Sequential Analysis, 31: 238–262, 2012

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474946.2012.665730

Sequential Bonferroni Methods for Multiple Hypothesis Testing with Strong Control of Family-Wise Error Rates I and II

Shyamal K. De and Michael Baron
Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas, USA

Abstract: Sequential procedures are developed for simultaneous testing of multiple hypotheses in sequential experiments. Proposed stopping rules and decision rules achieve strong control of both family-wise error rates I and II. The optimal procedure is sought that minimizes the expected sample size under these constraints. Bonferroni methods for multiple comparisons are extended to sequential setting and are shown to attain an approximately 50% reduction in the expected sample size compared with the earlier approaches. Asymptotically optimal procedures are derived under Pitman alternative.

Keywords: Asymptotic optimality; Family wise error rate; Multiple comparisons; Pitman alternative; Sequential probability ratio test; Stopping rule.

Subject Classifications: 62L; 62F03; 62F05; 62H15.

1. INTRODUCTION

The problem of multiple hypothesis testing often arises in sequential experiments such as sequential clinical trials with more than one endpoint (see Glaspy et al., 2001, Jennison and Turnbull, 1993, O'Brien, 1984, Pocock et al., 1987, and others), multichannel change-point detection (Tartakovsky and Veeravalli, 2004, Tartakovsky et al., 2003), acceptance sampling with different criteria of acceptance (Baillie, 1987, Hamilton and Lesperance, 1991), etc. In such experiments, it is necessary to find a statistical answer to each posed question by testing *each individual hypothesis* instead of combining all the null hypotheses and giving one global answer to the resulting composite hypothesis. This article develops sequential

Received December 28, 2011; Revised January 29, 2012; Accepted February 6, 2012
Recommended by N. Mukhopadhyay
Address correspondence to Shyamal K. De, Department of Mathematical Sciences, FO 35, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021, USA; E-mail: shyamalkd@gmail.com

methods that result in accepting or rejecting each null hypothesis at the final stopping time.

Non-sequential methods of multiple comparisons are already well developed. Proposed Procedures include Bonferroni and Sidak testing, Holm and Hommel step-down methods controlling family-wise error rate, Benjamini-Hochberg step-up and Guo-Sarkar method controlling false discovery rate, and others, see Sidak (1967), Holm (1979), Benjamini and Hochberg (1995), Benjamini et al. (2004), and Sarkar (2007).

There is a growing demand for *sequential* analogues of these procedures. Furthermore, sequential tests of single hypotheses can be designed to control both probabilities of Type I and Type II errors. Generalizing to multiple testing, we develop sequential schemes to control both family-wise error rates I and II in the strong sense.

In the literature of sequential multiple comparisons, two types of testing problems are commonly understood as sequential multiple hypothesis testing. One of them is to compare effects $\theta_1, \ldots, \theta_d$ of $d > 1$ treatments with the goal of finding the best treatment, as in Edwards and Hsu (1983), Edwards (1987), Hughes (1993), Betensky (1996), and (Jennison and Turnbull, 2000, Chap. 16). This involves testing of a composite null hypothesis $H_0 : \theta_1 = \cdots = \theta_d$ against $H_A : H_0$ not true.

The other type of problems is to classify sequentially observed data into one of the desired models; see Armitage (1950), Baum and Veeravalli (1994), Novikov (2009), and Darkhovsky (2011). In this case, the parameter of interest θ is classified into one of d alternative hypotheses, namely, $H_1 : \theta \in \Theta_1$ vs. , $\ldots$, vs. $H_d : \theta \in \Theta_d$.

This article deals with a different and more general problem of testing d individual hypotheses sequentially.

1.1. Problem Formulation

Suppose a sequence of iid random vectors $\{\boldsymbol{X}_n, n = 1, 2, \ldots\}$ is observed, where n is a sampled unit and $\boldsymbol{X}_n = \left(X_n^{(1)}, \ldots, X_n^{(d)}\right) \in \mathbb{R}^d$ is a set of d measurements, possibly dependent, made on unit n. Let the jth measurement on each sampled unit have a marginal density $f_j(x \mid \theta^{(j)})$ with respect to a reference measure μ_j, $j = 1, \ldots, d$.

While vectors $\boldsymbol{X}_n$ are observed *sequentially*, the d tests are conducted simultaneously

$$\begin{array}{c} H_0^{(1)} : \theta^{(1)} \in \Theta_{01} \quad \text{vs.} \quad H_A^{(1)} : \theta^{(1)} \in \Theta_{A1}, \\ \cdots \\ H_0^{(d)} : \theta^{(d)} \in \Theta_{0d} \quad \text{vs.} \quad H_A^{(d)} : \theta^{(d)} \in \Theta_{Ad}. \end{array} \tag{1.1}$$

Denote the index set of true null hypotheses by $J_0 \in \{1, \ldots, d\}$ and the set of false null hypotheses by $J_A = \{1, \ldots, d\} \backslash J_0$. Further, let $\psi_j = \psi_j(\{\boldsymbol{X}_n\})$ equal 1 if $H_0^{(j)}$ is rejected and equal 0 otherwise. Then *the family-wise error rate I*

$$\text{FWER I}(J_0) = P\left(\sum_{j \in J_0} \psi_j > 0 \mid J_0 \neq \varnothing\right)$$

is defined as the probability to reject at least one true null hypothesis (and therefore commit at least one Type I error), *the family-wise error rate II*

$$\text{FWER II}(J_A) = P\left(\sum_{j\in J_A}(1-\psi_j) > 0 \,|\, J_A \neq \varnothing\right),$$

is the probability of accepting at least one false null hypothesis (and committing at least one Type II error), and *the family-wise power* is the probability of rejecting all the false null hypotheses, FP $= 1 -$ FWER II (e.g., Lee, 2004, Chap. 14).

