Generalized Sequential Probability Ratio Test for Separate Families of Hypotheses

Xiaoou Li, Jingchen Liu, and Zhiliang Ying
Department of Statistics, Columbia University,
New York, NY 10027, USA

Abstract: In this paper, we consider the problem of testing two separate families of hypotheses via a generalization of the sequential probability ratio test. In particular, the generalized likelihood ratio statistic is considered and the stopping rule is the first boundary crossing of the generalized likelihood ratio statistic. We show that this sequential test is asymptotically optimal in the sense that it achieves asymptotically the shortest expected sample size as the maximal type I and type II error probabilities tend to zero.

Keywords: boundary crossing; generalized likelihood ratio test; sequential test; testing separate families of hypotheses.

Subject Classifications: 62F12.

1. INTRODUCTION

Sequential analysis starts with testing a simple null hypothesis against a simple alternative hypothesis. The fixed sample size problem of this classic test is solved by Neyman and Pearson (1933) who laid the theoretical foundation of likelihood-based hypothesis testing. The sequential probability ratio test (SPRT), formulated via the boundary crossing of the likelihood ratio statistic, is proved to be optimal in terms of minimal expected sample size for fixed type I and type II error probabilities (Wald, 1945; Wald and Wolfowitz, 1948). In this paper, we consider a natural extension of this classical problem to testing two families of composite hypotheses, that is,

$$H_0 : f \in \{g_\theta : \theta \in \Theta\} \quad \text{against} \quad H_A : f \in \{h_\gamma : \gamma \in \Gamma\}, \tag{1.1}$$

where the two families are completely separated from each other. Motivated by the optimality of the sequential probability ratio test, we consider a sequential test based on the generalized likelihood ratio statistic. The sampling stops after the nth observation if the generalized likelihood ratio crosses either of the two boundaries $L_n > e^A$ or $L_n < e^{-B}$ for some positive constants A and B, where

$$L_n = \frac{\sup_{\gamma\in\Gamma} \prod_{i=1}^n h_\gamma(X_i)}{\sup_{\theta\in\Theta} \prod_{i=1}^n g_\theta(X_i)}.$$

The null hypothesis is rejected if $L_n > e^A$ and is accepted otherwise where A and B are determined by the type I and type II error probabilities. We call this procedure the generalized sequential probability ratio test (generalized SPRT).

Address correspondence to Jingchen Liu, Department of Statistics, Columbia University, New York, NY 10027, USA; E-mail: jcliu@stat.columbia.edu

The generalized sequential probability ratio test is a very natural generalization of the sequential probability ratio test in terms of both the problem formulation and the stopping rule. However, to the authors' best knowledge, there has not been rigorous discussion on this sequential procedure in the literature. The results in this paper fill in this void by providing asymptotic descriptions of the type I and type II error probabilities in terms of the levels A and B, the expected sample size (stopping time), and its asymptotic optimality in terms of expected sample size. As a corollary of these results, the generalized SPRT is asymptotically optimal in the following sense. As the maximal type I and type II error probabilities tend to zero, possibly with different rates, the expected stopping time of the generalized SPRT achieves its asymptotic lower bound. For the test as general as (1.1) with a fixed sample size, the uniformly most powerful test usually does not exist. Therefore, we do not expect the optimal sequential test in terms of expected sample size (as optimal as SPRT) for (1.1) to exist. The asymptotic optimality is naturally the next level of optimality to consider. The current result for the generalized SPRT is parallel to the optimality result for SPRT.

From the technical point of view, the challenges mainly lie in the fact that the generalized likelihood ratio statistic is the ratio of two maximized likelihood functions. Usual techniques, such as large deviations theory for independent and identically distributed random variables, exponential tilting for random walks, and Bayesian arguments employed by Wald and Wolfowitz (1948), are no longer applicable. The technical contribution of this paper is the proposal of a set of tools for the large deviations studies of the generalized likelihood ratio statistic. A key element is the construction of a change of measure for developing approximations of the type I and type II error probabilities. This change of measure is not of the traditional exponential tilting form and therefore is nonstandard. Similar change of measure techniques for the computation of small probabilities have been employed under various settings by Shi et al. (2007); Naiman and Priebe (2001); Adler et al. (2012).

Testing separate families of hypotheses, originally introduced by Cox (1961, 1962), is an important and fundamental problem in statistics. Cox recently revisited this problem in Cox (2013) that mentions several applications such as the one-hit and two-hit models of binary dose-response and testing of interactions in a balanced 2^k factorial experiment. Furthermore, this problem has been studied in econometrics (Vuong, 1989). Another application is in educational testing. Under the one-dimensional item response theory models, each examinee is assigned with a scalar θ indicating this person's ability. The so-called mastery test is interested in testing whether $\theta < \theta_-$ or $\theta > \theta_+$. Item response theory usually employs logistic models that fall into the exponential family for which there is a vast literature (Lai and Shih, 2004; Bartroff et al., 2008; Bartroff and Lai, 2008; Shih et al., 2010). However, some more complicated models go beyond exponential family, for which existing results do not apply. For instance, the normal ogive model is not of the canonical form and the three-parameter logistic model includes a guessing parameter. The current results fill in this void. For more applications of testing separate families of hypotheses, see Berrington de González and Cox (2007), Braganca Pereira (2005), and the references therein.

There is a vast literature on sequential tests starting with seminal works Wald (1945); Wald and Wolfowitz (1948); Kiefer and Weiss (1957); Hoeffding (1960) for testing simple null hypothesis against simple alternative hypothesis. An important generalization to SPRT is the 2-SPRT by Lorden (1976). For composite hypotheses, a univariate or multivariate exponential family is usually assumed. Under such a setting, sequential testing procedures for two separate families of hypotheses are discussed by Pollak and Siegmund (1975); Lai (1988); Lai and Zhang (1994). For testing non-exponential families, random walk based sequential procedures are discussed in the textbook Bartroff et al. (2013). Another relevant work is given by Pavlov (1987, 1990) who considers testing/selecting among multiple composite hypotheses. The author establishes asymptotic efficiency of a different sequential procedure (similar to 2-SPRT). The efficiency results are similar to those in this paper. Therefore, the generalized sequential probability ratio test admits the same asymptotic efficiency as that in Pavlov's papers. Recent applications of sequential tests are included in Lai and Shih (2004); Bartroff et al. (2008). Additional references can be found in the textbook Bartroff et al. (2013).

The rest of this paper is organized as follows. The generalized sequential probability ratio test and its asymptotic properties are described in Section 2. Possible relaxation of some technical conditions are provided in Section 3. Numerical examples are given in Section 4. Proofs of the theorems are provided in Section 5.

2. MAIN RESULTS

2.1. Generalized Sequential Probability Ratio Test

Let X_1,...,X_n,... be independent and identically distributed random variables following a density f with respect to a baseline measure μ. We consider the problem of testing two separate families of hypotheses

$$H_0 : f \in \{g_\theta : \theta \in \Theta\} \quad \text{and} \quad H_A : f \in \{h_\gamma : \gamma \in \Gamma\}, \tag{2.1}$$

where g_θ and h_γ are density functions with respect to a common measure μ. To avoid singularity, we assume that g_θ and h_γ are mutually absolutely continuous for all θ and γ. The generalized sequential probability ratio test is based on the generalized likelihood ratio statistic

$$L_n = \frac{\sup_{\gamma\in\Gamma} \prod_{i=1}^n h_\gamma(X_i)}{\sup_{\theta\in\Theta} \prod_{i=1}^n g_\theta(X_i)}. \tag{2.2}$$

For two positive numbers A and B, we define stopping time

$$\tau = \inf\{n : L_n > e^A \text{ or } L_n < e^{-B}\}. \tag{2.3}$$

Under very mild conditions, τ is almost surely finite for any distribution within the two families. The null hypothesis is rejected if $L_\tau > e^A$ and is not rejected if $L_\tau < e^{-B}$. We define notation for the Kullback-Leibler divergence

$$D_g(\theta|\gamma) = E_{g_\theta}\{\log g_\theta(X) - \log h_\gamma(X)\} \quad \textit{and} \quad D_h(\gamma|\theta) = E_{h_\gamma}\{\log h_\gamma(X) - \log g_\theta(X)\},$$

where E_{g_θ} and E_{h_γ} are expectations under the corresponding distributions. The following technical conditions will be used.

- A1 The two families are completely separate, that is, $\inf_{\theta,\gamma} D_g(\theta|\gamma) > \varepsilon_0$ and $\inf_{\theta,\gamma} D_h(\gamma|\theta) > \varepsilon_0$ for some $\varepsilon_0 > 0$. In addition, for each θ and γ, the solutions to the minimizations $\inf_\theta D_h(\gamma|\theta)$ and $\inf_\gamma D_g(\theta|\gamma)$ are unique. Lastly, both $D_g(\theta|\gamma)$ and $D_h(\gamma|\theta)$ are twice continuously differentiable with respect to θ and γ.

- A2 The parameter spaces $\Theta \subset R^{d_1}$ and $\Gamma \subset R^{d_2}$ are compact.

