θ_i.

Theorem 3.1. *If there exist an increasing positive function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and a positive function $I(\theta, \vartheta)$, satisfying (14), such that right-tail conditions (15) hold, then for every $0 < \varepsilon < 1$*

$$\lim_{\alpha_{\max}\to 0} \inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{T > F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})\right\} = 1 \quad \text{for all } \theta \in \Theta_i \text{ and } i \in \mathscr{N}_0 \tag{18}$$

and for every $0 < \varepsilon < 1$

$$\lim_{\alpha_{\max}\to 0} \inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{T > \min_{0\leqslant i\leqslant N} F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})\right\} = 1 \quad \text{for all } \theta \in \Theta_{\text{in}}. \tag{19}$$

Therefore, for all $r \geqslant 1$ as $\alpha_{\max} \to 0$

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \geqslant [F_{i,\theta}(\boldsymbol{\alpha})]^r \,(1 + o(1)) \quad \text{for all } \theta \in \Theta_i \text{ and } i \in \mathscr{N}_0; \tag{20}$$

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \geqslant \left[\min_{i\in\mathscr{N}_0} F_{i,\theta}(\boldsymbol{\alpha})\right]^r (1 + o(1)) \quad \text{for all} \quad \theta \in \Theta_{\text{in}}. \tag{21}$$

Proof. Let $\mathsf{D} = (T, d)$ be an arbitrary test from class $\mathbb{C}(\boldsymbol{\alpha})$. It suffices to consider tests that terminate almost surely, $\mathsf{P}_\theta(T < \infty) = 1$, since otherwise $\mathsf{E}_\theta[T^r] = \infty$ and the statement follows trivially. Changing the measure $\mathsf{P}_\vartheta \to \mathsf{P}_\theta$ and using Wald's likelihood ratio identity, we obtain that for any $s \geqslant 1$, $C > 0$, and any two distinct points θ and ϑ

$$\begin{aligned}
\mathsf{P}_\vartheta(d = i) &= \mathsf{E}_\theta\big\{\mathbb{1}_{\{d=i\}} \Lambda_{\theta,\vartheta}(T)^{-1}\big\} \\
&\geqslant \mathsf{E}_\theta\big\{\mathbb{1}_{\{d=i, T\leqslant s, \Lambda_{\theta,\vartheta}(T)<e^C\}} \Lambda_{\theta,\vartheta}(T)^{-1}\big\} \\
&\geqslant e^{-C} \mathsf{P}_\theta\Big(d = i, T \leqslant s, \max_{1\leqslant n\leqslant s} \Lambda_{\theta,\vartheta}(n) < e^C\Big) \\
&\geqslant e^{-C}\Big\{\mathsf{P}_\theta(d = i, T \leqslant s) - \mathsf{P}_\theta\Big(\max_{1\leqslant n\leqslant s} \lambda_{\theta,\vartheta}(n) \geqslant C\Big)\Big\}.
\end{aligned} \tag{22}$$

The last inequality follows from the Boole inequality $\mathsf{P}(\mathcal{A} \cap \mathcal{B}) \geqslant \mathsf{P}(\mathcal{A}) - \mathsf{P}(\mathcal{B}^c)$, where $\mathcal{B}^c$ is a complement of the event $\mathcal{B}$, if we set $\mathcal{A} = \{d = i, T \leqslant s\}$ and $\mathcal{B} = \{\max_{1\leqslant n\leqslant s} \lambda_{\theta,\vartheta}(n) < C\}$. It follows that

$$\mathsf{P}_\theta(d = i, T \leqslant s) \leqslant \mathsf{P}_\vartheta(d = i)\, e^C + \mathsf{P}_\theta\left\{\max_{1\leqslant n\leqslant s} \lambda_{\theta,\vartheta}(n) \geqslant C\right\}, \tag{23}$$

and since, by Boole's inequality,

$$\mathsf{P}_\theta(d=i, T \leqslant s) \geqslant \mathsf{P}_\theta(T \leqslant s) - \mathsf{P}_\theta(d \neq i),$$

we obtain

$$\mathsf{P}_\theta(T \leqslant s) \leqslant \mathsf{P}_\theta(d \neq i) + \mathsf{P}_\vartheta(d=i)\, e^C + \mathsf{P}_\theta\left\{\max_{1\leqslant n\leqslant s} \lambda_{\theta,\vartheta}(n) \geqslant C\right\}. \tag{24}$$

Let $\theta = \theta_i \in \Theta_i$ and $\vartheta \notin \Theta_i$ and set

$$\begin{aligned} s &= s_{ij}(\varepsilon, \theta_i, \alpha_{ji}) = \Psi\left((1-\varepsilon)\frac{|\log \alpha_{ji}|}{I(\theta_i,\vartheta)}\right), \\ C &= C_{ij}(\varepsilon, \alpha_{ji}) = (1+\varepsilon) I(\theta_i,\vartheta)\psi(s_{ij}) = (1-\varepsilon^2)|\log \alpha_{ji}|. \end{aligned}$$

Using (24) we obtain that for all $j \in \mathscr{N}_0 \setminus i$, $\theta_i \in \Theta_i$ and $i = 0, 1, \dots, N$

$$\begin{aligned} \mathsf{P}_{\theta_i}\left\{T \leqslant s_{ij}(\varepsilon,\theta_i,\alpha_{ji})\right\} &\leqslant \sum_{k\in\mathscr{N}_0\setminus i} \alpha_{ik} + \alpha_{ji}^{\varepsilon^2} \\ &+ \mathsf{P}_{\theta_i}\left\{\max_{1\leqslant n\leqslant s_{ij}} \lambda_{\theta_i,\vartheta}(n) \geqslant (1+\varepsilon) I(\theta_i,\vartheta)\psi(s_{ij})\right\}. \end{aligned}$$

Since the right-hand side does not depend on D, we have

$$\begin{aligned} \inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_{\theta_i}\left\{T > s_{ij}(\varepsilon,\theta_i,\alpha_{ji})\right\} &\geqslant 1 - \sum_{k\in\mathscr{N}_0\setminus i} \alpha_{ik} - \alpha_{ji}^{\varepsilon^2} \\ &- \mathsf{P}_{\theta_i}\left\{\frac{1}{\psi(s_{ij})}\max_{1\leqslant n\leqslant s_{ij}} \lambda_{\theta_i,\vartheta}(n) \geqslant (1+\varepsilon) I(\theta_i,\vartheta)\right\}. \end{aligned}$$

The second and third terms on the right-hand side in the above inequality go to 0 as $\alpha_{\max} \to 0$. The fourth term also goes to 0 for all $0 < \varepsilon < 1$ by condition (15). Hence, for all $\theta_i \in \Theta_i$, $\vartheta \notin \Theta_i$ and all $j \in \mathscr{N}_0 \setminus i$

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_{\theta_i}\left\{T > \Psi\left((1-\varepsilon)\frac{|\log \alpha_{ji}|}{I(\theta_i,\vartheta)}\right)\right\} \xrightarrow[\alpha_{\max}\to 0]{} 1,$$

which implies (18).

Next, we prove (19) for the indifference zone, $\theta \in \Theta_{\text{in}}$. For any $\vartheta = \theta_j \in \Theta_j$, let

$$K_\theta(\boldsymbol{\alpha}) = \min_{i\in\mathscr{N}_0} \max_{j\in\mathscr{N}_0\setminus i} \frac{|\log \alpha_{ji}|}{I(\theta,\theta_j)}.$$

By (22), for every $\theta \in \Theta_{\rm in}$ and $\theta_j \in \Theta_j$,

$$\mathsf{P}_\theta(d=i, T \leqslant s) \leqslant \mathsf{P}_{\theta_j}(d=i) e^C + \mathsf{P}_\theta\Big\{\max_{1\leqslant n\leqslant s} \lambda_{\theta,\theta_j}(n) \geqslant C\Big\}.$$

Setting $s = s_\theta(\varepsilon, \boldsymbol{\alpha}) = \Psi((1-\varepsilon)K_\theta(\boldsymbol{\alpha}))$ and

$$C = (1+\varepsilon) I(\theta,\theta_j)\psi(s_\theta(\varepsilon,\boldsymbol{\alpha})) = (1-\varepsilon^2) I(\theta,\theta_j) K_\theta(\boldsymbol{\alpha}),$$

we obtain that for all $j \in \mathscr{N}_0 \setminus i$ and $i = 0, 1, \dots, N$

$$\begin{aligned} \mathsf{P}_\theta\left\{d=i, T \leqslant s_\theta(\varepsilon,\boldsymbol{\alpha})\right\} &\leqslant \alpha_{ji}^{\varepsilon^2} \\ &+ \mathsf{P}_\theta\left\{\max_{1\leqslant n\leqslant s_\theta(\varepsilon,\boldsymbol{\alpha})} \lambda_{\theta,\theta_j}(n) \geqslant (1+\varepsilon) I(\theta,\theta_j)\psi(s_\theta(\varepsilon,\boldsymbol{\alpha}))\right\}, \end{aligned}$$

so that for all $i = 0, 1, \dots, N$

$$\mathsf{P}_\theta\left\{d=i, T \leqslant s_\theta(\varepsilon,\boldsymbol{\alpha})\right\} \leqslant \beta_i^{\varepsilon^2} + \gamma_i(\theta,\boldsymbol{\alpha},\varepsilon),$$

where $\beta_i = \max_{j\in\mathscr{N}_0\setminus i} \alpha_{ji}$ and

$$\gamma_i(\theta,\boldsymbol{\alpha},\varepsilon) = \max_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\frac{1}{\psi(s_\theta(\varepsilon,\boldsymbol{\alpha}))} \max_{1\leqslant n\leqslant s_\theta(\varepsilon,\boldsymbol{\alpha})} \lambda_{\theta,\theta_j}(n) \geqslant (1+\varepsilon) I(\theta,\theta_j)\right\}.$$

Consequently,

$$\mathsf{P}_\theta\left\{T \leqslant s_\theta(\varepsilon,\boldsymbol{\alpha})\right\} \leqslant \sum_{i=0}^N \left[\beta_i^{\varepsilon^2} + \gamma_i(\theta,\boldsymbol{\alpha},\varepsilon)\right],$$

and since the right-hand side does not depend on any test D, we obtain the inequality

$$\sup_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta\left\{T \leqslant \Psi\left((1-\varepsilon)\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i} \frac{|\log\alpha_{ji}|}{I(\theta,\theta_j)}\right)\right\} \leqslant \sum_{i=0}^N \left[\beta_i^{\varepsilon^2} + \gamma_i(\theta,\boldsymbol{\alpha},\varepsilon))\right],$$

where by condition (15) $\gamma_i(\theta,\boldsymbol{\alpha},\varepsilon) \to 0$ as $\alpha_{\max} \to 0$ for all $\theta \in \Theta_{\rm in}$ and $\varepsilon \in (0,1)$. It follows that for every $0 < \varepsilon < 1$ and $\theta \in \Theta_{\rm in}$,

$$\sup_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta\left\{T \leqslant \Psi\left((1-\varepsilon)\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i} \frac{|\log\alpha_{ji}|}{I_j(\theta)}\right)\right\} \to 0 \quad \text{as} \quad \alpha_{\max} \to 0,$$

which completes the proof of (19).

The asymptotic lower bounds (20) and (21) now follow from Chebyshev's inequality. Indeed, by the Chebyshev inequality, for any $r \geqslant 1$ and all $\theta \in \Theta_i$ and $i = 0, 1, \dots, N$,

$$\inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta \left[\left(\frac{T}{F_{i,\theta}(\boldsymbol{\alpha})} \right)^r \right] \geqslant \left(\frac{F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})}{F_{i,\theta}(\boldsymbol{\alpha})} \right)^r \inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{ T > F_{i,\theta}(\varepsilon, \boldsymbol{\alpha}) \right\}.$$

Since ε can be arbitrarily small and noting that, by condition (12),

$$\lim_{\varepsilon \to 0} \lim_{\alpha_{\max} \to 0} \frac{F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})}{F_{i,\theta}(\boldsymbol{\alpha})} = 1 \tag{25}$$

and that by (18)

$$\inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{ T > F_{i,\theta}(\varepsilon, \boldsymbol{\alpha}) \right\} \xrightarrow[\alpha_{\max} \to 0]{} 1,$$

taking the limit $\varepsilon \to 0$, we obtain that for all $r \geqslant 1$, $\theta \in \Theta_i$ and $i = 0, 1, \dots, N$,

$$\liminf_{\alpha_{\max} \to 0} \inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta \left[\left(\frac{T}{F_{i,\theta}(\boldsymbol{\alpha})} \right)^r \right] \geqslant 1,$$

which completes the proof of assertion (20).

Analogously, define

$$\tau_\theta(\varepsilon, \boldsymbol{\alpha}) = \frac{T}{\min_{i \in \mathscr{N}_0} F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})}.$$

By the Chebyshev inequality, for any $r \geqslant 1$ and all $\theta \in \Theta_{\mathrm{in}}$

$$\inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta [\tau_\theta(0, \boldsymbol{\alpha})^r] \geqslant \left(\frac{\min_{i \in \mathscr{N}_0} F_{i,\theta}(\varepsilon, \boldsymbol{\alpha})}{\min_{i \in \mathscr{N}_0} F_{i,\theta}(\boldsymbol{\alpha})} \right)^r \inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{ \tau_\theta(\varepsilon, \boldsymbol{\alpha}) > 1 \right\},$$

where by (19)

$$\inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta \left\{ \tau_\theta(\varepsilon, \boldsymbol{\alpha}) > 1 \right\} \xrightarrow[\alpha_{\max} \to 0]{} 1 \quad \text{for every } 0 < \varepsilon < 1.$$

So taking the limit $\varepsilon \to 0$ and accounting for (25), we obtain that for all $r \geqslant 1$ and $\theta \in \Theta_{\mathrm{in}}$

$$\liminf_{\alpha_{\max} \to 0} \inf_{\mathsf{D} \in \mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta [\tau_\theta(\varepsilon, \boldsymbol{\alpha})^r] \geqslant 1,$$

which completes the proof of the lower bound (21). $\square$

Remark 3.1. By Lemma 1 in [27], the right-tail condition (15) is satisfied whenever

$$\frac{\lambda_{\theta,\vartheta}(n)}{\psi(n)} \xrightarrow[n\to\infty]{\mathsf{P}_\theta-\text{a.s.}} I(\theta,\vartheta) \quad \text{for all} \quad \theta,\vartheta\in\Theta, \theta\neq\vartheta. \tag{26}$$

Therefore, assertions of Theorem 3.1 hold under the strong law for the LLR (26), which is a natural and more convenient condition. Furthermore, as the proof of Lemma 1 in [27] shows, for the right-tail condition (15) to hold it suffices to assume that

$$\mathsf{P}_\theta\left(\limsup_{n\to\infty}\frac{\lambda_{\theta,\vartheta}(n)}{\psi(n)} > I(\theta,\vartheta)\right) = 0.$$

Consider now the case of simple hypotheses when the parameter θ takes $N+1$ distinct values $\theta_0,\theta_1,\dots,\theta_N$ or, more generally, the distributions P_i under hypotheses H_i, $i=0,1,\dots,N$ have joint distinct densities defined in (7). Obviously, in this case, the indifference zone is empty $\Theta_{\text{in}}=\varnothing$ and the subsets $\Theta_i=\{i\}$, $i=0,1,\dots,N$. Also, the LR and LLR defined in (11) become the LR and LLR processes between the hypotheses H_i and H_j for the sample $\mathbf{X}_1^n$

$$\begin{aligned}\Lambda_{ij}(n) &= \frac{p_{i,n}(\mathbf{X}_1^n)}{p_{j,n}(\mathbf{X}_1^n)} = \prod_{t=1}^n \frac{f_{i,t}(X_t|\mathbf{X}_1^{t-1})}{f_{j,t}(X_t|\mathbf{X}_1^{t-1})},\\ \lambda_{ij}(n) &= \log\Lambda_{ij}(n) = \sum_{t=1}^n \log\left[\frac{f_{i,t}(X_t|\mathbf{X}_1^{t-1})}{f_{j,t}(X_t|\mathbf{X}_1^{t-1})}\right].\end{aligned} \tag{27}$$

Obviously, in this case, Theorem 3.1 implies the following corollary, whose proof can also be established directly using proofs of Lemma 2.1 and Theorem 2.2 in Tartakovsky [25]. Recall that class $\mathbb{C}_{\text{sim}}(\boldsymbol{\alpha})$ is defined in (9).