This article develops stopping rules and multiple decision rules that control both FWER I and FWER II in the *strong sense*, that is,

$$\text{FWER I} = \sup_{J_0\neq\varnothing} \text{FWER I}(J_0) \leq \alpha \quad \text{and} \quad \text{FWER II} = \sup_{J_A\neq\varnothing} \text{FWER II}(J_A) \leq \beta \quad (1.2)$$

for prespecified levels α and β. This task alone is not difficult to accomplish. For example, one can conduct d separate sequential probability ratio tests, one for each null hypothesis, at levels α/d and β/d each. However, this leaves room for optimization. Our goal is to control both family-wise error rates at *the minimum sampling cost*. Namely, an optimization problem of *minimizing the expected sample size* $E(T)$ under the strong control of FWER I and FWER II is investigated.

A battery of one-sided tests $\{H_0^{(j)} : \theta^{(j)} \leq \theta_0^{(j)}$ vs. $H_A^{(j)} : \theta^{(j)} > \theta_0^{(j)}, j = 1, \ldots, d\}$ is considered, which is typical for clinical trials of *safety* and *efficacy*. For example, in a case study of early detection of lung cancer (Lange et al., 1994, Chap. 15), 30,000 men at high risk of lung cancer were recruited sequentially over 12 years for testing three one-sided hypotheses: whether intensive screening improves detectability of lung cancer, whether it reduces mortality due to lung cancer, and whether a surgical treatment is more efficient for the reduction of mortality rate than a nonsurgical treatment. In another study, a crossover trial of chronic respiratory diseases (Pocock et al., 1987, Tang and Geller, 1999) involved 17 patients with asthma or chronic obstructive airways disease to test any harmful effect of an inhaled drug based on four measures of the standard respiratory function: FEV_1, FVC, PEFR, and PI.

1.2. Existing Approaches and Our Objective

Sequential methods for simultaneous testing of multiple hypotheses have recently received a new wave of attention. Jennison and Turnbull (2000, Chap. 15) develop group sequential methods of testing several outcome measures of a clinical trial. Their sampling strategy is to stop at the first time when one of the tests is accepted or rejected. Protecting FWER I conservatively, all of the null hypotheses that are not rejected at the stopping time are accepted. While strongly controlling the rate of falsely rejected null hypotheses, this testing scheme sacrifices power significantly.

Tang and Geller (1999) introduce a closed testing procedure for group sequential clinical trials with multiple endpoints. Their method controls FWER I and yields power that is at least the power obtained by the Jennison–Turnbull procedure.

In a recent work by Bartroff and Lai (2010), conventional methods of multiple testing such as Holm's step-down procedure and closed testing technique have been combined to develop a multistage step-down procedure that also controls FWER I. Their method does not assume any special structure of the hypotheses such as closedness or any correlation structure among different test statistics and can be applied to any group sequential, fully sequential, or truncated sequential experiments.

All of these methods mainly aim at controlling FWER I. To the best of our knowledge, no theory has been developed to control both FWER I and FWER II and also to design a sequential procedure that yields an optimal expected sampling cost under these constraints.

The next section extends Bonferroni methods for multiple comparisons to sequential experiments. Existing stopping rules are evaluated and modified to achieve lower family-wise error rates or lower expected sampling cost. In Section 3, asymptotically optimal testing procedures are derived under Pitman alternative. Controlling FWER I and FWER II in the strong sense, they attain the asymptotically optimal rate of the expected sample size. Finite-sample performances of the proposed schemes are evaluated in a simulation study in Section 4. All lengthy proofs are moved to the Appendix.

2. SEQUENTIAL BONFERRONI METHODS

In this section, we propose and compare four sequential approaches for testing d hypotheses controlling FWER I and FWER II that are based on the Bonferroni inequality.

Assuming the *monotone likelihood ratio (MLR)* property of the likelihood ratios and introducing a practical region of indifference $(\theta_0^{(j)}, \theta_A^{(j)})$ for parameter $\theta^{(j)}$, the one-sided testing considered in Section 1 becomes equivalent to the test of

$$H_0^{(j)} : \theta^{(j)} = \theta_0^{(j)} \quad \text{vs.} \quad H_A^{(j)} : \theta^{(j)} = \theta_A^{(j)} \quad \text{for all } j = 1, \ldots, d.$$

Following Neyman–Pearson arguments for the uniformly most powerful test and Wald's sequential probability ratio test, our test procedures are based on log-likelihood ratios

$$\Lambda_n^{(j)} = \sum_{i=1}^{n} \ln \frac{f_j(X_i^{(j)} \mid \theta_A^{(j)})}{f_j(X_i^{(j)} \mid \theta_0^{(j)})} = \sum_{i=1}^{n} Z_i^{(j)} \quad \text{for } j = 1, \ldots, d. \tag{2.1}$$

With stopping boundaries a_j and b_j, after n observations, we *reject* $H_0^{(j)}$ if $\Lambda_n^{(j)} \geq a_j$, *accept* $H_0^{(j)}$ if $\Lambda_n^{(j)} \leq b_j$, and *continue sampling* if $\Lambda_n^{(j)} \in (b_j, a_j)$. The SPRT stopping time for the jth test is therefore

$$T_j = \inf \left\{ n : \Lambda_n^{(j)} \notin (b_j, a_j) \right\}. \tag{2.2}$$

If Type I and Type II errors for the jth test are α_j and β_j, Wald's approximation gives the approximate stopping boundaries as

$$a_j \approx \ln \left(\frac{1 - \beta_j}{\alpha_j} \right) \quad \text{and} \quad b_j \approx \ln \left(\frac{\beta_j}{1 - \alpha_j} \right). \tag{2.3}$$