- A3 Let $\xi(\theta,\gamma) = \log h_\gamma(X) - \log g_\theta(X)$. There exists $\alpha > 1$ and x_0 such that for all θ, γ, and $x > x_0$

$$P_{g_\theta}(\sup_{\gamma\in\Gamma} |\nabla_\gamma \xi(\theta,\gamma)| > x) \le e^{-|\log x|^\alpha} \quad \textit{and} \quad P_{h_\gamma}(\sup_{\theta\in\Theta} |\nabla_\theta \xi(\theta,\gamma)| > x) \le e^{-|\log x|^\alpha}.$$

Condition A1 is important for the analysis that guarantees the exponential decay of error probabilities as a function of the expected sample size. A sufficient condition for the complete separation is that the Hellinger distances between any two distributions in the two families are strictly positive. Condition A2 can be further relaxed and replaced by some other conditions that will be discussed subsequently. Condition A3 imposes certain tail restrictions on the score function that has a tail decaying faster than any polynomial.

2.2. The Main Theorems

We start the discussion with a simple null $H_0 : f = g_0$ against a composite alternative $H_A : f \in \{h_\gamma : \gamma \in \Gamma\}$. In this case, the generalized likelihood ratio statistic is given by

$$L_n = \frac{\sup_{\gamma\in\Gamma} \prod_{i=1}^n h_\gamma(X_i)}{g_0(X_i)}. \tag{2.4}$$

The definition of the stopping time τ remains. The following theorem provides the asymptotic type I and type II error probabilities of the generalized sequential probability ratio test under this setting.

Theorem 2.1. *In the case of the simple null hypothesis against composite hypothesis, consider the generalized probability ratio test with stopping time* (2.3) *and the generalized likelihood ratio statistic given by* (2.4)*. Under Conditions A1-3, the type I and maximal type II error probabilities admit the following approximations*

$$\log P_{g_0}(L_\tau > e^A) \sim -A, \quad \sup_{\gamma\in\Gamma} \log P_{h_\gamma}(L_\tau < e^{-B}) \sim -B \quad \text{as } A,\ B \to \infty.$$

The analysis technique of Theorem 2.1 and its intermediate results are central to all the analyses. For the general case of composite null hypothesis against composite alternative hypothesis, we establish similar asymptotic results that are given by the following theorem.

Theorem 2.2. *Consider the composite null hypothesis against composite alternative hypothesis given as in* (2.1)*. The generalized sequential probability ratio test admits stopping time* (2.3) *and the generalized likelihood ratio statistic* (2.2)*. Under Conditions A1-3, the maximal type I and type II error probabilities are approximated by*

$$\sup_{\theta\in\Theta} \log P_{g_\theta}(L_\tau > e^A) \sim -A, \quad \sup_{\gamma\in\Gamma} \log P_{h_\gamma}(L_\tau < e^{-B}) \sim -B \quad \text{as } A,\ B \to \infty. \tag{2.5}$$

In the power calculation of SPRT for the simple null hypothesis versus simple alternative hypothesis, if the likelihood ratio has zero overshoot, then we have the following equalities $A = \log \frac{1-\alpha_2}{\alpha_1}$ and $B = \log \frac{1-\alpha_1}{\alpha_2}$ where α_1 is the type I error probability and α_2 is the type II error probability. They have exactly the same asymptotic decay rate as (2.5). Lastly, we provide the asymptotic approximations of the expected stopping time.

Theorem 2.3. *Under the setting and the conditions of Theorem 2.2, the expected stopping time admits the following asymptotic approximation*

$$E_{g_\theta}(\tau) \sim \frac{B}{\inf_{\gamma\in\Gamma} D_g(\theta|\gamma)}, \quad E_{h_\gamma}(\tau) \sim \frac{A}{\inf_{\theta\in\Theta} D_h(\gamma|\theta)}, \quad \text{as } A,\ B \to \infty \text{ for all } \theta \text{ and } \gamma.$$

Based on the results of Theorems 2.2 and 2.3, we now discuss the asymptotic optimality of the generalized SPRT. Consider type I and type II error probabilities α_1 and α_2 that approach zero possibly with different rates. Theorem 2.2 suggests that we need to choose $A \sim -\log \alpha_1$ and $B \sim -\log \alpha_2$ for the generalized SPRT to achieve such levels of error probabilities. Then, the corresponding expected stopping time is given by Theorem 2.3. In what follows, we show that the expected stopping time in Theorem 2.3 is asymptotically the shortest. Consider an arbitrarily chosen sequential procedure testing between the g-family and the h-family with stopping time τ'. The two types of error probabilities of this test are less than or equal to α_1 and α_2 respectively. Then, its expected stopping time is bounded from below by

$$E_{g_\theta}(\tau') \geq (1+o(1))E_{g_\theta}(\tau) \quad \text{and} \quad E_{h_\gamma}(\tau') \geq (1+o(1))E_{h_\gamma}(\tau)$$

for all θ and γ.

We establish the above asymptotic inequalities by making use of the optimality results of SPRT. For a fixed pair θ and γ, we consider the testing problem of the simple null $H_0 : f = g_\theta$ against the simple alternative $H_A : f = h_\gamma$. We further consider SPRT for this test with stopping boundaries $e^{\tilde{A}}$ and $e^{-\tilde{B}}$. We choose $\tilde{A}$ and $\tilde{B}$ such that the type I error and type II error probabilities of SPRT for the simple (g_θ) versus simple (h_γ) test are (or slightly larger than, but of the same order as) α_1 and α_2 respectively. According to Theorem 2.2 and standard results of SPRT (Wald and Wolfowitz, 1948), we have that $A \sim \tilde{A} \sim -\log \alpha_1$ and $B \sim \tilde{B} \sim -\log \alpha_2$ if the overshoot is of order $O(1)$. Let $\tilde{\tau}$ be the stopping time of SPRT. According to classic results on random walks, it can be shown that

$$E_{g_\theta}(\tilde{\tau}) \sim B/D_g(\theta|\gamma) \quad \textit{and} \quad E_{h_\gamma}(\tilde{\tau}) \sim A/D_h(\gamma|\theta).$$

Furthermore, we view the test with stopping time τ' in the previous paragraph as a testing procedure for the simple null (g_θ) versus simple alternative (h_γ) problem. According to the definition of α_1 and α_2, the type I and type II error probabilities of this test for the simple versus simple problem are bounded from the above by α_1 and α_2. Therefore, according to the optimality of SPRT we have that

$$E_{g_\theta}(\tau') \geq E_{g_\theta}(\tilde{\tau}) = (1+o(1))B/D_g(\theta|\gamma) \quad \textit{and} \quad E_{h_\gamma}(\tau') \geq E_{h_\gamma}(\tilde{\tau}) = (1+o(1))A/D_h(\gamma|\theta).$$

For the first inequality, the left-hand side does not depend on γ and furthermore Γ is a compact set. Thus, the $o(1)$ is uniformly small for $\gamma \in \Gamma$. We maximize the right-hand side with respect to γ and obtain that

$$E_{g_\theta}(\tau') \geq (1+o(1))\frac{B}{\inf_\gamma D_g(\theta|\gamma)}.$$

Note that the right-hand side of the above inequality is precisely the asymptotic expected stopping time in Theorem 2.3. With the same argument, we have that

$$E_{h_\gamma}(\tau') \geq (1+o(1))\frac{A}{\inf_\theta D_h(\gamma|\theta)}.$$

Summarizing the above discussion, we have the following corollary.

Corollary 2.1. *Let $\mathcal{T}(\alpha_1, \alpha_2)$ be the class of sequential tests with their type I and type II errors bounded above by α_1 and α_2, respectively. Each test in $\mathcal{T}(\alpha_1, \alpha_2)$ corresponds to a stopping time τ' and a decision function D'. Let $\alpha_1^{A,B} = \sup_\theta P_{g_\theta}(L_\tau > e^A)$ and $\alpha_2^{A,B} = \sup_\gamma P_{h_\gamma}(L_\tau < e^{-B})$. Then, under the setting of Theorem 2.2 and under Conditions A1-3, the generalized sequential probability test is asymptotically optimal in the sense that*

$$E_{g_\theta}(\tau) \sim \inf_{(\tau', D') \in \mathcal{T}(\alpha_1^{A,B}, \alpha_2^{A,B})} E_{g_\theta}(\tau')$$

as $A \to \infty$ and $B \to \infty$.

3. FURTHER DISCUSSION ON THE CONDITIONS

In this section, we provide further discussion on Condition A1, A2, and A3 and possible relaxations. Condition A1 requires that the two families of hypotheses are completely separate. This condition is crucial for the exponential decay of the error probabilities in Theorems 2.1 and 2.2. The uniqueness of the minimization of the Kullback-Leibler divergence ensures the convergence of the maximum likelihood estimators and validity of the stopping time analysis. Therefore, Condition A1 is necessary for the theorems. In what follows, we provide further discussions on Conditions A2 and A3.

3.1. Relaxing Condition A2 and Analysis for Non-compact Spaces

When the parameter spaces Θ and Γ are non-compact, the expected stopping time of the generalized sequential probability ratio test can usually be approximated similarly as that of Theorem 2.3 with mild regularity conditions such as almost sure convergence of the maximum likelihood estimators. For the asymptotic decay rate of the type I and type II error probabilities, the generalization to non-compact spaces is not straightforward and additional conditions are necessary. We start the discussion with a counterexample in which Theorem 2.2 fails when the parameter spaces are non-compact.