Corollary 3.1. *If there exist an increasing function $\psi(t)$, $\psi(\infty)=\infty$, satisfying condition (12), and positive and finite numbers I_{ij}, $i,j=0,1,\dots,N$, $i\neq j$ such that for all $\varepsilon>0$ and all $i,j=0,1,\dots,N$ $(i\neq j)$*

$$\lim_{L\to\infty}\mathsf{P}_i\left\{\frac{1}{\psi(L)}\max_{1\leqslant n\leqslant L}\lambda_{ij}(n)\geqslant(1+\varepsilon)I_{ij}\right\}=0, \tag{28}$$

then for all $i=0,1,\dots,N$ and every $0<\varepsilon<1$

$$\lim_{\alpha_{\max}\to 0}\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})}\mathsf{P}_i\left\{T>\Psi\left((1-\varepsilon)\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_{ij}}\right)\right\}=1, \tag{29}$$

and therefore, for all $r \geqslant 1$ *and* $i = 0, 1, \dots, N$

$$\inf_{\mathsf{D} \in \mathbb{C}_{\rm sim}(\boldsymbol{\alpha})} \mathsf{E}_i[T^r] \geqslant \left[\Psi\left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{|\log \alpha_{ji}|}{I_{ij}}\right)\right]^r (1 + o(1)) \quad as \quad \alpha_{\max} \to 0. \tag{30}$$

Remark 3.2. The right-tail condition (28) is satisfied whenever

$$\mathsf{P}_i\left(\limsup_{n \to \infty} \frac{\lambda_{ij}(n)}{\psi(n)} \leqslant I_{ij}\right) = 1,$$

and therefore, if normalized LLRs $\lambda_{ij}(n)/\psi(n)$ converge P_i-a.s. to I_{ij} as $n \to \infty$.

4. Asymptotic optimality of likelihood ratio based sequential multi-decision rules for simple hypotheses

We begin with the case of $N + 1$ simple hypotheses $\mathsf{H}_i : \mathsf{P} = \mathsf{P}_i$, $i = 0, 1, \dots, N$ with joint densities (7).

4.1. The matrix sequential probability ratio test

Recall that $\mathscr{N}_0 = \{0, 1, \dots, N\}$. For a threshold matrix $(A_{ij})_{i,j \in \mathscr{N}_0}$, with $A_{ij} > 0$ and the A_{ii} are immaterial (0, say), define the Matrix SPRT (MSPRT) $\mathsf{D}_* = (T_*, d_*)$, built on $(N+1)N$ one-sided SPRTs between the hypotheses H_i and H_j, as follows: Stop at the first $n \geqslant 1$ such that, for some $i \in \mathscr{N}_0$, $\Lambda_{ij}(n) \geqslant A_{ji}$ for all $j \in \mathscr{N}_0 \setminus i$ and accept the unique H_i that satisfies these inequalities. Obviously, with $a_{ji} = \log A_{ji}$, introducing the Markov accepting times for the hypotheses H_i, $i \in \mathscr{N}_0$ as

$$\begin{aligned} T_i &= \inf\left\{n \geqslant 1 : \lambda_{ij}(n) \geqslant a_{ji} \text{ for all } j \in \mathscr{N}_0 \setminus i\right\}, \\ &= \inf\left\{n \geqslant 1 : \min_{j \in \mathscr{N}_0 \setminus i} [\lambda_{ij}(n) - a_{ji}] \geqslant 0\right\}, \quad i = 0, 1, \dots, N, \end{aligned} \tag{31}$$

the MSPRT $\mathsf{D}_* = (T_*, d_*)$ can be written as

$$T_* = \min_{k \in \mathscr{N}_0} T_k, \qquad d_* = i \quad \text{if} \quad T_* = T_i. \tag{32}$$

Thus, in the MSPRT, each component SPRT is extended until, for some $i \in \mathscr{N}_0$, all N SPRTs involving H_i accept H_i. Note that for $N = 1$ the MSPRT coincides with Wald's SPRT.

Using Wald's likelihood ratio identity, it can be easily shown that the error probabilities of the MSPRT $\alpha_{ij}(\mathsf{D}_*) = \mathsf{P}_i(d_* = j)$ satisfy the inequalities

$$\alpha_{ij}(\mathsf{D}_*) \leqslant \exp(-a_{ij}) \quad \text{for all} \quad i,j = 0,1,\dots,N, \; i \neq j, \tag{33}$$

so selecting $a_{ij} = |\log \alpha_{ij}|$ implies $\mathsf{D}_* \in \mathbb{C}(\boldsymbol{\alpha})$. See, e.g., Lemma 4.1.1 (page 192) in Tartakovsky *et al.* [28].

In his ingenious paper, Lorden [14] showed that with a special design that includes accurate estimation of thresholds accounting for overshoots, the MSPRT is nearly optimal in the third-order sense – it minimizes expected sample sizes for all hypotheses up to an additive disappearing term, i.e., $\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_i[T] = \mathsf{E}_i[T_*] + o(1)$ as $\alpha_{\max} \to 0$. This result holds only for i.i.d. models with the finite second moment $\mathsf{E}_i[\lambda_{ij}(1)^2] < \infty$. In the non-i.i.d. case, it is practically impossible to obtain such a result, so we will focus on the first-order optimality (8).

To establish the asymptotic optimality property of the MSPRT we use the ideas of the groundbreaking paper of Lai [6] where he proved first-order asymptotic optimality of Wald's SPRT in the general non-i.i.d. case.

By the SLLN in the i.i.d. case, the LLR $\lambda_{ij}(n)$ has the following stability property

$$n^{-1}\lambda_{ij}(n) \xrightarrow[n\to\infty]{\mathsf{P}_i-\text{a.s.}} I_{ij}, \quad i,j = 0,1,\dots,N, \; i \neq j, \tag{34}$$

where

$$I_{ij} = \mathsf{E}_i[\lambda_{ij}(1)] = \int \log\left[\frac{f_i(x)}{f_j(x)}\right] f_i(x)\mathrm{d}\mu(x)$$

is the Kullback-Leibler (K-L) information number, which characterizes the distance between the hypotheses H_i and H_j. This allows one to conjecture that if in the general non-i.i.d. case the LLR is also stable in the sense that the almost sure convergence conditions (34) are satisfied with some positive and finite numbers I_{ij}, then the MSPRT is approximately optimal. In the general case, these numbers represent the local K-L information in the sense that often (while not always) $I_{ij} = \lim_{n\to\infty} n^{-1}\mathsf{E}_i[\lambda_{ij}(n)]$.

Having said that, in what follows, we will assume that the normalized LLRs $\lambda_{ij}(n)/\psi(n)$ converge almost surely to finite and positive numbers I_{ij} under P_i:

$$\frac{\lambda_{ij}(n)}{\psi(n)} \xrightarrow[n\to\infty]{\mathsf{P}_i-\text{a.s.}} I_{ij}, \quad i,j = 0,1,\dots,N, \; i \neq j, \tag{35}$$

where $\psi(t)$ is an increasing function ($\psi(\infty) = \infty$) as in Section 3.

A standard approach for proving asymptotic optimality is to show that the lower bounds (30) in Corollary 3.1 are attained for the MSPRT under certain conditions.

4.2. Asymptotic optimality of the MSPRT

While the almost sure convergence (35) for the LLRs guarantees lower bounds (30), in the general non-i.i.d. case, this condition is not sufficient for asymptotic optimality of the MSPRT since it does not even guarantee the finiteness of the moments $\mathsf{E}_i[T_*^r]$ of the MSPRT's stopping time. To establish the optimality of the MSPRT a strengthening is needed, such as a certain convergence rate in the strong law.

Lai [6] proved the asymptotic optimality of Wald's SPRT for testing two hypotheses under the r-quick version of the SLLN for the LLR $\lambda(n)/n$, i.e., for the models with dependent and asymptotically stationary observations. Tartakovsky [24] generalized Lai's result to the case of multiple hypotheses with asymptotically non-stationary observations proving that the MSPRT is asymptotically optimal as long as LLRs $\lambda_{ij}(n)/\psi(n)$ converge r-quickly to finite numbers I_{ij}. Below we relax the r-quick convergence condition used in [6, 24] by the r-complete convergence.

Definition 4.1 (r-***Complete Convergence***). We say that the sequence of random variables $\{Y_n\}_{n\geqslant 1}$ converges to a random variable Y *r-completely* as $n \to \infty$ under probability measure P and write $Y_n \xrightarrow[n\to\infty]{\mathsf{P}\text{-}r\text{-completely}} Y$ if

$$\lim_{n\to\infty} \sum_{t=n}^{\infty} t^{r-1} \mathsf{P}(|Y_t - Y| > \varepsilon) = 0 \quad \text{for every } \varepsilon > 0,$$

which is equivalent to

$$\sum_{n=1}^{\infty} n^{r-1} \mathsf{P}(|Y_n - Y| > \varepsilon) < \infty \quad \text{for every } \varepsilon > 0.$$

As we will see below, a sufficient condition for asymptotic optimality of the MSPRT is the r-complete convergence of $\lambda_{ij}(n)/\psi(n)$ to numbers I_{ij}, $0 < I_{ij} < \infty$ as $n \to \infty$ under P_i, i.e., that for all $i, j = 0, 1, \dots, N$ $(i \neq j)$

$$\lim_{n\to\infty} \sum_{t=n}^{\infty} t^{r-1} \mathsf{P}_i\left(\left|\frac{\lambda_{ij}(t)}{\psi(t)} - I_{ij}\right| > \varepsilon\right) = 0 \quad \text{for every } \varepsilon > 0. \tag{36}$$

The following theorem provides the asymptotic upper bounds for moments of the sample sizes of the MSPRT which along with the lower bounds (30) allow us to conclude that the MSPRT minimizes moments of the sample sizes up to order r in class $\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})$.

Theorem 4.1. *Assume that there exist an increasing function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and positive and finite numbers I_{ij}, $i, j = 0, 1, \dots, N$, $i \neq j$ such that for all $i, j = 0, 1, \dots, N$ $(i \neq j)$ and some $r \geqslant 1$*

$$\lim_{n\to\infty} \sum_{t=n}^{\infty} t^{r-1} \mathsf{P}_i \left(\frac{\lambda_{ij}(t)}{\psi(t)} - I_{ij} < -\varepsilon \right) = 0 \quad \text{for every } \varepsilon > 0. \tag{37}$$

If the thresholds in the MSPRT are so selected that $\alpha_{ij}(\mathsf{D}_*) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log \alpha_{ji}^{-1}$ as $\alpha_{\max} \to 0$, in particular as $a_{ji} = \log \alpha_{ji}^{-1}$, then for all $i = 0, 1, \dots, N$

$$\mathsf{E}_i[T_*^r] \leqslant \left[\Psi \left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{|\log \alpha_{ji}|}{I_{ij}} \right) \right]^r (1 + o(1)) \quad \text{as} \quad \alpha_{\max} \to 0. \tag{38}$$

Proof. For $i, j = 0, 1, \dots, N$ and $0 < \varepsilon < \max_{i,j \in \mathscr{N}_0} I_{ij}$, define $\mathbf{a} = (a_{ij})_{i,j \in \mathscr{N}_0}$ and

$$M_i(\mathbf{a}, \varepsilon) = 1 + \Psi \left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{a_{ji}}{I_{ij} - \varepsilon} \right).$$

Noting that $\mathsf{E}_i[T_*^r] \leqslant \mathsf{E}_i[T_i^r]$ and setting $\tau = T_i$ and $N = M_i(\mathbf{a}, \varepsilon)$ in Lemma A.1 (Appendix A, page 239) in Tartakovsky [26] we obtain the inequalities

$$\mathsf{E}_i[T_*^r] \leqslant \mathsf{E}_i[T_i^r] \leqslant M_i^r + r 2^{r-1} \sum_{n=M_i}^{\infty} n^{r-1} \mathsf{P}_i \left(T_i > n \right), \quad i = 0, 1, \dots, N. \tag{39}$$

By the definition of the Markov time T_i (see (31)), we have

$$\begin{aligned} \mathsf{P}_i(T_i > n) &= \mathsf{P}_i \left\{ \max_{1 \leqslant t \leqslant n} \min_{j \in \mathscr{N}_0 \setminus i} [\lambda_{ij}(t) - a_{ji}] < 0 \right\} \\ &\leqslant \sum_{j \in \mathscr{N}_0 \setminus i} \mathsf{P}_i \left\{ \lambda_{ij}(n) < a_{ji} \right\}, \end{aligned}$$

where for $n \geqslant M_i(\mathbf{a}, \varepsilon)$ the probability $\mathsf{P}_i \left\{ \lambda_{ij}(n) < a_{ji} \right\}$ does not exceed the probability $\mathsf{P}_i \left\{ \lambda_{ij}(n)/\psi(n) < I_{ij} - \varepsilon \right\}$, and therefore, for all sufficiently large n

$$\mathsf{P}_i(T_i > n) \leqslant \sum_{j \in \mathscr{N}_0 \setminus i} \mathsf{P}_i \left\{ \frac{\lambda_{ij}(n)}{\psi(n)} < I_{ij} - \varepsilon \right\}. \tag{40}$$

Substituting (40) into (39) yields

$$\mathsf{E}_i[T_*^r] \leqslant M_i(\mathbf{a}, \varepsilon)^r + r 2^{r-1} \sum_{j \in \mathscr{N}_0 \setminus i} \sum_{n = M_i(\mathbf{a}, \varepsilon)}^{\infty} n^{r-1} \mathsf{P}_i \left\{ \frac{\lambda_{ij}(n)}{\psi(n)} < I_{ij} - \varepsilon \right\}.$$

Since $M_i(\mathbf{a}, \varepsilon) \to \infty$ as $a_{\min} \to \infty$, by condition (37), the second term goes to 0 and, hence, for all $i \in \mathscr{N}_0$

$$\mathsf{E}_i[T_*^r] \leqslant M_i(\mathbf{a}, \varepsilon)^r + o(1) \quad \text{as } a_{\min} \to \infty,$$

where ε can be arbitrarily small, so taking the limit $\varepsilon \to 0$ and noticing that due to condition (12)

$$\lim_{\varepsilon \to 0} \lim_{a_{\min} \to 0} [M_i(\mathbf{a}, \varepsilon) / M_i(\mathbf{a}, 0)] = 1,$$

we obtain the asymptotic inequality

$$\mathsf{E}_i[T_*^r] \leqslant \left[\Psi \left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{a_{ji}}{I_{ij}} \right) \right]^r (1 + o(1)) \quad \text{as } a_{\min} \to \infty. \tag{41}$$

Setting $a_{ji} = |\log \alpha_{ji}|$ or, more generally, $a_{ji} \sim |\log \alpha_{ji}|$ (assuming $\alpha_{ij}(\mathsf{D}_*) \leqslant \alpha_{ij}$), gives the required inequality (38). □

Corollary 3.1 and Theorem 4.1 give the following first-order asymptotic optimality result.

Theorem 4.2. *Assume that for some $0 < I_{ij} < \infty$, $i, j = 0, 1, \dots, N$, $i \neq j$ and increasing function $\psi(n)$, $\psi(\infty) = \infty$, satisfying condition (12), the normalized LLRs $\lambda_{ij}(n)/\psi(n)$ converge r-completely to I_{ij} under P_i as $n \to \infty$, i.e., for all $i, j = 0, 1, \dots, N$ ($i \neq j$) and some $r \geqslant 1$ condition (36) holds. If the thresholds in the MSPRT are so selected that $\alpha_{ij}(\mathsf{D}_*) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log \alpha_{ji}^{-1}$ as $\alpha_{\max} \to 0$, in particular as $a_{ji} = \log \alpha_{ji}^{-1}$, then for all $i = 0, 1, \dots, N$*

$$\mathsf{E}_i[T_*^r] \sim \left[\Psi \left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{|\log \alpha_{ji}|}{I_{ij}} \right) \right]^r \sim \inf_{\mathsf{D} \in \mathbb{C}_{\mathrm{sim}}(\boldsymbol{\alpha})} \mathsf{E}_i[T^r] \quad \text{as } \alpha_{\max} \to 0. \tag{42}$$

Proof. Obviously, the r-complete convergence condition implies both conditions (28) in Corollary 3.1 and (37) in Theorem 4.1. Hence, using (30) in Corollary 3.1, we obtain

$$\mathsf{E}_i[T_*^r] \geqslant \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_{ij}}\right)\right]^r(1+o(1)) \quad \text{as} \quad \alpha_{\max}\to 0.$$

This inequality along with the reverse inequality (38) gives the asymptotic approximation

$$\mathsf{E}_i[T_*^r] = \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_{ij}}\right)\right]^r(1+o(1)) \quad \text{as} \quad \alpha_{\max}\to 0,$$

which, by the lower bound (30), is the best one can do in class $\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})$. Thus, both asymptotic equalities in (42) follow and the proof is complete. ☐

Remark 4.1. Inequalities (33) for error probabilities imply that the MSPRT belongs to $\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})$ with $\alpha_{ij}=\exp\{-a_{ij}\}$, so Corollary 3.1 implies asymptotic lower bounds

$$\mathsf{E}_i[T_*^r] \geqslant \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_{ij}}\right)\right]^r(1+o(1)) \quad \text{as} \quad a_{\min}\to \infty,$$

while the upper bounds (41) yield reverse inequalities. Therefore, the following approximations for moments of the sample sizes of the MSPRT as functions of thresholds $\mathbf{a}=(a_{ij})$ hold regardless of the probabilities of errors

$$\mathsf{E}_i[T_*^r] = \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_{ij}}\right)\right]^r(1+o(1)) \quad \text{as} \quad a_{\min}\to \infty.$$

These approximations can be useful in different problem settings, e.g., in Bayes problems.