Now, consider any stopping time τ based on the likelihood ratios $\Lambda_n^{(j)}$ and some stopping boundaries (b_j, a_j) such that Type I and Type II errors of the jth test are controlled at levels α_j and β_j respectively for $j = 1, \ldots, d$. The *strong control* of the family-wise error rates I and II at levels α and β follows from the Bonferroni inequality. That is,

$$\text{FWER I} \leq \sum_{j=1}^{d} P(\text{Type I error on jth test } \mid J_0) = \sum_{j=1}^{d} \alpha_j \leq \alpha,$$

$$\text{FWER II} \leq \sum_{j=1}^{d} P(\text{Type II error on jth test } \mid J_0) = \sum_{j=1}^{d} \beta_j \leq \beta$$

for any index set J_0 of true null hypotheses.

2.1. Choice of Stopping and Decision Rules

In this subsection, we discuss stopping and decision rules that appeared in the sequential multiple testing literature and propose their improvements. One multiple decision sequential rule is proposed in Jennison and Turnbull (2000, Chap. 15).

1. **Tmin rule.** Jennison and Turnbull proposed a Bonferroni adjustment for the sequential testing of multiple hypotheses. According to their procedure, sampling is terminated as soon as at least one test statistics $\Lambda_n^{(j)}$ leaves the continue-sampling region (b_j, a_j). This stopping rule that we call *Tmin* is therefore

$$T^{\min} = \min(T_1, \ldots, T_d),$$

where stopping times T_j are defined in (2.2). If $\Lambda_n^{(k)}$ falls in its continue-sampling region at time $T^{\min}$ for some $k \neq j$, then $H_0^{(k)}$ is considered accepted. For the stopping boundaries (2.3), the Tmin rule controls FWER I conservatively. However, due to an early stopping the Tmin rule fails to control FWER II and, hence, does not yield a high family-wise power.

In order to control both FWER I and II, sampling must continue beyond $T^{\min}$.

2. **Incomplete Tmax rule.** For the control of both Type I and Type II error probabilities for testing each parameter $\theta^{(j)}$, we propose to continue sampling the jth coordinate at $n = T_j$ until $\Lambda_n^{(j)} \notin (b_j, a_j)$ and to reject $H_0^{(j)}$ if and only if $\Lambda_{T_j}^{(j)} \geq a_j$. Based on Wald's approximation (2.3), this rule yields approximate control of both Type I and Type II error probabilities at α_j and β_j for the jth test for each $j = 1, \ldots, d$.

This rule that we call *incomplete Tmax* continues sampling until results for all tests are obtained,

$$T^{\max} = \max(T_1, \ldots, T_d) \geq T^{\min}.$$

It is incomplete because for all j with $T_j < T^{\max}$, only a portion of the available data $(X_1^{(j)}, \ldots, X_{T_j}^{(j)})$ is utilized to reach a decision.

Sobel and Wald (1949) used a similar approach for a different sequential problem of classifying the unknown mean of a normal distribution into one of three regions. According to their rule, testing based on the jth likelihood ratio ($j = 1, 2$) stops as soon as it leaves its continue-sampling region.

Using the incomplete Tmax rule, the number of used measurements taken on each sampling unit decreases in the course of this sequential scheme, although the cost of each unit (e.g., patient) is roughly the same. Typically, it is of no extra cost to ask patients to complete the entire questionnaire instead of a portion of it.

Moreover, it may be important to include the entire new observed vectors because it may be necessary to retest some of the hypotheses. Indeed, if one or several test statistics $\Lambda_n^{(j)}$ return to the continue-sampling region after crossing a stopping boundary once, and the corresponding hypothesis is retested based on a larger sample size n, it may further reduce the rate of false discoveries and false non-discoveries.

Therefore, we propose an improvement of the incomplete Tmax decision rule that is based on the same stopping time $T^{\max}$ but utilizes all of the available data.

3. **Complete Tmax rule.** As a plausible improvement over the incomplete Tmax rule, we propose *the complete Tmax* rule that continues sampling *each coordinate* of $\boldsymbol{X}$ and testing the corresponding hypothesis until time $T^{\max}$. The sample size requirement for both complete and incomplete Tmax rules are the same although the complete rule utilizes $(T^{\max} - T_j)$ additional observations for testing the jth hypothesis.

Based on all the available data at time $T^{\max}$, different versions of the terminal decision rules can be considered. Following the SPRT, one can accept or reject $H_0^{(j)}$ depending on whether $\Lambda_{T^{\max}}^{(j)}$ is below the acceptance boundary b_j or above the rejection boundary a_j. However, some of the log-likelihood ratios may fall in the continue-sampling region (b_j, a_j) at $T^{\max}$. An additional boundary $c_j \in [b_j, a_j]$ has to be pre-chosen to decide on the jth test. In Section 4, complete and incomplete Tmax rules are compared by a simulation study; superiority of the complete rule is established with a certain choice of c_j.

Alternatively, we propose a randomized terminal decision rule that is based on a sufficient statistics for $\boldsymbol{\theta} = \left(\theta^{(1)}, \ldots, \theta^{(d)}\right)$ at $T^{\max}$ (unlike the incomplete Tmax rule) and show that it strongly controls FWER I and FWER II.

