Asymptotically optimal multistage tests for iid data

Yiming Xing, *Student, IEEE*, and Georgios Fellouris, *Member, IEEE*

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834797

***Abstract*—The problem of testing two simple hypotheses about the distribution of iid random elements is considered. In particular, the focus is on multistage tests that control the two error probabilities below arbitrary, user-specified levels. A novel multistage test is proposed, analyzed, and shown to achieve the optimal expected sample size under both hypotheses, in the class of all *sequential* tests with the same error control, to a first-order approximation as the two target error probabilities go to zero at *arbitrary* rates. The proposed test is compared, both theoretically and numerically, with a multistage test that enjoys the same asymptotic optimality property under one of the two hypotheses, while performing much worse under the other.**

***Index Terms*—multistage tests, asymptotic optimality, sequential testing, sequential thresholding, 3-stage test.**

I. Introduction

In contrast to traditional fixed-sample-size tests, tests that allow for stopping and making a decision after each observation is taken can achieve, exactly or in an asymptotic sense, the optimal expected sample size under both hypotheses. This was shown for the first time in the context of testing two simple hypotheses about iid data [1], [2], and since then, for a variety of distributional setups (see, e.g., [3]). However, the continuous monitoring of the sampling process can often be an expensive and even infeasible task. In such cases, it may be much more convenient to apply a *multistage test*, i.e., a testing procedure in which the sampling process can only be terminated at a small number of deterministic time instances.

Various multistage tests have been proposed and analyzed in the literature, mainly in the context of testing two simple hypotheses for iid data [4], [5], [6], [7], [8], [9], [10], [11]. In these works, the main problem is to minimize the expected sample size under one of the two hypotheses, or a mixture of them, subject to a constraint on the number of stages and/or the requirement of equal sizes per stage. For general textbook references, we refer to [12], [13].

A different approach was taken in [14], where a specific test with at most 3 stages was proposed and was shown to achieve the optimal expected sample size under both hypotheses, in the class of all *sequential* tests with the same error control, to a first-order asymptotic approximation as the two target error error probabilities go to 0, *but not very asymmetrically*.

More recently, a different multistage test was proposed in [15], in the context of a sparse signal detection problem. The number of stages depends on the target error probabilities, but it is finite and deterministic. Moreover, if the number of stages is selected appropriately as a function of the error probabilities, this test was shown to achieve a similar asymptotic optimality property as in [14], without any constraints on the decay rates of the two error probabilities, *but only under the null hypothesis*. Our first contribution in this work is that we revisit the test in [15] and we show that while it enjoys asymptotic optimality under the null, its expected sample size under the alternative is much larger than even that of the corresponding optimal fixed-sample-size test.

The latter result implies that the test in [15] is competitive only if one is predominantly interested in having a small sample size under the null (such as in the sparse detection setup considered in that paper). As we mentioned earlier however, there is a test in the literature that requires at most 3 stages and achieves asymptotic optimality under *both* hypotheses, as long as the two target error probabilities go to zero at certain, *not very asymmetric* rates. Our second and main contribution in this paper is that we introduce and analyze a novel multistage test, which generalizes the one in [14]. Moreover, we show that with a specific, concrete selection for the number of stages, which increases with the asymmetry of the target error probabilities, it is asymptotically optimal under both hypotheses as the two error probabilities go to zero *at arbitrary rates*.

The remainder of this work is organized as follow: In Section II we formulate the testing problem. In Section III we review the optimal fixed-sample-size test, which is the building block for the multistage tests under consideration. In Section IV we revisit the test in [15], and in Section V we introduce and analyze the proposed multistage test. In Section VI, we present two simulation studies in which we illustrate the above theoretical results.

Y. Xing and G. Fellouris are with Department of Statistics of the University of Illinois at Urbana-Champaign. Email: {yimingx4, fellouri}@illinois.edu.

This research was supported by the US National Science Foundation under grant ATD-1737962 through the University of Illinois at Urbana-Champaign.