5. Asymptotic optimality of likelihood ratio based sequential multi-decision rules for composite hypotheses

Consider now the problem of testing $N+1$ composite hypotheses $\mathsf{H}_i:\mathsf{P}=\mathsf{P}_\theta$, $\theta\in\Theta_i$, $i=0,1,\dots,N$ with joint densities (3). In this case, we focus on the class of tests $\mathbb{C}(\boldsymbol{\alpha})$ defined in (4) that confines the maximal error probabilities $\sup_{\theta\in\Theta_i}\mathsf{P}_\theta(d=j)$.

Recall that we suppose that the parameter space Θ is split into $N+2$ disjoint subsets $\Theta_0,\Theta_1,\dots,\Theta_N$, $\Theta_{\rm in}$, so $\Theta=\sum_{i=0}^N\Theta_i+\Theta_{\rm in}$, where $\Theta_{\rm in}$ is the indifference zone.

5.1. The matrix mixture sequential probability ratio test

Let π be a mixing measure (prior distribution for θ) on Θ, $\pi(\theta) > 0$ for all $\theta \in \Theta$ and $\int_\Theta \pi(\mathrm{d}\theta) = 1$.[1] For $i \in \mathscr{N}_0$ and $n \geqslant 1$, define

$$g_n(\mathbf{X}_1^n) = \int_\Theta p_{\theta,n}(\mathbf{X}_1^n)\pi(\mathrm{d}\theta) = \int_\Theta \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})\pi(\mathrm{d}\theta);$$

$$\hat{g}_{i,n}(\mathbf{X}_1^n) = \sup_{\theta\in\Theta_j} p_{\theta,n}(\mathbf{X}_1^n) = \sup_{\theta\in\Theta_i} \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})$$

and introduce the statistics

$$\Lambda_i^\pi(n) = \frac{\int_\Theta \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})\pi(\mathrm{d}\theta)}{\sup_{\theta\in\Theta_i}\prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})} = \frac{g_n(\mathbf{X}_1^n)}{\hat{g}_{i,n}(\mathbf{X}_1^n)}, \tag{43}$$

$$\lambda_i^\pi(n) = \log\Lambda_i^\pi(n) = \log g_n(\mathbf{X}_1^n) - \log \hat{g}_{i,n}(\mathbf{X}_1^n). \tag{44}$$

For a $(N+1)\times(N+1)$ matrix $(A_{ij})_{i,j\in\mathscr{N}_0}$ of thresholds, with $A_{ij} > 0$ and the A_{ii} are immaterial, define the Matrix Mixture SPRT (MMSPRT) $\mathsf{D}_*^\pi = (T_*^\pi, d_*^\pi)$ as follows: Stop at the first $n \geqslant 1$ such that, for some $i \in \mathscr{N}_0$, $\Lambda_j^\pi(n) \geqslant A_{ji}$ for all $j \neq i$ and accept the unique H_i that satisfies these inequalities. Setting $a_{ji} = \log A_{ji}$ and introducing the Markov accepting times for the hypotheses H_i, $i = 0, 1, \dots, N$ as

$$\begin{aligned} T_i^\pi &= \inf\left\{n \geqslant 1 : \lambda_j^\pi(n) \geqslant a_{ji} \text{ for all } j \in \mathscr{N}_0 \setminus i\right\} \\ &= \inf\left\{n \geqslant 1 : \min_{j\in\mathscr{N}_0\setminus i}\left[\lambda_j^\pi(n) - a_{ji}\right] \geqslant 0\right\}, \quad i = 0, 1, \dots, N, \end{aligned} \tag{45}$$

the MMSPRT can be written as

$$T_*^\pi = \min_{k\in\mathscr{N}_0} T_k^\pi, \qquad d_*^\pi = i \quad \text{if} \quad T_*^\pi = T_i^\pi. \tag{46}$$

5.2. Error Probabilities of the MMSPRT

The following lemma provides upper bounds for the error probabilities $\alpha_{ij}(\mathsf{D}_*^\pi, \theta) = \mathsf{P}_\theta(d_*^\pi = j)$, $\theta \in \Theta_i$ of the MMSPRT D_*^π.

[1]The results also hold for improper measures with minor constraints if Θ is a compact set.

Lemma 5.1. *The following upper bounds on the error probabilities of the MMSPRT hold:*

$$\sup_{\theta\in\Theta_i} \alpha_{ij}(\mathsf{D}_*^\pi,\theta) \leqslant \exp\{-a_{ij}\} \quad \text{for } i,j=0,1,\dots,N,\ i\neq j. \tag{47}$$

Therefore, if $a_{ij}=\log(1/\alpha_{ij})$ *then* $\mathsf{D}_*^\pi \in \mathbb{C}(\boldsymbol{\alpha})$.

Proof. Observe that $\{d_*^\pi = j\} = \{T_*^\pi = T_j^\pi\}$ implies $\{T_j^\pi < \infty\}$ and $\Lambda_i^\pi(T_j^\pi) \geqslant A_{ij} = e^{a_{ij}}$ on $\{T_j^\pi < \infty\}$. So, for all $\theta\in\Theta_i$ $(i\neq j)$, we have the following chain of equalities and inequalities

$$\begin{aligned}\alpha_{ij}(\mathsf{D}_*^\pi,\theta) &= \mathsf{E}_\theta\left[\mathbb{1}_{\{d_*^\pi=j\}}\right] \leqslant \mathsf{E}_\theta\left[\mathbb{1}_{\{T_j^\pi<\infty\}}\right] \\ &= \mathsf{E}_\theta\left[\mathbb{1}_{\{T_j^\pi<\infty\}}\Lambda_i^\pi(T_j^\pi)/\Lambda_i^\pi(T_j^\pi)\right] \\ &\leqslant \exp\{-a_{ij}\}\,\mathsf{E}_\theta\left[\mathbb{1}_{\{T_j^\pi<\infty\}}\Lambda_i^\pi(T_j^\pi)\right].\end{aligned}$$

Since, obviously, for any $n\geqslant 1$ and all $\theta\in\Theta_i$

$$\Lambda_i^\pi(n) \leqslant \int_\Theta \Lambda_{\vartheta,\theta}(n)\pi(\mathrm{d}\vartheta)$$

and, by the Wald likelihood ratio identity,

$$\mathsf{E}_\theta\left[\mathbb{1}_{\{T_j^\pi<\infty\}}\Lambda_{\vartheta,\theta}(T_j^\pi)\right] = \mathsf{P}_\vartheta\left(T_j^\pi<\infty\right),$$

we obtain that for all $i\in\mathscr{N}_0\setminus j$

$$\begin{aligned}\sup_{\theta\in\Theta_i}\alpha_{ij}(\mathsf{D}_*^\pi,\theta) &\leqslant \exp\{-a_{ij}\}\,\sup_{\theta\in\Theta_i}\int_\Theta \mathsf{E}_\theta\left[\mathbb{1}_{\{T_j^\pi<\infty\}}\Lambda_{\vartheta,\theta}(T_j^\pi)\right]\pi(\mathrm{d}\vartheta) \\ &= \exp\{-a_{ij}\}\int_\Theta \mathsf{P}_\vartheta\left(T_j^\pi<\infty\right)\pi(\mathrm{d}\vartheta) \leqslant \exp\{-a_{ij}\},\end{aligned}$$

which proves the upper bound (47) and gives the Lemma. □

5.3. First-order uniform asymptotic optimality of the MMSPRT

For general non-i.i.d. models, the SLLN (26) for the LLR does not guarantee the optimality of the MMSPRT. As for simple hypotheses, a sort of r-complete convergence suffices for the asymptotic optimality. The following r-complete convergence-type conditions for left-tail probabilities are sufficient. For the convenience sake, we write θ_i for θ when θ belongs to Θ_i.

For $\theta \notin \Theta_j$, define

$$\widetilde{\lambda}_{\theta,j}(n) = \log\left[\frac{\prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}{\sup_{\theta\in\Theta_j}\prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}\right] = \log\left[\frac{p_{\theta,n}(\mathbf{X}_1^n)}{\sup_{\theta\in\Theta_j} p_{\theta,n}(\mathbf{X}_1^n)}\right],$$

and for $\theta \in \Theta$,

$$\Upsilon_{\theta,j,\delta,\varepsilon,r}(n) = \sum_{t=n}^{\infty} t^{r-1}\mathsf{P}_\theta\left\{\inf_{\vartheta\in\Gamma_{\delta,\theta}}\widetilde{\lambda}_{\vartheta,j}(t) < (I_j(\theta)-\varepsilon)\psi(t)\right\},$$

where $I_j(\theta) = \inf_{\vartheta\in\Theta_j} I(\theta,\vartheta)$ is the “distance” between the point θ and the subset Θ_j, which is assumed strictly positive for all $\theta \notin \Theta_j$ (see (14)), and $\Gamma_{\delta,\theta} = \{\vartheta\in\Theta : |\vartheta-\theta| < \delta\}$ is the δ-neighborhood of the point θ.

C2. *Left-tail Condition.* There exist a positive continuous function $I(\theta,\vartheta)$, satisfying condition (14), such that for any $\varepsilon > 0$ and some $r \geqslant 1$

$$\lim_{n\to\infty}\limsup_{\delta\to 0}\max_{\substack{i,j\in\mathscr{N}_0\\ j\neq i}}\sup_{\theta_i\in\Theta_i}\Upsilon_{\theta_i,j,\delta,\varepsilon,r}(n) = 0 \tag{48}$$

and

$$\lim_{n\to\infty}\limsup_{\delta\to 0}\max_{j\in\mathscr{N}_0}\sup_{\theta\in\Theta_{\rm in}}\Upsilon_{\theta,j,\delta,\varepsilon,r}(n) = 0. \tag{49}$$

Recall that $F_{i,\theta}(\varepsilon,\boldsymbol{\alpha})$ and $F_{i,\theta}(\boldsymbol{\alpha})$ are defined in (16) and (17), respectively.

The following theorem establishes the asymptotic upper bounds for moments of the MMSPRT sample sizes which together with the lower bounds (20) and (21) imply asymptotic optimality of the MMSPRT in class $\mathbb{C}(\boldsymbol{\alpha})$.

Theorem 5.1. *Assume that there exist an increasing function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and a function $I(\theta,\vartheta)$, satisfying (14), such that for some $r \geqslant 1$ left-tail condition* **C**2 *holds. If the thresholds in the MMSPRT are so selected that $\sup_{\theta_i\in\Theta_i}\alpha_{ij}(\mathsf{D}_*^\pi,\theta_i) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log\alpha_{ji}^{-1}$ as $\alpha_{\max}\to 0$, in particular as $a_{ji} = \log\alpha_{ji}^{-1}$, then as $\alpha_{\max}\to 0$*

$$\mathsf{E}_\theta[(T_*^\pi)^r] \leqslant [F_{i,\theta}(\boldsymbol{\alpha})]^r\,(1+o(1)) \quad \textit{for all } \theta\in\Theta_i \textit{ and } i\in\mathscr{N}_0; \tag{50}$$

$$\mathsf{E}_\theta[(T_*^\pi)^r] \leqslant \left[\min_{0\leqslant i\leqslant N} F_{i,\theta}(\boldsymbol{\alpha})\right]^r (1+o(1)) \quad \textit{for all } \theta\in\Theta_{\rm in}. \tag{51}$$

Proof. By the definition of the Markov time T_i^π, we have

$$\begin{aligned}
\mathsf{P}_\theta(T_i^\pi > n) &= \mathsf{P}_\theta\left\{\max_{1\leqslant t\leqslant n}\min_{j\in\mathscr{N}_0\setminus i}\left[\lambda_j^\pi(t) - a_{ji}\right] < 0\right\} \\
&\leqslant \sum_{j\in\mathscr{N}_0\setminus i}\mathsf{P}_\theta\left\{\lambda_j^\pi(n) < a_{ji}\right\} \\
&= \sum_{j\in\mathscr{N}_0\setminus i}\mathsf{P}_\theta\left\{\log g_n(\mathbf{X}_1^n) - \log \hat{g}_{j,n}(\mathbf{X}_1^n) < a_{ji}\right\}.
\end{aligned}$$

Since

$$\log g_n(\mathbf{X}_1^n) \geqslant \log \int_{\Gamma_{\delta,\theta}} p_{\vartheta,n}(\mathbf{X}_1^n)\pi(\mathrm{d}\vartheta) \geqslant \inf_{\vartheta\in\Gamma_{\delta,\theta}} \log p_{\vartheta,n}(\mathbf{X}_1^n) + \log \pi(\Gamma_{\delta,\theta})$$

it follows that

$$\mathsf{P}_\theta(T_i^\pi > n) \leqslant \sum_{j\in\mathscr{N}_0\setminus i}\mathsf{P}_\theta\left\{\frac{\inf_{\vartheta\in\Gamma_{\delta,\theta}}\widetilde{\lambda}_{\vartheta,j}(n)}{\psi(n)} < \frac{a_{ji} - \log\pi(\Gamma_{\delta,\theta})}{\psi(n)}\right\}. \tag{52}$$

Let $\mathbf{a} = (a_{ij})$ denote the matrix of finite thresholds a_{ij}, $i, j = 0, 1, \dots, N$ (a_{ii} are immaterial) of the MMSPRT and let

$$M_{i,\theta}(\mathbf{a},\varepsilon) = 1 + \Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_j(\theta) - \varepsilon}\right) \quad \text{for } \theta\in\Theta. \tag{53}$$

It is easily seen that for $n \geqslant M_{i,\theta}(\mathbf{a},\varepsilon)$ the right-hand side in (52) does not exceed the sum of probabilities

$$\sum_{j\in\mathscr{N}_0\setminus i}\mathsf{P}_\theta\left\{\frac{1}{\psi(n)}\inf_{\vartheta\in\Gamma_{\delta,\theta}}\widetilde{\lambda}_{\vartheta,j}(n) < I_j(\theta) - \varepsilon + \frac{1}{\psi(n)}|\log\pi(\Gamma_{\delta,\theta})|\right\}.$$

Therefore, for all sufficiently large n and $a_{\min}$, for which $|\log\pi(\Gamma_{\delta,\theta})|/\psi(n) < \varepsilon/2$, we obtain

$$\mathsf{P}_\theta(T_i^\pi > n) \leqslant \sum_{j\in\mathscr{N}_0\setminus i}\mathsf{P}_\theta\left\{\frac{1}{\psi(n)}\inf_{\vartheta\in\Gamma_{\delta,\theta}}\widetilde{\lambda}_{\vartheta,j}(n) < I_j(\theta) - \varepsilon/2\right\}. \tag{54}$$

Noting that by the definition of the stopping time T_*^π in (46) $\mathsf{E}_\theta[(T_*^\pi)^r] \leqslant \mathsf{E}_\theta[(T_i^\pi)^r]$ for all $i \in \mathscr{N}_0$ and setting $\tau = T_i^\pi$ and $N = M_{i,\theta}(\mathbf{a},\varepsilon/2)$ in

Lemma A.1 (Appendix A, page 239) in Tartakovsky [26], we obtain the inequalities

$$\text{(55)} \qquad \mathsf{E}_\theta[(T_*^\pi)^r] \leqslant \mathsf{E}_\theta[(T_i^\pi)^r] \leqslant M_{i,\theta}^r + r2^{r-1} \sum_{n=M_{i,\theta}}^{\infty} n^{r-1} \mathsf{P}_\theta\left(T_i^\pi > n\right),$$

which hold for all $i = 0, 1, \ldots, N$.

Substituting (54) in (55) gives

$$\begin{aligned} \mathsf{E}_\theta[(T_*^\pi)^r] &\leqslant \mathsf{E}_\theta[(T_i^\pi)^r] \leqslant M_{i,\theta}(\mathbf{a}, \varepsilon/2)^r \\ &\quad + r2^{r-1} N \max_{j \in \mathscr{N}_0 \setminus i} \sum_{n=M_{i,\theta}(\mathbf{a},\varepsilon/2)}^{\infty} n^{r-1} \mathsf{P}_\theta\left\{\inf_{\vartheta \in \Gamma_{\delta,\theta}} \widetilde{\lambda}_{\vartheta,j}(n) < (I_j(\theta) - \varepsilon/2)\psi(n)\right\} \\ \text{(56)} \qquad &= M_{i,\theta}(\mathbf{a}, \varepsilon/2)^r + r2^{r-1} N \max_{j \in \mathscr{N}_0 \setminus i} \Upsilon_{\theta,j,\delta,\varepsilon,r}(M_{i,\theta}(\mathbf{a}, \varepsilon/2)). \end{aligned}$$

Consider the case where $\theta = \theta_i \in \Theta_i$ ($i = 0, 1, \ldots, N$). Inequality (56) implies that for all $i = 0, 1, \ldots, N$

$$\mathsf{E}_{\theta_i}[(T_*^\pi)^r] \leqslant M_{i,\theta_i}(\mathbf{a}, \varepsilon/2)^r + r2^{r-1} N \max_{j \in \mathscr{N}_0 \setminus i} \sup_{\theta_i \in \Theta_i} \Upsilon_{\theta_i,j,\delta,\varepsilon/2,r}(M_{i,\theta_i}(\mathbf{a}, \varepsilon/2)).$$

Since $M_{i,\theta_i}(\mathbf{a}, \varepsilon/2) \to \infty$ as $a_{\min} \to \infty$, by the left-tail condition (48), the second term goes to 0, so that for all $i = 0, 1, \ldots, N$

$$\mathsf{E}_{\theta_i}[(T_*^\pi)^r] \leqslant M_{i,\theta_i}(\mathbf{a}, \varepsilon/2)^r + o(1) \quad \text{as } a_{\min} \to \infty.$$

Since ε can be arbitrarily small, taking the limit $\varepsilon \to 0$ and noticing that due to condition (12)

$$\lim_{\varepsilon \to 0} \lim_{a_{\min} \to 0} [M_{i,\theta_i}(\mathbf{a}, \varepsilon/2)/M_{i,\theta_i}(\mathbf{a}, 0)] = 1,$$

we obtain the following asymptotic upper bound

$$\text{(57)} \qquad \mathsf{E}_{\theta_i}[(T_*^\pi)^r] \leqslant \left[\Psi\left(\max_{j \in \mathscr{N}_0 \setminus i} \frac{a_{ji}}{I_j(\theta_i)}\right)\right]^r (1 + o(1)) \quad \text{as } a_{\min} \to \infty,$$

which holds for all $\theta_i \in \Theta_i$ and all $i = 0, 1, \ldots, N$.