Theorem 2.1. *The randomized decision rule $\mathscr{D}_j^*$ that rejects $H_0^{(j)}$ at time $T^{\max}$ with probability $p_j^* = P\left(\Lambda_{T_j}^{(j)} \geq a_j \mid \boldsymbol{\Lambda}_{T^{\max}}, T^{\max}\right)$ strongly controls Type I and Type II error probabilities at levels α_j and β_j, respectively.*

Proof. Let $\psi_j = I\left(\Lambda_{T_j}^{(j)} \geq a_j\right)$ be the jth test function for the incomplete Tmax decision rule $\mathscr{D}_j$ and $L\left(\theta^{(j)}, \mathscr{D}_j\right) = I\left(\theta^{(j)} = \theta_0^{(j)}, \text{reject } H_0^{(j)}\right) + I\left(\theta^{(j)} = \theta_A^{(j)}, \text{accept } H_0^{(j)}\right)$ be the loss function for jth test. Also, note that $\boldsymbol{\Lambda}_n = \left(\Lambda_n^{(1)}, \ldots, \Lambda_n^{(d)}\right)$ is a sufficient statistic for $\boldsymbol{\theta}$ for any n, and therefore, $(\boldsymbol{\Lambda}_{T^{\max}}, T^{\max})$ is a sufficient statistic (Govindarajulu, 1987, Sect. 4.3). Hence, $p_j^* = E\left(\psi_j \mid \boldsymbol{\Lambda}_{T^{\max}}, T^{\max}\right)$ is free of $\boldsymbol{\theta}$. By the Rao–Blackwell theorem for randomized rules in Berger (1985, Sect. 36), we have $EL\left(\theta^{(j)}, \mathscr{D}_j^*\right) = EL\left(\theta^{(j)}, \mathscr{D}_j\right)$. Since the incomplete Tmax decision rule $\mathscr{D}_j$ controls Type I and Type II error probabilities at levels α_j and β_j, the randomized rule $\mathscr{D}_j^*$ controls error probabilities at the same levels. □

4. **Intersection rule.** To avoid sampling termination in the continue-sampling region of some coordinate and yet to use all coordinates of the already sampled

random vectors, we propose *the intersection rule*. According to it, sampling continues until *all* log-likelihood ratios leave their respective continue-sampling regions,

$$T^{\text{int}} = \inf \left\{ n : \bigcap_{j=1}^{d} \Lambda_n^{(j)} \notin (b_j, a_j) \right\}.$$

This is the first moment when SPRT-type decisions are obtained for all d tests simultaneously. Clearly, $T^{\text{int}} \geq T^{\max}$ a.s. The difference, however, is likely to be small, and it appears only when a log-likelihood ratio statistic for some j, having crossed the test boundary once, curved back into the continue-sampling region (b_j, a_j).

The following lemma gives sufficient conditions for T^{int} to be a proper stopping time.

Lemma 2.1. *If for each $j = 1, \ldots, d$, the Kullback–Leibler information numbers*

$$K(\theta^{(j)}, \theta^{(j')}) = E_{\theta^{(j)}} \log \left\{ f_j(X_j \mid \theta^{(j)}) / f_j(X_j \mid \theta^{(j')}) \right\}$$

are strictly positive and finite for $\theta^{(j)} \neq \theta^{(j')}$, then $P(T^{int} < \infty) = 1$.

This lemma follows from the weak law of large numbers; see De and Baron (2012) for the complete proof. Consequently, $T^{\max}$ is also a proper stopping time.

The strong control of FWER I and II by the intersection rule is shown in the following two statements.

Lemma 2.2. *Let τ be any stopping time with respect to $\sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ satisfying*

$$P(\Lambda_\tau^{(j)} \in (b_j, a_j)) = 0 \quad \textit{for } a_j > 0, b_j < 0 \quad \textit{for all } j \leq d.$$

with the terminal decision rule of rejecting $H_0^{(j)}$ if and only if $\Lambda_\tau^{(j)} \geq a_j$. For such a test

$$P(\textit{Type I error on the jth test}) = P(\Lambda_\tau^{(j)} \geq a_j \mid H_0^{(j)}) \leq e^{-a_j} \quad \textit{for all } j = 1, \ldots, d,$$

$$P(\textit{Type II error on the jth test}) = P(\Lambda_\tau^{(j)} \leq b_j \mid H_A^{(j)}) \leq e^{b_j} \quad \textit{for all } j = 1, \ldots, d.$$

The proof is given in De and Baron (2012).

Corollary 2.1. *With stopping boundaries $a_j = -\ln \alpha_j$ and $b_j = \ln \beta_j$ for $j = 1, \ldots, d$, the intersection rule strongly controls FWER I and FWER II at levels $\alpha = \sum \alpha_j$ and $\beta = \sum \beta_j$, respectively.*

Proof. As a result of Lemma 2.2, Type I and Type II error probabilities are controlled at levels α_j and β_j for $a_j = -\ln \alpha_j$ and $b_j = \ln \beta_j$. The corollary is established by applying the Bonferroni inequality and choosing (α_j, β_j) such that $\sum_{j=1}^{d} \alpha_j \leq \alpha$ and $\sum_{j=1}^{d} \beta_j \leq \beta$. □

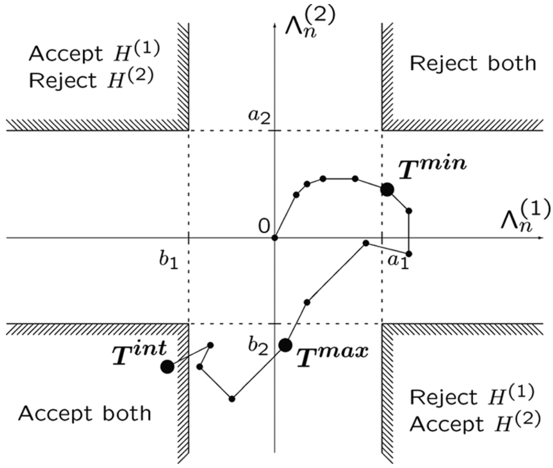

Figure 1. Graphical display of Tmin, Tmax (complete and incomplete), and intersection rule, $d = 2$.

