Statistica Sinica Preprint No: SS-2022-0235	
Title	Asymptotically Optimal Multistage Tests for Non-iid Data
Manuscript ID	SS-2022-0235
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0235
Complete List of Authors	Yiming Xing and Georgios Fellouris
Corresponding Authors	Georgios Fellouris
E-mails	fellouri@illinois.edu

Statistica Sinica

ASYMPTOTICALLY OPTIMAL MULTISTAGE TESTS FOR NON-I.I.D. DATA

Yiming Xing and Georgios Fellouris

University of Illinois at Urbana-Champaign

Abstract: Given a fixed-sample-size test that controls the error probabilities under two specific arbitrary distributions, we propose and analyze a 3-stage test and two 4-stage tests. For each test, we specify a novel, concrete, non-conservative design, and establish a first-order asymptotic approximation for the expected sample size under the two prescribed distributions as the error probabilities go to zero. As a corollary, we show that the proposed multistage tests can asymptotically achieve the optimal expected sample size under these two distributions in the class of all sequential tests with the same error control. Furthermore, the tests are shown to be more robust than Wald's sequential probability ratio test when applied to one-sided testing problems and the error probabilities under control are small. We apply these general results to testing problems in the independent and identically distributed setup and beyond, such as testing the correlation coefficient of a first-order autoregressive process or testing the transition matrix of a finite-state Markov chain, and illustrate them in various numerical studies.

Key words and phrases: asymmetric error probabilities, asymptotic optimality, group-sequential tests, large-deviation theory, multistage tests, sequential testing

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

1. Introduction

A typical motivation for employing a *sequential* test, that is, a testing procedure with a sample size that depends on the collected observations, is that its average sample size can be much smaller than that of the corresponding *fixed-sample-size* test. One of the first tests of this kind was the double sampling procedure in Dodge and Romig (1929), a precursor to the sequential probability ratio test (SPRT) and the field of "sequential analysis" (Wald (1947)). However, implementing the SPRT and most sequential tests in the literature (see, e.g., Tartakovsky et al. (2014)) requires continuous monitoring of the data collection process. This is often inconvenient, if not infeasible, in application areas such as sampling inspection (Dodge and Romig (1929)), clinical trials (Jennison and Turnbull (1999), Bartroff et al. (2012)), and educational assessment (Wang et al. (2016)). As a result, such applications focus on multistage tests, also known as group-sequential tests, where the implementation requires collecting only a small number of groups of samples.

Works on multistage tests, such as Armitage et al. (1969), Pocock (1977), O'Brien and Fleming (1979), Pocock (1982), Wang and Tsiatis (1987), Emerson and Fleming (1989), Eales and Jennison (1992), Pampallona and Tsiatis (1994), Barber and Jennison (2002), typically focus on testing the mean of independent and identically distributed (i.i.d.) Gaussian observations with a known variance, are designed to control predetermined type-I and type-II error probabilities under two specific distributions, and require equal stage sizes. Free parameters, if any, as in Wang and Tsiatis (1987), are selected to optimize the expected sample size under a certain distribution, for example, the one under which the type-II error probability is controlled. This optimization is performed using dynamic programming in Eales and Jennison (1992) and Barber and Jennison (2002).

Multistage tests with unequal and random stage sizes have been considered by Lan and DeMets (1983), Kim and DeMets (1987), Jennison (1987), and Lai and Shih (2004). The latter work also studies more general testing problems related to the parameters of an exponential family.

In all of the aforementioned works, the stage sizes are treated as user-specified inputs. However, Lorden (1983) showed that a 3-stage test with properly designed stages can achieve the optimal expected sample size under both hypotheses among all sequential tests with the same or smaller error probabilities, asymptotically as the latter go to zero. In the case of two simple hypotheses for i.i.d. data, this was shown for multistage tests with *deterministic* stage sizes (Section 2 of Lorden (1983)). In the case of composite hypotheses for the one-sided testing problem in a one-parameter exponential family, this was shown for multistage tests with *adaptive* stage sizes, that is, they can depend on data from previous stages (Section 3 of Lorden (1983)). Such multistage tests are also considered in Bartroff and Lai (2008a;b), who propose a less conservative design. These asymptotic optimality results all require certain assumptions on the decay rates of the prescribed error probabilities, which are not allowed to go to zero very asymmetrically.

In the present work, we focus on the design and analysis of multistage tests with *deterministic* stage sizes, and strengthen, extend, and generalize the results in Section 2 of Lorden (1983). First, unlike the previously mentioned works, we do not require i.i.d. observations. Instead, we assume a fixed-sample-size test is given that can control the type-I and type-II error probabilities under two specific distributions below arbitrary levels. Given this, we introduce and analyze a *3-stage test*, that generalizes the one in Section 2 of Lorden (1983), and two novel *4-stage tests*. For each test, we propose a concrete design that guarantees non-asymptotic and non-conservative error control. The designs require knowledge of the number of observations and the threshold the fixed-sample-size test requires in order to achieve certain error control. While there are not, in general, explicit formulae for these quantities, they can be estimated via simulation. For this task, we also propose an efficient importance sampling approach, which is necessary in the case of small error probabilities, where a plain Monte Carlo approach may be inefficient or even infeasible (see, e.g., Bucklew (2010)).

In order to obtain theoretical insights about the proposed multistage tests, we impose some structure on the above general setup. Specifically, we assume that there exist thresholds for which the error probabilities of the fixed-sample-size test under the two prescribed distributions decay exponentially fast in the sample size. Based on this assumption, we establish first-order asymptotic approximations for the expected sample sizes of the proposed multistage tests under the distributions where the error probabilities are controlled, as the latter go to zero. For the 3-stage test, the relative decay of the error probabilities is allowed to be much more asymmetric than the one required in Section 2 of Lorden (1983). Even more asymmetric decay is allowed for the two 4-stage tests.

When the given fixed-sample-size test is the likelihood ratio test, the proposed multistage tests are shown, similarly to the SPRT, to achieve the optimal expected sample size under the two prescribed distributions in the family of all sequential tests with the same or smaller error probabilities, to a first-order asymptotic approximation as the latter go to zero. The difference is that the asymptotic optimality of the multistage tests, unlike

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

that of the SPRT, requires certain restrictions on how asymmetrically the error probabilities decay (which are less strict for the 4-stage tests than for the 3-stage test).

In order to obtain a more complete picture for this comparison, we establish a *distribution-free* asymptotic upper bound on the expected sample sizes of the proposed multistage tests as at least one of the two prescribed error probabilities goes to zero. This reveals that, when the prescribed error probabilities are small, these multistage tests are much more robust than the SPRT, whose expected sample size can be inflated when the true distribution is "between" the prescribed ones (see, e.g., Bechhofer (1960)).

We illustrate the proposed methodology and the above asymptotic results in numerical studies for various testing problems. Indeed, the distributional assumptions for our asymptotic analysis can be shown to hold, using the Gärtner–Ellis theorem from large deviation theory (see, e.g., Dembo and Zeitouni (1998)), for various testing problems beyond the i.i.d. setup. Two specific examples, used in our numerical studies, are testing the correlation coefficient of a first-order autoregression series and testing the transition matrix of an irreducible and recurrent finite-state Markov chain.

The remainder of this paper is organized as follows. In Section 2, we formulate the testing setup. In Section 3, we introduce and design the proposed multistage tests. In Section 4, we establish our asymptotic theory. In Section 5, we conclude and discuss potential extensions. In Section S1 of the Supplementary Material, we state sufficient conditions for the asymptotic analysis of Section 4. In Section S2, we develop an importance sampling approach for the efficient implementation of the proposed design when the error probabilities are small. In Section S3, we illustrate the general theory in three specific testing problems. In Section S4, we present our numerical studies. All proofs are presented in Section S5.

Next, we introduce some notations. We denote by $\mathbb{N}$ the set of positive integers, that is, $\mathbb{N} \equiv \{1, 2, \ldots\}$, and by $\mathbb{R}$ the set of real numbers. For $x, y \in \mathbb{R}$, we set $x \wedge y \equiv \min\{x, y\}$ and $x \vee y \equiv \max\{x, y\}$. For positive sequences $(x_n), (y_n)$, we write $x_n \sim y_n$ for $\lim_n (x_n/y_n) = 1$, $x_n \gtrsim y_n$ for $\underline{\lim}_n\, (x_n/y_n) \geq 1$, $x_n \lesssim y_n$ for $\overline{\lim}_n\, (x_n/y_n) \leq 1$, $x_n \ll y_n$ for $x_n/y_n \to 0$, and $x_n \gg y_n$ for $x_n/y_n \to \infty$.