Setting $a_{ji} = |\log \alpha_{ji}|$ in (57) or, more generally, $a_{ji} \sim |\log \alpha_{ji}|$ (assuming $\sup_{\theta_i \in \Theta_i} \alpha_{ij}(\mathsf{D}_*^\pi, \theta_i) \leqslant \alpha_{ij}$), we obtain the inequalities (50) for all $\theta \in \Theta_i$ and all $i = 0, 1, \ldots, N$.

It remains to prove the inequality (51) for the indifference zone $\Theta_{\rm in}$. By inequality (56), for $\theta \in \Theta_{\rm in}$ and any $i \in \mathscr{N}_0$ we have the inequalities

$$\mathsf{E}_\theta[(T_i^\pi)^r] \leqslant M_{i,\theta}(\mathbf{a},\varepsilon/2)^r + r2^{r-1}N \max_{j\in\mathscr{N}_0\setminus i} \Upsilon_{\theta,j,\delta,\varepsilon/2,r}(M_{i,\theta}(\mathbf{a},\varepsilon/2)).$$

Now, note that $\mathsf{E}_\theta[(T_*^\pi)^r] \leqslant \min_{i\in\mathscr{N}_0} \mathsf{E}_\theta[T_i^r]$, and therefore,

$$\begin{aligned}\mathsf{E}_\theta[(T_*^\pi)^r] &\leqslant \min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a},\varepsilon/2)^r \\ &\qquad + r2^{r-1}N \min_{i\in\mathscr{N}_0} \max_{j\in\mathscr{N}_0\setminus i} \sup_{\theta\in\Theta_{\rm in}} \Upsilon_{\theta,j,\delta,\varepsilon/2,r}(M_{i,\theta}(\mathbf{a},\varepsilon/2)),\end{aligned}$$

where, by the left-tail condition (49), the second term goes to 0. Hence,

$$\mathsf{E}_\theta[(T_*^\pi)^r] \leqslant \min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a},\varepsilon/2)^r + o(1) \quad \text{as } a_{\min}\to\infty.$$

Since ε can be arbitrarily small, taking the limit $\varepsilon\to 0$ and noticing that due to condition (12)

$$\lim_{\varepsilon\to 0}\lim_{a_{\min}\to 0}\left[\min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a},\varepsilon/2)/\min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a},0)\right] = 1,$$

we obtain the asymptotic upper bound for $\theta\in\Theta_{\rm in}$

$$\mathsf{E}_\theta[T_*^r] \leqslant \left[\Psi\left(\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_j(\theta)}\right)\right]^r (1+o(1)) \quad \text{as } a_{\min}\to\infty. \tag{58}$$

Setting $a_{ji} = |\log\alpha_{ji}|$ in (58) or, more generally, $a_{ji}\sim|\log\alpha_{ji}|$ (assuming $\sup_{\theta_i\in\Theta_i}\alpha_{ij}(\mathsf{D}_*^\pi,\theta_i)\leqslant\alpha_{ij}$), we obtain the inequality (51) for $\theta\in\Theta_{\rm in}$. □

Theorems 3.1 and 5.1 give the following first-order asymptotic optimality result.

Theorem 5.2. *Assume that there exist an increasing function $\psi(t)$, $\psi(\infty)=\infty$, satisfying condition (12), and a function $I(\theta,\vartheta)$, satisfying (14), such that the SLLN for the LLR (26) holds, i.e., the normalized LLR $\lambda_{\theta,\vartheta}(n)/\psi(n)$ converges almost surely to $I(\theta,\vartheta)$ under P_θ as $n\to\infty$. Assume, in addition, that for some $r\geqslant 1$ left-tail condition* **C**2 *holds. If the thresholds in the MMSPRT are so selected that $\sup_{\theta\in\Theta_i}\alpha_{ij}(\mathsf{D}_*^\pi,\theta)\leqslant\alpha_{ij}$ and $a_{ji}\sim\log\alpha_{ji}^{-1}$ as $\alpha_{\max}\to 0$, in particular as $a_{ji}=\log\alpha_{ji}^{-1}$, then as $\alpha_{\max}\to 0$*

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})}\mathsf{E}_\theta[T^r]\sim[F_{i,\theta}(\boldsymbol{\alpha})]^r\sim\mathsf{E}_\theta[(T_*^\pi)^r] \quad \textit{for all } \theta\in\Theta_i \textit{ and } i\in\mathscr{N}_0; \tag{59}$$

$$(60) \qquad \inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \sim \left[\min_{0\leqslant i\leqslant N} F_{i,\theta}(\boldsymbol{\alpha})\right]^r \sim \mathsf{E}_\theta[(T_*^\pi)^r] \quad \textit{for all } \theta\in\Theta_{\mathrm{in}}.$$

Proof. Since the SLLN (26) implies the right-tail condition (15) we can use Theorem 3.1 to obtain the asymptotic lower bounds (as $\alpha_{\max}\to 0$)

$$\mathsf{E}_\theta[(T_*^\pi)^r] \geqslant \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_i, i\in\mathscr{N}_0;$$
$$\mathsf{E}_\theta[(T_*^\pi)^r] \geqslant \left[\Psi\left(\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_{\mathrm{in}}.$$

These inequalities along with the reverse inequalities (50) and (51) in Theorem 5.1 yield the asymptotic approximations

$$\mathsf{E}_\theta[(T_*^\pi)^r] = \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_i, i\in\mathscr{N}_0;$$
$$\mathsf{E}_\theta[(T_*^\pi)^r] = \left[\Psi\left(\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i}\frac{|\log\alpha_{ji}|}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_{\mathrm{in}},$$

so the MMSPRT attains the best lower bounds (20) and (21) in class $\mathbb{C}(\boldsymbol{\alpha})$. This completes the proof. ☐

Remark 5.1. It follows from inequalities (57) and (58) and the fact that the MMSPRT belongs to $\mathbb{C}(\boldsymbol{\alpha})$ with $\alpha_{ij}=\exp\{-a_{ij}\}$ that the following approximations (as $a_{\min}\to\infty$) for moments of the sample sizes of the MMSPRT as functions of thresholds $\mathbf{a}=(a_{ij})$ hold regardless of the probabilities of errors

$$\mathsf{E}_\theta[(T_*^\pi)^r] = \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_i,\ i\in\mathscr{N}_0;$$
$$\mathsf{E}_\theta[(T_*^\pi)^r] = \left[\Psi\left(\min_{i\in\mathscr{N}_0}\max_{j\in\mathscr{N}_0\setminus i}\frac{a_{ji}}{I_j(\theta)}\right)\right]^r(1+o(1)) \quad \text{for all } \theta\in\Theta_{\mathrm{in}}.$$

These approximations may be useful in various problem settings.

Remark 5.2. An alternative proof for Theorem 5.2 is to use r-complete convergence of the normalized to $\psi(n)$ decision statistics $\lambda_{ij}^\pi(n)$ defined in (44) to $I_j(\theta)=\inf_{\vartheta\in\Theta_j} I(\theta,\vartheta)$ as $n\to\infty$ under hypotheses H_i and to $\max_{j\in\mathscr{N}_0} I_j(\theta)$ for $\theta\in\Theta_{\mathrm{in}}$. That is, replacing condition **C**2 by the condition,

$$\lim_{n\to\infty}\sum_{t=n}^\infty n^{r-1}\mathsf{P}_\theta\left\{\left|\frac{1}{\psi(n)}\lambda_{ij}^\pi(n)-I_j(\theta)\right|>\varepsilon\right\} \quad \text{for } \theta\in\Theta_i \text{ and all } \varepsilon>0,$$

where $I_j(\theta)$ is replaced by $\max_{j\in\mathscr{N}_0} I_j(\theta)$ for $\theta \in \Theta_{\rm in}$. This latter "direct" sufficient condition for asymptotic optimality can be verified in certain interesting examples.

Remark 5.3. The proposed MMSPRT can be modified by replacing the statistics Λ_i^π defined in (43) with the statistics

$$\Lambda_{ji}^\pi(n) = \frac{\int_{\Theta_j} \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})\pi(\mathrm{d}\theta)}{\sup_{\theta\in\Theta_i} \prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}, \quad i,j=0,1,\dots,N, \; i\neq j.$$

This alternative version of the MMSPRT is also uniformly asymptotically optimal to first order for all hypotheses and parameter values $\theta_i \in \Theta_i$, $i=0,1,\dots,N$.

Remark 5.4. The above results are also satisfied for improper prior distributions as long as Θ is a compact set. This is important, for example, in invariant testing problems. See Section 6.3.

5.4. Adaptive matrix sequential probability ratio test

Let $\hat\theta_n = \hat\theta_n(\mathbf{X}_1^n)$ be an ($\mathscr{F}_n$-measurable) estimator of θ. If in conditional density $f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})$ for the t-th observation we replace the parameter with the estimator $\hat\theta_{t-1}(\mathbf{X}_1^{t-1})$ built upon the sample $\mathbf{X}_1^{t-1}$ of size $t-1$ that includes $t-1$ observations, then $f_{\hat\theta_{t-1},t}(X_t|\mathbf{X}_1^{t-1})$ and $\prod_{t=1}^n f_{\hat\theta_{t-1},t}(X_t|\mathbf{X}_1^{t-1})$ are still viable probability densities in contrast to the popular generalized likelihood ratio approach when maximization over θ is performed in the density $p_\theta(\mathbf{X}_1^n)$ for the whole sample $\mathbf{X}_1^n$ containing all n available observations, so that $p_{\hat\theta_n}(\mathbf{X}_1^n)$ is not a probability density anymore.[2]

For $n\geqslant 1$, introduce the adaptive LR statistic

$$\widehat\Lambda_\vartheta(n) = \prod_{t=1}^n \frac{f_{\hat\theta_{t-1},t}(X_t|\mathbf{X}_1^{t-1})}{f_{\vartheta,t}(X_t|\mathbf{X}_1^{t-1})}, \quad \widehat\Lambda_\vartheta(0)=1.$$

It satisfies the recursion

$$\widehat\Lambda_\vartheta(n) = \widehat\Lambda_\vartheta(n-1) \times \frac{f_{\hat\theta_{n-1},n}(X_n|\mathbf{X}^{n-1})}{f_{\vartheta,n}(X_n|\mathbf{X}^{n-1})}, \quad n\geqslant 1, \; \widehat\Lambda_\vartheta(0)=1 \tag{61}$$

with $f_{\hat\theta_0,1}(X_1|\mathbf{X}^0) = p_{\theta_0}(X_1)$, where the initial value of the estimator $\hat\theta_0=\theta_0$ is a design parameter. Since

$$\mathsf{E}_\vartheta[\widehat\Lambda_\vartheta(n)|\mathbf{X}^{n-1}] = \widehat\Lambda_\vartheta(n-1), \quad n\geqslant 2, \quad \mathsf{E}_\vartheta[\widehat\Lambda_\vartheta(1)]=1$$

[2]In the latter case, the $\hat\theta_n$ is the maximum likelihood estimator.

the adaptive LR $(\widehat{\Lambda}_\vartheta(n), \mathscr{F}_n)_{n\geq 1}$ is a P_ϑ-martingale with unit expectation. This important property allows us to deduce simple upper bounds for error probabilities of the adaptive multi-hypothesis sequential test, introduced below, using the Wald-Doob likelihood ratio identity, as we will see in the next subsection.

Introduce the statistics

$$\widehat{\Lambda}_i^*(n) = \frac{\prod_{t=1}^n f_{\hat{\theta}_{t-1},t}(X_t|\mathbf{X}_1^{t-1})}{\sup_{\theta\in\Theta_i}\prod_{t=1}^n f_{\theta,t}(X_t|\mathbf{X}_1^{t-1})}, \quad i=0,1,\dots,N, \tag{62}$$

and based on these statistics for a $(N+1)\times(N+1)$ matrix $(A_{ij})_{i,j\in\mathscr{N}_0}$ of boundaries, with $A_{ij}>0$ and the A_{ii} are immaterial, define the Adaptive Matrix SPRT (AMSPRT) $\widehat{\mathsf{D}} = (\widehat{T}, \hat{d})$ as follows: Stop at the first $n \geqslant 1$ such that, for some $i \in \mathscr{N}_0$, $\widehat{\Lambda}_j^*(n) \geqslant A_{ji}$ for all $j \neq i$ and accept the unique H_i that satisfies these inequalities. Let $\widehat{\lambda}_i^*(n) = \log \widehat{\Lambda}_i^*(n)$. Setting $a_{ji} = \log A_{ji}$ and introducing the Markov accepting times for the hypotheses H_i as

$$\begin{aligned}\widehat{T}_i &= \inf\left\{n\geqslant 1 : \widehat{\lambda}_i^*(n) \geqslant a_{ji} \text{ for all } j\in\mathscr{N}_0\setminus i\right\} \\ &= \inf\left\{n\geqslant 1 : \min_{j\in\mathscr{N}_0\setminus i}\left[\widehat{\lambda}_i^*(n) - a_{ji}\right]\geqslant 0\right\}, \quad i=0,1,\dots,N,\end{aligned} \tag{63}$$

the AMSPRT $\widehat{\mathsf{D}} = (\widehat{T}, \hat{d})$ can be written as

$$\widehat{T} = \min_{k\in\mathscr{N}_0} \widehat{T}_k, \qquad \hat{d} = i \quad \text{if} \quad \widehat{T} = \widehat{T}_i. \tag{64}$$

Because of the simple recursive structure of the adaptive LR (61), the AMSPRT is a very attractive alternative to the MMSPRT. Robbins and Siegmund [19, 20] were the first who suggested the idea of using the adaptive LR in the context of the so-called power 1 tests for i.i.d. models.

5.5. Probabilities of errors of the AMSPRT

The following lemma provides simple upper bounds for the error probabilities $\alpha_{ij}(\widehat{\mathsf{D}}, \theta) = \mathsf{P}_\theta(\hat{d} = j)$, $\theta\in\Theta_i$ of the AMSPRT (63)-(64).