II. Problem formulation

Let $\{X_n : n \in \mathbb{N}\}$ be a iid sequence with density f with respect to a σ-finite measure ν. We consider two simple hypotheses about f,

$$H_0 : f = f_0 \quad \text{versus} \quad H_1 : f = f_1, \tag{1}$$

where our only assumption about f_0 and f_1 throughout this work is that their Kullback-Leibler divergences are positive and finite, i.e.,

$$\begin{aligned} I_0 &\equiv \int \log(f_0/f_1) f_0 \, d\nu \in (0, \infty) \\ I_1 &\equiv \int \log(f_1/f_0) f_1 \, d\nu \in (0, \infty). \end{aligned} \tag{2}$$

We denote by P (resp. P_i) probability under f (resp. f_i), and by E (resp. E_i) expectation under P (resp. P_i), where $i = 0, 1$. Moreover, we denote by $\{\mathcal{F}_n\}$ the natural filtration of $\{X_n\}$, i.e., $\mathcal{F}_n \equiv \sigma(X_1, \ldots, X_n)$, $n \in \mathbb{N}$. We define a *test* for (1) to be a pair that consists of

- an $\{\mathcal{F}_n\}$-stopping time, T, representing the total number of utilized observations,
- and an $\mathcal{F}_T$-measurable Bernoulli random variable, D, representing the hypothesis that is accepted upon stopping.

Thus, if (T, D) is a test, the event $\{T = n, D = i\}$, on which H_i is selected after taking n observations, belongs to $\mathcal{F}_n$ for every $n \in \mathbb{N}$ and $i \in \{0, 1\}$. We denote by $\mathcal{E}$ the family of all tests, and for any α, β in $(0, 1)$ we denote by $\mathcal{E}(\alpha, \beta)$ the subfamily of tests that control the type-I and type-II error probabilities below α and β respectively, i.e.,

$$\begin{aligned} \mathcal{E}(\alpha, \beta) \equiv \{(T, D) \in \mathcal{E} : \mathsf{P}_0(D = 1) \leq \alpha \\ \text{and} \quad \mathsf{P}_1(D = 0) \leq \beta\}. \end{aligned}$$

For any $\alpha, \beta \in (0, 1)$ and $i \in \{0, 1\}$ we introduce the optimal expected sample size under H_i in $\mathcal{E}(\alpha, \beta)$:

$$\mathcal{L}_i(\alpha, \beta) \equiv \inf\{\mathsf{E}_i[T] : (T, D) \in \mathcal{E}(\alpha, \beta)\}.$$

As it was shown in [2], both $\mathcal{L}_0(\alpha, \beta)$ and $\mathcal{L}_1(\alpha, \beta)$ are attained simultaneously by the Sequential Probability Ratio Test (SPRT), i.e.,

$$\begin{aligned} T^* &\equiv \inf\{n \in \mathbb{N} : \Lambda_n \notin (-A, B)\}, \\ D^* &\equiv 1\{\Lambda_{T^*} \geq B\}, \end{aligned}$$

where Λ_n is the log-likelihood ratio of the first n observations, i.e.,

$$\Lambda_n \equiv \sum_{i=1}^{n} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \tag{3}$$

when the thresholds A and B are selected to satisfy the error constraints with equalities. Moreover, it is well known (see, e.g., [3]) that as $\alpha, \beta \to 0$, then

$$\mathcal{L}_0(\alpha, \beta) \sim \frac{|\log \beta|}{I_0}, \quad \mathcal{L}_1(\alpha, \beta) \sim \frac{|\log \alpha|}{I_1}, \tag{4}$$

where $\sim, <<$ or $>>$ in the above and later notations means that the ratio of the two quantities converges to $1, 0$ or ∞ as $\alpha, \beta \to 0$.

III. The fixed-sample-size test

For any $\alpha, \beta \in (0, 1)$ we denote by $n^*(\alpha, \beta)$ the smallest sample size n for which there exists a threshold $t \in \mathbb{R}$ so that the test that rejects H_0 if and only if the average log-likelihood ratio of the first n observations, $\bar{\Lambda}_n \equiv \Lambda_n / n$, exceeds t belongs to $\mathcal{E}(\alpha, \beta)$, i.e.,

$$\begin{aligned} n^*(\alpha, \beta) \equiv \min\{n \in \mathbb{N} : \exists\, t \in \mathbb{R} \text{ such that} \\ \mathsf{P}_0(\bar{\Lambda}_n > t) \leq \alpha \quad \text{and} \quad \mathsf{P}_1(\bar{\Lambda}_n \leq t) \leq \beta\}. \end{aligned}$$