Example 3.1. Consider the null hypothesis being the lognormal distributions

$$g_\theta(x) = x^{-1}(2\pi\theta)^{-1/2}e^{-\frac{(\log x)^2}{2\theta}}$$

and the alternative hypothesis being the exponential distributions

$$h_\gamma(x) = \gamma^{-1}e^{-x/\gamma}.$$

Both distributions live on the positive real line. The maximum likelihood estimators for the parameters based on n observations are $\hat\theta_n = \frac{1}{n}\sum_{i=1}^n(\log X_i)^2$ and $\hat\gamma_n = \frac{1}{n}\sum_{i=1}^n X_i$. The generalized log-likelihood ratio statistic based on one sample is $\log h_{\hat\gamma_1}(X_1) - \log g_{\hat\theta}(X_1) = \log|\log X_1| - \frac{1}{2} + \frac{1}{2}\log(2\pi)$ and $L_1 = \sqrt{2\pi/e}\times|\log(X_1)|$. The type I error probability is bounded from below by

$$\sup_{\theta\in\Theta} P_{g_\theta}(L_\tau > e^A) \geq \sup_{\theta\in\Theta} P_{g_\theta}(L_1 > e^A) \geq \lim_{\theta\to\infty} P_{g_\theta}\{\sqrt{2\pi/e}\times|\log(X_1)| > e^A\} = 1$$

regardless of the choice of A. The last equality holds because $\log(X_1)$ follows a normal distribution with mean 0 and variance $2\pi\theta/e$.

Therefore, additional conditions are certainly needed to generalize the results of Theorem 2.2 to non-compact parameter spaces and to rule out cases such as Example 3.1. Let $\xi_i(\theta,\gamma)$, $i = 1, 2...$ be i.i.d. copies of $\xi(\theta,\gamma)$. The log-likelihood ratio based on n observations is defined as

$$S_n(\theta,\gamma) = \sum_{i=1}^n \xi_i(\theta,\gamma). \tag{3.1}$$

We further define $S_n = \sup_\gamma \inf_\theta \sum_{i=1}^n \xi_i(\theta,\gamma)$ and $\tau = \inf\{n : S_n < -B \text{ or } S_n > A\}$. To rule out the cases such as Example 3.1, we need to carefully go through the proof of Theorem 2.2 (Section 5) that consists of the development of an upper and a lower bound of the error probabilities. The lower bound does not require the compactness of the parameter spaces and is generally applicable.

It is the development of the upper bound where the compactness plays an important role in the analysis. Define

$$H_{A,\theta} = \sum_{n=1}^{\infty} \int_{\Gamma} P_{g_\theta}(S_n(\theta,\gamma) > A) d\gamma. \tag{3.2}$$

The condition for non-compact parameter spaces is

A2′ Let $H_{A,\theta}$ be defined as in (3.2) and we require $\limsup_{A\to\infty} \sup_{\theta\in\Theta} \frac{1}{A} \log H_{A,\theta} \le -1$. Symmetrically, we define

$$G_{B,\gamma} = \sum_{n=1}^{\infty} \int_{\Theta} P_{h_\gamma}(S_n(\theta,\gamma) < -B) d\theta$$

that satisfies $\limsup_{B\to\infty} \sup_{\theta\in\Theta} \frac{1}{B} \log G_{B,\theta} \le -1$.

Condition A2′ is usually difficult to check. Therefore, we provide a set of sufficient conditions for A2′.

Lemma 3.1. *Assume that the following conditions hold.*

B1 For each θ, let $\gamma_\theta = \arg\inf_{\gamma\in\Gamma} D_h(\gamma|\theta)$. There exist ε and δ positive such that

$$D_h(\gamma|\theta) \ge D_h(\gamma_\theta|\theta) + \delta|\gamma - \gamma_\theta|^l, \quad \text{for some } l > (d+1)/2, \text{ all } \theta \in \Theta, \text{ and all } |\gamma - \gamma_\theta| > \varepsilon$$

where d is the dimension of Γ.

B2 The log-likelihood ratio $\xi(\theta,\gamma)$ has bounded variance under h_γ for all $\theta \in \Theta$ and $\gamma \in \Gamma$.

B3 There exists $\varepsilon > 0$ such that $\varepsilon < D_g(\theta|\gamma)/D_h(\gamma|\theta) < \varepsilon^{-1}$ for all θ and γ.

Then, $\limsup_{A\to\infty} \sup_{\theta\in\Theta} \frac{1}{A} \log H_{A,\theta} \le -1$.

For the two families of distributions in Example 3.1, Condition A2′ is not satisfied. With Condition A2′ in addition to Conditions A1 and A3, we expect to obtain similar approximation results as in Theorem 2.2. Given that the techniques are similar but substantially more tedious, we do not provide the details.

3.2. Relaxing Condition A3

We now consider the situation in which Condition A3 is violated. For instance, if the alternative hypothesis h_γ is the exponential distributions, then the partial derivative $\partial_\gamma \xi(\theta,\gamma)$ is infinity when $\gamma \to 0$. For these types of families, we need to replace Condition A3 by some localization condition. Let $\hat\gamma_n$ be the maximum likelihood estimator based on n i.i.d. samples. The localization condition to replace A3 is as follows.

A3′ There exists a family of sets $\Gamma'_A \subset \Gamma$ indexed by A such that $P_{g_\theta}(\hat\gamma_n \notin \Gamma'_A) \le e^{-(n+1)A}$ and for some $\alpha > 1$, $\beta \in (\alpha^{-1}, 1)$ and all $\theta \in \Theta$

$$P_{g_\theta}(\sup_{\gamma\in\Gamma'_A} |\partial_\gamma \xi(\theta,\gamma)| > e^{A^\beta} x) \le e^{-|\log x|^\alpha}.$$

Similarly, there exists $\Theta'_B \subset \Theta$ such that $P_{h_\gamma}(\hat\theta_n \notin \Theta'_B) \le e^{-(n+1)B}$ and

$$P_{h_\gamma}(\sup_{\theta\in\Theta'_b} |\partial_\theta \xi(\theta,\gamma)| > e^{A^\beta} x) \le e^{-|\log x|^\alpha}.$$

For the two hypotheses in Example 3.1, we have $\alpha = 2$. For some $1/2 < \beta < 1$ let

$$\Gamma' = [e^{-A^{\beta'}}, \infty) \quad \text{where } 1/2 < \beta' < \beta.$$

Then, we can verify that such a choice of Γ' satisfies Condition A3′. We summarize the discussion in this section as follows.

Theorem 3.1. *Under Conditions A1, A2′, and A3′, the approximations in* (2.5) *holds.*

Given that the proof of the above theorem is basically identical to that of Theorem 2.2 and therefore we do not provide the details.

4. NUMERICAL EXAMPLES

4.1. Poisson Family of Distributions against Geometric Family of Distributions

In this section, we provide numerical examples to illustrate the results of the theorems. We start with the Poisson distribution against the geometric distribution. Let

$$g_\theta(x) = \frac{e^{-\theta}\theta^x}{x!} \quad \Theta = [0.5, 2], \qquad h_\gamma(x) = \frac{\gamma^x}{(1+\gamma)^{x+1}} \quad \Gamma = [0.5, 2]$$

where x takes non-negative integer values and $1/(1+\gamma)$ is the success probability of the geometric trials. We truncate the parameter spaces from above for Condition A2 and from below to make these two families of distributions completely separated for Condition A1. The test statistic is

$$L_n = \frac{\prod_{i=1}^n h_{\hat\gamma}(X_i)}{\prod_{i=1}^n g_{\hat\theta}(X_i)} \qquad \text{where } \hat\theta = \hat\gamma = \max\{\min(\frac{1}{n}\sum_{i=1}^n X_i, 2), 0.5\}.$$

For B fixed at 4, we compute the type I error probabilities for different values of A via Monte Carlo. Figure 1 plots the logarithm of the type I error probabilities against the boundary parameter A. For fixed $A = 4$, we compute the expected sample size under distributions $g_{0.5}(x)$, $g_1(x)$, and $g_{1.5}(x)$ for different values of B as shown in Figure 2. Similarly, for fixed $B = 4$, we compute the expected sample sizes under distributions $h_{0.5}(x)$, $h_1(x)$, and $h_{1.5}(x)$ for different values of A as shown in Figure 3.

The slope of the fitted line in Figure 1 is -1.02. The fitted slopes in Figure 2 are 35.50, 12.12, and 6.81. The fitted slopes in Figure 3 are 26.22, 8.22, and 4.61. From Theorems 2 and 3, the theoretical values of the slope in Figure 1 is -1, and the theoretical values of the slopes in Figure 2 are $\{\inf_\gamma D(g_{0.5}|h_\gamma)\}^{-1} = 36.85$, $\{\inf_\gamma D(g_1|h_\gamma)\}^{-1} = 12.28$, and $\{\inf_\gamma D(g_{1.5}|h_\gamma)\}^{-1} = 6.99$. The theoretical slopes in Figure 3 are $\{\inf_\theta D(g_\theta|h_{0.5})\}^{-1} = 26.97$, $\{\inf_\theta D(g_\theta|h_1)\}^{-1} = 8.23$, and $\{\inf_\theta D(g_\theta|h_{1.5})\}^{-1} = 4.30$. It is clear that the numerically fitted values are close to the theoretical ones.