Lemma 5.2. *The following upper bounds on the error probabilities of the AMSPRT hold:*

$$\sup_{\theta\in\Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \exp\{-a_{ij}\} \quad \textit{for } i,j=0,1,\dots,N,\ i\neq j.$$

Therefore, if $a_{ij} = \log(1/\alpha_{ij})$ *then* $\widehat{\mathsf{D}} \in \mathbb{C}(\boldsymbol{\alpha})$ *.*

Proof. Observe that the event $\{\hat{d} = j\} = \{\widehat{T} = \widehat{T}_j\}$ implies the event $\{\widehat{T}_j < \infty\}$ and, by the definition of the Markov time $\widehat{T}_j$, $\widehat{\lambda}_j^*(\widehat{T}_j) \geqslant a_{ij}$ on $\{\widehat{T}_j < \infty\}$. So, for all $\theta \in \Theta_i$ $(i \neq j)$, we obtain

$$\begin{aligned}
\alpha_{ij}(\widehat{\mathsf{D}}, \theta) &= \mathsf{E}_\theta \left[\mathbb{1}_{\{\hat{d}=j\}} \right] \leqslant \mathsf{E}_\theta \left[\mathbb{1}_{\{\widehat{T}_j < \infty\}} \right] \\
&= \mathsf{E}_\theta \left[\mathbb{1}_{\{\widehat{T}_j < \infty\}} \widehat{\Lambda}_j^*(\widehat{T}_j) \exp\left\{ -\widehat{\lambda}_j^*(\widehat{T}_j) \right\} \right] \\
&\leqslant \exp\left\{-a_{ij}\right\} \, \mathsf{E}_\theta \left[\mathbb{1}_{\{\widehat{T}_j < \infty\}} \widehat{\Lambda}_j^*(\widehat{T}_j) \right].
\end{aligned}$$

Since $\widehat{\Lambda}_j^*(n) \leqslant \widehat{\Lambda}_\theta(n)$ for any $n \geqslant 1$ and all $\theta \in \Theta_i$, $i \neq j$ we obtain that for all $\theta \in \Theta_i$ and $i = 0, 1, \dots, N$

$$\alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \exp\left\{-a_{ij}\right\} \, \mathsf{E}_\theta \left[\mathbb{1}_{\{\widehat{T}_j < \infty\}} \widehat{\Lambda}_\theta(\widehat{T}_j) \right],$$

where, as established in the previous section, the adaptive LR $(\widehat{\Lambda}_\theta(n), \mathscr{F}_n)_{n \geq 1}$ is the P_θ-martingale with expectation $\mathsf{E}_\theta[\widehat{\Lambda}_\theta(n)] = 1$. Thus, the adaptive LR $\{\widehat{\Lambda}_\theta(n)\}_{n \geqslant 1}$ is a viable likelihood ratio process, i.e., for any $\theta \in \Theta$, there exists a probability measure $\widehat{\mathsf{P}}_\theta$ such that

$$\widehat{\Lambda}_\theta(n) = \frac{\mathrm{d}\widehat{\mathsf{P}}_\theta^n}{\mathrm{d}\mathsf{P}_\theta^n}, \quad n \geqslant 1.$$

Applying Wald's likelihood ratio identity yields

$$\mathsf{E}_\theta \left[\mathbb{1}_{\{\widehat{T} < \infty\}} \widehat{\Lambda}_\theta(\widehat{T}_j) \right] = \widehat{\mathsf{P}}_\theta(\widehat{T}_j < \infty),$$

and hence,

$$\sup_{\theta \in \Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \exp\left\{-a_{ij}\right\} \sup_{\theta \in \Theta_i} \widehat{\mathsf{P}}_\theta(\widehat{T}_j < \infty) \leqslant \exp\left\{-a_{ij}\right\},$$

which gives the lemma. □

5.6. First-order uniform asymptotic optimality of the AMSPRT

Recall that $I_j(\theta) = \inf_{\vartheta \in \Theta_j} I(\theta, \vartheta)$. For $\theta \in \Theta$, define

$$\widehat{\Upsilon}_{\theta,j,\varepsilon,r}(n) = \sum_{t=n}^{\infty} t^{r-1} \mathsf{P}_\theta \left\{ \widehat{\lambda}_j^*(t) < (I_j(\theta) - \varepsilon)\psi(t) \right\}.$$

To obtain asymptotic upper bounds for moments of the stopping time distribution of the AMSPRT we will use the following left-tail condition.

C3. *Adaptive Left-tail Condition.* There exists a positive continuous function $I(\theta, \vartheta)$, satisfying condition (14), such that for any $\varepsilon > 0$ and some $r \geqslant 1$

$$\lim_{n\to\infty} \max_{\substack{i,j\in\mathscr{N}_0 \\ j\neq i}} \sup_{\theta_i\in\Theta_i} \widehat{\Upsilon}_{\theta_i,j,\varepsilon,r}(n) = 0 \tag{65}$$

and

$$\lim_{n\to\infty} \max_{j\in\mathscr{N}_0} \sup_{\theta\in\Theta_{\rm in}} \widehat{\Upsilon}_{\theta,j,\varepsilon,r}(n) = 0. \tag{66}$$

The following theorem provides the asymptotic upper bounds for moments of the AMSPRT stopping time distribution which together with the lower bounds (20) and (21) imply asymptotic optimality of the AMSPRT in class $\mathbb{C}(\boldsymbol{\alpha})$.

Theorem 5.3. *Assume that there exist an increasing function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and a function $I(\theta, \vartheta)$, satisfying (14), such that for some $r \geqslant 1$ left-tail condition* **C**3 *holds. If the thresholds in AMSPRT are so selected that $\sup_{\theta\in\Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log \alpha_{ji}^{-1}$ as $\alpha_{\max} \to 0$, in particular as $a_{ji} = \log \alpha_{ji}^{-1}$, then as $\alpha_{\max} \to 0$*

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant [F_{i,\theta}(\boldsymbol{\alpha})]^r\,(1+o(1)) \quad \textit{for all } \theta\in\Theta_i \textit{ and } i\in\mathscr{N}_0; \tag{67}$$

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant \left[\min_{0\leqslant i\leqslant N} F_{i,\theta}(\boldsymbol{\alpha})\right]^r (1+o(1)) \quad \textit{for all } \theta\in\Theta_{\rm in}. \tag{68}$$

Proof. By the definition of the Markov time $\widehat{T}_i$, we have

$$\begin{aligned}
\mathsf{P}_\theta(\widehat{T}_i > n) &= \mathsf{P}_\theta\left\{\max_{1\leqslant t\leqslant n} \min_{j\in\mathscr{N}_0\setminus i}\left[\widehat{\lambda}_i^*(t) - a_{ji}\right] < 0\right\} \\
&\leqslant \mathsf{P}_\theta\left\{\min_{j\in\mathscr{N}_0\setminus i}\left[\widehat{\lambda}_i^*(n) - a_{ji}\right] < 0\right\} \\
&\leqslant \sum_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\widehat{\lambda}_i^*(n)/\psi(n) < a_{ji}/\psi(n)\right\}.
\end{aligned}$$

Let $\mathbf{a} = (a_{ij})$ denote the matrix of finite thresholds a_{ij}, $i, j = 0, 1, \ldots, N$ (a_{ii} are immaterial) of AMSPRT and let $M_{i,\theta}(\mathbf{a}, \varepsilon)$ be as in (53). Obviously, for $n \geqslant M_{i,\theta}(\mathbf{a},\varepsilon)$

$$\sum_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\widehat{\lambda}_j^*(n)/\psi(n) < a_{ji}/\psi(n)\right\} \leqslant \sum_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\frac{1}{\psi(n)}\widehat{\lambda}_j^*(n) < I_j(\theta)-\varepsilon\right\},$$

and hence, for all sufficiently large n and $a_{\min}$, we obtain the inequality

$$\mathsf{P}_\theta(\widehat{T}_i > n) \leqslant \sum_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\widehat{\lambda}_j^*(n) < (I_j(\theta)-\varepsilon)\psi(n)\right\}. \tag{69}$$

Similarly to (55), for all $i = 0,1,\dots,N$,

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant \mathsf{E}_\theta[\widehat{T}_i^r] \leqslant M_{i,\theta}(\mathbf{a},\varepsilon)^r + r2^{r-1} \sum_{n=M_{i,\theta}(\mathbf{a},\varepsilon)}^{\infty} n^{r-1}\mathsf{P}_\theta(\widehat{T}_i > n). \tag{70}$$

Using (69) and (70) yields

$$\begin{aligned}\mathsf{E}_\theta[\widehat{T}^r] &\leqslant \mathsf{E}_\theta[\widehat{T}_i^r] \leqslant M_{i,\theta}(\mathbf{a},\varepsilon)^r \\ &\quad + r2^{r-1}N \sum_{n=M_{i,\theta}(\mathbf{a},\varepsilon)}^{\infty} n^{r-1} \max_{j\in\mathscr{N}_0\setminus i} \mathsf{P}_\theta\left\{\widehat{\lambda}_j^*(n) < (I_j(\theta)-\varepsilon)\psi(n)\right\} \\ &= M_{i,\theta}(\mathbf{a},\varepsilon)^r + r2^{r-1}N \max_{j\in\mathscr{N}_0\setminus i} \widehat{\Upsilon}_{\theta,j,\varepsilon,r}(M_{i,\theta}(\mathbf{a},\varepsilon)). \end{aligned} \tag{71}$$

Let $\theta = \theta_i \in \Theta_i$ ($i = 0,1,\dots,N$). Then inequality (71) implies that for all $i = 0,1,\dots,N$

$$\mathsf{E}_{\theta_i}[\widehat{T}^r] \leqslant M_{i,\theta_i}(\mathbf{a},\varepsilon)^r + r2^{r-1}N \max_{j\in\mathscr{N}_0\setminus i} \sup_{\theta_i\in\Theta_i} \widehat{\Upsilon}_{\theta_i,j,\varepsilon,r}(M_{i,\theta_i}(\mathbf{a},\varepsilon)).$$

Since $M_{i,\theta_i}(\mathbf{a},\varepsilon) \to \infty$ as $a_{\min} \to \infty$, by the left-tail condition (65), the second term goes to 0, so that for all $i = 0,1,\dots,N$

$$\mathsf{E}_{\theta_i}[\widehat{T}^r] \leqslant M_{i,\theta_i}(\mathbf{a},\varepsilon)^r + o(1) \quad \text{as } a_{\min} \to \infty.$$

Since ε can be arbitrarily small, taking the limit $\varepsilon \to 0$ and using condition (12) we obtain the following asymptotic upper bound

$$\mathsf{E}_{\theta_i}[\widehat{T}^r] \leqslant \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i} \frac{a_{ji}}{I_j(\theta_i)}\right)\right]^r (1+o(1)) \quad \text{as } a_{\min} \to \infty, \tag{72}$$

which holds for all $\theta_i \in \Theta_i$ and all $i = 0,1,\dots,N$.

Setting $a_{ji} = |\log \alpha_{ij}|$ in (72) or, more generally, $a_{ji} \sim |\log \alpha_{ij}|$ (assuming $\sup_{\theta\in\Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \alpha_{ij}$), we obtain the inequalities (67) for all $\theta \in \Theta_i$ and all $i = 0, 1, \dots, N$.

Now, let $\theta \in \Theta_{\rm in}$. Then, by inequality (71), for any $i \in \mathscr{N}_0$

$$\mathsf{E}_\theta[\widehat{T}_i^r] \leqslant M_{i,\theta}(\mathbf{a}, \varepsilon)^r + r2^{r-1} N \max_{j\in\mathscr{N}_0\setminus i} \Upsilon_{\theta,j,\varepsilon,r}(M_{i,\theta}(\mathbf{a}, \varepsilon)).$$

Now, note that $\mathsf{E}_\theta[\widehat{T}^r] \leqslant \min_{i\in\mathscr{N}_0} \mathsf{E}_\theta[\widehat{T}_i^r]$, and therefore,

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant \min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a}, \varepsilon)^r + r2^{r-1} N \min_{i\in\mathscr{N}_0} \max_{j\in\mathscr{N}_0\setminus i} \sup_{\theta\in\Theta_{\rm in}} \Upsilon_{\theta,j,\varepsilon,r}(M_{i,\theta}(\mathbf{a}, \varepsilon)),$$

where, by the left-tail condition (66), the second term goes to 0. Hence,

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant \min_{i\in\mathscr{N}_0} M_{i,\theta}(\mathbf{a}, \varepsilon)^r + o(1) \quad \text{as } a_{\min} \to \infty.$$

Since ε can be arbitrarily small, taking the limit $\varepsilon \to 0$ and using condition (12) we obtain the asymptotic upper bound for $\theta \in \Theta_{\rm in}$

$$\mathsf{E}_\theta[\widehat{T}^r] \leqslant \left[\Psi\left(\min_{i\in\mathscr{N}_0} \max_{j\in\mathscr{N}_0\setminus i} \frac{a_{ji}}{I_j(\theta)}\right)\right]^r (1 + o(1)) \quad \text{as } a_{\min} \to \infty.$$

Setting $a_{ji} = |\log \alpha_{ij}|$ in this inequality or, more generally, $a_{ji} \sim |\log \alpha_{ij}|$ (assuming $\sup_{\theta\in\Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \alpha_{ij}$) gives inequality (68) for $\theta \in \Theta_{\rm in}$. ☐

Theorems 3.1 and 5.1 give the following first-order asymptotic optimality result.

Theorem 5.4. *Assume that there exist an increasing function $\psi(t)$, $\psi(\infty) = \infty$, satisfying condition (12), and a function $I(\theta, \vartheta)$, satisfying (14), such that the SLLN for the LLR (26) holds, i.e., the normalized LLR $\lambda_{\theta,\vartheta}(n)/\psi(n)$ converges almost surely to $I(\theta, \vartheta)$ under P_θ as $n \to \infty$. Assume, in addition, that for some $r \geqslant 1$ left-tail condition* **C**3 *holds. If the thresholds in the AMSPRT are so selected that $\sup_{\theta\in\Theta_i} \alpha_{ij}(\widehat{\mathsf{D}}, \theta) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log \alpha_{ji}^{-1}$ as $\alpha_{\max} \to 0$, in particular as $a_{ji} = \log \alpha_{ji}^{-1}$, then as $\alpha_{\max} \to 0$*

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \sim [F_{i,\theta}(\boldsymbol{\alpha})]^r \sim \mathsf{E}_\theta[\widehat{T}^r] \quad \textit{for all } \theta \in \Theta_i \textit{ and } i \in \mathscr{N}_0; \tag{73}$$

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \sim \left[\min_{i\in\mathscr{N}_0} F_{i,\theta}(\boldsymbol{\alpha})\right]^r \sim \mathsf{E}_\theta[\widehat{T}^r] \quad \textit{for all } \theta \in \Theta_{\rm in}. \tag{74}$$

The proof is elementary, similar to the proof of Theorem 5.2, and is omitted.

Remark 5.5. The same reasoning as in Remark 5.1 gives the following asymptotic approximations (as $a_{\min} \to \infty$) for moments of the sample sizes of the AMSPRT as functions of thresholds $\mathbf{a} = (a_{ij})$ regardless of the error probabilities

$$\mathsf{E}_\theta[\widehat{T}^r] \sim \left[\Psi\left(\max_{j\in\mathscr{N}_0\setminus i} \frac{a_{ji}}{I_j(\theta)}\right)\right]^r \quad \text{for all } \theta \in \Theta_i,\ i \in \mathscr{N}_0;$$
$$\mathsf{E}_\theta[\widehat{T}^r] \sim \left[\Psi\left(\min_{i\in\mathscr{N}_0} \max_{j\in\mathscr{N}_0\setminus i} \frac{a_{ji}}{I_j(\theta)}\right)\right]^r \quad \text{for all } \theta \in \Theta_{\text{in}}.$$

6. Examples

In this section, we consider several examples that are useful for certain practical applications. The substantially non-stationary model for observations in the first example turns out to be adequate for sequential detection of epidemics, as discussed in [10, 30] in the context of quickest change-point detection.

6.1. Example 1: Testing for the mean of normal autoregressive non-stationary process

This example has many applications. In particular, (a) in sensor systems such as radars, acoustic systems, and electro-optic imaging systems where it is required to detect signals with unknown intensities from objects in clutter and sensor noise (see, e.g., [31, 28]) and (b) in the detection of epidemics, e.g., Covid-19 (see, e.g., [10, 30]).

Observations are of the form

$$X_n = \theta\, S_n + \xi_n, \quad n \geqslant 1, \tag{75}$$

where S_n is a deterministic function (e.g., a signal) observed in additive noise ξ_n and $\theta \in \Theta = (-\infty, +\infty)$ is an unknown parameter. In many applications, noise $\{\xi_n\}_{n\geqslant 1}$ can be adequately modeled by the p-th order Gaussian autoregressive process AR(p) that satisfies the recursion

$$\xi_n = \sum_{t=1}^{p} \rho_t \xi_{n-t} + w_n, \quad n \geqslant 1, \tag{76}$$

where $\{w_n\}_{n\geqslant 1}$ is an i.i.d. normal $\mathcal{N}(0,\sigma^2)$ sequence ($\sigma>0$). For simplicity, let us set zero initial conditions $\xi_{1-p}=\xi_{2-p}=\cdots=\xi_0=0$. The coefficients $\rho_1,\dots,\rho_p$ and the variance σ^2 are known and all roots of the equation $z^p-\rho_1 z^{p-1}-\cdots-\rho_p=0$ are in the interior of the unit circle, so that the AR(p) process is stable.

For $n\geqslant 1$, define the p_n-th order residuals

$$\widetilde{S}_n = S_n - \sum_{t=1}^{p_n}\rho_t S_{n-t}, \quad \widetilde{X}_n = X_n - \sum_{t=1}^{p_n}\rho_t X_{n-t},$$

where $p_n=p$ if $n>p$ and $p_n=n$ if $1\leqslant n\leqslant p$. The conditional density has the form

$$f_{\theta,n}(X_n|\mathbf{X}^{n-1}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\widetilde{X}_n-\theta\widetilde{S}_n)^2}{2\sigma^2}\right\},$$

and therefore, the LLR

$$\lambda_{\theta,\vartheta}(n) = \frac{\theta-\vartheta}{\sigma^2}\sum_{t=1}^{n}\widetilde{S}_t\widetilde{X}_t - \frac{\theta^2-\vartheta^2}{2\sigma^2}\sum_{t=1}^{n}\widetilde{S}_t^2. \tag{77}$$

Since under measure P_θ the random variables $\{\widetilde{X}_n\}_{n\geqslant 1}$ are independent normal random variables $\mathcal{N}(\theta\widetilde{S}_n,\sigma^2)$, the LLR $\{\lambda_{\theta,\vartheta}(n)\}_{n\geqslant 1}$ is a P_θ-Gaussian process (with independent but non-identically distributed increments) with mean and variance

$$\mathsf{E}_\theta[\lambda_{\theta,\vartheta}(n)] = \frac{1}{2}\mathsf{Var}_\theta[\lambda_{\theta,\vartheta}(n)] = \frac{(\theta-\vartheta)^2}{2\sigma^2}\sum_{t=1}^{n}\widetilde{S}_t^2,$$

Assume that

$$\frac{1}{\sigma^2}\lim_{n\to\infty}\frac{1}{\psi(n)}\sum_{t=1}^{n}\widetilde{S}_t^2 = Q^2, \tag{78}$$

where $0<Q^2<\infty$. In a variety of signal processing applications, this condition holds with $\psi(n)=n$, e.g., in radar applications where the signal S_n is the sequence of harmonic pulses, in which case θ^2Q^2 is the so-called signal-to-noise ratio (SNR). In some applications such as detection, recognition, and tracking of objects on ballistic trajectories that can be approximated by polynomials of order $m = 2 - 3$, the function $\psi(n) = n^m$, $m > 1$. Under condition (78)

$$\frac{1}{\psi(n)} \lambda_{\theta,\vartheta}(n) \xrightarrow[n\to\infty]{\mathsf{P}_\theta-\text{a.s.}} \frac{1}{2}(\theta - \vartheta)^2 Q^2 = I(\theta, \vartheta) \quad \text{for all } \theta \in (-\infty, +\infty),$$

so that the SLLN (26) and the right-tail condition **C**1 hold.