We further denote by $c^*(\alpha, \beta)$ any of the corresponding thresholds. By the Chernoff bound we know that for every $n \in \mathbb{N}$ and $t \in [-I_0, I_1]$ we have

$$\begin{aligned} \mathsf{P}_0(\bar{\Lambda}_n > t) &\leq \exp\{-n\psi_0(t)\}, \\ \mathsf{P}_1(\bar{\Lambda}_n \leq t) &\leq \exp\{-n\psi_1(t)\}, \end{aligned} \tag{5}$$

where $\psi_1(t) \equiv \psi_0(t) - t$ and

$$\psi_0(t) \equiv \sup_{\theta \in [0,1]} \left\{\theta t - \log\left(\int f_1^\theta f_0^{1-\theta} \, d\nu\right)\right\}.$$

We also introduce the Chernoff information of f_1 and f_0:

$$\mathcal{C} \equiv -\log\left(\inf_{\theta \in [0,1]} \int f_1^\theta f_0^{1-\theta} \, d\nu\right).$$

From (4), (5), as well as the fact that $\psi_0(-I_0) = I_0$ and $\psi_1(I_1) = I_1$, we have the following implications about $n^*(\alpha, \beta)$, which we state next without proof.

Lemma III.1. *(i) For any $\alpha, \beta \in (0, 1)$ we have*

$$n^*(\alpha, \beta) \leq \frac{|\log(\alpha \wedge \beta)|}{\mathcal{C}} + 1.$$

(ii) If $\alpha, \beta \to 0$ so that $|\log \alpha| << |\log \beta|$, then

$$n^*(\alpha, \beta) \sim \frac{|\log \beta|}{I_0} \sim \mathcal{L}_0(\alpha, \beta).$$

(iii) If $\alpha, \beta \to 0$ so that $|\log \alpha| >> |\log \beta|$, then

$$n^*(\alpha, \beta) \sim \frac{|\log \alpha|}{I_1} \sim \mathcal{L}_1(\alpha, \beta).$$

IV. Generalized sequential thresholding

In this section we revisit a multistage test introduced in [15], entitled "generalized sequential thresholding". The main features of this test are that (i) it allows for accepting the null at each stage but for rejecting the null only at the last stage, and (ii) it discards all past observations at the beginning of each new stage.

To be more specific, this test consists of at most K stages and requires the specification of $2K$ parameters, $\{m_i, d_i : i \in [K]\}$, where m_i denotes the sample size of the i-th stage and d_i is the critical value used for making a decision in the i-th stage.

To describe it, we denote by Λ'_i the average log-likelihood ratio of the observations collected in the i-th stage, i.e., $\Lambda'_1 \equiv \Lambda_{m_1}/m_1$ and

$$\Lambda'_i \equiv (\Lambda_{M_i} - \Lambda_{M_{i-1}})/m_i, \quad 2 \leq i \leq K$$
$$M_i \equiv m_1 + \ldots + m_i, \quad 2 \leq i \leq K.$$

Then, assuming that $K > 2$, according to this test

- m_1 observations are initially collected and if $\Lambda'_1 \leq d_1$, H_0 is accepted.
- Otherwise, m_2 additional observations are collected and if $\Lambda'_2 \leq d_2$, H_0 is accepted.
- This is repeated for a total of at most $K-1$ times (including the previous ones), and if H_0 has not been accepted yet, then m_K additional observations are collected and H_0 is rejected if and only if $\Lambda'_k > d_K$.

The definition of the test when $K = 2$ is analogous. We denote by T' the resulting sample size and by D' the resulting decision rule. By the definition of this test it follows that , for any selection of its parameters, its probability to reject H_0 is

$$\mathsf{P}(D' = 1) = \prod_{i=1}^{K} \mathsf{P}\left(\Lambda'_i > d_i\right), \tag{6}$$

and its expected sample size is

$$\mathsf{E}[T'] = m_1 + \sum_{i=2}^{K} m_i \prod_{j=1}^{i-1} \mathsf{P}\left(\Lambda'_j > d_j\right). \tag{7}$$

From (6) we can easily then obtain design parameters that guarantee the desired error control.