4.2. Gaussian Scale Family against Laplace Scale Family

We proceed to testing Gaussian scale family against Laplace scale family, both of which are non-compact. Specifically, let

$$g_\theta(x) = (2\pi\theta)^{-1/2}e^{-x^2/(2\theta)}, \quad \Theta = (0,\infty) \quad \text{and} \quad h_\gamma(x) = (2\gamma)^{-1}e^{-|x|/\gamma} \quad \Gamma = (0,\infty).$$

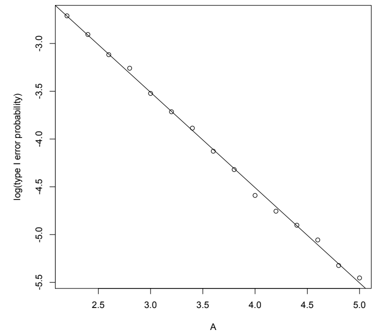

***Figure 1*.** Logarithm of the type I error probabilities (y-coordinate) against boundary parameter A (x-coordinate) for Poisson against Geometric with $B = 4$.

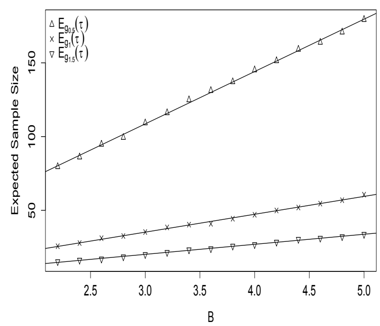

***Figure 2*.** $E_{g_{0.5}}(\tau)$, $E_{g_1}(\tau)$ and $E_{g_{1.5}}(\tau)$ (y-coordinate) against boundary parameter B (x-coordinate) for Poisson against Geometric with $A = 4$.

The generalized likelihood statistic is

$$L_n = \frac{\prod_{i=1}^n h_{\hat{\gamma}(X_i)}}{\prod_{i=1}^n g_{\hat{\theta}}(X_i)}, \text{ where } \hat{\gamma} = \tfrac{1}{n}\textstyle\sum_{i=1}^n |X_i| \text{ and } \hat{\theta} = \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i^2.$$

For B fixed at 4 and different A values, we compute the type I error probabilities of the generalized sequential probability ratio test. Figure 4 plots the logarithm of the type I error probabilities against the boundary parameter A. Furthermore, for fixed $A = 4$ and different B values, we calculate the expected sample size under g_1 and for fixed $B = 4$ with different A values we calculate the expeted sample size under h_2. Figure 5 plots the expected sample size against B, and Figure 6 is the plot for expected sample size against A. We fit a straight line to each of the three plots via the least squares. The slopes of the fitted lines in Figures 4, 5, and 6 are -1.00, 20.60, and 14.42 respectively. The theoretical values of these three slopes should be -1, $\{\inf_{\gamma\in\Gamma} D(g_1|h_\gamma)\}^{-1} = 20.65$ and $\{\inf_{\theta\in\Theta} D(g_\theta|h_2)\}^{-1} = 13.82$ that are close to the numerically fitted values.

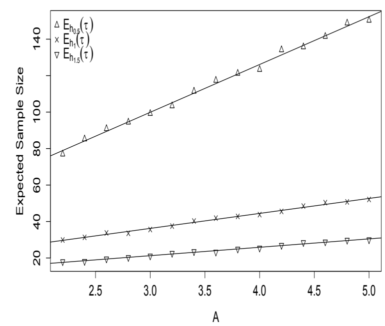

Figure 3. $E_{h_{0.5}}(\tau)$, $E_{h_1}(\tau)$ and $E_{h_{1.5}}(\tau)$ (y-coordinate) against boundary parameter A (x-coordinate) for Poisson against Geometric with B fixed to be 4.

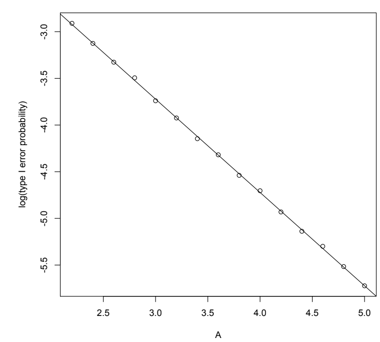

Figure 4. Logarithm of the type I error probabilities (y-coordinate) against boundary parameter A (x-coordinate) for Gaussian against Laplace with $B = 4$.

4.3. Lognormal against Exponential

We proceed to the lognormal distribution against exponential distribution

$$g_\theta(x) = \frac{1}{x\sqrt{2\pi\theta}} e^{-\frac{(\log x)^2}{2\theta}} \quad \Theta = [0,1], \qquad h_\gamma(x) = \frac{1}{\gamma} e^{-\frac{x}{\gamma}} \quad \Gamma = [0,1]$$

As explained in Example 3.1, we consider θ and γ on compact sets for Condition A2. The generalized likelihood ratio statistic is

$$L_n = \frac{\prod_{i=1}^n h_{\hat\gamma}(X_i)}{\prod_{i=1}^n g_{\hat\theta}(X_i)}, \qquad \text{where } \hat\gamma = \min(\frac{1}{n}\sum_{i=1}^n X_i, 1), \quad \hat\theta = \min\Big\{\frac{1}{n}\sum_{i=1}^n (\log X_i)^2, 1\Big\}.$$

For a fixed B= 4 and different values of A, we compute the type I error probabilities of the generalized sequential probability ratio test under the distribution $g_1(x)$. Figure 7 is the scatter plot for the logarithm of the type I error probabilities against the boundary parameter A. Furthermore, for a fixed A and different B values, we compute the expected sample sizes under $g_{0.5}(x)$ and $g_1(x)$ via Monte Carlo. For a fixed B and different A, we also compute the expected sample sizes under probability measure $h_{0.5}$ and $h_1(x)$. Figure 8 plots the expected sample sizes under

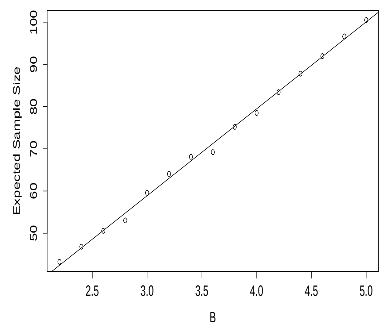

Figure 5. Expected sample size $E_{g_1}(\tau)$ (y-coordinate) against boundary parameter B (x-coordinate) for Gaussian against Laplace with $A = 4$.

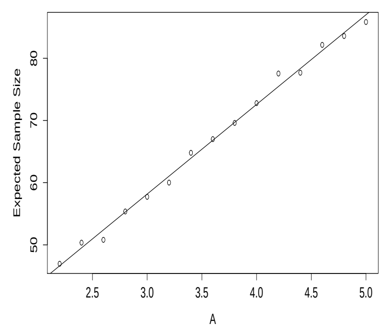

Figure 6. Expected sample size $E_{h_2}(\tau)$ (y-coordinate) against boundary parameter A (x-coordinate) for Gaussian against Laplace with $B = 4$.

probability measure $g_{0.5}$ and g_1 against B. Figure 9 plots the expected sample sizes under $h_{0.5}$ and h_1 against A. We fit a straight line to each of the three plots via the least squares. The slope of the fitted line in Figure 7 is -0.92. The slopes of the regression lines in Figure 8 are 4.67, and 4.75. . The slopes of the regression lines in Figure 9 are 1.08, and 3.28. From Theorems 2.2 and 2.3, the theoretical value of the slope in Figure 7 should be -1, and the slopes in Figure 8 are $\{\inf_{\gamma\in\Gamma} D(g_{0.5}|h_\gamma)\}^{-1} = 4.72$, and $\{\inf_{\gamma\in\Gamma} D(g_1|h_\gamma)\}^{-1} = 4.54$. The theoretical value of slopes in Figure 9 are $\{\inf_{\theta\in\Theta} D(g_\theta|h_{0.5})\}^{-1} = 1.03$ and $\{\inf_{\theta\in\Theta} D(g_\theta|h_1)\}^{-1} = 3.02$.

5. TECHNICAL PROOFS

5.1. Proof of Theorem 2.1

We write $a_n \cong b_n$ if $\log a_n \sim \log b_n$ as $n \to \infty$. To make the discussion smooth, we delay the proofs of the supporting lemmas to the appendix.