Furthermore, $\lambda_{\theta,\vartheta}(n)/\psi(n) \to I(\theta, \vartheta)$ r-completely for all $r \geqslant 1$. Indeed, under P_θ, the "whitened" observations can be written as $\widetilde{X}_n = \theta \widetilde{S}_n + w_n$ and the LLR as

$$\lambda_{\theta,\vartheta}(n) = (\theta - \vartheta) W_n + \frac{(\theta - \vartheta)^2}{2\sigma^2} \sum_{t=1}^n \widetilde{S}_t^2,$$

where $W_n = \sigma^{-2} \sum_{t=1}^n \widetilde{S}_t w_t$ is a weighted sum of i.i.d. normal $\mathcal{N}(0, \sigma^2)$ random variables w_t. Let

$$\eta_n = \frac{W_n}{\sqrt{\sigma^{-2} \sum_{t=1}^n \widetilde{S}_t^2}} \quad \text{and} \quad b_n(\varepsilon) = \frac{\varepsilon \psi(n)}{\sqrt{\sigma^{-2} \sum_{t=1}^n \widetilde{S}_t^2}}. \tag{79}$$

Note that η_n is a standard normal $\mathcal{N}(0, 1)$ random variable and that

$$\mathsf{P}\left\{|W_n| > \varepsilon \psi(n)\right\} = \mathsf{P}\left\{|\eta_n| > b_n(\varepsilon)\right\},$$

and consequently, for arbitrary $b_n(\varepsilon) > 1$

$$\mathsf{P}\left\{|W_n| > \varepsilon \psi(n)\right\} = \sqrt{\frac{2}{\pi}} \int_{b_n(\varepsilon)}^\infty \exp\left\{-\frac{u^2}{2}\right\} \mathrm{d}u \leqslant \exp\left\{-\frac{b_n(\varepsilon)^2}{2}\right\}.$$

It follows from assumption (78) that $b_n^2(\varepsilon) \sim \varepsilon^2 \psi(n)/Q^2$ as $n \to \infty$, and therefore, for sufficiently large n

$$\mathsf{P}\left\{|W_n| > \varepsilon \psi(n)\right\} \leqslant O\left(\exp\left\{-\varepsilon^2 \psi(n)/2Q^2\right\}\right).$$

Recall that, by assumption (13), $\lim_{n\to\infty} [\psi(n)/\log n] = \infty$, which along with the previous inequality implies that

$$\lim_{n\to\infty} n^m \mathsf{P}\left\{|W_n| > \varepsilon \psi(n)\right\} = 0 \quad \text{for all } m > 0.$$

Hence,

$$\sum_{n=1}^{\infty} n^{r-1} \mathsf{P}\left\{|W_n| > \varepsilon\psi(n)\right\} < \infty \quad \text{for all } \varepsilon > 0 \text{ and all } r \geqslant 1. \tag{80}$$

In other words, $W_n/\psi(n)$ converges to 0 r-completely, which implies that for all $\varepsilon > 0$ and all $r \geqslant 1$

$$\sum_{n=1}^{\infty} n^{r-1} \mathsf{P}_\theta\left\{|\lambda_{\theta,\vartheta}(n) - I(\theta,\vartheta)| > \varepsilon\psi(n)\right\} < \infty, \tag{81}$$

where $I(\theta,\vartheta) = (\theta-\vartheta)^2 Q^2/2$.

If we are interested in testing simple hypotheses $\mathsf{H}_i : \theta = \theta_i$, $i = 0, 1, \dots, N$, then inequality (81) implies r-complete convergence condition (36) for the LLRs $\lambda_{ij}(n) = \lambda_{\theta_i,\theta_j}(n)$ with $I_{ij} = (\theta_i - \theta_j)^2 Q^2/2$, $i, j \in \mathscr{N}_0$, $i \neq j$. Hence, by Theorem 4.2, the MSPRT $\mathsf{D}_* = (T_*, d_*)$ is asymptotically optimal, minimizing all positive moments of the sample size to first-order, and asymptotic approximations (42) hold for all θ_i, $i = 0, 1, \dots, N$ as long as thresholds a_{ij} in the MSPRT are so selected that $\alpha_{ij}(\mathsf{D}_*) \leqslant \alpha_{ij}$ and $a_{ji} \sim \log \alpha_{ji}^{-1}$ as $\alpha_{\max} \to 0$.

Next, consider composite hypotheses, and for simplicity, let us focus on two hypotheses $\mathsf{H}_0 : \theta \leqslant \theta_0$ and $\mathsf{H}_1 : \theta \geqslant \theta_1$ $(\theta_0 < \theta_1)$ with the indifference interval $\Theta_{\rm in} = (\theta_0, \theta_1)$ when $N = 1$. Generalization for multiple hypotheses is straightforward, but the argument is more cumbersome. The case of two hypotheses is of special interest in object detection and epidemics detection applications. In the case of two hypotheses, the MMSPRT will be referred to as the M-2-SPRT.

To establish the optimality of the M-2-SPRT we need to show that the left-tail r-complete convergence condition $\mathbf{C}2$ holds. This follows from the argument analogous to that used for establishing r-complete convergence (81). The details are omitted. Thus, by Theorem 5.2, the M-2-SPRT D_*^π minimizes as $\alpha_{\max} \to 0$ all positive moments of the sample size and asymptotic formulas (59) and (60) hold for $i, j = 0, 1$ with

$$\begin{aligned} I_1(\theta) &= \inf_{\vartheta \geqslant \theta_1} I(\theta,\vartheta) = \frac{(\theta_1-\theta)^2 Q^2}{2} \quad \text{for } \theta < \theta_1, \\ I_0(\theta) &= \inf_{\vartheta \leqslant \theta_0} I(\theta,\vartheta) = \frac{(\theta-\theta_0)^2 Q^2}{2} \quad \text{for } \theta > \theta_0. \end{aligned} \tag{82}$$

Note that in the case of two hypotheses $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$, where $\alpha_0 = \alpha_{01}$, $\alpha_1 = \alpha_{10}$, and asymptotic formulas (59) and (60) yield

$$\inf_{\mathsf{D} \in \mathbb{C}(\alpha_0,\alpha_1)} \mathsf{E}_\theta[T^r] \sim \mathsf{E}_\theta[(T_*^\pi)^r]$$

$$\sim \begin{cases} [\Psi\left(|\log \alpha_0|/I_0(\theta)\right)]^r & \text{for } \theta \geqslant \theta_1 \\ [\Psi\left(|\log \alpha_1|/I_1(\theta)\right)]^r & \text{for } \theta \leqslant \theta_0 \\ [\Psi\left(\min_{i=0,1} |\log \alpha_i|/I_i(\theta)\right)]^r & \text{for } \theta \in (\theta_0, \theta_1) \end{cases}. \tag{83}$$

To prove the asymptotic optimality of the AMSPRT we need to verify condition **C**3. In the case of two hypotheses, the AMSPRT will be referred to as the A-2-SPRT.

Let

$$\hat{\theta}_n = \frac{\sum_{t=1}^n \widetilde{S}_t \widetilde{X}_t}{\sum_{t=1}^n \widetilde{S}_t^2}$$

be the unconditional MLE of θ and let $\hat{\theta}_{n,1} = \max(\theta_1, \hat{\theta}_n)$ and $\hat{\theta}_{n,0} = \min(\theta_0, \hat{\theta}_n)$ be MLEs restricted to the sets $\Theta_1 = [\theta_1, \infty)$ and $\Theta_0 = (-\infty, \theta_0]$, respectively. Then the statistics $\widehat{\lambda}_i^*(n)$ can be written as

$$\widehat{\lambda}_i^*(n) = \frac{1}{\sigma^2} \sum_{t=1}^n (\hat{\theta}_{t-1} - \hat{\theta}_{n,i}) \widetilde{S}_t \widetilde{X}_t - \frac{1}{2\sigma^2} \sum_{t=1}^n (\hat{\theta}_{t-1}^2 - \hat{\theta}_{n,i}^2) \widetilde{S}_t^2, \quad i = 0, 1.$$

In analogy with the argument that has led to (80), it can be shown that r-completely under P_θ

$$\begin{gathered} \hat{\theta}_n \to \theta, \quad \hat{\theta}_{n,1} \to \max(\theta_1, \theta), \quad \hat{\theta}_{n,0} \to \min(\theta_0, \theta), \\ \hat{\theta}_n^2 \to \theta^2, \quad \hat{\theta}_{n,1}^2 \to \max(\theta_1^2, \theta^2), \quad \hat{\theta}_{n,0}^2 \to \min(\theta_0^2, \theta^2), \\ \frac{1}{\psi(n)} \sum_{t=1}^n \hat{\theta}_{t-1}^2 \widetilde{S}_t^2 \to \theta^2 \sigma^2 Q^2, \quad \frac{1}{\psi(n)} \sum_{t=1}^n \hat{\theta}_{t-1} \widetilde{S}_t \widetilde{X}_t \to \theta^2 \sigma^2 Q^2. \end{gathered} \tag{84}$$

Indeed, we have

$$\mathsf{P}_\theta\left\{\left|\hat{\theta}_n - \theta\right| > \varepsilon\right\} = \mathsf{P}_\theta\left\{|\eta_n| > b_n(\varepsilon)/\psi(n)\right\},$$

where $\eta_n \sim \mathcal{N}(0,1)$ and $b_n(\varepsilon)$ are defined in (79). The same argument that has led to (80) yields

$$\sum_{n=1}^\infty n^{r-1} \mathsf{P}_\theta\left\{\left|\hat{\theta}_n - \theta\right| > \varepsilon \psi(n)\right\} < \infty \quad \text{for all } \varepsilon > 0 \text{ and all } r \geqslant 1,$$

so

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-completely}} \theta. \tag{85}$$

The rest of r-complete convergences in (84) are established analogously to (85). Using (84), after some manipulations we obtain that for all $r \geqslant 1$

$$\frac{1}{\psi(n)}\widehat{\lambda}_1^*(n) \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-completely}} I_1(\theta) \quad \text{for } \theta < \theta_1,$$
$$\frac{1}{\psi(n)}\widehat{\lambda}_0^*(n) \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-completely}} I_0(\theta) \quad \text{for } \theta > \theta_0,$$

where $I_i(\theta)$'s are given by (82). Hence, condition **C3** holds.

By Theorem 5.4, the A-2-SPRT is asymptotically optimal, minimizing all positive moments of the sample size: for all $r \geqslant 1$ as $\alpha_{\max} \to 0$ the asymptotics (83) hold with $\widehat{T}$. These asymptotic formulas can be also written as

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{E}_\theta[T^r] \sim \mathsf{E}_\theta[\widehat{T}^r] \sim \begin{cases} \Psi(2|\log\alpha_1|/[(\theta_1-\theta)^2Q^2])^r & \text{if } \theta \leqslant \theta^* \\ \Psi(2|\log\alpha_0|/[(\theta-\theta_0)^2Q^2])^r & \text{if } \theta \geqslant \theta^*, \end{cases}$$

where $\theta^* = (\theta_1\sqrt{c}+\theta_0)/(1+\sqrt{c})$ is the solution of the equation

$$|\log\alpha_0|/(\theta-\theta_0)^2 = |\log\alpha_1|/(\theta_1-\theta)^2$$

and $c \sim (\log\alpha_0)/(\log\alpha_1)$ as $\alpha_{\max} = \max(\alpha_0,\alpha_1) \to 0$ (see (10) in Remark 2.1).

In particular, if $S_n = S$, then $\psi(n) = n$, $\Psi(t) = t$, and $Q^2 = (1-\rho_1-\cdots-\rho_p)^2S^2/\sigma^2$. In this case, a higher-order approximation to the expected sample size $\mathsf{E}_\theta[\widehat{T}]$ of the A-2-SPRT up to an additive vanishing term $o(1)$ has been obtained in [29].

6.2. Example 2: Testing for covariance in Gaussian autoregressive models

Consider the problem of testing hypotheses regarding the covariance of the AR(p) process $\{X_n\}_{n\geqslant 1}$ which satisfies the recursion

$$X_n = \sum_{t=1}^{p} \rho_t\, X_{n-t} + w_n\,, \quad n \geqslant 1, \tag{86}$$

where $\{w_n\}_{n\geqslant 1}$ are i.i.d. standard normal $\mathcal{N}(0,1)$ random variables and coefficients $\rho_1,\dots,\rho_p$ are unknown. In this case, the parameter θ is p-dimensional, $\theta=(\rho_1,\dots,\rho_p)^\top$ where hereafter $\top$ denotes transpose. For $s\geqslant \ell\geqslant 1$, write $\mathbf{X}_\ell^s=(X_\ell,\dots,X_s)$.

The conditional density $f_\theta(X_n|\mathbf{X}_1^{n-1})=f_\theta(X_n|\mathbf{X}_{n-p}^{n-1})$ is

$$f_\theta(X_n|\mathbf{X}_{n-p}^{n-1})=\frac{1}{(2\pi)^{p/2}}\,\exp\left\{-\frac{(\eta_\theta(X_n,\mathbf{X}_{n-p}^{n-1}))^2}{2}\right\},$$

where $\eta_\theta(y,x)=y-(\theta)^\top x$ ($y\in\mathbb{R}$, $x=(x_1,\dots,x_p)\in\mathbb{R}^p$). Thus, for any $\theta\in\Theta=\mathbb{R}^p$, the LLR $\lambda_{\theta,\vartheta}(n)=\sum_{t=1}^n\lambda^*_{\theta,\vartheta}(t)$, where

$$\lambda^*_{\theta,\vartheta}(t)=\log\frac{f_\theta(X_t|\mathbf{X}_{t-p}^{t-1})}{f_\vartheta(X_t|\mathbf{X}_{t-p}^{t-1})}=X_t(\theta-\vartheta)^\top\mathbf{X}_{t-p}^{t-1}+\frac{1}{2}\left[(\vartheta^\top\mathbf{X}_{t-p}^{t-1})^2-(\theta^\top\mathbf{X}_{t-p}^{t-1})^2\right].$$

The process (86) is not Markov, but the p-dimensional process

$$Y_n=(X_n,\dots,X_{n-p+1})^\top\in\mathbb{R}^p$$

is Markov.

Next, for any $\theta=(\rho_1,\dots,\rho_p)\in\mathbb{R}^p$, define the matrix

$$L(\theta)=\begin{pmatrix}\rho_1 & \rho_2 & \dots & \rho_p\\ 1 & 0 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots 1 & 0\end{pmatrix}$$

and notice that

$$Y_n=L(\theta)Y_{n-1}+\tilde{w}_n,\quad n\geqslant 1, \tag{87}$$

where $\tilde{w}_n=(w_n,0,\dots,0)^\top\in\mathbb{R}^p$. Obviously,

$$\mathsf{E}[\tilde{w}_n\,\tilde{w}_n^\top]=B=\begin{pmatrix}1 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & 0\end{pmatrix}.$$

Assume that θ belongs to the set Θ_{st} for which all eigenvalues $\mathbf{e}_\ell(\Lambda)$ of the matrix $L(\theta)$ in modules are less than 1:

$$\Theta_{\mathrm{st}}=\{\theta\in\mathbb{R}^p:\max_{1\leqslant\ell\leqslant p}|\mathbf{e}_\ell(L(\theta))|<1\}. \tag{88}$$

Using (87) it can be shown that in this case the process $\{Y_n\}_{n\geqslant 1}$ is ergodic with stationary normal distribution $\mathcal{N}(0, \mathsf{F}(\theta))$, where

$$\mathsf{F}(\theta) = \sum_{n=0}^{\infty} (\Lambda(\theta))^n B(\Lambda^\top(\theta))^n.$$

Since $\sup_{t\geqslant 1} \mathsf{E}_\theta |X_t|^r < \infty$ for any $r \geqslant 1$ and $\theta \in \mathbb{R}^p$ it follows that $\sup_{t\geqslant 1} \mathsf{E}_\theta |\lambda_{\theta,\vartheta}(t)|^r < \infty$ for any $r \geqslant 1$ and $\theta \in \mathbb{R}^p$, and therefore, the SLLN (26) holds with $I_{\theta,\vartheta} = (\theta-\vartheta)^\top \mathsf{F}(\theta)(\theta-\vartheta)/2$. Also, using techniques developed in [17, 18] it can be shown that the left-tail condition $\mathbf{C}2$ is satisfied with

$$I_j(\theta) = \inf_{\vartheta\in\Theta_j} \frac{1}{2}(\theta-\vartheta)^\top \mathsf{F}(\theta)(\theta-\vartheta) \quad \text{for } \theta \in \Theta_i \;\text{ and } \theta \in \Theta_{\text{in}}. \tag{89}$$

In particular, in the Markov scalar case where $p = 1$ and $\theta = \rho_1 = \rho$ in (86), we have

$$I_j(\rho) = \inf_{\rho^*\in\Theta_j} \frac{(\rho-\rho^*)^2}{2(1-\rho^2)} \quad \text{for } \rho \in \Theta_i \;\text{ and } \rho \in \Theta_{\text{in}}.$$

By Theorem 5.2, the MMSPRT D_*^π minimizes as $\alpha_{\max} \to 0$ all positive moments of the sample size and asymptotic formulas (59) and (60) hold with $I_j(\theta)$ specified in (89) for any compact subset of $\Theta = \Theta_{\text{st}}$ defined in (88).