2. Problem formulation

We consider a sequence of $\mathbb{S}$-valued random elements, $X \equiv \{X_n, n \in \mathbb{N}\}$, where $(\mathbb{S}, \mathcal{S})$ is an arbitrary measurable space. For any $n \in \mathbb{N}$, we denote by $\mathcal{F}_n$ the σ-algebra generated by the first n terms of this sequence, that is, $\mathcal{F}_n \equiv \sigma(X_1, \ldots, X_n)$. We denote by P the distribution of X, assume that it belongs to some family, $\mathcal{P}$, and consider the following hypothesis testing problem:

$$H_0: \mathsf{P} \in \mathcal{P}_0 \qquad \text{versus} \qquad H_1: \mathsf{P} \in \mathcal{P}_1, \tag{2.1}$$

where $\mathcal{P}_0$ and $\mathcal{P}_1$ are disjoint subsets of $\mathcal{P}$.

We assume that the data can be collected sequentially, and that it is possible to decide after each observation whether or not to stop sampling. Thus, if τ is the total sample size of a testing procedure and d is its decision, with H_i being selected when $d = i$ for $i \in \{0, 1\}$, we say that $\chi \equiv (\tau, d)$ is a *test* for (2.1) if τ is a stopping time with respect to the filtration $\{\mathcal{F}_n,\, n \in \mathbb{N}\}$ and d is an $\mathcal{F}_\tau$-measurable Bernoulli random variable, that is, $\{\tau = n\}$, $\{\tau = n,\, d = i\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$ and $i \in \{0, 1\}$.

We refer to a test as a *fixed-sample-size* test if τ is deterministic and as a *multistage* test if τ can take a small number of values. We denote by $\mathcal{C}$ the family of all tests, and we further introduce a subfamily of tests that control the two types of error probabilities under two specific, but arbitrary, distributions. Specifically, we fix $\mathsf{P}_0 \in \mathcal{P}_0$ and $\mathsf{P}_1 \in \mathcal{P}_1$, and, for any $\alpha, \beta \in (0, 1)$, we denote by $\mathcal{C}(\alpha, \beta)$ the family of tests whose type-I error probability under P_0 does not exceed α and whose type-II probability under P_1 does not exceed β, that is,

$$\mathcal{C}(\alpha, \beta) \equiv \{(\tau, d) \in \mathcal{C} :\ \mathsf{P}_0(d = 1) \leq \alpha \quad \text{and} \quad \mathsf{P}_1(d = 0) \leq \beta\}. \tag{2.2}$$

For each $i \in \{0,1\}$, we denote by E_i the expectation under P_i, and by $\mathcal{L}_i(\alpha,\beta)$ the optimal expected sample size in $\mathcal{C}(\alpha,\beta)$ under P_i, that is,

$$\mathcal{L}_i(\alpha,\beta) \equiv \inf\left\{\mathsf{E}_i[\tau] : (\tau,d) \in \mathcal{C}(\alpha,\beta)\right\}. \tag{2.3}$$

First, we aim to introduce 3-stage and 4-stage tests with deterministic stage sizes that can be designed to belong to $\mathcal{C}(\alpha,\beta)$ for any given α, $\beta \in (0,1)$. For this design, we only require the existence of a fixed-sample-size test that can provide such error guarantees. Thus, our only standing assumption throughout this paper is that there is a sequence of test statistics, $T \equiv \{T_n,\, n \in \mathbb{N}\}$, such that, for every $n \in \mathbb{N}$, T_n is $\mathcal{F}_n$-measurable and, *for any $\alpha,\beta \in (0,1)$, there exist $n \in \mathbb{N}$ and $\kappa \in \mathbb{R}$ so that the fixed-sample-size test that rejects H_0 if and only if $T_n > \kappa$ belongs to $\mathcal{C}(\alpha,\beta)$.* Suppressing the dependence on T, we denote by $n^*(\alpha,\beta)$ the smallest such sample size, that is,

$$\begin{aligned} n^*(\alpha,\beta) \equiv \min\big\{n \in \mathbb{N} : \exists\, \kappa \in \mathbb{R} \;\text{ so that }\; & \mathsf{P}_0(T_n > \kappa) \leq \alpha \\ \text{and }\; & \mathsf{P}_1(T_n \leq \kappa) \leq \beta\big\}, \end{aligned} \tag{2.4}$$

and by $\kappa^*(\alpha,\beta)$ any of the corresponding thresholds. In Section S2 of the Supplementary Material, we discuss the computation of these quantities in practice when they do not admit closed-form expressions.

Second, we aim to show that, when the test statistic T is selected appropriately, the proposed multistage tests achieve the optimal expected sample

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

size in $\mathcal{C}(\alpha, \beta)$ under both P_0 and P_1, that is, $\mathcal{L}_0(\alpha, \beta)$ and $\mathcal{L}_1(\alpha, \beta)$, to a first-order asymptotic approximation as $\alpha, \beta \to 0$. For this asymptotic optimality result, we need some additional distributional assumptions, which we state in Section 4.

We end this section by illustrating the above testing formulation using the generic one-sided testing problem, which we use in all our examples and numerical studies in Sections S3–S4 of the Supplementary Material.

2.1 The one-sided testing problem

Suppose that the family of plausible distributions, $\mathcal{P}$, is parametrized by a scalar parameter, μ, taking values in an open interval $M \subseteq \mathbb{R}$. That is, if we denote by $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ the distribution and expectation, respectively, of X when the true parameter is μ, then $\mathcal{P} = \{\mathbb{P}_\mu : \mu \in M\}$. Moreover, suppose the testing problem of interest is whether the true parameter μ is smaller or larger than some user-specified value, $\mu_* \in M$, that is,

$$H_0 : \mu < \mu_* \qquad \text{versus} \qquad H_0 : \mu > \mu_*, \tag{2.5}$$

or equivalently, $\mathcal{P}_0 = \{\mathbb{P}_\mu : \mu < \mu_*\}$ and $\mathcal{P}_1 = \{\mathbb{P}_\mu : \mu > \mu_*\}$. Suppose further that the type-I error probability must be controlled below α when $\mu = \mu_0$, and the type-II error probability must be below β when $\mu = \mu_1$, where $\alpha, \beta \in (0, 1)$ and $\mu_0, \mu_1 \in M$, $\mu_0 < \mu_* < \mu_1$. Then, this is a special

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

case of the framework introduced in this section, with $\mathsf{P}_i = \mathbb{P}_{\mu_i}$, $i \in \{0, 1\}$.

Remark 1. In the context of the above one-sided testing problem, a test $\chi \equiv (\tau, d)$ in $\mathcal{C}(\alpha, \beta)$ should ideally control the type-I error probability below α for *every* $\mu \leq \mu_0$ and the type-II error probability below β for *every* $\mu \geq \mu_1$, that is,

$$\begin{aligned} &\mathbb{P}_\mu(d = 1) \leq \alpha \text{ for every } \mu \leq \mu_0 \\ &\mathbb{P}_\mu(d = 0) \leq \beta \text{ for every } \mu \geq \mu_1. \end{aligned} \tag{2.6}$$

This is obviously the case for the fixed-sample-size test that rejects H_0 if and only if $T_n > \kappa$ when

$$\mathbb{P}_{\mu_0}(T_n > \kappa) = \sup_{\mu \leq \mu_0} \mathbb{P}_\mu(T_n > \kappa), \quad \mathbb{P}_{\mu_1}(T_n \leq \kappa) = \sup_{\mu \geq \mu_1} \mathbb{P}_\mu(T_n \leq \kappa). \tag{2.7}$$

If the monotonicity property (2.7) holds for every $n \in \mathbb{N}$ and $\kappa \in \mathbb{R}$, then the uniform error control in (2.6) will also hold for the proposed multistage tests in this work.

3. The multistage tests

In this section, we introduce and analyze the multistage tests that we consider in this work.

3.1 The 3-stage test

We first introduce and analyze a test that offers two opportunities to accept the null hypothesis and two to reject it. Its implementation requires the specification of three positive integers, n_0, n_1, N, and three real thresholds, κ_0, κ_1, K, so that

$$n_0 \vee n_1 \leq N \quad \text{and} \quad \kappa_0 \leq \kappa_1 \ \text{ if } \ n_0 = n_1. \tag{3.1}$$

Specifically, n_0 (resp. n_1) is the number of observations collected by the first opportunity to accept (resp. reject) H_0, and N is the maximum number of observations that can be collected.