Proof of Theorem 2.1. Define the log-likelihood ratio of a single observation

$$\xi(\gamma) = \log h_\gamma(X) - \log g_0(X)$$

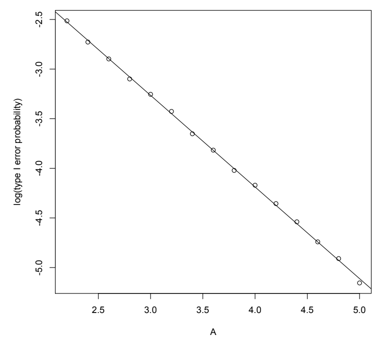

Figure 7. Logarithm of the type I error probabilities (y-coordinate) against boundary parameter A (x-coordinate) for lognormal distribution against exponential distribution where B is fixed to be 4

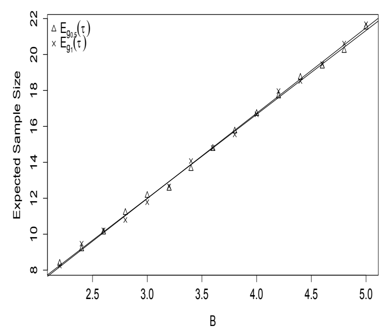

Figure 8. Expected sample size $E_{g_{0.5}}(\tau)$ and $E_{g_1}(\tau)$ (y-coordinate) against boundary parameter B (x-coordinate) for lognormal distribution against exponential distribution, where A is fixed to be 4.

and let $\xi_i(\gamma) = \log h_\gamma(X_i) - \log g_0(X_i)$ be i.i.d. copies of it. The log-likelihood ratio becomes

$$S_n(\gamma) = \sum_{i=1}^{n} \xi_i(\gamma).$$

The generalized log-likelihood ratio statistic is $\log L_n = S_n = \sup_{\gamma\in\Gamma} S_n(\gamma)$. The stopping time can be equivalently written as $\tau = \inf\{n : S_n < -B \text{ or } S_n > A\}$. We reject the null hypothesis if $S_\tau > A$ and do not reject otherwise. Let $\gamma_* = \arg\sup_\gamma E_{g_0}\{\xi(\gamma)\}$, $-\mu_g^\gamma = E_{g_0}\{\xi(\gamma)\} = D_{g_0}(0|\gamma)$, and $\mu_h^\gamma = E_{h_\gamma}\{\xi(\gamma)\} = -D_{h_\gamma}(\gamma|0)$. We now proceed to the computation of the type I and type II error probabilities. The decay rate of the type I error probability is given by the following lemmas that is the key result of the remaining derivations.

Lemma 5.1. *Under the setting and conditions of Theorem 2.1, the type I error probability is approximated by*

$$e^{-(1+o(1))A} \le P_g(S_\tau > A) \le P_g(\sup_n \sup_\gamma S_n(\gamma) > A) \le \kappa A^{\alpha_0} H_A$$

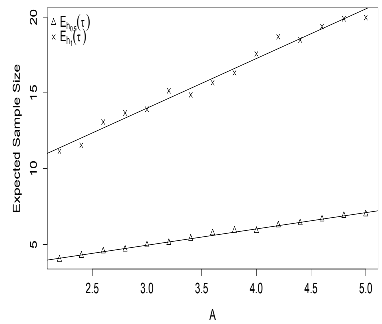

Figure 9. Expected sample size $E_{h_{0.5}}(\tau)$ and $E_{h_1}(\tau)$ (y-coordinate) against boundary parameter A (x-coordinate) for lognormal distribution against exponential distribution, where B is fixed to be 4.

for some ε_0, α_0, and $\kappa > 0$ and

$$H_A = \sum_{n=1}^{\infty} \int_{\Gamma} P(S_n(\gamma) > A - 1) d\gamma.$$

The constant κ depends on the dimension of Γ and α_0 depends on α in Condition A3.

Lemma 5.2. *Let $mes(\Gamma) = \int I(t \in \Gamma)dt$ be the Lebesgue measure of the parameter set Γ and let $D_h(\gamma|0) = E_{h_\gamma}\{\log h_\gamma(X) - \log g_0(X)\}$ be the Kullback-Leibler divergence. Under the setting and conditions of Theorem 2.1, there exists some $\kappa_0 > 0$ such that for A sufficiently large H_A defined as in Lemma 5.1 admits the following bound*

$$H_A \leq \frac{\kappa_0 mes(\Gamma) A e^{-A}}{\min_\gamma D_h(\gamma|0)}.$$

Therefore, we finished the analysis of the type I error probability. We focus on the type II error computation $\alpha_2 = \sup_{\gamma \in \Gamma} P_{h_\gamma}(S_\tau < -B)$. For each γ_0, notice that $S_n \geq S_n(\gamma_0)$ and thus

$$P_{h_{\gamma_0}}(S_\tau < -B) < P_{h_{\gamma_0}}(S_{\tau(\gamma_0)}(\gamma_0) < -B) \leq e^{-B}$$

where $\tau(\gamma_0) = \inf\{n : S_n(\gamma_0) < -B \text{ or } S_n(\gamma_0) > A\}$. The last step of the above display is a classical large deviations result of random walk. This provides an upper bound of α_2. We now show that this upper bound is achieved in the sense of "$\cong$". In particular, we wish to show that

$$\liminf_{A,B \to \infty} \frac{\log P_{h_{\gamma_*}}(S_\tau < -B)}{B} \geq -1. \tag{5.1}$$

We establish the above inequality via contradiction. Suppose that (5.1) is not true, that is, there exist two sequences A_i, $B_i \to \infty$ as $i \to \infty$ and $\varepsilon_0 > 0$ such that

$$\frac{\log P_{h_{\gamma_*}}(S_\tau < -B_i)}{B_i} < -1 - \varepsilon_0$$

and equivalently $P_{h_{\gamma_*}}(S_\tau < -B_i) < e^{-(1+\varepsilon_0)B_i}$. Recall that, from the type I error computation, we have that $P_{g_0}(S_\tau > A_i) \cong e^{-A_i}$.

Now we consider the simple null $f = g_0$ against the simple alternative $f = h_{\gamma_*}$ and SPRT with stopping time

$$\tilde{\tau}_i = \inf\{n : S_n(\gamma_*) < -\tilde{B}_i \text{ or } S_n(\gamma_*) > \tilde{A}_i\}.$$

The threshold $\tilde{A}_i$ and $\tilde{B}_i$ is chosen such that the SPRT has exactly the same (or slightly larger) type I and type II error probability as the generalized SPRT, that is,

$$e^{-\tilde{A}_i} \cong P_{g_0}(S_{\tilde{\tau}_i}(\gamma_*) > \tilde{A}_i) \cong P_{g_0}(S_\tau > A_i) \cong e^{-A_i}$$

and

$$e^{-\tilde{B}_i} \cong P_{h_{\gamma_*}}(S_{\tilde{\tau}_i}(\gamma_*) < -\tilde{B}_i) \cong P_{h_{\gamma_*}}(S_\tau < -B_i) < e^{-(1+\varepsilon_0)B_i}.$$

Therefore, we have that $\tilde{A}_i \sim A_i$ and $\tilde{B}_i > (1 + \varepsilon_0/2)B_i$. Furthermore, notice that the expected stopping time for SPRT is

$$E_g(\tilde{\tau}_i) \sim \tilde{B}_i/\mu_g^{\gamma_*}, \quad E_{h_{\gamma_*}}(\tilde{\tau}_i) \sim \tilde{A}_i/\mu_h^{\gamma_*}.$$

Note that $\mu_g^{\gamma_*} = \inf_{\gamma\in\Gamma} D_g(\theta|\gamma)$. According to Theorem 2.3 (whose proof is independent of the current one), we have that $E_g(\tilde{\tau}_i) > E_g(\tau) \sim B_i/\mu_g^{\gamma_*}$ that contradicts the optimality result of SPRT (Wald and Wolfowitz, 1948). Thus, (5.1) must be true and we establish that

$$\alpha_2 = \sup_{\gamma\in\Gamma} P_{h_\gamma}(S_\tau < -B) \cong e^{-B} \quad \text{as } A,\, B \to \infty.$$

□

5.2. Proof of Theorem 2.2

With the above proof, Theorem 2.2 can be obtained rather easily. This proof also requires some intermediate results in the proof of Theorem 2.1.

Proof of Theorem 2.2. Let $S_n(\theta, \gamma)$ be defined as in (3.1). We define notation

$$S_n = \sup_\gamma \inf_\theta \sum_{i=1}^n \xi_i(\theta, \gamma), \quad \tau = \inf\{n : S_n < -B \text{ or } S_n > A\}.$$

As the two types of errors are completely symmetric, we only derive the type I error. We start with the upper bound. For each θ, by slightly abusing the notation, define

$$S_n(\theta) = \sup_\gamma S_n(\theta, \gamma), \quad \tau_1(\theta) = \inf\{n : S_n(\theta) < -B \textit{ or } S_n(\theta) > A\}.$$

Then, an upper bound is given by

$$P_{g_\theta}(S_\tau > A) \le P_{g_\theta}(S_{\tau_1(\theta)}(\theta) > A) \le \kappa A^{\alpha_0} \sum_{n=1}^\infty \int_\Gamma P_{g_\theta}(S_n(\theta, \gamma) > A - 1) d\gamma. \tag{5.2}$$

The last step follows from the fact that the right-hand side is precisely the type I error probability of the simple null g_θ versus composite alternative $\{h_\gamma : \gamma \in \Gamma\}$. We now consider the lower bound. For each given γ and $\theta_* = \arg\inf_{\theta\in\Theta} D_h(\gamma|\theta)$, we have that

$$P_{g_{\theta_*}}(\sup_{\gamma'} \inf_{\theta'} S_\tau(\theta', \gamma') > A) \ge P_{g_{\theta_*}}(\inf_\theta S_{\tau_2(\gamma)}(\theta, \gamma) > A) \cong e^{-A}$$

where $\tau_2(\gamma) = \inf\{n : \inf_\theta S_n(\theta, \gamma) < -B \textit{ or } \sup_\gamma S_n(\theta, \gamma) > A\}$. Once again, the last step is thanks to the type II error proof in Theorem 2.1. □

5.3. Proof of Theorem 2.3

The proof of this theorem uses a change of measure. Suppose that $\xi(x)$ is a stochastic process living on some d-dimensional compact parameter space $x \in \mathcal{X} \subset R^d$. A generic probability measure is denoted by P. The following change of measure helps to compute the tail probability of $\sup_x \xi(x)$. In particular, this change of measure is introduced in two ways. We first define it through the Radon-Nikodym derivative

$$\frac{dQ_b}{dP} = H_b^{-1} \int_{\mathcal{X}} I(\xi(x) > b-1)\mu(dx)$$

where $I(\cdot)$ is the indicator function, μ is a positive measure, and

$$H_b = \int_{\mathcal{X}} P(\xi(x) > b-1)\mu(dx).$$

To better understand this measure Q_b, we provide a procedure generating sample paths of $\xi(x)$ under Q_b. This provides an alternative distributional description of $\xi(x)$ under Q_b. The corresponding sample path generation is given as follows.