The three stopping times compared in this subsection are displayed in Figure 1, based on a sample path of a two-dimensional log-likelihood ratio random walk. As seen in this figure, stopping time $T^{\min}$ occurs when *one* of the coordinates of $\boldsymbol{\Lambda}_n$ crosses the corresponding stopping boundary, $T^{\max}$ occurs when *all* of the coordinates of $\boldsymbol{\Lambda}_n$ have crossed their boundaries at least once, and, finally, T^{int} occurs when all of the coordinates are beyond the boundaries simultaneously. In the case of $T^{\min}$ and T^{int}, the crossed boundaries determine the decision rules. For the data in Figure 1, hypothesis $H_0^{(1)}$ is rejected at time $T^{\min}$, but later, both $H_0^{(1)}$ and $H_0^{(2)}$ are accepted at T^{int}. At time $T^{\max}$, the incomplete Tmax rule rejects $H_0^{(1)}$ and accepts $H_0^{(2)}$ based on the first crossings. However, since $\Lambda_{T^{\max}}^{(1)} \in (b_1, a_1)$ lands in the continue-sampling region, other decision rules can also be chosen, including the complete Tmax rule with the decision boundary c_j or the randomized rule D^*.

3. ASYMPTOTIC OPTIMALITY UNDER PITMAN ALTERNATIVE

In this section, we give an asymptotically optimal solution to the problem considered in the previous sections. Namely, we find the optimal allocation of individual error probabilities α_j and β_j that yields the asymptotically lowest rate of the expected sample size subject to the constraints on family-wise error rates I and II in the strong sense. Asymptotic analysis of the considered stopping times is based on the asymptotics of likelihood ratio and its expectation (Kullback–Leibler information) under Pitman alternative.

The rigorous formulation of this optimization problem under *Pitman alternative* is as follows. Consider an array of sequences of observed random vectors

$$\begin{gathered} \boldsymbol{X}_{11}, \boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{i1} \ldots \\ \ldots \\ \boldsymbol{X}_{1v}, \boldsymbol{X}_{2v}, \ldots, \boldsymbol{X}_{iv} \ldots \\ \ldots \end{gathered}$$

where $\boldsymbol{X}_{iv} = \left(X_{iv}^{(1)}, \ldots, X_{iv}^{(d)}\right)$. Observations $X_{iv}^{(j)}$ in the vth row have a marginal density (pmf or p.d.f.) $f_j(\,.\,|\,\theta_v^{(j)})$ for all $j = 1, \ldots, d$. A battery of d simultaneous tests

$$H_{0v}^{(j)} : \theta_v^{(j)} = \theta_0^{(j)} \quad \text{vs.} \quad H_{Av}^{(j)} : \theta_v^{(j)} = \theta_0^{(j)} + \epsilon_{jv} \quad \text{for } j = 1, \ldots, d,$$

is considered, where $\min_{j \le d} \epsilon_{jv} \to 0$ as $v \to \infty$. The jth log-likelihood ratio statistic for the vth row is then

$$\Lambda_{n,v}^{(j)} = \sum_{i=1}^{n} \ln \frac{f_j(X_{iv}^{(j)} \mid \theta_0^{(j)} + \epsilon_{jv})}{f_j(X_{iv}^{(j)} \mid \theta_0^{(j)})} = \sum_{i=1}^{n} Z_{i,v}^{(j)} \quad \text{for } j = 1, \ldots, d.$$

Based on this statistic, let $T_{jv} = \inf\left\{n : \Lambda_{n,\epsilon_{jv}}^{(j)} \notin (B_{jv}, A_{jv})\right\}$ be the single-hypothesis Wald's SPRT stopping time for testing $H_{0v}^{(j)}$ vs. $H_{Av}^{(j)}$ with stopping boundaries

$$A_{jv} = \ln \frac{1 - \beta_{jv}}{\alpha_{jv}} \quad \text{and} \quad B_{jv} = \ln \frac{\beta_{jv}}{1 - \alpha_{jv}},$$

where α_{jv} and β_{jv} are the individual Type I and Type II error probabilities allocated to the jth test. Also, let $T_v^{\max} = (T_{1v}, \ldots, T_{dv})$ be the corresponding stopping time for incomplete and complete Tmax rules.

In this section, we study the asymptotic behavior of stopping times $T_{1v}, \ldots, T_{dv}$, and $T_v^{\max}$ as $v \to \infty$ and also find the asymptotically optimal allocation of FWER I and FWER II among the d tests in order to minimize the *weighted expected stopping time*

$$E_\pi\left(T_v^{\max}\right) = \sum_{J_0 \subset \{1,\ldots,d\}} \pi(J_0) E\left(T_v^{\max} \mid J_0\right), \tag{3.1}$$

where $\pi(J_0)$ are weights or prior probabilities assigned to all possible combinations J_0 of true null hypotheses.

This problem implies the asymptotic optimization of stopping boundaries $\left\{(B_{jv}, A_{jv}) : j = 1, \ldots, d\right\}$, or, equivalently, the optimal allocation of individual error probabilities $\left\{(\alpha_{jv}, \beta_{jv}) : j = 1, \ldots, d\right\}$ among d tests. Asymptotic analysis presented here requires the following regularity conditions on the families of distributions $\left\{f_j(.\,|\,\theta^{(j)}) : 1 \le j \le d\right\}$.