Given these parameters, the test proceeds as follows:

(i) $n_0 \wedge n_1$ observations are initially collected.

- If $n_0 \leq n_1$ and $T_{n_0} \leq \kappa_0$, then H_0 is accepted.
- If $n_1 \leq n_0$ and $T_{n_1} > \kappa_1$, then H_0 is rejected.

(ii) If no decision has been reached yet, $(n_0 \vee n_1) - (n_0 \wedge n_1)$ additional observations are collected.

- If $n_0 \leq n_1$ and $T_{n_1} > \kappa_1$, then H_0 is rejected.
- If $n_1 \leq n_0$ and $T_{n_0} \leq \kappa_0$, then H_0 is accepted.

(iii) If no decision has been reached yet, $N - (n_0 \vee n_1)$ additional observations are collected, and H_0 is rejected if and only if $T_N > K$.

To avoid a possible overlap between acceptance and rejection regions when $n_0 \vee n_1 = N$, we include the convention that whenever the test reaches its final stage, the only effective threshold is K.

This testing procedure can be implemented by collecting at most three samples of deterministic sizes. In what follows, we refer to it as the *3-stage test* and denote it by $\tilde{\chi} \equiv (\tilde{\tau}, \tilde{d})$.

Remark 2. This test was first proposed in Section 2 of Lorden (1983), where X is an i.i.d. sequence and the test statistic, T, is the average log-likelihood ratio between P_1 and P_0. Our setup here is essentially universal, because the only assumption in this section about X and T is that the corresponding fixed-sample-size test can control the error probabilities below arbitrary, user-specified levels, that is, that $n^*(\alpha, \beta)$ is finite for every $\alpha, \beta \in (0, 1)$. Moreover, we next propose a concrete and non-asymptotic specification of the design parameters, which is novel and practically useful, even in the specific setup of Section 2 of Lorden (1983).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

3.1.1 Error control

By the definition of the 3-stage test, it follows that, for any selection of its parameters and any $\mathsf{P} \in \mathcal{P}$,

$$\mathsf{P}(\tilde{d} = 1) \leq \mathsf{P}\left(T_{n_1} > \kappa_1\right) + \mathsf{P}\left(T_N > K\right), \tag{3.2}$$

$$\mathsf{P}(\tilde{d} = 0) \leq \mathsf{P}\left(T_{n_0} \leq \kappa_0\right) + \mathsf{P}\left(T_N \leq K\right). \tag{3.3}$$

Consequently, if the sample size and the threshold are

$$n_0 = n^*(\gamma, \beta) \quad \text{and} \quad \kappa_0 = \kappa^*(\gamma, \beta) \quad \text{for some} \quad \gamma \in [\alpha \vee \beta, 1) \tag{3.4}$$

in the first opportunity to accept H_0,

$$n_1 = n^*(\alpha, \delta) \quad \text{and} \quad \kappa_1 = \kappa^*(\alpha, \delta) \quad \text{for some} \quad \delta \in [\alpha \vee \beta, 1) \tag{3.5}$$

in the first opportunity to reject H_0, and

$$N = n^*(\alpha, \beta) \quad \text{and} \quad K = \kappa^*(\alpha, \beta) \tag{3.6}$$

in the final stage, then by (3.2) with $\mathsf{P} = \mathsf{P}_0$, and by (3.3) with $\mathsf{P} = \mathsf{P}_1$, we have $\mathsf{P}_0(\tilde{d} = 1) \leq 2\alpha$ and $\mathsf{P}_1(\tilde{d} = 0) \leq 2\beta$. This observation leads to the following theorem.

Theorem 1. *Let $\alpha, \beta \in (0, 1)$. If the design parameters are selected according to (3.4)–(3.6), with α and β replaced by $\alpha/2$ and $\beta/2$, respectively, then (3.1) is satisfied and $\tilde{\chi} \in \mathcal{C}(\alpha, \beta)$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

Proof. Condition (3.1) can be verified using the following straightforward observations:

$$n^*(\alpha_1, \beta_1) \le n^*(\alpha_2, \beta_2) \quad \text{if} \quad \alpha_1 \ge \alpha_2 \ \text{ and } \ \beta_1 \ge \beta_2, \tag{3.7}$$

$$\kappa^*(\alpha_1, \beta_1) \le \kappa^*(\alpha_2, \beta_2) \quad \text{if} \quad n^*(\alpha_1, \beta_1) = n^*(\alpha_2, \beta_2) \ \text{ and } \ \alpha_1 \ge \alpha_2, \ \beta_1 \le \beta_2.$$

The proof that $\tilde{\chi} \in \mathcal{C}(\alpha, \beta)$ follows by the discussion preceding this theorem.

□

Theorem 1 specifies a design for $\tilde{\chi} \in \mathcal{C}(\alpha, \beta)$ up to two *free* parameters, $\gamma, \delta \in [(\alpha \vee \beta)/2, 1)$. Increasing the value of γ (resp. δ) reduces the number of observations until the first opportunity to accept (resp. reject) H_0, but increases the probability of continuing to the final stage. To solve this trade-off, in Subsection (3.1.3), we propose selecting γ (resp. δ) to minimize *an upper bound on* $\mathsf{E}_0[\tilde{\tau}]$ *(resp.* $\mathsf{E}_1[\tilde{\tau}]$*) that is independent of* δ *(resp.* γ*).*

3.1.2 The expected sample size

By the definition of the 3-stage test, it follows that, for any $\mathsf{P} \in \mathcal{P}$,

- if $n_0 \le n_1 < N$, then

$$\mathsf{E}[\tilde{\tau}] = n_0 + (n_1 - n_0) \cdot \mathsf{P}\left(T_{n_0} > \kappa_0\right) + (N - n_1) \cdot \mathsf{P}\begin{pmatrix} T_{n_0} > \kappa_0 \\ T_{n_1} \le \kappa_1 \end{pmatrix}, \tag{3.8}$$

- if $n_1 \leq n_0 < N$, then

$$\mathsf{E}[\tilde{\tau}] = n_1 + (n_0 - n_1) \cdot \mathsf{P}\left(T_{n_1} \leq \kappa_1\right) + (N - n_0) \cdot \mathsf{P}\begin{pmatrix} T_{n_1} \leq \kappa_1 \\ T_{n_0} > \kappa_0 \end{pmatrix}, \quad (3.9)$$

where E is the expectation under P.

Applying to these identities the inequality

$$\max\{0, \mathsf{P}(A) - \mathsf{P}(B^c)\} \leq \mathsf{P}(A \cap B) \leq \mathsf{P}(A),$$

we obtain, for any selection of the design parameters, the following bounds:

$$\begin{aligned} \mathsf{E}[\tilde{\tau}] &\geq \; n_0 \cdot \mathsf{P}(T_{n_1} \leq \kappa_1) + (N - n_0) \cdot (\mathsf{P}(T_{n_0} > \kappa_0) - \mathsf{P}(T_{n_1} > \kappa_1))^+ \\ \mathsf{E}[\tilde{\tau}] &\leq \; n_0 + (N - n_0) \cdot \mathsf{P}(T_{n_0} > \kappa_0) \end{aligned} \quad (3.10)$$

and

$$\begin{aligned} \mathsf{E}[\tilde{\tau}] &\geq n_1 \cdot \mathsf{P}(T_{n_0} > \kappa_0) + (N - n_1) \cdot \big(\mathsf{P}(T_{n_1} \leq \kappa_1) - \mathsf{P}(T_{n_0} \leq \kappa_0)\big)^+ \\ \mathsf{E}[\tilde{\tau}] &\leq \; n_1 + (N - n_1) \cdot \mathsf{P}(T_{n_1} \leq \kappa_1). \end{aligned} \quad (3.11)$$

When, in particular, the design parameters are selected as in Theorem 1, by (3.10) with $\mathsf{P} = \mathsf{P}_0$, we obtain

$$\begin{aligned} & n_0 \cdot (1 - \alpha/2) + (N - n_0) \cdot (\gamma - \alpha/2) \leq \mathsf{E}_0[\tilde{\tau}] \leq n_0 + (N - n_0) \cdot \gamma, \\ & \text{where } \; \gamma \in \big[(\alpha \vee \beta)/2, 1\big), \quad n_0 = n^*(\gamma, \beta/2), \quad N = n^*(\alpha/2, \beta/2), \end{aligned} \quad (3.12)$$

and by (3.11) with $\mathsf{P} = \mathsf{P}_1$, we obtain

$$\begin{aligned} &n_1 \cdot (1-\beta/2) + (N-n_1)\cdot(\delta-\beta/2) \le \mathsf{E}_1[\tilde{\tau}] \le n_1 + (N-n_1)\cdot\delta, \\ &\text{where } \ \delta \in \big[(\alpha \vee \beta)/2, 1\big), \quad n_1 = n^*(\alpha/2,\delta), \quad N = n^*(\alpha/2,\beta/2). \end{aligned} \tag{3.13}$$

We can see that, at least when α (resp. β) is small, the upper bound in (3.12) (resp. (3.13)) is approximately equal to the lower bound and, thus, it provides an accurate approximation to $\mathsf{E}_0[\tilde{\tau}]$ (resp. $\mathsf{E}_1[\tilde{\tau}]$). This observation motivates the method for selecting the free parameters, γ and δ, which we present next.