1. Sample a random index $x_* \in \mathcal{X}$ according to the density function (with respect to measure μ)
$$q_n(x_*) = P(\xi(x_*) > b-1)/H_b.$$
2. Conditional on the realized x_*, sample $\xi(x_*)$ conditional on $\xi(x_*) > b-1$ under the measure P.
3. Sample the rest of the process $\{\xi(x) : x \neq x_*\}$ conditional on the realization $\xi(x_*)$ under the original measure P.

It is not hard to verify that the above three-step sample path generation is consistent with the Radnon-Nikodym derivative. Some variations of this change of measure will be used in the proof of other lemmas.

Proof of Theorem 2.3. Without loss of generality, we derive the approximation for $E_{g_0}(\tau)$, that is, the true θ is 0. Using the notation in the proof of Theorem 2.2, we consider the limiting process of $S_n(\theta,\gamma)$. To start with, we consider a large constant $M > 0$ and split the expected stopping time

$$E_{g_0}(\tau/B) = E_{g_0}(\tau/B; \tau/B \le M) + E_{g_0}(\tau/B; \tau/B > M). \tag{5.3}$$

Let $\hat\theta_n = \arg\inf_\theta S_n(\theta,\gamma)$ and $\hat\gamma_n = \arg\sup_\gamma S_n(\theta,\gamma)$. Then, as $n \to \infty$, we have the following almost sure convergence, $\hat\theta \to 0$ and $\hat\gamma \to \gamma_0 \triangleq \arg\inf_\gamma D_g(0|\gamma)$. Thus, we have the following weak convergence

$$\{S_{\lfloor Bt \rfloor}(\hat\theta_n, \hat\gamma_n)/B : t \in [0, M]\} \Rightarrow \{-t \times \inf_\gamma D_g(\theta|\gamma) : t \in [0, M]\}$$

where "$\Rightarrow$" is weak convergence. Thus, the first term is approximated by

$$E_{g_0}(\tau/B; \tau/B \le M) \to 1/E_{g_0}\{\xi(\gamma_0)\} = 1/\inf_\gamma D_g(0|\gamma) \quad \text{as } B \to \infty. \tag{5.4}$$

In what follows, we show that the second term $E_{g_\theta}(\tau/B; \tau/B > M) \to 0$ as $B \to \infty$ for M sufficiently large. Let $\tau' = \inf\{n : \sup_\gamma S_n(0,\gamma) < -B\}$. We observe that $\tau' \ge \tau$ and thus it is sufficient to bound $E_{g_\theta}(\tau'/B; \tau'/B > M)$. For each $\lambda > 0$, we consider the probability $P_{g_0}(\tau' > \lambda B)$. Notice that $S_n(0,\gamma)$ has a negative drift that is bounded from the above by $-\varepsilon$ and thus $\sup_\gamma E\{S_{\lambda B}(0,\gamma)\} < -\varepsilon\lambda B$. For $\lambda > M$ with M sufficiently large, we have that $\sup_\gamma E\{S_{\lambda B}(0,\gamma)\} < -B - \varepsilon\lambda B/2$. Note that

$$P_{g_0}(\tau' > \lambda B) \leq P_{g_0}(\sup_\gamma S_{\lambda B}(0,\gamma) \geq -B).$$

The last issue is to provide a bound of $P_{g_0}(\sup_\gamma S_{\lambda B}(0,\gamma) \geq -B)$. We consider the change of measure

$$\frac{dQ_{-B}}{dP_{g_0}} = H_{-B}^{-1} \int_\Gamma I(S_{\lambda B}(0,\gamma) \geq -B-1) d\gamma$$

where $H_{-B} = \int P_{g_0}(S_{\lambda B}(0,\gamma) \geq -B-1) d\gamma$ and thus

$$P_{g_0}\left(\sup_\gamma S_{\lambda B}(0,\gamma) \geq -B\right) \leq P_{g_0}\left(\sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| \geq e^{(\lambda B)^\beta}\right) + P_{g_0}\left(S_{\lambda B}(0,\gamma) \geq -B; \sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| < e^{(\lambda B)^\beta}\right).$$

The first term is bounded by

$$P_{g_0}\left(\sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| \geq e^{(\lambda B)^\beta}\right) \leq \lambda B e^{-(\lambda B)^{\alpha\beta}}.$$

We use the change of measure for the second term

$$\begin{aligned} & P_{g_0}\left(S_{\lambda B}(0,\gamma) < -B; \sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| < e^{-(\lambda B)^\beta}\right) \\ = \ & H_{-B} \times E^{Q_{-B}}\left[\int_\Gamma I(S_{\lambda B}(0,\gamma) \geq -B-1) d\gamma; S_{\lambda B}(0,\gamma) \geq -B; \sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| < e^{(\lambda B)^\beta}\right]. \end{aligned}$$

By means of standard large deviations analysis,

$$H_{-B} \leq e^{-\varepsilon_0 \lambda B}.$$

For the expectation, on the set $\{S_{\lambda B}(0,\gamma) \geq -B\}$, there exists at least one γ_0 such that $S_{\lambda B}(0,\gamma_0) \geq -B$. In addition, the derivative of $S_{\lambda B}(0,\gamma)$ is bounded by $\lambda B e^{(\lambda B)^\beta}$. Thus, we have a lower bound

$$\int_\Gamma I(S_{\lambda B}(0,\gamma) \geq -B-1) d\gamma \geq \delta_0 \lambda^d B^d e^{d(\lambda B)^\beta}.$$

Plugging the above bound back, we have that

$$P_{g_0}\left(S_{\lambda B}(0,\gamma) < -B; \sup_{k\leq\lambda B} |\partial\xi(0,\gamma)| < e^{-(\lambda B)^\beta}\right) \leq e^{-\varepsilon_0 \lambda B/2}$$

and with λ sufficiently large

$$P_{g_0}(\tau' > \lambda B) \leq e^{-\varepsilon_0 \lambda B/2}.$$

With the above bound, we have that

$$E_{g_\theta}(\tau'/B; \tau'/B > M) = o(1)$$

as $B \to 0$. Together with the approximation in (5.4), we put this estimate back to (A.3) and conclude the proof. □

6. CONCLUSION

In this paper, we study the asymptotic properties of the generalized sequential probability ratio test for the composite null hypothesis against composite alternative hypothesis. We derived the exponential decay rate of the maximal type I and type II error probabilities as the crossing levels tend to infinity. In particular, we show that these two probabilities decay to zero at rate e^{-A} and e^{-B}, respectively, which are the same as those of the classic sequential probability ratio test. With such approximations, we are able to establish the asymptotic optimality of the generalized SPRT, that is, it admits asymptotically the shortest expected sample size among all the sequential tests with the same maximal type I and type II error probabilities. These results serve as a natural extension to those of the classic optimality results for the sequential probability ratio test.

7. ACKNOWLEDGEMENTS

We are grateful to the editors and the referee for providing valuable comments. This work is supported in part by NSF SES-1323977, DMS-1308566, NIH R37GM047845, and Army Research Laboratory W911NF-14-1-0020.

REFERENCES

R.J. Adler, J.H. Blanchet, and J. Liu. Efficient monte carlo for large excursions of gaussian random fields. *Annals of Applied Probability*, 22(3):1167–1214, 2012.

Jay Bartroff and Tze Leung Lai. Efficient adaptive designs with mid-course sample size adjustment in clinical trials. *Statistics in medicine*, 27(10):1593–1611, 2008.

Jay Bartroff, Matthew Finkelman, and Tze Leung Lai. Modern sequential analysis and its applications to computerized adaptive testing. *Psychometrika*, 73(3):473–486, 2008.

Jay Bartroff, Tze Leung Lai, and Mei-Chiung Shih. *Sequential Experimentation in Clinical Trials*, volume 298. Springer, 2013.

Amy Berrington de González and David R Cox. Interpretation of interaction: A review. *The Annals of Applied Statistics*, 1(2):371–385, 2007.

Basilio Braganca Pereira. Separate families of hypotheses. *Encyclopedia of Biostatistics*, 2005.