3.1. Regularity Conditions

Simplifying notations, we omit index j from $\theta^{(j)}, H^{(j)}, f_j, \mu_j$, etc., and consider the following problem. Suppose $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) random variables with density $f(.\,|\,\theta)$. Consider testing of $H_0 : \theta = \theta_0$ vs. $H_A : \theta \ge \theta_A$, where $\theta_A = \theta_0 + \epsilon$. Define a log-likelihood ratio for one variable as $Z_\epsilon = \ln\{f(X \mid \theta_0 + \epsilon)/f(X \mid \theta_0)\}$. Assume that

R1. The density $f(.\,|\,\theta)$ is identifiable; that is, $\theta \ne \theta'$ implies $P_\theta\{f(X_1 \mid \theta) \ne f(X_1 \mid \theta')\} > 0$.

R2. Identity $\int f(x \mid \theta) dx = 1$ can be differentiated twice under the integral sign.

R3. Derivatives $\ell''(x, \theta)$ and $\ell'''(x, \theta)$ of log-likelihood $\ell(x, \theta) = \ln f(x \mid \theta)$ are bounded by integrable functions in some neighborhood of θ_0. That is, there are functions $L_1(x)$ and $L_2(x)$ and $\delta > 0$ such that $|\theta - \theta_0| < \delta$ implies $E_\theta L_j(X_1) < \infty$ for $j = 1, 2$, and

$$|\ell''(x, \theta)|^2 \leq L_1(x) \quad \text{and} \quad |\ell'''(x, \theta)| \leq L_2(x) \quad \mu\text{-a.s.}$$

R4. The moment generating function $M(s, \epsilon) = M_{Z_\epsilon}(s) = E \exp\{s Z_\epsilon\}$ of Z_ϵ exists for some $s > 0$, and for this s, $\frac{\partial^3}{\partial \epsilon^3} M(s, \epsilon)$ is bounded for all ϵ within a neighborhood of 0.

Conditions R1, R2, and the first part of R3 correspond to conditions A_0, RR3, and R of Borovkov (1997, Chap. 2, Sects. 16, 26, and 31).

3.2. Asymptotics of Log-Likelihood Ratio

Under regularity conditions R1–R4, the log-likelihood ratio Z_ϵ is expanded as

$$Z_\epsilon = \ln f(X \mid \theta_0 + \epsilon) - \ln f(X \mid \theta_0) = \epsilon \ell'(X, \theta_0) + \frac{\epsilon^2}{2} \ell''(X, \theta_0) + R_{2,\epsilon},$$

by the Taylor theorem, where the Lagrange form of the remainder is

$$R_{2,\epsilon} = \frac{\epsilon^3}{6} \ell'''(X, \tilde{\theta}) \quad \text{and} \quad \tilde{\theta} \in [\theta_0, \theta_A].$$

By regularity condition R3,

$$E\left(R_{2,\epsilon}\right) \leq \frac{\epsilon^3}{6} E\left(L_2(X)\right) = O(\epsilon^3).$$

Therefore, under H_0,

$$\begin{aligned} E_{\theta_0}(Z_\epsilon) &= \epsilon E_{\theta_0}\left(\ell'(X, \theta_0)\right) + \frac{\epsilon^2}{2} E_{\theta_0}\left(\ell''(X, \theta_0)\right) + O(\epsilon^3) \\ &= -\frac{\epsilon^2}{2} I(\theta_0) + O(\epsilon^3). \end{aligned} \tag{3.2}$$

Under H_A, using conditions R1–R4,

$$\begin{aligned} E_{\theta_A}(Z_\epsilon) &= -E_{\theta_A}\left[\ln \frac{f(X \mid \theta_A - \epsilon)}{f(X \mid \theta_A)}\right] \\ &= \epsilon E_{\theta_A}\left(\ell'(X, \theta_A)\right) - \frac{\epsilon^2}{2} E_{\theta_A}\left(\ell''(X, \theta_A)\right) + O(\epsilon^3) \\ &= \frac{\epsilon^2}{2} I(\theta_A) + O(\epsilon^3). \end{aligned} \tag{3.3}$$

Next, $E_{\theta_0}(Z_\epsilon) = -K(\theta_0, \theta_0 + \epsilon)$ and $E_{\theta_A}(Z_\epsilon) = K(\theta_A, \theta_A - \epsilon)$, where

$$K(\theta_0, \theta_A) = \int \ln \frac{f(x \mid \theta_0)}{f(x \mid \theta_A)} f(x \mid \theta_0) dx$$

is the Kullback–Leibler information distance between $f(. \mid \theta_0)$ and $f(. \mid \theta_A)$. Thus, the first-order Taylor expansion of Z_ϵ is

$$Z_\epsilon = \epsilon \ell'(X, \theta_0) + \frac{\epsilon^2}{2} \ell''(X, \tilde{\theta}_1) \quad \text{where } \tilde{\theta}_1 \in [\theta_0, \theta_A].$$