3.1.3 Specification of the free parameters

For any given $\alpha, \beta \in (0,1)$, we suggest selecting γ (resp. δ) to minimize the upper bound in (3.12) (resp. (3.13)) over a grid $\tilde{L}_{\alpha,\beta}$ of $[(\alpha\vee\beta)/2, 1)$, that is, as follows:

$$\begin{aligned} \tilde{\gamma} &\equiv \underset{\gamma \in \tilde{L}_{\alpha,\beta}}{\operatorname{argmin}} \left\{ n^*(\gamma,\beta/2) + \big(n^*(\alpha/2,\beta/2) - n^*(\gamma,\beta/2)\big)\cdot\gamma \right\} \\ \tilde{\delta} &\equiv \underset{\delta \in \tilde{L}_{\alpha,\beta}}{\operatorname{argmin}} \left\{ n^*(\alpha/2,\delta) + \big(n^*(\alpha/2,\beta/2) - n^*(\alpha/2,\delta)\big)\cdot\delta \right\}, \end{aligned} \tag{3.14}$$

where we suppress the dependence of $\tilde{\gamma}$ and $\tilde{\delta}$ on α and β to lighten the notation, and we allow ties to be solved in an arbitrary way. In practice, the grid $\tilde{L}_{\alpha,\beta}$ should, of course, be selected as fine as possible, given the computational constraints involved with the evaluation of the function n^*.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

Nevertheless, as show in the next section, it suffices to have a grid length that goes to zero as fast as $|\log(\alpha \wedge \beta)|^{-1}$ as $\alpha, \beta \to 0$ in order to achieve asymptotic optimality under both P_0 and P_1 for a large class of testing problems.

3.2 The 4-stage tests

Next, we introduce and analyze two novel tests, $\hat{\chi} \equiv (\hat{\tau}, \hat{d})$ and $\check{\chi} \equiv (\check{\tau}, \check{d})$. These tests differ from that of the previous subsection only in that the first (resp. second) one allows for stopping and accepting (resp. rejecting) the null hypothesis if the value of the test statistic after collecting N_0 (resp. N_1) observations is smaller (resp. larger) than K_0 (resp. K_1). Here, N_0, N_1 are additional positive integers and K_0, K_1 are additional real thresholds such that

$$\begin{aligned} n_0 \leq N_0 \leq N \quad &\text{and} \quad K_0 \leq \kappa_1 \;\text{ if }\; N_0 = n_1, \\ n_1 \leq N_1 \leq N \quad &\text{and} \quad \kappa_0 \leq K_1 \;\text{ if }\; n_0 = N_1. \end{aligned} \tag{3.15}$$

Both tests can be implemented by collecting at most four samples of deterministic sizes, and, thus, we refer to them as *4-stage tests.* To avoid repetition, we present a detailed analysis for $\hat{\chi}$, and only state the corresponding results for $\check{\chi}$.

Specifically, given the above parameters, $\hat{\chi}$ proceeds as follows:

(i) $n_0 \wedge n_1$ observations are initially collected.

- If $n_0 \leq n_1$ and $T_{n_0} \leq \kappa_0$, then H_0 is accepted.
- If $n_1 \leq n_0$ and $T_{n_1} > \kappa_1$, then H_0 is rejected.

(ii) If no decision has been reached yet, $\big((n_0 \vee n_1) \wedge N_0\big) - (n_0 \wedge n_1)$ additional observations are collected.

- If $n_0 \leq n_1 \leq N_0$ and $T_{n_1} > \kappa_1$, then H_0 is rejected.
- If $n_0 \leq N_0 \leq n_1$ and $T_{N_0} \leq K_0$, then H_0 is accepted.
- If $n_1 \leq n_0 \leq N_0$ and $T_{n_0} \leq \kappa_0$, then H_0 is accepted.

(iii) If no decision has been reached yet, $(n_1 \vee N_0) - \big((n_0 \vee n_1) \wedge N_0\big)$ additional observations are collected.

- If $n_1 \leq N_0$ and $T_{N_0} \leq K_0$, then H_0 is accepted.
- If $N_0 \leq n_1$ and $T_{n_1} > \kappa_1$, then H_0 is rejected.

(iv) If no decision has been reached yet, $N - (n_1 \vee N_0)$ additional observations are collected and H_0 is rejected if and only if $T_N > K$.

Similarly to the 3-stage test, to avoid possible overlap between acceptance and rejection regions when $n_1 \vee N_0 = N$, we include the convention that when the test reaches its final stage, K is the only effective threshold.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

3.2.1 Error control

By the definition of $\hat{\chi}$, it follows that, for any selection of its parameters and any $\mathsf{P} \in \mathcal{P}$,

$$\mathsf{P}(\hat{d}=1) \leq \mathsf{P}\left(T_{n_1} > \kappa_1\right) + \mathsf{P}\left(T_N > K\right), \tag{3.16}$$

$$\mathsf{P}(\hat{d}=0) \leq \mathsf{P}\left(T_{n_0} \leq \kappa_0\right) + \mathsf{P}\left(T_{N_0} \leq K_0\right) + \mathsf{P}\left(T_N \leq K\right). \tag{3.17}$$

Therefore, if we select n_0, n_1, N, κ_0, κ_1, K as in (3.4)–(3.6), and also

$$N_0 = n^*(\gamma', \beta) \quad \text{and} \quad K_0 = \kappa^*(\gamma', \beta) \quad \text{for some} \quad \gamma' \in [\alpha \vee \beta, \gamma], \tag{3.18}$$

then by (3.7), it follows that conditions (3.1) and (3.15) are satisfied. Moreover, by (3.16) with $\mathsf{P} = \mathsf{P}_0$, and by (3.17) with $\mathsf{P} = \mathsf{P}_1$, it follows that $\mathsf{P}_0(\hat{d}=1) \leq 2\alpha$ and $\mathsf{P}_1(\hat{d}=0) \leq 3\beta$.

Using a similar analysis, if n_0, n_1, N, κ_0, κ_1, K are selected as in (3.4)–(3.6) and

$$N_1 = n^*(\alpha, \delta') \quad \text{and} \quad K_1 = \kappa^*(\alpha, \delta') \quad \text{for some} \quad \delta' \in [\alpha \vee \beta, \delta], \tag{3.19}$$

then conditions (3.1) and (3.15) are satisfied, and $\mathsf{P}_0(\check{d}=1) \leq 3\alpha$ and $\mathsf{P}_1(\check{d}=0) \leq 2\beta$. Thus, we have shown the following theorem.

Theorem 2. *Let* $\alpha, \beta \in (0,1)$.

(i) If the design parameters of $\hat{\chi}$ *are selected according to* (3.4)–(3.6)

and (3.18)*, with* α *and* β *replaced by* $\alpha/2$ *and* $\beta/3$*, respectively, then conditions* (3.1) *and* (3.15) *are satisfied and* $\hat{\chi} \in \mathcal{C}(\alpha, \beta)$*.*

(ii) If the design parameters of $\check{\chi}$ *are selected according to* (3.4)–(3.6) *and* (3.19)*, with* α *and* β *replaced by* $\alpha/3$ *and* $\beta/2$*, respectively, then conditions* (3.1) *and* (3.15) *are satisfied and* $\check{\chi} \in \mathcal{C}(\alpha, \beta)$*.*

Theorem 2 specifies designs for $\hat{\chi}$ and $\check{\chi}$ that guarantee the desired error control up to three *free* parameters, γ, γ', δ and γ, δ, δ', respectively. We next propose a specific selection for these parameters, similar to the one for the 3-stage test in Subsection 3.1.3.