Proposition IV.1. *For any $K \geq 1$ and $\alpha, \beta \in (0, 1)$ let the parameters of the test be selected as*

$$m_i = n^*(\alpha_i, \beta_i), \; d_i = c^*(\alpha_i, \beta_i), \; i \in [K], \tag{8}$$

for any $\{\alpha_i, \beta_i, i \in [K]\}$ such that

$$\prod_{i=1}^{K} \alpha_i \leq \alpha \quad \text{and} \quad \prod_{i=1}^{K} (1 - \beta_i) \geq 1 - \beta. \tag{9}$$

Then: $(T', D') \in \mathcal{E}(\alpha, \beta)$.

In view of the previous proposition, to complete the specification of this test we need to determine K and $\{\alpha_i, \beta_i, i \in [K]\}$ so that (9) is satisfied. However, since this test allows for rejecting the null only at the last stage and discards past data at the beginning of each stage, this selection depends heavily on whether the goal is to have good performance under H_0 or under H_1. In the latter case, we should clearly set $K = 1$, in which case (T', D') reduces to a fixed-sample-size test. On the other hand, with an appropriate selection of K, this test can achieve the optimal expected sample size *under the null*, $\mathcal{L}_0(\alpha, \beta)$, to a first-order asymptotic approximation as α and β go to 0 *at arbitrary rates*, when $\{\alpha_i, \beta_i\}$ are selected as follow:

$$\alpha_i = \alpha^{1/K} \quad \text{and} \quad \beta_i = (\beta/2)^i, \quad i \in [K]. \tag{10}$$

This result was shown in [15] under a second-moment assumption on the log-likelihood ratio statistic. In the next theorem we prove it under the sole assumption of the finiteness of the Kullback-Leibler divergences, (2), which is our standing assumption throughout this paper. However, at the same time we show that the corresponding expected sample size under H_1 is of larger order of magnitude compared to that of the optimal sequential test.

Theorem IV.1. *For any K, α, β, let the parameters of (T', D') be selected according to (8) and (10). Then:*

$$\mathsf{E}_0[T'] \sim \mathcal{L}_0(\alpha, \beta), \quad \mathsf{E}_1[T'] >> \mathcal{L}_1(\alpha, \beta)$$

as $\alpha, \beta \to 0$ and K is selected so that

$$\frac{|\log \alpha|}{|\log \beta|} << K << |\log \alpha|. \tag{11}$$

Proof. Note first of all that condition (9) is satisfied when $\{\alpha_i, \beta_i\}$ are selected according to (10). Note also that (11) is equivalent to

$$\alpha^{1/K} \to 0 \qquad \text{and} \qquad \frac{|\log \alpha^{1/K}|}{|\log \beta|} \to 0, \tag{12}$$

and, in view of (10), it implies that as $\alpha, \beta \to 0$, then $\alpha_i, \beta_i \to 0$ such that

$$|\log \alpha_i| << |\log \beta_i| \sim i\,|\log \beta| \quad \forall\, i \in [K].$$

Consequently, by Lemma III.1(ii) we obtain that as $\alpha, \beta \to 0$ then

$$m_i = n^*\left(\alpha_i, \beta_i\right) \sim i \frac{|\log \beta|}{I_0} \quad \forall\, i \in [K]. \tag{13}$$

From this asymptotic approximation and (7), (8), (10) we therefore have

$$\begin{aligned}
\mathsf{E}_0[T'] &= n^*\left(\alpha_1, \beta_1\right) + \sum_{i=2}^{K} n^*\left(\alpha_i, \beta_i\right)\, \alpha^{(i-1)/K} \\
&\lesssim \frac{|\log \beta|}{I_0} \left(1 + \sum_{i=2}^{K} i\, \alpha^{(i-1)/K}\right) \\
&\leq \frac{|\log \beta|}{I_0} \left(1 - \alpha^{1/K}\right)^{-2}.
\end{aligned}$$