Therefore, for θ in ϵ-neighborhood of θ_0,

$$\begin{aligned} E_\theta\left(Z_\epsilon^2\right) &= \epsilon^2 \left\{ E_\theta \left(\ell'(X, \theta_0)\right)^2 + \epsilon E_\theta \left(\ell'(X, \theta_0) \ell''(X, \tilde{\theta}_1)\right) + \frac{\epsilon^2}{4} E_\theta \left(\ell''(X, \tilde{\theta}_1)\right)^2 \right\} \\ &\leq \epsilon^2 \left\{ E_\theta \left(\ell'(X, \theta_0)\right)^2 + \epsilon \sqrt{E_\theta \left(\ell'(X, \theta_0)\right)^2 E_\theta (\ell''(X, \tilde{\theta}_1))^2} + \frac{\epsilon^2}{4} E_\theta \left(L_1(X)\right) \right\} \\ &\leq \epsilon^2 \left\{ E_\theta \left(\ell'(X, \theta_0)\right)^2 + O(\epsilon) \right\} \leq c_1 \epsilon^2 \end{aligned} \tag{3.4}$$

for some constant $c_1 > 0$. The boundedness of $E_\theta \left(\ell'(X, \theta_0)\right)^2$ follows from condition R3 because

$$\begin{aligned} E_\theta \left(\ell'(X, \theta_0)\right)^2 &\leq E_\theta \left(\ell'(X, \theta)\right)^2 + |\theta - \theta_0| E_\theta \left(\ell''(X, \tilde{\theta})\right)^2 \\ &\leq |E_\theta \ell''(X, \theta)| + \epsilon E_\theta \left(\ell''(X, \tilde{\theta})\right)^2 \\ &\leq E_\theta \sqrt{L_1(X)} + \epsilon E_\theta L_1(X) \\ &\leq \sqrt{E_\theta L_1(X)} + \epsilon E_\theta L_1^2(X), \end{aligned}$$

by the Jensen inequality. Let $\Lambda_{n,\epsilon} = \sum_{i=1}^n Z_{i,\epsilon}$ be the log-likelihood ratio of n variables. Note that

$$\begin{aligned} E(\Lambda_{n,\epsilon}^2) &= \sum_{i=1}^n E(Z_{i,\epsilon}^2) + \sum_{i \neq j} E(Z_{i,\epsilon}) E(Z_{j,\epsilon}) \\ &\leq n c_1 \epsilon^2 + n(n-1) \left(E Z_\epsilon\right)^2 \\ &\leq n \epsilon^2 \left(c_1 + (n-1) k \epsilon^2\right), \end{aligned}$$

for some positive constant k. The last inequality holds as (3.2) and (3.3) hold. Therefore,

$$E(\Lambda_{n,\epsilon}^2) \leq c n \epsilon^2, \tag{3.5}$$

for any $c \geq |c_1 + (n-1) k \epsilon^2|$. Inequalities (3.2), (3.3), (3.4), and (3.5) are frequently used to prove asymptotic results in the later sections.

The last auxiliary result establishes the upper bounds for Chernoff entropy $\rho_\epsilon = \inf_s E\left(e^{s Z_\epsilon}\right)$ under both null and alternative hypotheses. Its proof is given in the Appendix.

Lemma 3.1. *There exist positive constant ϵ_0, K_0, and K_A such that for all $\epsilon \leq \epsilon_0$,*

(i) $\inf_{0<s<1} E_{\theta_0}\left(e^{sZ_\epsilon}\right) \leq 1 - K_0\epsilon^2$,
(ii) $\inf_{-1<s<0} E_{\theta_A}\left(e^{sZ_\epsilon}\right) \leq 1 - K_A\epsilon^2$.

3.3. Asymptotics of SPRT Stopping Time

Next, we derive the asymptotic expressions for the single-hypothesis SPRT stopping time

$$T_\epsilon^{\mathrm{SPRT}} = \inf_n \left\{\Lambda_{n,\epsilon} \notin (B, A)\right\}$$

for testing $H_0 : \theta = \theta_0$ vs. $H_A : \theta \geq \theta_0 + \epsilon$ with stopping boundaries A and B. If Type I and Type II error probabilities are given as α and β respectively, Wald's approximation for the expected stopping time gives under H_0,

$$E_{\theta_0}(T_\epsilon^{\mathrm{SPRT}}) \approx \frac{\alpha \ln\left(\frac{1-\beta}{\alpha}\right) + (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right)}{-K(\theta_0, \theta_0+\epsilon)}, \tag{3.6}$$

and under H_A,

$$E_{\theta_A}(T_\epsilon^{\mathrm{SPRT}}) \approx \frac{(1-\beta)\ln\left(\frac{1-\beta}{\alpha}\right) + \beta\ln\left(\frac{\beta}{1-\alpha}\right)}{K(\theta_A, \theta_A-\epsilon)}. \tag{3.7}$$

If the underlying density f satisfies regularity conditions R1-R4, inequalities (3.2) and (3.3) yield, for $\epsilon = \theta_A - \theta_0$,

$$\begin{aligned} K(\theta_0, \theta_A) &= \frac{\epsilon^2}{2} I(\theta_0) + O(\epsilon^3) \quad \text{and} \\ K(\theta_A, \theta_0) &= \frac{\epsilon^2}{2} I(\theta_A) + O(\epsilon^3) = \frac{\epsilon^2}{2} I(\theta_0) + O(\epsilon^3). \end{aligned} \tag{3.8}$$

Therefore, as $\epsilon \to 0$ under H_0,

$$\begin{aligned} E_{\theta_0}(T_\epsilon^{\mathrm{SPRT}}) &\approx \frac{\alpha \ln\left(\frac{1-\beta}{\alpha}\right) + (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right)}{-\epsilon^2 I(\theta_0)/2}(1+o(1)) \\ &= \frac{k_0(T_\epsilon^{\mathrm{SPRT}})}{\epsilon^2}(1+o(1)), \end{aligned} \tag{3.9}$$

and under H_A,

$$\begin{aligned} E_{\theta_A}(T_\epsilon^{\mathrm{SPRT}}) &\approx \frac{(1-\beta)\ln\left(\frac{1-\beta}{\alpha}\right) + \beta\ln\left(\frac{\beta}{1-\alpha}\right)}{\epsilon^2 I(\theta_0)/2}(1+o(1)) \\ &= \frac{k_A(T_\epsilon^{\mathrm{SPRT}})}{\epsilon^2}(1+o(1)). \end{aligned} \tag{3.10}$$

As seen from (3.9) and (3.10), under both null and alternative hypotheses, $E(T_\epsilon^{\mathrm{SPRT}}) \to \infty$ as $\epsilon \to 0$. The next lemma establishes that the stopping time $T_\epsilon^{\mathrm{SPRT}}$ converges to ∞ in probability at the rate of ϵ^{-2}.