3.2.2 The expected sample size

By the definition of $\hat{\chi}$, it follows that, for any $\mathsf{P} \in \mathcal{P}$,

- if $n_0 \leq n_1 \leq N_0 \leq N$, then

$$\begin{aligned} \mathsf{E}[\hat{\tau}] = n_0 + (n_1 - n_0) \cdot \mathsf{P}\left(T_{n_0} > \kappa_0\right) + (N_0 - n_1) \cdot \mathsf{P} \begin{pmatrix} T_{n_0} > \kappa_0 \\ T_{n_1} \leq \kappa_1 \end{pmatrix} \\ + (N - N_0) \cdot \mathsf{P} \begin{pmatrix} T_{n_0} > \kappa_0 \\ T_{n_1} \leq \kappa_1 \\ T_{N_0} > K_0 \end{pmatrix}, \end{aligned} \tag{3.20}$$

- if $n_0 \leq N_0 \leq n_1 \leq N$, then

$$\mathsf{E}[\hat{\tau}] = n_0 + (N_0 - n_0) \cdot \mathsf{P}\left(T_{n_0} > \kappa_0\right) + (n_1 - N_0) \cdot \mathsf{P}\begin{pmatrix} T_{n_0} > \kappa_0 \\ T_{N_0} > K_0 \end{pmatrix}$$
$$+ (N - n_1) \cdot \mathsf{P}\begin{pmatrix} T_{n_0} > \kappa_0 \\ T_{N_0} > K_0 \\ T_{n_1} \leq \kappa_1 \end{pmatrix}, \tag{3.21}$$

- if $n_1 \leq n_0 \leq N_0 \leq N$, then

$$\mathsf{E}[\hat{\tau}] = n_1 + (n_0 - n_1) \cdot \mathsf{P}\left(T_{n_1} \leq \kappa_1\right) + (N_0 - n_0) \cdot \mathsf{P}\begin{pmatrix} T_{n_1} \leq \kappa_1 \\ T_{n_0} > \kappa_0 \end{pmatrix}$$
$$+ (N - N_0) \cdot \mathsf{P}\begin{pmatrix} T_{n_1} \leq \kappa_1 \\ T_{n_0} > \kappa_0 \\ T_{N_0} > K_0 \end{pmatrix}. \tag{3.22}$$

Applying to these identities the inequalities

$$\max\{\mathsf{P}(A) - \mathsf{P}(B^c) - \mathsf{P}(C^c), 0\} \leq \mathsf{P}(A \cap B \cap C) \leq \mathsf{P}(A),$$

we obtain, for any selection of the design parameters, the following bounds:

$$\begin{aligned} \mathsf{E}[\hat{\tau}] \geq\ & n_0 \cdot \mathsf{P}(T_{n_1} \leq \kappa_1) + (N_0 - n_0) \cdot \big(\mathsf{P}(T_{n_0} > \kappa_0) - \mathsf{P}(T_{n_1} > \kappa_1)\big) \\ & + (N - N_0) \cdot (\mathsf{P}(T_{N_0} > K_0) - \mathsf{P}(T_{n_0} \leq \kappa_0) - \mathsf{P}(T_{n_1} > \kappa_1))^+ , \\ \mathsf{E}[\hat{\tau}] \leq\ & n_0 + (N_0 - n_0) \cdot \mathsf{P}(T_{n_0} > \kappa_0) + (N - N_0) \cdot \mathsf{P}(T_{N_0} > K_0) \end{aligned} \tag{3.23}$$

and

$$\begin{aligned}\mathsf{E}[\hat{\tau}] \geq\ & n_1 \cdot \big(\mathsf{P}(T_{n_0} > \kappa_0) - \mathsf{P}(T_{N_0} \leq K_0)\big) \\ & + (N - n_1) \cdot \big(\mathsf{P}(T_{n_1} \leq \kappa_1) - \mathsf{P}(T_{n_0} \leq \kappa_0) - \mathsf{P}(T_{N_0} \leq K_0)\big) \qquad (3.24) \\ \mathsf{E}[\hat{\tau}] \leq\ & n_1 + (N - n_1) \cdot \mathsf{P}(T_{n_1} \leq \kappa_1).\end{aligned}$$

When, in particular, the parameters of $\hat{\chi}$ are selected as in Theorem 2(i), by (3.23) with $\mathsf{P} = \mathsf{P}_0$, we obtain

$$\begin{aligned}\mathsf{E}_0[\hat{\tau}] &\leq n_0 + (N_0 - n_0) \cdot \gamma + (N - N_0) \cdot \gamma', \\ \mathsf{E}_0[\hat{\tau}] &\geq n_0 \cdot (1 - \alpha/2) + (N_0 - n_0) \cdot (\gamma - \alpha/2) \\ &\quad + (N - N_0) \cdot ((1 - \alpha/2) - (1 - \gamma) - (1 - \gamma'))^+, \qquad (3.25)\end{aligned}$$

$$\text{where} \quad (\alpha/2) \vee (\beta/3) \leq \gamma' \leq \gamma < 1, \quad n_0 = n^*(\gamma, \beta/3),$$
$$N_0 = n^*(\gamma', \beta/3), \quad N = n^*(\alpha/2, \beta/3),$$

and by (3.24) with $\mathsf{P} = \mathsf{P}_1$, we obtain

$$n_1 \cdot (1 - 2\beta/3) + (N - n_1) \cdot (\delta - 2\beta/3) \leq \mathsf{E}_1[\hat{\tau}] \leq n_1 + (N - n_1) \cdot \delta,$$
$$\text{where } \delta \in \big[(\alpha/2) \vee (\beta/3), 1\big), \ n_1 = n^*(\alpha/2, \delta), \ N = n^*(\alpha/2, \beta/3). \quad (3.26)$$

A similar analysis shows that when the parameters of $\check{\chi}$ are selected according to Theorem 2(ii), then

$$n_0 \cdot (1 - 2\alpha/3) + (N - n_0) \cdot (\gamma - 2\alpha/3) \leq \mathsf{E}_0[\check{\tau}] \leq n_0 + (N - n_0) \cdot \gamma,$$
$$\text{where } \gamma \in \big[(\alpha/3) \vee (\beta/2), 1\big), \ n_0 = n^*(\gamma, \beta/2), \ N = n^*(\alpha/3, \beta/2), \quad (3.27)$$

and

$$\begin{aligned}
&\mathsf{E}_1[\check{\tau}] \leq n_1 + (N_1 - n_1) \cdot \delta + (N - N_1) \cdot \delta', \\
&\mathsf{E}_1[\check{\tau}] \geq n_1 (1 - \beta/2) + (N_1 - n_1)(\delta - \beta/2) \\
&\qquad + (N - N_1) \cdot ((1 - \beta/2) - (1 - \delta) - (1 - \delta'))^+, \qquad (3.28) \\
&\text{where} \quad (\alpha/3) \vee (\beta/2) \leq \delta' \leq \delta < 1, \quad n_1 = n^*(\alpha/3, \delta), \\
&\qquad N_1 = n^*(\alpha/3, \delta'), \quad N = n^*(\alpha/3, \beta/2).
\end{aligned}$$

Remark 3. Compared with the corresponding bounds for the 3-stage test, with the same selection of δ (resp. γ), $\mathsf{E}_1[\hat{\tau}]$ is close to $\mathsf{E}_1[\tilde{\tau}]$ (resp. $\mathsf{E}_0[\check{\tau}]$ is close to $\mathsf{E}_0[\tilde{\tau}]$) when β (resp. α) is small. Indeed, the additional stage in $\hat{\chi}$ (resp. $\check{\chi}$) is useful mainly for reducing the expected sample size under P_0 (resp. P_1). This is illustrated in Figure 4 of the Supplementary Material.

3.2.3 Specification of the free parameters

We start with the specification of the free parameters of $\hat{\chi}$. For any $\alpha, \beta \in (0, 1)$, we suggest selecting (γ, γ') (resp. δ) to minimize the upper bound in (3.25) (resp. (3.26)) in the following way:

$$\begin{aligned}
(\hat{\gamma}, \hat{\gamma'}) \equiv \underset{\gamma, \gamma' \in \hat{L}_{\alpha,\beta},\, \gamma' \leq \gamma}{\operatorname{argmin}} &\Big\{ n^*(\gamma, \beta/3) + \big(n^*(\gamma', \beta/3) - n^*(\gamma, \beta/3)\big) \cdot \gamma \\
&+ \big(n^*(\alpha/2, \beta/3) - n^*(\gamma', \beta/3)\big) \cdot \gamma' \Big\}, \qquad (3.29)
\end{aligned}$$

$$\hat{\delta} \equiv \underset{\delta \in \hat{L}_{\alpha,\beta}}{\operatorname{argmin}} \Big\{ n^*(\alpha/2, \delta) + \big(n^*(\alpha/2, \beta/3) - n^*(\alpha/2, \delta)\big) \cdot \delta \Big\},$$

where $\hat{L}_{\alpha,\beta}$ is a grid of $[(\alpha/2) \vee (\beta/3), 1)$.