6.3. Example 3: Testing for the mean of Gaussian data with unknown variance

6.3.1. Multi-hypothesis invariant sequential *t*-test. The model discussed in Example 1 has focused on Gaussian data of known variability. A more common practical scenario is when the variability of data is unknown. In this section, we discuss the simplest i.i.d. model, although the results can be extended for more general non-i.i.d. situations.

Let $\{X_n\}_{n\geqslant 1}$ be the sequence of i.i.d. normal $\mathcal{N}(\mu,\sigma^2)$ random variables with unknown mean μ and unknown variance σ^2, where the variance σ^2 is a nuisance parameter. Let $\theta = \mu/\sigma$. We are interested in testing the hypotheses $\mathsf{H}_i : \theta = \theta_i$, $i = 0, 1, \dots, N$, where $\theta_0, \theta_1, \dots, \theta_N$ are given distinct numbers. Lai [6] considered this problem for testing two hypotheses in the context of invariant tests relative to the unknown variance σ^2. Lai proved that the invariant sequential t-test (t-SPRT) is first-order asymptotically optimal among all tests invariant to σ^2. Lai's result can be easily extended to multiple hypotheses. Below we show that the proposed MSPRT is also asymptotically optimal to first order, minimizing all positive moments of the sample size for all hypotheses in class $\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})$ among all tests invariant under scale changes.

The hypothesis testing problem is invariant under the group of scale changes, i.e., under the transformation which transforms $X_1, X_2, \dots, X_n$ into $cX_1, cX_2, \dots, cX_n$ for an arbitrary non-zero constant c. Under this group of transformations, the maximal invariant is $\mathbb{M}_n = (1, X_2/X_1, \dots, X_n/X_1)$. For $n \geqslant 1$, let $Y_n = X_n/X_1$,

$$\overline{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t, \quad v_n^2 = \frac{1}{n}\sum_{t=1}^{n} Y_t^2, \quad t_n = \frac{\overline{Y}_n}{v_n} = \frac{n^{-1}\sum_{t=1}^{n} X_t}{[n^{-1}\sum_{t=1}^{n} X_t^2]^{1/2}}.$$

Straightforward calculation shows that the density of the maximal invariant under the hypothesis H_i is

$$p_i(\mathbb{M}_n) = \frac{1}{\sqrt{2\pi(n-1)n v_n^{2(n-1)}}} \int_0^\infty u^{-1} \exp\{n f(u, \theta_i t_n)\}\, \mathrm{d}u, \tag{90}$$

where $f(u,z) = -u^2/2 + zu + \log u$. Therefore, the invariant LLRs are given by

$$\lambda_{ij}(n) = \log\left[\frac{\int_0^\infty u^{-1}\exp\{n f(u,\theta_i t_n)\}\,\mathrm{d}u}{\int_0^\infty u^{-1}\exp\{n f(u,\theta_j t_n)\}\,\mathrm{d}u}\right], \quad i,j = 0,1,\dots,N, \ i \neq j.$$

The invariant MSPRT is defined as in (31)-(32) with these invariant LLRs. Note that the statistic t_n is the famous Student t-statistic which is the basis for Student's t-test in the fixed sample size setting. For this reason, the invariant MSPRT based on $\lambda_{ij}(n, t_n)$ will be referred to as the t-MSPRT.

Define

$$J_n(z) = \int_0^\infty u^{-1}\exp\{n f(u,z)\}\ \mathrm{d}u,$$

so the LLRs for the maximal invariant are of the form

$$\lambda_{ij}(n) = \log[J_n(\theta_i t_n)/J_n(\theta_j t_n)], \quad i,j = 0,1,\dots,N, \ i \neq j.$$

The invariant LLRs $\lambda_{ij}(n)$ are too complicated for direct use. However, it is possible to replace $\lambda_{ij}(n)$ with a suitable approximation, $\lambda_{ij}(n) \approx \tilde{\lambda}_{ij}(n)$. If $\mathsf{P}_i(|\lambda_{ij}(n) - \tilde{\lambda}_{ij}(n)| < C) = 1$ for $n \geqslant n_0$ ($n_0 \geqslant 1$) with C a constant, then the r-complete convergence of $n^{-1}\tilde{\lambda}_{ij}(n)$ to I_{ij} under P_i implies the r-complete convergence $n^{-1}\lambda_{ij}(n) \to I_{ij}$ under P_i.

Specifically, using the uniform version of the Laplace asymptotic integration method (cf. Wijsman [36]), it can be shown that uniformly in t_n

$$|\lambda_{ij}(n) - n\, g_{ij}(t_n) - \Delta_{ij}(t_n)| \to 0 \quad \text{as } n \to \infty,$$

where the term $\Delta_{ij}(t_n)$ is bounded by a finite positive constant C_{ij} and

$$g_{ij}(t_n) = \phi(\theta_i t_n) - \phi(\theta_j t_n) - \tfrac{1}{2}\,(\theta_i^2 - \theta_j^2),$$
$$\phi(t_n) = \tfrac{1}{4}\, t_n \left(t_n + \sqrt{4+t_n^2}\right) + \log\left(t_n + \sqrt{4+t_n^2}\right).$$

Consequently,

$$\left|n^{-1}\lambda_{ij}(n) - g_{ij}(t_n)\right| \leqslant C_{ij}/n, \quad n \geqslant 1,$$

Since $\mathsf{E}_i\, |X_1|^r < \infty$ for all $r \geqslant 1$, it follows that

$$t_n \xrightarrow[n\to\infty]{\mathsf{P}_i - r - \text{completely}} \frac{\mathsf{E}_i[X_1]}{\sqrt{\mathsf{E}_i[X_1^2]}} = \frac{\theta_i}{\sqrt{1+\theta_i^2}} = Q_i \quad \text{for all } r \geqslant 1,$$

and therefore, the normalized LLR $n^{-1}\lambda_{ij}(n)$ converges r-completely to $g_{ij}(Q_i)$ under P_i, so that the r-complete convergence conditions (36) hold for all $r \geqslant 1$ with $\psi(n) = n$ and $I_{ij} = g_{ij}\,(Q_i)$.

It remains to verify that $I_{ij} > 0$. To this end, note that for any fixed $|t| \leqslant 1$, the maximum of the function $\tilde{\phi}(\theta, t) = \phi(\theta t) - \theta^2/2$ over θ is attained at $\theta^* = t/(1-t^2)^{1/2}$, so that $\theta^* = \theta_i$ if $t = Q_i = \theta_i/(1+\theta_i)^{1/2}$. Hence,

$$g_{ij}\,(Q_i) = \tilde{\phi}(\theta_i, Q_i) - \tilde{\phi}(\theta_j, Q_i) > 0.$$

By Theorem 4.2, the t-MSPRT asymptotically minimizes all positive moments of the stopping time and, as $\alpha_{\max} \to 0$,

$$\inf_{\mathsf{D}\in\mathbb{C}_{\rm sim}(\boldsymbol{\alpha})} \mathsf{E}_i[T^r] \sim \left(\max_{j\in\mathscr{N}_0\setminus i} \frac{|\log \alpha_{ji}|}{g_{ij}(Q_i)}\right)^r \sim \mathsf{E}_i[T_*^r], \quad i = 0, 1, \dots, N.$$

In the case of two hypotheses ($N = 1$), the above results are identical to those obtained by Lai [6] for the t-SPRT.

6.3.2. Adaptive sequential test. We continue considering the same model as in Subsection 6.3.1 but now in the context of the adaptive SPRT. So again $X_n \sim \mathcal{N}(\mu, \sigma^2)$, $n = 1, 2, \dots$ are i.i.d. normal random variables with unknown mean μ and unknown variance σ^2, but now we focus on the two composite hypotheses $\mathsf{H}_0 : \mu \leqslant \mu_0, \sigma^2 > 0$ and $\mathsf{H}_1 : \mu \geqslant \mu_1, \sigma^2 > 0$, where μ_1, μ_0 are given numbers, $\mu_1 > \mu_0$, and σ^2 is an unknown nuisance parameter. If this model is treated in the context of invariant tests when the hypotheses are $\mathsf{H}_i : \mu/\sigma = q_i$, $i = 0, 1$, where q_0 and q_1 are given numbers, then the results in the previous subsection show that the invariant t-SPRT is asymptotically optimal in the class of invariant tests. However, for values of $q = \mu/\sigma$ different from q_i, this test is not optimal. It performs especially poorly in the indifference zone (q_0, q_1). To overcome this drawback Tartakovsky *et al.* [28] construct an invariant t-2-SPRT, which minimizes the expected sample size at the worst point $q^* \in (q_0, q_1)$. But this test is also not optimal for any other point and performs not great at the points located far from q^*. On the other hand, the AMSPRT (which we refer to as the A-2-SPRT in the case of two hypotheses) is adaptive and asymptotically efficient at any point $q \in (-\infty, \infty)$. Furthermore, it is also invariant to scale transformations.

Let $\theta = (\mu, \sigma^2)$ and $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}^2)$. We now show that all conditions of Theorem 5.4 (with $N = 1$) are satisfied when $\{\hat{\theta}_n\}$ is a sequence of MLEs, which implies uniform asymptotic optimality of the A-2-SPRT with $a_{01} = a_0 = \log(1/\alpha_0)$ and $a_{10} = a_1 = \log(1/\alpha_1)$ in class $\mathbb{C}(\boldsymbol{\alpha}) = \mathbb{C}(\alpha_0, \alpha_1)$.

The LLR is given by

$$\lambda_{\theta,\tilde{\theta}}(n) = \frac{n}{2} \log\left(\frac{\tilde{\sigma}^2}{\sigma^2}\right) + \frac{\sigma^2 - \tilde{\sigma}^2}{2\tilde{\sigma}^2\sigma^2} \sum_{t=1}^{n} X_t^2 + \frac{\mu\tilde{\sigma}^2 - \tilde{\mu}\sigma^2}{\tilde{\sigma}^2\sigma^2} \sum_{t=1}^{n} X_t - \frac{\mu^2\tilde{\sigma}^2 - \tilde{\mu}^2\sigma^2}{2\tilde{\sigma}^2\sigma^2} n,$$

and the Kullback–Leibler "distance" is

$$I(\theta, \tilde{\theta}) = \mathsf{E}_\theta[\lambda_{\theta,\tilde{\theta}}(1)] = \frac{1}{2}\left\{\frac{(\mu - \tilde{\mu})^2 + \sigma^2}{\tilde{\sigma}^2} + \log\frac{\tilde{\sigma}^2}{\sigma^2} - 1\right\}.$$

By the SLLN,

$$n^{-1}\lambda_{\theta,\tilde{\theta}}(n) \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-a.s.}} I(\theta, \tilde{\theta}).$$

Thus, condition **C**1 holds. It remains to verify positiveness of

$$I_1(\mu, \sigma^2) = \inf_{\tilde{\mu} \geqslant \mu_1, \tilde{\sigma}^2 > 0} I(\mu, \sigma^2; \tilde{\mu}, \tilde{\sigma}^2) \quad \text{for } \mu < \mu_1, \sigma^2 > 0$$

and

$$I_0(\mu, \sigma^2) = \inf_{\tilde{\mu} \leqslant \mu_0, \tilde{\sigma}^2 > 0} I(\mu, \sigma^2; \tilde{\mu}, \tilde{\sigma}^2) \quad \text{for } \mu > \mu_0, \sigma^2 > 0$$

and the left-tail condition **C**3.

Let $q = \mu/\sigma$ and $q_i = \mu_i/\sigma$. Let $\mathbb{Q} = (-\infty, +\infty)$ denote the q-parameter space and let $\mathbb{Q}_0 = (-\infty, q_0]$, $\mathbb{Q}_1 = [q_1, \infty)$, $\mathbb{Q}_{\rm in} = (q_0, q_1)$.

The minimum value $\min_{\tilde\sigma>0} I(\theta, \tilde\theta) = \frac{1}{2}\log\left[1 + (\mu - \tilde\mu)^2/\sigma^2\right]$ is achieved at the point $\tilde\sigma^2 = \sigma^2 + (\mu - \tilde\mu)^2$ and $I_i(\mu, \sigma^2)$ are given by

$$
\begin{aligned}
I_1(\theta) &= \inf_{\substack{\tilde\mu \geqslant \mu_1, \\ \tilde\sigma^2 > 0}} I(\theta, \tilde\theta) = I_1(q) = \frac{1}{2}\log[1 + (q_1 - q)^2] \quad \text{for } q < q_1, \\
I_0(\theta) &= \inf_{\substack{\tilde\mu \leqslant \mu_0, \\ \tilde\sigma^2 > 0}} I(\theta, \tilde\theta) = I_0(q) = \frac{1}{2}\log[1 + (q - q_0)^2] \quad \text{for } q > q_0.
\end{aligned} \tag{91}
$$

Clearly, $I_0(q) > 0$ for $q \in \mathbb{Q}_1 + \mathbb{Q}_{\rm in} = (q_0, \infty)$ and $I_1(q) > 0$ for $q \in \mathbb{Q}_0 + \mathbb{Q}_{\rm in} = (-\infty, q_1)$, and hence, $\min[I_0(q), I_1(q)] > 0$ for $q \in \mathbb{Q}_{\rm in} = (q_0, q_1)$. Therefore, the conditions related to the minimal Kullback–Leibler "distances" for the corresponding sets hold and it remains to deal with the left-tail r-complete convergence condition **C**3.

The unrestricted MLE

$$
\hat\theta_n = (\hat\mu_n, \hat\sigma_n^2) = \arg \sup_{\substack{\mu \in (-\infty, \infty), \\ \sigma^2 > 0}} \lambda_{\theta;\tilde\theta}(n)
$$

is a combination of the sample mean and sample variance,

$$
\hat\mu_n = \overline{X}_n = n^{-1}\sum_{t=1}^n X_t, \quad \hat\sigma_n^2 = v_n^2 = n^{-1}\sum_{t=1}^n (X_t - \overline{X}_n)^2.
$$

Let

$$
\hat\mu_{n,1} = \max\{\mu_1, \overline{X}_n\}, \quad \hat\mu_{n,0} = \min\{\mu_0, \overline{X}_n\}, \quad \hat\sigma_{n,i}^2 = n^{-1}\sum_{t=1}^n (X_t - \hat\mu_{n,i})^2
$$

be the restricted MLEs of μ and σ^2 conditioned on the hypotheses H_1 and H_0, respectively,

$$
\begin{aligned}
\hat\mu_{n,1} &= \arg \sup_{\mu \geqslant \mu_1} \lambda_{\theta;\tilde\theta}(n), \quad \hat\mu_{n,0} = \arg \sup_{\mu \leqslant \mu_0} \lambda_{\theta;\tilde\theta}(n), \\
\hat\sigma_{n,1}^2 &= \arg \sup_{\substack{\mu \geqslant \mu_1, \\ \sigma^2 > 0}} \lambda_{\theta;\tilde\theta}(n), \quad \hat\sigma_{n,0}^2 = \arg \sup_{\substack{\mu \leqslant \mu_0, \\ \sigma^2 > 0}} \lambda_{\theta;\tilde\theta}(n).
\end{aligned}
$$

Straightforward calculation shows that the decision statistics in the A-2-SPRT are $\hat{\lambda}_i^*(n) = \ell(n) - \ell_i(n)$ $(i = 0, 1)$, where

$$
\begin{aligned}
\ell(n) &= \frac{1}{2}\sum_{t=1}^{n}\left[\log\left(\frac{1}{v_{t-1}^2}\right) + \frac{1}{v_{t-1}^2}\left(2\overline{X}_{t-1}X_t - X_t^2 - \overline{X}_{t-1}^2\right)\right], \\
\ell_i(n) &= \frac{n}{2}\left[\log\left(\frac{1}{\hat{\sigma}_{n,i}^2}\right) - 1\right], \quad i = 0, 1.
\end{aligned} \tag{92}
$$

These statistics allow for an efficient recursive computation. Note that $\ell(n)$ requires an initial condition for the estimate $\hat{\theta}_0$. This condition is the design parameter that can be deterministic or random. In particular, we could set $\ell(0) = 0$.