Recalling (12) and comparing with (4), we conclude that $\mathsf{E}_0[T'] \sim \mathcal{L}_0(\alpha, \beta)$. On the other hand, from (8), (10) and (13) we have

$$\begin{aligned}
\mathsf{E}_1[T'] &\geq (1 - \beta) \sum_{i=1}^{K} n^*\left(\alpha_i, \beta_i\right) \\
&\gtrsim \frac{|\log \beta|}{I_0} \sum_{i=1}^{K} i = \frac{|\log \beta|}{I_0} \frac{K(K+1)}{2}.
\end{aligned}$$

If $|\log \beta| >> |\log \alpha|$, then this clearly implies $\mathsf{E}_1[T'] >> |\log \alpha|$. Otherwise, from (11) we have

$$\mathsf{E}_1[T'] >> |\log \beta| \left(\frac{|\log \alpha|}{|\log \beta|}\right)^2 \gtrsim |\log \alpha|.$$

Comparing with (4) then proves that $\mathsf{E}_1[T'] >> \mathcal{L}_1(\alpha, \beta)$. $\square$

V. A NOVEL MULTISTAGE TEST

In this section we introduce and analyze the proposed test, which generalizes the one in [14].

A. Description

The test in [14] requires specifying 3 positive integers n_0, n_1, N, where $n_i \leq N$, $i = 0, 1$, and three real thresholds, t_0, t_1, c. Then, it stops after collecting n_0 observations and $\bar{\Lambda}_{n_0} \leq t_0$, in which case it accepts H_0, or after collecting n_1 observations and $\bar{\Lambda}_{n_1} > t_1$, in which case it rejects H_0. If none of these happens, then N observations are collected in total and H_0 is rejected if and only if $\bar{\Lambda}_N > c$.

The proposed test adds M_0 opportunities to accept H_0 and M_1 opportunities to reject it. Thus, to describe it we need to specify $2M_0 + 2M_1$ additional parameters,

$$\{N_{i,j}, c_{ij} : j \in [M_i],\ i \in \{0, 1\}\},$$

where $n_i \leq N_{i,j} \leq N$ for every $j \in [M_i], i \in \{0, 1\}$. Then, in addition to the stopping rule of the 3-stage test,

- H_0 is accepted after collecting $N_{0,j}$ observations if $\bar{\Lambda}_{N_{0,j}} \leq c_{0,j}$, where $j \in [M_0]$,
- H_0 is rejected after collecting $N_{1,j}$ observations if $\bar{\Lambda}_{N_{1,j}} > c_{0,j}$, where $j \in [M_1]$.

In what follows, we denote by $(\hat{T}, \hat{D})$ the sample size and decision rule of this test.

B. Error bounds and expected sample size

For any selection of the above parameters, the type-I error probability of $(\hat{T}, \hat{D})$ can be upper bounded as follows:

$$\begin{aligned}\mathsf{P}_0(\hat{D} = 1) \leq \mathsf{P}_0\left(\bar{\Lambda}_{n_1} > t_1\right) + \mathsf{P}_0\left(\bar{\Lambda}_N > c\right) \\ + \sum_{i=1}^{M_1} \mathsf{P}_0\left(\bar{\Lambda}_{N_{1,i}} > c_{1,i}\right)\end{aligned} \tag{14}$$

and its expected sample size under P_0 can be bounded as follows:

$$\begin{aligned}\mathsf{E}_0[\hat{T}] \leq\ & n_0 + N_{0,1}\,\mathsf{P}_0(\bar{\Lambda}_{n_0} > t_0) \\ & + \sum_{i=2}^{M_0} N_{0,i}\,\mathsf{P}_0(\bar{\Lambda}_{N_{0,i-1}} > c_{0,i-1}) \\ & + N\,\mathsf{P}_0(\bar{\Lambda}_{N_{0,M_0}} > c_{0,M_0}).\end{aligned} \tag{15}$$

Similarly for the type-II error probability and the expected sample size under P_1:

$$\begin{aligned}\mathsf{P}_1(\hat{D} = 0) \leq \mathsf{P}_1\left(\bar{\Lambda}_{n_0} \leq t_0\right) + \mathsf{P}_1\left(\bar{\Lambda}_N \leq c\right) \\ + \sum_{i=1}^{M_0} \mathsf{P}_1\left(\bar{\Lambda}_{N_{0,i}} \leq c_{0,i}\right)\end{aligned} \tag{16}$$

$$\begin{aligned}\mathsf{E}_1[\hat{T}] \leq\ & n_1 + N_{1,1}\,\mathsf{P}_1(\bar{\Lambda}_{n_1} \leq t_1) \\ & + \sum_{i=2}^{M_1} N_{1,i}\,\mathsf{P}_1(\bar{\Lambda}_{N_{1,i-1}} \leq c_{1,i-1}) \\ & + N\,\mathsf{P}_1(\bar{\Lambda}_{N_{1,M_1}} \leq c_{1,M_1}),\end{aligned} \tag{17}$$