Similarly, we suggest selecting the free parameters of $\check{\chi}$, γ (resp. (δ, δ')) to minimize the upper bound in (3.27) (resp. (3.28)):

$$\begin{aligned}\check{\delta} &\equiv \underset{\delta \in \check{L}_{\alpha,\beta}}{\operatorname{argmin}} \Big\{ n^*(\gamma, \beta/2) + \big(n^*(\alpha/3, \beta/2) - n^*(\gamma, \beta/2)\big) \cdot \gamma \Big\}, \\ (\check{\delta}, \check{\delta}') &\equiv \underset{\delta, \delta' \in \check{L}_{\alpha,\beta},\, \delta' \leq \delta}{\operatorname{argmin}} \Big\{ n^*(\alpha/3, \delta) + \big(n^*(\alpha/3, \delta') - n^*(\alpha/3, \delta)\big) \cdot \delta \\ &\qquad + \big(n^*(\alpha/3, \beta/2) - n^*(\alpha/3, \delta')\big) \cdot \delta' \Big\},\end{aligned} \tag{3.30}$$

where $\check{L}_{\alpha,\beta}$ is a grid of $[(\alpha/3) \vee (\beta/2), 1)$.

As in the 3-stage test, the grids should ideally be as fine as possible, subject to computational constraints related to the evaluation of the function n^*. However, we will show that letting the grid length go to zero as fast as $|\log(\alpha \wedge \beta)|^{-1}$ as $\alpha, \beta \to 0$ suffices to achieve asymptotic optimality under P_0 and P_1 for a large class of testing problems.

4. Asymptotic analysis

In this section, we obtain asymptotic bounds and approximations for the expected sample sizes of the multistage tests of the previous section as $\alpha, \beta \to 0$. For this analysis, we need to impose some structure on the almost universal setup we have considered so far.

4.1 Assumptions on the testing problem

Throughout this section, we assume that P_0 and P_1 are mutually absolutely continuous when restricted to $\mathcal{F}_n$, for any $n \in \mathbb{N}$, and we denote by $\Lambda \equiv \{\Lambda_n,\, n \in \mathbb{N}\}$ and $\bar{\Lambda} \equiv \{\bar{\Lambda}_n,\, n \in \mathbb{N}\}$ the corresponding log-likelihood ratio and average log-likelihood ratio statistics, respectively,

$$\Lambda_n \equiv \log \frac{d\mathsf{P}_1}{d\mathsf{P}_0}(\mathcal{F}_n) \quad \text{and} \quad \bar{\Lambda}_n \equiv \frac{1}{n}\Lambda_n, \quad n \in \mathbb{N}. \tag{4.1}$$

Moreover, we assume there are numbers $I_0, I_1 > 0$ so that

$$\mathsf{P}_0(\bar{\Lambda}_n \to -I_0) = \mathsf{P}_1(\bar{\Lambda}_n \to I_1) = 1, \tag{4.2}$$

$$\text{for any } \epsilon > 0, \quad \sum_{n=1}^{\infty} \mathsf{P}_0(\bar{\Lambda}_n > -I_0 + \epsilon) + \sum_{n=1}^{\infty} \mathsf{P}_1(\bar{\Lambda}_n \leq I_1 - \epsilon) < \infty. \tag{4.3}$$

These assumptions imply (see, e.g., (Tartakovsky et al.; 2014,Lemma 3.4.1 and Theorem 3.4.2)) asymptotic approximations for the optimal expected sample sizes $\mathcal{L}_0(\alpha, \beta)$ and $\mathcal{L}_1(\alpha, \beta)$, defined in (2.3), as well as the asymptotic optimality under P_0 and P_1 of Wald's SPRT $\chi' \equiv (\tau', d')$, where

$$\tau' \equiv \inf\{n \in \mathbb{N} : \Lambda_n \notin (-A, B)\} \quad \text{and} \quad d' \equiv 1\{\Lambda_{\tau'} \geq B\}, \tag{4.4}$$

with A and B selected, for example, as $A = |\log \beta|$ and $B = |\log \alpha|$. Specifically, under the above assumptions, as $\alpha, \beta \to 0$,

$$\mathsf{E}_0[\tau'] \sim \mathcal{L}_0(\alpha, \beta) \sim \frac{|\log \beta|}{I_0} \quad \text{and} \quad \mathsf{E}_1[\tau'] \sim \mathcal{L}_1(\alpha, \beta) \sim \frac{|\log \alpha|}{I_1}. \tag{4.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

4.1.1 The i.i.d. setup

When X is an i.i.d. sequence with common density f_i under P_i with respect to some dominating measure ν, for $i \in \{0,1\}$, and the Kullback–Leibler divergences are positive and finite, that is,

$$\begin{aligned} D(f_0\|f_1) &\equiv \int \log(f_0/f_1) f_0 \, d\nu \in (0,\infty) \\ D(f_1\|f_0) &\equiv \int \log(f_1/f_0) f_1 \, d\nu \in (0,\infty), \end{aligned} \tag{4.6}$$

then the log-likelihood ratio statistic in (4.1) becomes

$$\Lambda_n = \sum_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)}, \quad n \in \mathbb{N}, \tag{4.7}$$

and assumptions (4.2)–(4.3) hold with $I_0 = D(f_0\|f_1)$ and $I_1 = D(f_1\|f_0)$ (for more details, see Subsection S1.3 of the Supplementary Material).

4.2 Assumptions on the test statistic

With respect to the test statistic, T, throughout this section, we assume there are real numbers J_0, J_1, with $J_0 < J_1$, so that

$$\mathsf{P}_0(T_n \to J_0) = \mathsf{P}_1(T_n \to J_1) = 1, \tag{4.8}$$

and, for every $\kappa \in (J_0, J_1)$, the error probabilities of the fixed-sample-size test that rejects H_0 if and only if $T_n > \kappa$ go to zero exponentially fast in n. Specifically, we assume there are nonnegative, convex, continuous functions $\psi_0, \psi_1 : \mathbb{R} \to [0,\infty]$, so that

- $[J_0, J_1]$ is a subset of the effective domains of ψ_0 and ψ_1,

- $\psi_0(J_0) = 0$ and ψ_0 is strictly increasing in $[J_0, J_1]$,

- $\psi_1(J_1) = 0$ and ψ_1 is strictly decreasing in $[J_0, J_1]$,

- for every $\kappa \in (J_0, J_1)$,

$$\lim_n \frac{1}{n} \log \mathsf{P}_0(T_n > \kappa) = -\psi_0(\kappa), \tag{4.9}$$

$$\lim_n \frac{1}{n} \log \mathsf{P}_1(T_n \leq \kappa) = -\psi_1(\kappa). \tag{4.10}$$

Remark 4. 1) When $T_n = \bar{\Lambda}_n$, (4.8) follows from (4.2) and (4.9)–(4.10) imply (4.3), with $J_0 = -I_0$ and $J_1 = I_1$.

2) In Section S1 of the Supplementary Material we state sufficient conditions for the existence of functions ψ_0 and ψ_1 that satisfy (4.9)–(4.10), which we also specify. In Section S3, we show that these sufficient conditions are satisfied in various testing problems and for different test statistics. The graphs of ψ_0 and ψ_1 in each of these examples are plotted in Figures 1a, 1c and 1e of the Supplementary Material.

3) In the i.i.d. setup of Subsection 4.1.1, the above assumptions hold when $T = \bar{\Lambda}$, as long as (4.6) holds (see Subsection S1.3).

The above assumptions suffice for obtaining first-order asymptotic *upper* bounds on the expected sample sizes of the proposed multistage tests under P_0 and P_1 as $\alpha, \beta \to 0$. When $T = \bar{\Lambda}$, they also suffice for obtaining matching *lower* bounds. However, in order to establish such lower bounds when $T \neq \bar{\Lambda}$, we need to additionally assume that

$$\begin{aligned} &\exists \text{ a neighborhood of } J_1 \text{ in which } \psi_0 \text{ is finite and (4.9) holds} \\ &\exists \text{ a neighborhood of } J_0 \text{ in which } \psi_1 \text{ is finite and (4.10) holds.} \end{aligned} \tag{4.11}$$

In Section S1, we also state sufficient conditions for (4.11), which hold for all the test statistics, different from $\bar{\Lambda}$, that we consider in Section S3.