Since $X_1, X_2, \dots$ are i.i.d. and $\mathsf{E}_\theta[|X_1|^r] < \infty$ for all $r \geqslant 1$, it can be shown that the following r-complete convergence conditions hold as $n \to \infty$ under P_θ:

$$
\begin{aligned}
&\overline{X}_n \to \mu, \quad \overline{X}_n^2 \to \mu^2, \quad v_n^2 \to \sigma^2 \quad \text{for all } \mu \in (-\infty, +\infty),\ \sigma^2 > 0; \\
&\hat{\mu}_{n,1} \to \begin{cases} \mu & \text{if } \mu \geqslant \mu_1 \\ \mu_1 & \text{if } \mu < \mu_1 \end{cases}, \quad \hat{\mu}_{n,0} \to \begin{cases} \mu & \text{if } \mu \leqslant \mu_0 \\ \mu_0 & \text{if } \mu > \mu_0 \end{cases}, \\
&\hat{\sigma}_{n,1}^2 \to \begin{cases} \sigma^2 & \text{if } \mu \geqslant \mu_1 \\ \sigma^2 + (\mu - \mu_1)^2 & \text{if } \mu < \mu_1 \end{cases}, \quad \hat{\sigma}_{n,0}^2 \to \begin{cases} \sigma^2 & \text{if } \mu \leqslant \mu_0 \\ \sigma^2 + (\mu - \mu_0)^2 & \text{if } \mu > \mu_0 \end{cases}.
\end{aligned}
$$

Using these relations along with (92), it can be verified that r-completely under P_θ as $n \to \infty$

$$
\begin{aligned}
n^{-1}\ell(n) &\to \frac{1}{2}\left(\log\frac{1}{\sigma^2} - 1\right) \quad \text{for all } \mu \in (-\infty, +\infty),\ \sigma^2 > 0; \\
n^{-1}\ell_1(n) &\to \begin{cases} \frac{1}{2}(\log\frac{1}{\sigma^2} - 1) & \text{if } \mu \geqslant \mu_1,\ \sigma^2 > 0 \\ \frac{1}{2}(\log\frac{1}{\sigma^2 + (\mu_1 - \mu)^2} - 1) & \text{if } \mu < \mu_1,\ \sigma^2 > 0 \end{cases}, \\
n^{-1}\ell_0(n) &\to \begin{cases} \frac{1}{2}(\log\frac{1}{\sigma^2} - 1) & \text{if } \mu \leqslant \mu_0,\ \sigma^2 > 0 \\ \frac{1}{2}(\log\frac{1}{\sigma^2 + (\mu - \mu_0)^2} - 1) & \text{if } \mu > \mu_0,\ \sigma^2 > 0 \end{cases}.
\end{aligned}
$$

Combining these formulas yields (for all $r \geqslant 1$)

$$
n^{-1}\widehat{\lambda}_i^*(n) \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-completely}} I_i(\theta) = I_i(q) \quad \text{for } q \in \mathbb{Q} \setminus \mathbb{Q}_i,\ i = 0, 1,
$$

where $I_i(q)$ are given by (91).

Therefore, condition **C3** is satisfied with $I_i(\theta) = I_i(q)$. By Theorem 5.4, the A-2-SPRT is asymptotically optimal, minimizing all positive moments of the sample size in the class of tests $\mathbb{C}(\alpha_0, \alpha_1)$: for all $r \geqslant 1$ as $\alpha_{\max} \to 0$

$$\inf_{\mathsf{D}\in\mathbb{C}(\alpha_0,\alpha_1)} \mathsf{E}_\theta[T^r] \sim \mathsf{E}_\theta[\widehat{T}^r] \sim \begin{cases} (2|\log\alpha_1|/\log[1+(q_1-q)^2])^r & \text{if } q \leqslant q^* \\ (2|\log\alpha_0|/\log[1+(q-q_0)^2])^r & \text{if } q \geqslant q^*, \end{cases}$$

where q^* is the solution of the equation

$$\left[1+(q_1-q)^2\right]^c = 1+(q-q_0)^2$$

($c \sim (\log\alpha_0)/(\log\alpha_1)$ as $\alpha_{\max} \to 0$). In particular, $q^* = (q_0+q_1)/2$ if $c = 1$.

It is also worth noting that the mixture test M-2-SPRT with the mixing improper prior density $\pi(\mu,\sigma) = \sigma^{-1}\mathrm{d}\sigma\mathrm{d}\mu$ is uniformly asymptotically optimal. To see this, consider for simplicity testing $\mu = q_0 = 0$ against $\mu \neq 0$ without the indifference zone. Introduce the probability measure

$$\mathsf{P}^\pi = \int_{-\infty}^{\infty}\int_0^{\infty} \sigma^{-1}\mathsf{P}_{\mu,\sigma}\mathrm{d}\sigma\mathrm{d}\mu.$$

Then the LR Λ_n^π of $(X_1,\dots,X_n)$ under P^π relative to $\mathsf{P}_{0,1}$ is the same as the LR of the maximal invariant $\mathbb{M}_n$ under $\bar{\mathsf{P}} = (2\pi)^{-1/2}\int_{-\infty}^{\infty}\mathsf{P}_q\mathrm{d}q$, i.e.,

$$\Lambda_n^\pi = \frac{\mathrm{d}\mathsf{P}^\pi(\mathbf{X}_1^n)}{\mathrm{d}\mathsf{P}_{0,1}(\mathbf{X}_1^n)} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{p_q(\mathbb{M}_n)}{p_0(\mathbb{M}_n)}\mathrm{d}q,$$

where $p_q(\mathbb{M}_n)$ is as in (90) with $\theta_i = q$. Direct calculation shows that

$$\lambda_n^\pi = \log\Lambda_n^\pi = \frac{n}{2}\log\left(1+\frac{\overline{X}_n^2}{v_n^2}\right) - \frac{1}{2}\log n$$

See [21], page 117. Clearly,

$$n^{-1}\lambda_n^\pi \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-completely}} I(\theta) = I(q) = \frac{1}{2}\log(1+q^2).$$

Since $I(q) \equiv I_1(q) \equiv I_0(q)$ when $q_0 = q_1 = 0$ (see (91)), it follows that asymptotic performance of the A-2-SPRT and the M-2-SPRT are the same.

7. Concluding remarks

1. Analogous near-optimality results can be obtained for the multi-hypothesis generalized likelihood ratio SPRT.

2. In the non-i.i.d. cases when observations are severely non-stationary and dependent, computing the LLRs is typically time-consuming since there are no recursive formulas as in the i.i.d. case. Computing the mixtures is even more time-consuming. One way to overcome this difficulty is to use window-limited versions when computing the corresponding statistics in a fixed moving time window, following the idea proposed by Lai [9] for change-point detection problems. To ensure the tests exhibit asymptotic optimality properties, the size of the window, denoted as $\tau = \tau(\alpha_{\max})$, must be a function of the specified error rate $\alpha_{\max}$. Moreover, it should approach infinity approximately as the maximal value of the optimum expected sample size

$$\tau(\alpha_{\max}) \sim \max\left\{\max_{i\in\mathscr{N}_0}\sup_{\theta\in\Theta_i} F_{i,\theta}(\boldsymbol{\alpha}), \min_{i\in\mathscr{N}_0}\sup_{\theta\in\Theta_{\rm in}} F_{i,\theta}(\boldsymbol{\alpha})\right\}.$$

3. The results of uniform optimality in Theorems 5.2 and 5.4 can be extended to establish the first-order minimax asymptotic optimality of the MMSPRT $\mathsf{D}_* = (T_*, d_*)$ and the AMSPRT $\widehat{\mathsf{D}} = (\widehat{T}, \hat{d})$. That is, both tests solve the asymptotic version of the Kiefer–Weiss problem [4] of minimizing the expected sample size in the worst scenario with respect to the parameter θ, or more generally, minimizing higher moments of the stopping time in the worst case. Specifically, let

$$\tilde{I}(\theta) = \begin{cases} \min\limits_{j\in\mathscr{N}_0\setminus i} I_j(\theta) & \text{if } \theta\in\Theta_i \\ \max\limits_{0\leqslant i\leqslant N}\min\limits_{j\in\mathscr{N}_0\setminus i} I_j(\theta) & \text{if } \theta\in\Theta_{\rm in} \end{cases}.$$

It can be shown that if instead of conditions (14) we will require a stronger separability condition $\inf_{\theta\in\Theta}\tilde{I}(\theta) > 0$ then as $\alpha_{\max}\to 0$

$$\inf_{\mathsf{D}\in\mathbb{C}(\boldsymbol{\alpha})}\sup_{\theta\in\Theta}\mathsf{E}_\theta[T^r] \sim \left[\max\left\{\max_{i\in\mathscr{N}_0}\sup_{\theta\in\Theta_i} F_{i,\theta}(\boldsymbol{\alpha}), \sup_{\theta\in\Theta_{\rm in}}\min_{i\in\mathscr{N}_0} F_{i,\theta}(\boldsymbol{\alpha})\right\}\right]^r,$$

and the right-hand side is attained for $\sup_{\theta\in\Theta}\mathsf{E}_\theta[T_*^r]$ and $\sup_{\theta\in\Theta}\mathsf{E}_\theta[\widehat{T}^r]$. In the i.i.d. case, this problem has been addressed by Lai [5], Lorden [13], Huffman [3], Pavlov [16] among others and in the non-i.i.d. case by Tartakovsky *et al.* [28].

4. As stated in previous sections, the almost sure convergence of the normalized LLR $\lambda_{\theta,\vartheta}(n)/\psi(n) \to I(\theta,\vartheta)$ as $n \to \infty$ under P_θ is not sufficient for the optimality of proposed sequential tests in the sense of minimizing the expected sample size or moments of the sample size. However, the following weak (in the almost sure sense) asymptotic optimality holds under this almost sure convergence condition for the MMSPRT (T_*, d_*) and the AMSPRT $(\widehat{T}, \hat{d})$: for all $\theta \in \Theta_i$ and $i = 0, 1, \dots, N$

$$\lim_{\alpha_{\max}\to 0} \sup_{T\in\mathbb{C}(\boldsymbol{\alpha})} \mathsf{P}_\theta\left\{\tau \geqslant (1+\varepsilon)T\right\} = 0 \quad \text{for every } \varepsilon > 0$$

and

$$\frac{\tau}{F_{i,\theta}(\boldsymbol{\alpha})} \xrightarrow[\alpha_{\max}\to 0]{\mathsf{P}_\theta-\text{a.s.}} 1,$$

where $\tau = T_*$ or $\tau = \widehat{T}$. Lai [6] established this result for the SPRT in the problem of testing two simple hypotheses and an asymptotically stationary case ($\psi(n) = n$).

Acknowledgements

This research draws inspiration from Tze Lai's seminal paper [6], which explored the asymptotic optimality of the SPRT for general non-i.i.d. models. The findings presented in Section 4, regarding the first-order asymptotic optimality of the multi-hypothesis (matrix) SPRT, directly extend Lai's contributions. I am deeply grateful to Tze Lai for the insightful conversations we shared between 1993 and 2023, which significantly contributed to this work.

References

[1] H. P. Chan and T. L. Lai, Asymptotic approximations for error probabilities of sequential or fixed sample size tests in exponential families, *Annals of Statistics* **28** (2000), 1638–1669.

[2] G. Fellouris and A. G. Tartakovsky, Multichannel sequential detection – Part I: Non-i.i.d. data, *IEEE Transactions on Information Theory* **63** (2017), 4551–4571. https://doi.org/10.1109/TIT.2017.2689785.

[3] M. D. Huffman, An efficient approximate solution to the Kiefer–Weiss problem, *Annals of Statistics* **11** (1983), 306–316.

[4] J. Kiefer and L. Weiss, Properties of generalized sequential probability ratio tests, *Annals of Mathematical Statistics* **28** (1957), 57–74.

[5] T. L. Lai, Optimal stopping and sequential tests which minimize the maximum expected sample size, *Annals of Statistics* **1** (1973), 659–673.

[6] T. L. Lai, Asymptotic optimality of invariant sequential probability ratio tests, *Annals of Statistics* **9** (1981), 318–333.

[7] T. L. Lai, Nearly optimal sequential tests of composite hypotheses, *Annals of Statistics* **16** (1988), 856–886.

[8] T. L. Lai and L. Zhang, A modification of Schwarz's sequential likelihood ratio tests in mulitvariate sequential analysis, *Sequential Analysis* **13** (1994), 79–96.

[9] T. L. Lai, Information bounds and quick detection of parameter changes in stochastic systems, *IEEE Transactions on Information Theory* **44** (1998), 2917–2929.

[10] Y. Liang, A. G. Tartakovsky and V. V. Veeravalli, Quickest change detection with non-stationary post-change observations, *IEEE Transactions on Information Theory* **69** (2023), 3400–3414. https://doi.org/0.1109/TIT.2022.3230583.

[11] G. Lorden, Integrated risk of asymptotically Bayes sequential tests, *Annals of Mathematical Statistics* **38** (1967), 1399–1422.

[12] G. Lorden, Open-ended tests for Koopman–Darmois families, *Annals of Statistics* **1** (1973), 633–643.

[13] G. Lorden, 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size, *Annals of Statistics* **4** (1976), 281–291.

[14] G. Lorden, Nearly-optimal sequential tests for finitely many parameter values, *Annals of Statistics* **5** (1977), 1–21.

[15] G. Lorden, Nearly optimal sequential tests for exponential families, *Unpublished Manuscript* (1977). Available from http://jaybartroff.com/research/gary.pdf

[16] I. V. Pavlov, Sequential procedure of testing composite hypotheses with applications to the Kiefer-Weiss problem, *Theory of Probability and its Applications* **35** (1990), 280–292.

[17] S. Pergamenchtchikov and A. G. Tartakovsky, Asymptotically optimal

pointwise and minimax quickest change-point detection for dependent data, *Statistical Inference for Stochastic Processes* **21** (2018), 217–259.

[18] S. M. Pergamenchtchikov, A. G. Tartakovsky and V. S. Spivak, Minimax and pointwise sequential changepoint detection and identification for general stochastic models, *Journal of Multivariate Analysis* **190** (2022), 1–22.

[19] H. Robbins and D. Siegmund, A class of stopping rules for testing parameter hypotheses, In L. M. Le Cam, J. Neyman, and E. L. Scott, editors, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4: Biology and Health, pages 37–41. University of California Press, Berkeley, CA, USA, 1972, 37–41.

[20] H. Robbins and D. Siegmund, The expected sample size of some tests of power one, *Annals of Statistics* **2** (1974), 415–436.

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

[22] D. Siegmund, Change-points: from sequential detection to biology and back, *Sequential Analysis* **32** (2013), 2–14.

[23] G. Schwarz, Asymptotic shapes of Bayes sequential testing regions, *Annals of Mathematical Statistics* **33** (1962), 224–236.

[24] A. G. Tartakovsky, Asymptotically optimal sequential tests for nonhomogeneous processes, *Sequential Analysis* **17** (1998), 33–62.

[25] A. G. Tartakovsky, Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case, *Statistical Inference for Stochastic Processes* **1** (1998), 265–295.

[26] A. G. Tartakovsky, *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*, Monographs on Statistics and Applied Probability 165, Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2020.

[27] A. G. Tartakovsky, Quick and complete convergence in the law of large numbers with applications to statistics, *MDPI Mathematics* **11** (2023), 2687.

[28] A. G. Tartakovsky, I. V. Nikiforov and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Monographs on Statistics and Applied Probability 136, Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2015.

[29] A. G. Tartakovsky, G. Sokolov and Y. Bar-Shalom, Nearly optimal adaptive sequential tests for object detection, *IEEE Transactions on Signal Processing* **68** (2020), pp. 3371–3384. DOI: 10.1109/TSP.2020.2986542.

[30] A. G. Tartakovsky and V. Spivak, Quickest changepoint detection in general multistream stochastic models: recent results, applications and future challenges, *Matrix Annals*, 2024 (to be published).

[31] A. G. Tartakovsky and J. Brown, Adaptive spatial-temporal filtering methods for clutter removal and target tracking, *IEEE Transactions on Aerospace and Electronic Systems* **44** (2008), 1522–1537.

[32] N. V. Verdenskaya and A. G. Tartakovsky, Asymptotically optimal sequential testing of multiple hypotheses for nonhomogeneous Gaussian processes in an asymmetric situation, *Theory of Probability and its Applications* **36** (1991), 536–547.

[33] A. Wald, Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, **16** (1945), 117–186.

[34] A. Wald, *Sequential Analysis*, John Wiley & Sons, Inc. New York, USA, 1947.

[35] A. Wald and J. Wolfowitz, Optimum character of the sequential probability ratio test, *Annals of Mathematical Statistics* **19** (1948), 326–339.

[36] R. A. Wijsman, Exponentially bounded stopping time of sequential probability ratio tests for composite hypotheses, *Annals of Mathematical Statistics* **42** (1971), 1859–1869.

[37] S. P. Wong, Asymptotically optimum properties of certain sequential tests, *Annals of Mathematical Statistics* **39** (1968), 1244–1263.

Alexander G. Tartakovsky
AGT StatConsult, Los Angeles, CA
USA
E-mail address: alexg.tartakovsky@gmail.com

Received 29 April 2024