C. Selection of parameters

Let the sample sizes and thresholds in the first opportunity to accept/reject H_0 be selected as follows:

$$\begin{aligned}& n_0 = n^*(\gamma, \beta), \quad t_0 = c^*(\gamma, \beta) \\ & n_1 = n^*(\alpha, \delta), \quad t_1 = c^*(\alpha, \delta) \\ & \text{for some } \gamma \in (\alpha, 1) \quad \text{and} \quad \delta \in (\beta, 1),\end{aligned} \tag{18}$$

and let the sample size and threshold in the last stage of the test be selected as follows:

$$N = n^*(\alpha, \beta), \quad c = c^*(\alpha, \beta). \tag{19}$$

Moreover, let M_0 and M_1 be selected as follows:

$$\begin{aligned}M_0 \equiv \max\{i \geq 0 : n^*(\beta^i, \beta^i) < n^*(\alpha, \beta)\}, \\ M_1 \equiv \max\{i \geq 0 : n^*(\alpha^i, \alpha^i) < n^*(\alpha, \beta)\},\end{aligned} \tag{20}$$

and note that at most one of them is non-zero. If $M_0 > 0$, let the sample sizes and thresholds in the additional M_0 opportunities to accept H_0 be selected as

$$N_{0,i} = n^*\left(\beta^i, \beta^i\right),\ c_{0,i} = c^*\left(\beta^i, \beta^i\right), \tag{21}$$

for $i \in [M_0]$. Similarly, if $M_1 > 0$, let the sample sizes and thresholds in the additional M_1 opportunities to reject H_0 be selected as

$$N_{1,i} = n^*\left(\alpha^i, \alpha^i\right),\ c_{1,i} = c^*\left(\alpha^i, \alpha^i\right), \tag{22}$$

for $i \in [M_1]$. Then, by (14) and (16) we conclude

$$\mathsf{P}_0(\hat{D} = 1) \leq 4\alpha \quad \text{and} \quad \mathsf{P}_1(\hat{D} = 0) \leq 4\beta$$

when $\alpha, \beta \leq 0.5$, which leads to the following proposition.

Proposition V.1. *If the parameters of* $(\hat{T}, \hat{D})$ *are selected according to* (18)-(22) *with* α *and* β *replaced by* $\alpha/4$ *and* $\beta/4$ *respectively, then* $(\hat{T}, \hat{D}) \in \mathcal{E}(\alpha, \beta)$.

The parameter specification of Proposition V.1 first of all provides a concrete expression for the number of stages of the test, which is a function of α and β. Specifically, when $\alpha = \beta$, we have $M_0 = M_1 = 0$, in which case we recover the 3-stage test in [14]. On the other hand, at most one of M_0 and M_1 is zero when α and β differ, which means that, in comparison to the 3-stage, we either add M_0 opportunities to accept H_0 or M_1 opportunities to reject it.

The remaining parameters are specified completely up to γ and δ in (18). We propose selecting γ to minimize the upper bound in (15) and δ to minimize the upper bound in (17). This specification is convenient from a practical point of view, as it requires the minimization of two deterministic functions, each with respect to a single parameter. This choice however is not necessary for achieving asymptotic optimality. Indeed, in the next theorem we show that, under quite general conditions on γ and δ,

$(\hat{T}, \hat{D})$ achieves asymptotic optimality under both hypotheses as α and β go to zero at arbitrary rates.

Theorem V.1. *Let the parameters of $(\hat{T}, \hat{D})$ be selected according to (18)-(22) with α and β replaced by $\alpha/4$ and $\beta/4$ respectively.*

(i) If $\gamma \to 0$ as $\beta \to 0$ such that

$$|\log \gamma| << |\log \beta|, \tag{23}$$

then $\mathsf{E}_0[\hat{T}] \sim \mathcal{L}_0(\alpha, \beta)$ as $\alpha, \beta \to 0$.