David R Cox. Tests of separate families of hypotheses. In *Proceedings: Fourth Berkeley Symposium on Mathematical Statistics and Probability*, page 105. University of California Press, 1961.

David R Cox. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 406–424, 1962.

David R. Cox. A return to an old paper:tests of separate families of hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Wassily Hoeffding. Lower bounds for the expected sample size and the average risk of a sequential procedure. *The Annals of Mathematical Statistics*, 31(2):352–368, 1960.

Jack Kiefer and Lionel Weiss. Some properties of generalized sequential probability ratio tests. *The Annals of Mathematical Statistics*, 28(1):57–74, 1957.

Tze Leung Lai. Boundary crossing problems for sample means. *Annals of probability*, 16(1):375–396, 1988.

Tze Leung Lai and Mei-Chiung Shih. Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika*, 91(3):507–528, 2004.

Tze Leung Lai and Limin Zhang. A modification of schwarz's sequential likelihood ratio tests in multivariate sequential analysis. *Sequential Anal.*, 13(2):79–96, 1994.

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

D.Q. Naiman and C. Priebe. Computing scan statistic p values using importance sampling, with applications to genetics and medical image analysis. *Journal of Computational and Graphical Statistics*, 10:296–328, 2001.

Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 231(694–706):289–337, 1933.

I.V. Pavlov. A sequential procedure for testing many composite hypotheses. *Theory of Probability and its Applications*, 21(1):138–142, 1987.

I.V. Pavlov. Sequential procedure of testing composite hypotheses with applications to the kiefer-weiss problem. *Theory of Probability and its Applications*, 35(2):280–292, 1990.

M. Pollak and D. Siegmund. Approximations to the expected sample size of certain sequential tests. *The Annals of Statistics*, 3(6):1267–1282, 1975.

J. Shi, D. Siegmund, and B. Yakir. Importance sampling for estimating p values in linkage analysis. *Journal of the American Statistical Association*, 102(479):929–937, 2007.

Mei-Chiung Shih, Tze Leung Lai, Joseph F Heyse, and Jie Chen. Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Statistics in medicine*, 29(26):2698–2708, 2010.

Quang H Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

Abraham Wald. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16 (2):117–186, 1945.

Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.

A. Other technical proofs

The proofs need some variations of the change of measure Q_b introduced in the previous section. Given that all the calculations for the rest of the proof are under the distribution g_0, we let $P = P_{g_0}$ through out this section. To start with, we introduce two measures that are special cases of the measure in the beginning of Section 5.3.

A change of measure. Define measure Q via the Radon-Nikodym

$$\frac{dQ}{dP} = \frac{1}{H_A}\sum_{n=1}^{\infty}\int_{\Gamma} I(S_n(\gamma) > A-1)d\gamma$$

where $I(\cdot)$ is the indicator function and

$$H_A = \sum_{n=1}^{\infty}\int_{\Gamma} P(S_n(\gamma) > A-1)d\gamma.$$

The measure Q depends on A. To simplify the notation, we omit the index A in notation Q. The sample path generation requires three steps.

1. Sample two random indices (n_*, γ_*) jointly according to the density/mass function

$$q(n_*, \gamma_*) = P(S_{n_*}(\gamma_*) > A-1)/H_A.$$

   Note that n_* is integer-valued and q as a function of n_* is a probability mass function. Furthermore, γ_* is a continuous variable and q as a function of γ_* is a density function.

2. Conditional on the realized n_* and γ_*, sample $S_{n_*}(\gamma_*)$ conditional on $S_{n_*}(\gamma_*) > A-1$ under the measure P.

3. Sample the rest of the process $\{S_n(\gamma) : n \neq n_*, \gamma \neq \gamma_*\}$ conditional on the realization $S_{n_*}(\gamma_*)$ under the original measure P.

A second change of measure. This change of measure is defined for $S_n(\gamma)$ with n fixed and $\gamma \in \Gamma$. Define measure Q_n via the Radon-Nikodym

$$\frac{dQ_n}{dP} = H_n^{-1}\int_{\Gamma} I(S_n(\gamma) > -1)d\gamma$$

where $I(\cdot)$ is the indicator function and

$$H_n = \int_{\Gamma} P(S_n(\gamma) > -1)d\gamma.$$

The corresponding sample path generation is given as follows.

1. Sample two random indices γ_* according to the density function

$$q_n(\gamma_*) = P(S_n(\gamma_*) > -1)/H_n.$$

2. Conditional on the realized γ_*, sample $S_n(\gamma_*)$ conditional on $S_n(\gamma_*) > -1$ under the measure P.

3. Sample the rest of the process $\{S_n(\gamma) : \gamma \neq \gamma_*\}$ conditional on the realization $S_n(\gamma_*)$ under the original measure P.

Proof of Lemma 5.1. We start the proof by deriving a lower bound. Notice that $S_\tau \geq S_\tau(\gamma)$ for all γ. Thus, we have

$$P(\sup_n \sup_\gamma S_n(\gamma) > A) \geq P(S_\tau > A) \geq P(S_{\tau(\gamma)} > A) \cong e^{-A}$$

where $\tau(\gamma) = \inf\{n : S_n(\gamma) < -B \text{ or } S_n(\gamma) > A\}$. The last step in the above display is a classic large deviations result. We now proceed to the derivation of an upper bound of $P(\sup_n \sup_\gamma S_n(\gamma) > A)$ and start with a localization on the set

$$L_A = \cup_{n=1}^{\infty}\{\sup_\gamma |\partial \xi_n(\gamma)| > n^\zeta e^{A^\beta}\}$$

for some $\alpha^{-1} < \beta < 1$ and ζ sufficiently large. According to Condition A3, we have that

$$P(L_A^c) \leq \sum_{n=1}^{\infty} P\{\sup_\gamma |\partial \xi_n(\gamma)| > n^\zeta e^{A^\beta}\} \leq \sum_{n=1}^{\infty} n^{-(\alpha-1)\zeta A^\beta} e^{-A^{\alpha\beta}} = o(e^{-A}).$$

Define

$$\tau_A = \inf\{n : \sup_\gamma S_n(\gamma) > A\}$$

and thus $\sup_n \sup_\gamma S_n(\gamma) > A$ if $\tau_A < \infty$. We now derive an upper bound for $P(\tau_A < \infty, L_A)$ via the change of measure Q as follows

$$P(\sup_n \sup_\gamma S_n(\gamma) > A, L_A) = H_A E^Q\left(\Big[\sum_{n=1}^{\infty}\int_\Gamma I\{S_n(\gamma) > A-1\}d\gamma\Big]^{-1}, \tau_A < \infty, L_A\right).$$

Note that on the set $\{\tau_A < \infty\}$, there exists at least one γ such that $S_{\tau_A}(\gamma) > A$. Furthermore, on the set L_A, the gradient $|\nabla S_{\tau_A}(\gamma)|$ is bound by $e^{A^\beta}\tau_A^{\zeta+1}$. Therefore, we have the following lower bound

$$\sum_{n=1}^{\infty}\int_\Gamma I\{S_n(\gamma) > A-1\}d\gamma \geq \int_\Gamma I\{S_{\tau_A}(\gamma) > A-1\}d\gamma \geq \{e^{dA^\beta}\tau_A^{(\zeta+1)d}\}^{-1}.$$

Thus,

$$P(\sup_n \sup_\gamma S_n(\gamma) > A, L_A) \leq e^{dA^\beta} H_A E^Q(\tau_A^{(\zeta+1)d}; \tau_A < \infty).$$

The last step is to control the moment $E^Q(\tau_A^{(\zeta+1)d})$. Let n_* and γ_* be the random indices generated from Step 1 of the three-step sample path generation from Q. Therefore, we split the expectation

$$\begin{aligned} E^Q(\tau_A^{(\zeta+1)d}; \tau_A < \infty) &\leq E^Q(\tau_A^{(\zeta+1)d}; \tau_A \leq n_*) + E^Q(\tau_A^{(\zeta+1)d}; \tau_A < \infty, n_* < \tau_A < \infty) \\ &\leq E^Q\{n_*^{(\zeta+1)d}\} + E^Q(\tau_A^{(\zeta+1)d}; \tau_A < \infty, n_* < \tau_A < \infty) \\ &\leq O(A^{(\zeta+1)d}) + E^Q(\tau_A^{(\zeta+1)d}; n_* < \tau_A < \infty). \end{aligned}$$

We now focus on the last term by starting with the probability

$$Q(\tau_A = n_* + k).$$

Note that $\tau_A > n_*$ implies that $A-1 < S_{n_*}(\gamma_*) < A$ and $S_n(\gamma) < A$ for all $n \le n_*$ and $\gamma \in \Gamma$. Therefore, we have

$$Q(\tau_A = n_* + k) \le P(\sup_\gamma S_k(\gamma) > 0).$$

To obtain an estimate of the above probability, we use the change of measure Q_n

$$\begin{aligned} & P\Big[\sup_\gamma S_k(\gamma) > 0; \cup_{n=1}^k \{\sup_\gamma |\partial \xi_n(\gamma)| > e^{k^\beta}\}\Big] \\ &\quad = H_k \times E^{Q_k}\left[(\int_\Gamma I(S_k(\gamma) > -1) d\gamma)^{-1}; \sup_\gamma S_k(\gamma) > 0, \cup_{n=1}^k \{\sup_\gamma |\partial \xi_n(\gamma)| > e^{k^\beta}\}\right] \end{aligned}$$

and

$$P(\cup_{n=1}^k \{\sup_\gamma |\partial \xi_n(\gamma)| > e^{k^\beta}\}) \le k e^{-k^{\alpha\beta}}. \tag{A.1}$$