4.3 Asymptotic analysis for multistage tests

We now focus on the multistage tests introduced in Section 3 and establish the main theoretical results of this work. They are based on asymptotic bounds and approximations for $n^*(\alpha, \beta)$ as $\alpha, \beta \to 0$, which are presented in Section S5.1 of the Supplementary Material.

4.3.1 An upper bound on the maximum sample size

By the definition of the multistage tests and the selection of their parameters according to Theorems 1 and 2, it follows that, for any $\alpha, \beta \in (0, 1)$ and any choice of the free parameters, $\tilde{\tau},\ \hat{\tau},\ \check{\tau} \leq n^*(\alpha/3, \beta/3)$. Consequently, in view of Theorem S2 of the Supplementary Material,

$$\tilde{\tau},\ \hat{\tau},\ \check{\tau} \lesssim \frac{|\log(\alpha \wedge \beta)|}{C} \quad \text{as} \quad \alpha, \beta \to 0, \tag{4.12}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

where C is defined as

$$C \equiv \sup_{\kappa \in (J_0, J_1)} \{\psi_1(\kappa) \wedge \psi_0(\kappa)\}, \tag{4.13}$$

which, in the i.i.d. setup of Subsection 4.1.1, is wellknown as the *Chernoff information* (see, e.g., (Dembo and Zeitouni; 1998,Corollary 3.4.6)).

On the other hand, even when X is an i.i.d. sequence, the SPRT not only does not have a bounded sample size, but even its expected sample size under some $\mathsf{P} \in \mathcal{P}$, in contrast to P_0 and P_1, can be much larger than $n^*(\alpha, \beta)$ when α and β are small. Indeed, consider a $\mathsf{P} \in \mathcal{P}$ under which Λ is a random walk whose increments have a *zero mean* and finite variance σ^2. The expected sample size under such a P of the SPRT, defined in (4.4), with $A = |\log \beta|$ and $B = |\log \alpha|$, is

$$\mathsf{E}[\tau'] \approx |\log \alpha| |\log \beta| / \sigma^2, \tag{4.14}$$

where $\approx$ becomes an equality in the absence of overshoot over the boundaries (see, e.g., (Tartakovsky et al.; 2014,Chapter 3.1.1.2)). Comparing this approximation with the upper bound in (4.12), all of the proposed multistage tests outperform the SPRT *under such a* P *when* α *and* β *are small.* This robustness property of the proposed multistage tests is illustrated in Figure 4 of the Supplementary Material.

4.3.2 Asymptotic analysis under P_0 and P_1

By the asymptotically optimal performance in (4.5), it follows that, as $\alpha, \beta \to 0$,

$$\mathsf{E}_1[\tilde{\tau}],\ \mathsf{E}_1[\hat{\tau}],\ \mathsf{E}_1[\check{\tau}] \gtrsim \frac{|\log\alpha|}{I_1} \quad \text{and} \quad \mathsf{E}_0[\tilde{\tau}],\ \mathsf{E}_0[\hat{\tau}],\ \mathsf{E}_0[\check{\tau}] \gtrsim \frac{|\log\beta|}{I_0},$$

for any selection of the free parameters and any choice of the test statistic. In the next lemma, we obtain a sharper asymptotic lower bound when T is not $\bar{\Lambda}$, but satisfies condition (4.11).

Lemma 1. *Suppose that $T \neq \bar{\Lambda}$ and (4.11) holds. Then, for any selection of the free parameters, as $\alpha, \beta \to 0$,*

$$\mathsf{E}_1[\tilde{\tau}],\ \mathsf{E}_1[\hat{\tau}],\ \mathsf{E}_1[\check{\tau}] \gtrsim \frac{|\log\alpha|}{\psi_0(J_1)} \quad \textit{and} \quad \mathsf{E}_0[\tilde{\tau}],\ \mathsf{E}_0[\hat{\tau}],\ \mathsf{E}_0[\check{\tau}] \gtrsim \frac{|\log\beta|}{\psi_1(J_0)}.$$

We next state the main results of this section, from which the previous asymptotic lower bounds are attained with an appropriate selection of the free parameters. To avoid repetition, we state these results only when $|\log\alpha| \gtrsim |\log\beta|$; analogous results hold when $|\log\alpha| \lesssim |\log\beta|$. Moreover, we denote by $\tilde{l}_{\alpha,\beta}$, $\hat{l}_{\alpha,\beta}$ and $\check{l}_{\alpha,\beta}$ the lengths of the grids $\tilde{L}_{\alpha,\beta}$, $\hat{L}_{\alpha,\beta}$ and $\check{L}_{\alpha,\beta}$, respectively, introduced in Subsections 3.1.3 and 3.2.3.

Theorem 3. *Suppose that $T = \bar{\Lambda}$. Let the free parameters be selected according to (3.14) for the 3-stage test $\tilde{\chi}$, and according to (3.29) and (3.30)* *for the 4-stage tests $\hat{\chi}$ and $\check{\chi}$, respectively. Moreover, suppose that, as* $\alpha, \beta \to 0$,

$$\tilde{l}_{\alpha,\beta},\ \hat{l}_{\alpha,\beta},\ \check{l}_{\alpha,\beta} \lesssim |\log(\alpha \wedge \beta)|^{-1}. \tag{4.15}$$

(i) If $\alpha, \beta \to 0$ so that $|\log \alpha| \gtrsim |\log \beta|$, then

$$\mathsf{E}_1[\tilde{\tau}] \sim \mathsf{E}_1[\hat{\tau}] \sim \mathsf{E}_1[\check{\tau}] \sim \frac{|\log \alpha|}{I_1} \sim \mathcal{L}_1(\alpha, \beta).$$

(ii) If, also, $|\log \alpha| \lesssim |\log \beta|/\beta^r$ for some $r > 0$, then

$$\mathsf{E}_0[\hat{\tau}] \sim \frac{|\log \beta|}{I_0} \sim \mathcal{L}_0(\alpha, \beta).$$

(iii) If, also, $|\log \alpha| \lesssim |\log \beta|^r$ for some $r \geq 1$, then

$$\mathsf{E}_0[\tilde{\tau}] \sim \mathsf{E}_0[\check{\tau}] \sim \frac{|\log \beta|}{I_0} \sim \mathcal{L}_0(\alpha, \beta).$$

Theorem 4. *Suppose that $T \neq \bar{\Lambda}$ and condition (4.11) holds. Let the free parameters be selected as in Theorem 3 and the grid lengths satisfy (4.15).*

(i) If $\alpha, \beta \to 0$ so that $|\log \alpha| \gtrsim |\log \beta|$, then

$$\mathsf{E}_1[\tilde{\tau}] \sim \mathsf{E}_1[\hat{\tau}] \sim \mathsf{E}_1[\check{\tau}] \sim \frac{|\log \alpha|}{\psi_0(J_1)} \sim \frac{I_1}{\psi_0(J_1)} \mathcal{L}_1(\alpha, \beta).$$

(ii) If also $|\log \alpha| \lesssim |\log \beta|/\beta^r$ for some $r > 0$, then

$$\mathsf{E}_0[\hat{\tau}] \sim \frac{|\log \beta|}{\psi_1(J_0)} \sim \frac{I_0}{\psi_1(J_0)} \mathcal{L}_0(\alpha, \beta).$$

(iii) If also $|\log \alpha| \lesssim |\log \beta|^r$ *for some* $r \geq 1$*, then*

$$\mathsf{E}_0[\tilde{\tau}] \sim \mathsf{E}_0[\check{\tau}] \sim \frac{|\log \beta|}{\psi_1(J_0)} \sim \frac{I_0}{\psi_1(J_0)} \mathcal{L}_0(\alpha, \beta).$$

Remark 5. 1) Condition (4.11) in Theorem 4 is used only to obtain the asymptotic lower bounds in Lemma 1; that is, it is not needed to establish the corresponding asymptotic upper bounds.

2) As shown in their proofs, Theorems 3 and 4 remain valid as long as the free parameters satisfy certain mild asymptotic relationships with the error probabilities, found in (S5.61), (S5.64) and (S5.66); that is, they do not have to be selected as the solutions to the minimization problems proposed in Subsections 3.1.3 and 3.2.3.