(ii) If $\delta \to 0$ as $\alpha \to 0$ such that

$$|\log \delta| << |\log \alpha|, \tag{24}$$

then $\mathsf{E}_1[\hat{T}] \sim \mathcal{L}_1(\alpha, \beta)$ as $\alpha, \beta \to 0$.

Proof. We only prove (i), as the proof of (ii) is analogous. By the assumption on the decay rate of γ, Lemma III.1(i)(ii) and (15) we obtain:

$$\begin{aligned}\mathsf{E}_0[\hat{T}] \lesssim & \frac{|\log \beta|}{I_0} + \gamma \frac{|\log \beta|}{\mathcal{C}} + \sum_{i=2}^{M_0} \left(\frac{\beta}{4}\right)^{i-1} i \frac{|\log \beta|}{\mathcal{C}} \\ & + \left(\frac{\beta}{4}\right)^{M_0} (M_0 + 1) \frac{|\log \beta|}{\mathcal{C}} \\ \leq & \frac{|\log \beta|}{I_0} \left(1 + \frac{I_0}{\mathcal{C}} \left(\gamma + \sum_{i=2}^{\infty} i \left(\frac{\beta}{4}\right)^{i-1}\right)\right).\end{aligned}$$

The series in the parenthesis goes to 0 as $\beta \to 0$ and comparing with (4) completes the proof. $\square$

As a final remark, we note that the maximum sample sizes of the proposed test, $n^*(\alpha/4, \beta/4)$, exceeds $n^*(\alpha, \beta)$ by a constant that does not depend on α and β. As a result, for small enough α and β, it will have much better expected sample size than even the SPRT when the true distribution is "between" the null and the alternative [3, p. 227].

VI. Numerical Studies

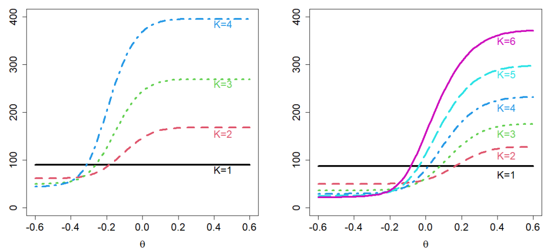
We consider the problem of testing the mean of a Gaussian sequence with unit variance, i.e., $f \in \{\mathrm{N}(\theta, 1), \theta \in \mathbb{R}\}$, with $f_0 = \mathrm{N}(-0.5, 1)$, $f_1 = \mathrm{N}(0.5, 1)$. Then, we have an explicit formula for the optimal fixed-sample-size test,

$$n^*(\alpha, \beta) = (z_\alpha + z_\beta)^2, \; c^*(\alpha, \beta) = \frac{z_\alpha - z_\beta}{2(z_\alpha + z_\beta)},$$

where z_α is the upper α-quantile of the standard Gaussian. We consider two cases for α and β, a completely symmetric and a quite asymmetric one: (i) $\alpha = \beta = 10^{-6}$ and (ii) $\alpha = 10^{-12}$, $\beta = 10^{-2}$, so that the optimal fixed-sample-size $n^*(\alpha, \beta)$ is equal to 90 and 88 respectively. In each of these two setups, we apply (a) the SPRT with thresholds $A = |\log \beta|$ and $B = |\log \alpha|$, (b) the test in Section IV, with parameters given by (8), (10), and different values of K, (c) the test in Section V with the thresholds described in Proposition V.1. In particular, in setup (i) we have $M_0 = M_1 = 0$ and the resulting test has at most 3 stages, whereas in setup (ii) we have $M_0 = 2$, $M_1 = 0$ and the resulting test has at most 5 stages. Moreover, we select the free parameters γ, δ, as suggested in Section IV, to minimize the upper bound in (15) and (17) respectively.