For the normalizing constant, we have that

$$H_k = O(e^{-\varepsilon_0 k}).$$

For the integral $\int_\Gamma I(S_k(\gamma) > -1)d\gamma$ inside the expectation, on the set $\{\sup_\gamma S_k(\gamma) > 0\}$, there exists at least one γ_0 such that $S_k(\gamma_0) > 0$. Furthermore, the derivative is bounded from the above by e^{k^β}. Thus, the integral is bounded from below by

$$\int_\Gamma I(S_k(\gamma) > -1) d\gamma \ge \delta_0 k^{-d} e^{-dk^\beta}.$$

Thus, we have

$$P(\sup_\gamma S_k(\gamma) > 0; \cup_{n=1}^k \{\sup_\gamma |\partial \xi_n(\gamma)| > e^{k^\beta}\}) = O(e^{-\varepsilon_0 k/2}). \tag{A.2}$$

We put together (A.1) and (A.2) and obtain that

$$Q(\tau_A = n_* + k) \le P(\sup_\gamma S_k(\gamma) > 0) = O(k e^{-k^{\alpha\beta}} + e^{-\varepsilon_0 k/2}).$$

Therefore, we have that

$$E^Q(\tau_A^{(\zeta+1)d}; n_* < \tau_A < \infty) = O(E(n_*^{(\zeta+1)d})) = O\{A^{(\zeta+1)d}\}.$$

Thereby, we conclude the proof. □

Proof of Lemma 5.2. We now prove an important fact that $H_A \cong e^{-A}$. Recall the notation $\xi(\gamma) = \log h_\gamma(X) - \log g_0(X)$. For each pair (n, γ), we consider the probability $P(S_n(\gamma) > A - 1)$. For each $\varepsilon > 0$ small enough but not changing with A, we approximate the tail probability via large deviations theory stated as follows. Let $\varphi_\gamma(\theta) = \log[E\{e^{\theta\xi(\gamma)}\}]$ and the rate function is

$$P\{S_n(\gamma) > A - 1\} \le e^{-nI(n,\gamma)}$$

where the rate function is $I(n,\gamma) = \theta_* \frac{A-1}{n} - \varphi_\gamma(\theta_*)$ and θ_* solves identity $\varphi'_\gamma(\theta_*) = \frac{A-1}{n}$. For each given γ, $n \times I(n,\gamma)$ is minimized at $n(\gamma) = (A-1)/E_{h_\gamma}\{\xi(\gamma)\}$ and $\min_n n \times I(n,\gamma) = A - 1$. Thus, we have that

$$P\{S_{n(\gamma)}(\gamma) > A - 1\} \le e^{-A+1}.$$

We switch the order of summation and integral by taking the sum with respect to n first. We derive the upper bound of H_A by splitting the summation (for some $M = \kappa_1 / \min_\gamma D_h(\gamma|0)$ and κ_1 large)

$$\sum_{n=1}^{\infty} P(S_n(\gamma) > A-1) = \sum_{n=1}^{MA} P(S_n(\gamma) > A-1) + \sum_{n=MA+1}^{\infty} P(S_n(\gamma) > A-1).$$

Therefore, the first term is bounded by

$$\sum_{n=1}^{MA} P(S_n(\gamma) > A-1) \leq MAe^{-A+1}.$$

Notice that, as $n/A \to \infty$, the rate function $I(n,\gamma) \to -\inf_\theta \varphi_\gamma(\theta) > 0$. Therefore, the large deviations approximation becomes

$$-\frac{1}{n} \log P\{S_n(\gamma) > A-1\} \to \inf_\theta \varphi_\gamma(\theta) > I(n(\gamma),\gamma)$$

as $n/A \to \infty$ and $A \to \infty$. Therefore, if we choose κ_1 sufficiently large depending and $M = \kappa_1 / \min_\gamma D_h(\gamma|0)$, then the second term is

$$\sum_{n=MA+1}^{\infty} P(S_n(\gamma) > A-1) \cong \sum_{n=MA+1}^{\infty} e^{-n \inf_\theta \varphi_\gamma(\theta)} = o(e^{-A})$$

and therefore $\sum_{n=1}^{\infty} P(S_n(\gamma) > A-1) \leq (MA+1)e^{-A}$. Since Γ is a compact set, then with κ_0 sufficiently large

$$H_A = \int_{\gamma\in\Gamma} \sum_{n=1}^{\infty} P(S_n(\gamma) > A-1) d\gamma \leq \kappa_0 mes(\Gamma) Ae^{-A} / \min_\gamma D_h(\gamma|0)$$

□

Proof of Lemma 3.1. We first switch the sum and integration

$$H_{A,\theta} = \int_\Gamma \sum_{n=1}^{\infty} P_{g_\theta}\{S_n(\theta,\gamma) > A\} d\gamma.$$

Furthermore, notice the following approximation (for some κ large)

$$P_{g_\theta}\{\sup_n S_n(\theta,\gamma) > A\} \leq \sum_{n=1}^{\infty} P_{g_\theta}\{S_n(\theta,\gamma) > A\} \leq A^\kappa P_{g_\theta}\{\sup_n S_n(\theta,\gamma) > A\}.$$

The first inequality is due to the inclusion and exclusion formula and the second step can be obtained by standard large deviations analysis, Condition B2 and B3. In addition, the choice of κ is independent of θ and γ. Then, it is sufficient to show that

$$\limsup_{A\to\infty} \sup_{\theta\in\Theta} \frac{1}{A} \log \Big[\int_\Gamma P_{g_\theta}\{\sup_n S_n(\theta,\gamma) > A\} d\gamma \Big] \leq -1.$$

We now consider the tail probability $P_{g_\theta}\{\sup_n S_n(\theta,\gamma) > A\}$ for each θ and γ. The tail probability has a universal upper bound

$$w(\theta,\gamma) \triangleq P_{g_\theta}\{\sup_n S_n(\theta,\gamma) > A\} \leq e^{-A}$$

and the equality holds only when the overshoot is zero. Therefore, we have split the integral for M sufficiently large

$$\int_{|\gamma-\gamma_\theta|<MA^{1/l}} w(\theta,\gamma)d\gamma \leq \kappa_d A^{d/l}e^{-A} + \int_{|\gamma-\gamma_\theta|\geq MA^{1/l}} w(\theta,\gamma)d\gamma \tag{A.3}$$

where κ_d is the volume of the d-dimensional unit ball. We now show that $w(\theta,\gamma)e^A \to 0$ as $|\gamma-\gamma_\theta| \to \infty$. Let $\tau_A = \inf\{n : S_n(\theta,\gamma) > A\}$. We choose M sufficiently large such that $E_{h_\gamma}\{\xi(\theta,\gamma)\} = D_h(\gamma|\theta) > 3A$. Then, the tail probability has the following upper bound

$$\begin{aligned} w(\theta,\gamma) &= E_{h_\gamma}\{e^{-S_{\tau_A}(\theta,\gamma)}; S_{\tau_A}(\theta,\gamma) > A\} \\ &\leq e^{-A}P_{h_\gamma}[\xi_1(\theta,\gamma) < \{D_h(\gamma|\theta)+1\}/2] + E_{h_\gamma}[e^{-S_{\tau_A}(\theta,\gamma)}; \xi_1(\theta,\gamma) > \{D_h(\gamma|\theta)+1\}/2]. \end{aligned}$$

The second term of the above inequality is bounded from the above by

$$e^{-\{D_h(\gamma|\theta)+1\}/2} \leq e^{-\{1+D_h(\gamma_\theta|\theta)+\delta|\gamma-\gamma_\theta|^l\}/2} \leq e^{-A-\varepsilon_0|\gamma-\gamma_\theta|^l}.$$

For the first term, notice that $\xi_1(\theta,\gamma)$ has mean $D_h(\gamma|\theta)$ and bounded second moment. By Chebyshev's inequality (noting that $E_{h_\gamma}\{\xi_1(\theta,\gamma)\} = D_h(\gamma|\theta)$), we have that

$$P_{h_\gamma}[\xi_1(\theta,\gamma) < \{D_h(\gamma|\theta)+1\}/2] = O(1)A^{-2}D_h(\gamma|\theta)^{-2} \leq O(1)A^{-2}|\gamma-\gamma_\theta|^{-2l}$$

Therefore, the integral has an upper bound

$$\int_{|\gamma-\gamma_*|\geq MA^{1/l}} w(\theta,\gamma)d\gamma \leq \int_{|\gamma-\gamma_*|\geq MA^{1/l}} O(1)A^{-2}e^{-A}|\gamma-\gamma_\theta|^{-2l}d\gamma.$$

Since $l > (d+1)/2$, the above integral is $O(A^{-2}e^{-A})$. We insert this bound back to (A.3) and obtain that $\int_\Gamma w(\theta,\gamma)d\gamma = O(A^{d/l}e^{-A})$ and conclude the proof. □