3) Part (i) in Theorems 3 and 4 states that under P_1, all multistage tests in this work achieve the asymptotically optimal performance when $T = \bar{\Lambda}$, and have the same asymptotic relative efficiency when $T \neq \bar{\Lambda}$ and condition (4.11) holds, as $\alpha, \beta \to 0$ so that $|\log \alpha| \gtrsim |\log \beta|$. On the other hand, parts (ii) and (iii) imply that the corresponding results under P_0 hold *as long as* α *does not go to zero much faster than* β, and that this constraint is much stricter for $\tilde{\chi}$ and $\check{\chi}$ than it is for $\hat{\chi}$. This suggests that $\hat{\chi}$ will outperform $\tilde{\chi}$ and $\check{\chi}$ under the null hypothesis when α is much smaller than β. This insight is supported by Figures 2, 3, and 4 in the numerical studies presented in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0235

4) Analogous results hold when $\alpha, \beta \to 0$ so that $|\log \alpha| \lesssim |\log \beta|$.

5) The asymptotic optimality under both P_0 and P_1 of the 3-stage test with $T = \bar{\Lambda}$ is established in Section 2 of Lorden (1983), in the i.i.d. setup of Subsection 4.1.1, as $\alpha, \beta \to 0$ so that $|\log \beta|/r \lesssim |\log \alpha| \lesssim r\,|\log \beta|$ for some $r \geq 1$. Therefore, apart from extending it to a more general distributional setup, we generalize this result even in the i.i.d. case. Indeed, from (i) and (iii) of Theorem 3 and the previous remark we conclude that the asymptotic optimality of the 3-stage test under both P_0 and P_1 holds as $\alpha, \beta \to 0$ so that $|\log \beta|^{1/r} \lesssim |\log \alpha| \lesssim |\log \beta|^r$ for some $r \geq 1$. At the same time, we show how adding one additional stage can further relax this asymptotic requirement. Specifically, from Theorem 3 and the previous remark we conclude that the 4-stage test $\hat{\chi}$ is asymptotically optimal under both P_0 and P_1 as $\alpha, \beta \to 0$ so that $|\log \beta|^{1/r} \lesssim |\log \alpha| \lesssim |\log \beta|/\beta^k$ for some $r \geq 1$ and $k > 0$. Similarly, the 4-stage test $\check{\chi}$ is asymptotically optimal under both P_0 and P_1 as $\alpha, \beta \to 0$ so that $|\log \alpha|^{1/r} \lesssim |\log \beta| \lesssim |\log \alpha|/\alpha^k$ for some $r \geq 1$ and $k > 0$.

5. Conclusion

Given a fixed-sample-size test that controls the error probabilities at two specific distributions, we design and analyze a 3-stage and two 4-stage tests, with *deterministic* stage sizes, that guarantee the same error control. Under general distributional assumptions, which hold for many testing problems beyond the i.i.d. setup, we also obtain asymptotic approximations for their expected sample sizes under the two distributions at which we control the error probabilities, as the latter go to zero. When, in particular, the test statistic is the average log-likelihood ratio between these two distributions, these tests attain asymptotically the optimal expected sample sizes under the two distributions in the family of all sequential tests with the same error control, similarly to the corresponding SPRT.

The above asymptotic optimality properties require certain constraints on how asymmetrically the two error probabilities go to zero. These constraints are removed in Xing and Fellouris (2022; 2023), in an i.i.d. setup, using a multistage test in which the test statistic is the corresponding log-likelihood ratio, and the number of stages is a function of the two user-specified error probabilities. Our results can be used to extend theirs beyond the i.i.d. setup and to general test statistics.

In order to design multistage tests that achieve asymptotic optimality under every plausible distribution, or in the presence of nuisance parameters, at least some stage sizes need to be adaptive, as in Section 3 of Lorden (1983), Hayre (1985), Bartroff (2007), Bartroff and Lai (2008a). In the first work, a uniform asymptotic optimality property is established for i.i.d. data whose distribution belongs to an exponential family and under the assumption of symmetric error probabilities. Ideas from the present work may be used to extend these results to more general distributional setups and more asymmetric error probabilities.

Finally, another direction of interest is the application of multistage tests, as considered in this work, in a multiple testing setup, similarly to Malloy and Nowak (2014) and Xing and Fellouris (2023).

Supplementary Material

The online Supplementary Material contains sufficient conditions for the asymptotic analysis, an importance sampling approach for the efficient implementation of the proposed tests when the error probabilities are small, three specific testing problems, two numerical studies that illustrate the general theory, and proofs of all main results and supporting lemmas.

Acknowledgments

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

References

Armitage, P., C. K. McPherson, and B. C. Rowe (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society. Series A (General) 132*(2), 235–244.

Barber, S. and C. Jennison (2002). Optimal asymmetric one-sided group sequential tests. *Biometrika 89*(1), 49–60.

Bartroff, J. (2007). Asymptotically optimal multistage tests of simple hypotheses. *The Annals of Statistics 35*(5), 2075–2105.

Bartroff, J. and T. L. Lai (2008a). Efficient adaptive designs with mid-course sample size adjustment in clinical trials. *Statistics in Medicine 27*(10), 1593–1611.

Bartroff, J. and T. L. Lai (2008b). Generalized likelihood ratio statistics and uncertainty adjustments in efficient adaptive design of clinical trials. *Sequential Analysis 27*(3), 254–276.

Bartroff, J., T. L. Lai, and M.-C. Shih (2012). *Sequential experimentation in clinical trials: design and analysis*, Volume 298. Springer Science & Business Media.

Bechhofer, R. (1960). A note on the limiting relative efficiency of the wald sequential probability ratio test. *Journal of the American Statistical Association 55*(292), 660–663.

Bucklew, J. (2010). *Introduction to Rare Event Simulation* (1st ed.). Springer Publishing Company, Incorporated.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications*. Springer.

Dodge, H. F. and H. G. Romig (1929). A method of sampling inspection. *The Bell System Technical Journal 8*(4), 613–631.

Eales, J. D. and C. Jennison (1992). An improved method for deriving optimal one-sided group sequential tests. *Biometrika 79*(1), 13–24.

Emerson, S. S. and T. R. Fleming (1989). Symmetric group sequential test designs. *Biometrics 45*(3), 905–923.

Hayre, L. S. (1985). Group sequential sampling with variable group sizes. *Journal of the Royal Statistical Society: Series B (Methodological) 47*(1), 90–97.

Jennison, C. (1987). Efficient group sequential tests with unpredictable group sizes. *Biometrika 74*(1), 155–165.

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

Kim, K. and D. L. DeMets (1987). Design and analysis of group sequential tests based on the type i error spending rate function. *Biometrika 74*(1), 149–154.

Lai, T. L. and M.-C. Shih (2004). Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika 91*(3), 507–528.

Lan, K. K. G. and D. L. DeMets (1983). Discrete sequential boundaries for clinical trials. *Biometrika 70*(3), 659–663.

Lorden, G. (1983). Asymptotic efficiency of three-stage hypothesis tests. *Annals of Statistics 11*, 129–140.

Malloy, M. L. and R. D. Nowak (2014). Sequential testing for sparse recovery. *IEEE Transactions on Information Theory 60*(12), 7862–7873.

O'Brien, P. C. and T. R. Fleming (1979). A multiple testing procedure for clinical trials. *Biometrics 35*(3), 549–556.

Pampallona, S. and A. A. Tsiatis (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *Journal of Statistical Planning and Inference 42*(1), 19–35.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika 64*(2), 191–199.

Pocock, S. J. (1982). Interim analyses for randomized clinical trials: The group sequential approach. *Biometrics 38*(1), 153–162.

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

Wald, A. (1947). *Sequential Analysis*. New York: John Wiley & Sons.

Wang, S., H. Lin, H.-H. Chang, and J. Douglas (2016). Hybrid computerized adaptive testing: From group sequential design to fully sequential design. *Journal of Educational Measurement 53*(1), 45–62.

Wang, S. K. and A. A. Tsiatis (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics 43*(1), 193–199.

Xing, Y. and G. Fellouris (2022). Asymptotically optimal multistage tests for iid data. In *2022*

IEEE International Symposium on Information Theory (ISIT), pp. 2985–2990.

Xing, Y. and G. Fellouris (2023). Signal recovery with multistage tests and without sparsity constraints. *IEEE Transactions on Information Theory 69*(11), 7220–7245.

Department of Statistics, University of Illinois at Urbana-Champaign, 605 E. Springfield Ave., Champaign, IL 61820, USA

E-mail: yimingx4@illinois.edu, fellouri@illinois.edu