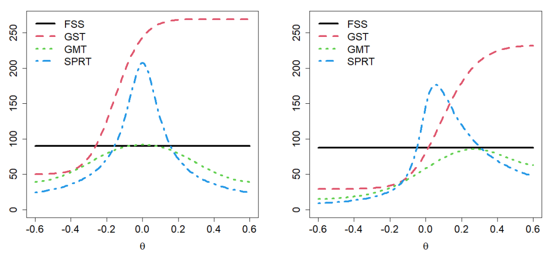
We evaluate the expected sample size of each of the above test, for each of the two cases (i) and (ii), when the true mean θ is between $(-0.6, 0.6)$. In Figure 1 we plot the expected sample size of the test in Section IV against the true mean for different values of K. In Figure 2 we plot the corresponding results for the SPRT, the proposed test, and the test in Section IV with the same number of stages.

In Figure 1 we see that the test in Section IV decreases/increases the expected sample under the null/alternative relative to the fixed-sample-size test. Moreover, as K increases, the additional decrease is minimal, whereas the increase is dramatic. Nevertheless, from Figure 2 we can see that even under the null the proposed test performs slighly better, at least when the number of stages is the same. From Figure 2 we can also see that the expected sample size of the proposed test is slightly larger than that of the SPRT under the two hypotheses and much smaller when the true mean is around 0, i.e., between the two hypotheses.



(a) (b)

Fig. 1: Expected sample size against the true mean for the test in Section IV with different K's.



(a) $\alpha = \beta = 10^{-6}$ (b) $\alpha = 10^{-12}$, $\beta = 10^{-2}$

Fig. 2: ESS of all four tests against θ. GST is the test in Section IV and GMT is the test in Section V. The number of stages in them are equal.

REFERENCES

[1] A. Wald, *Sequential Analysis*. New York: John Wiley & Sons, 1947.

[2] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Mathematical Statistics*, vol. 19, pp. 326–339, 1948.

[3] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, 1st ed. Chapman & Hall/CRC, 2014.

[4] P. Armitage, C. K. McPherson, and B. C. Rowe, "Repeated significance tests on accumulating data," *Journal of the Royal Statistical Society. Series A (General)*, vol. 132, no. 2, pp. 235–244, 1969. [Online]. Available: http://www.jstor.org/stable/2343787

[5] S. J. Pocock, "Group sequential methods in the design and analysis of clinical trials," *Biometrika*, vol. 64, no. 2, pp. 191–199, 1977. [Online]. Available: http://www.jstor.org/stable/2335684

[6] ——, "Interim analyses for randomized clinical trials: The group sequential approach," *Biometrics*, vol. 38, no. 1, pp. 153–162, 1982. [Online]. Available: http://www.jstor.org/stable/2530298

[7] P. C. O'Brien and T. R. Fleming, "A multiple testing procedure for clinical trials," *Biometrics*, vol. 35, no. 3, pp. 549–556, 1979. [Online]. Available: http://www.jstor.org/stable/2530245

[8] S. K. Wang and A. A. Tsiatis, "Approximately optimal one-parameter boundaries for group sequential trials," *Biometrics*, vol. 43, no. 1, pp. 193–199, 1987. [Online]. Available: http://www.jstor.org/stable/2531959

[9] K. K. G. Lan and D. L. DeMets, "Discrete sequential boundaries for clinical trials," *Biometrika*, vol. 70, no. 3, pp. 659–663, 1983. [Online]. Available: http://www.jstor.org/stable/2336502

[10] C. Jennison, "Efficient group sequential tests with unpredictable group sizes," *Biometrika*, vol. 74, no. 1, pp. 155–165, 1987. [Online]. Available: http://www.jstor.org/stable/2336030

[11] J. D. Eales and C. Jennison, "An improved method for deriving optimal one-sided group sequential tests," *Biometrika*, vol. 79, no. 1, pp. 13–24, 1992. [Online]. Available: http://www.jstor.org/stable/2337143

[12] C. Jennison and B. W. Turnbull, *Group sequential methods with applications to clinical trials*. CRC Press, 1999.

[13] J. Bartroff, T. L. Lai, and M.-C. Shih, "Sequential experimentation in clinical trials: Design and analysis," 2012.

[14] G. Lorden, "Asymptotic efficiency of three-stage hypothesis tests," *Annals of Statistics*, vol. 11, pp. 129–140, 1983.

[15] M. L. Malloy and R. D. Nowak, "Sequential testing for sparse recovery," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7862–7873, 2014.