Lemma 3.2. *Under regularity conditions R1–R4,*

$$P\left(\frac{1}{g(\epsilon)} \leq \epsilon^2 T_\epsilon^{\mathrm{SPRT}} \leq g(\epsilon)\right) \to 1 \quad as\ \epsilon \to 0,$$

for any function $g(\epsilon)$ such that $g(\epsilon) \to \infty$ as $\epsilon \to 0$.

Corollary 3.1. *$\epsilon^2 T_\epsilon^{\mathrm{SPRT}}$ is bounded away in probability from 0 and ∞ as $\epsilon \to 0$.*

The proofs of Lemma 3.2 and Corollary 3.1 are given in Appendix.

3.4. Asymptotics of Weighted Expected Stopping Time $T^{\max}$

From (3.9) and (3.10), we have $E_{\theta_j}(T_{jv}) \sim \frac{1}{\epsilon_{jv}^2}$ for $\theta_j \in \{\theta_0^{(j)}, \theta_A^{(j)}\}$, as $v \to \infty$, if α_{jv} and β_{jv} are bounded away from 0 for $j = 1, \ldots, d$. The following lemma provides a lower bound for weighted expected $T^{\max}$ defined in (3.1).

Lemma 3.3. *If α_{jv} and β_{jv} are bounded away from 0, there exist constants k_j such that*

$$E_w(T_v^{\max}) \geq \max_{j \leq d} \left\{ \frac{k_j}{\epsilon_{jv}^2}(1 + o(1)) \right\}.$$

Proof. Since $T_v^{\max} \geq T_{jv}$ a.s. for all $j \leq d$,

$$\begin{aligned} E_w\left(T_v^{\max}\right) &= \sum_{J_0 \subset \{1,\ldots,d\}} \pi\left(J_0\right) E_{J_0}\left(T_v^{\max}\right) \geq \sum_{J_0 \subset \{1,\ldots,d\}} \pi\left(J_0\right) E_{J_0}\left(T_{jv}\right) \\ &= \pi_j E_{H_{0v}^{(j)}}\left(T_{jv}\right) + \left(1 - \pi_j\right) E_{H_{Av}^{(j)}}\left(T_{jv}\right), \end{aligned} \tag{3.11}$$

where

$$\pi_j = \sum_{J_0 \subset \{1,\ldots,d\},\, j \in J_0} \pi\left(J_0\right)$$

is the marginal prior probability of $H_0^{(j)}$. Maximizing (3.11) for $j = 1, \ldots, d$,

$$E_w\left(T_v^{\max}\right) \geq \max_{1 \leq j \leq d} \left\{ \pi_j E_{H_{0v}^{(j)}}\left(T_{jv}\right) + \left(1 - \pi_j\right) E_{H_{Av}^{(j)}}\left(T_{jv}\right) \right\}. \tag{3.12}$$

Next, from (3.9) and (3.10), since α_{jv} and β_{jv} are bounded away from 0, there exist constant k_{0j} and k_{Aj} such that $E_{H_{0v}^{(j)}}\left(T_{jv}\right) = \frac{k_{0j}}{\epsilon_{jv}^2}(1 + o(1))$ and $E_{H_{Av}^{(j)}}\left(T_{jv}\right) = \frac{k_{Aj}}{\epsilon_{jv}^2}(1 + o(1))$. The proof is completed by substituting these expressions into (3.12). $\square$

As $v \to \infty$, $E_w(T_v^{\max})$ converges to ∞ at least at the rate of $\max_{j \leq d} \epsilon_{jv}^{-2}$. A natural question arises about the best distribution of error probabilities α_j and β_j, $\sum \alpha_j = \alpha$, $\sum \beta_j = \beta$, among d tests. Is equal error spending optimal? If not, is there an asymptotically optimal choice of error probabilities $(\alpha_{jv}, \beta_{jv})$ such that $E_w(T_v^{\max})$ attains its lower bound? The answers are given in the following two subsections.

3.5. Case 1: One Test More Difficult than Others

In general, by the *difficulty* of a test we understand the closeness between the tested null and alternative hypotheses that can be measured by Kullback–Leibler distance between the corresponding distributions. In view of (3.8), an equivalent measure is the absolute difference $|\epsilon|$ between the null and alternative parameters. Let the dth test be the *most difficult test in order* if $\epsilon_{dv} \ll \epsilon_{jv}$, that is, $\epsilon_{dv} = o(\epsilon_{jv})$ for all $j = 1, \ldots, d-1$.

Suppose the complete Tmax rule rejects $H_{0v}^{(j)}$ if and only if $\Lambda_{T_v^{\max},v}^{(j)} \geq c_j$ for some fixed boundary $c_j \in \left[\ln \frac{1-\beta_j}{\alpha_j}, \ln \frac{\beta_j}{1-\alpha_j}\right]$. Define Type I and II error probabilities of the jth test at $T_v^{\max}$ as

$$\alpha_{T_v^{\max}}^{(j)} = P(\Lambda_{T_v^{\max}}^{(j)} \geq c_j \mid H_{0v}^{(j)}) \quad \text{and} \quad \beta_{T_v^{\max}}^{(j)} = P(\Lambda_{T_v^{\max}}^{(j)} < c_j \mid H_{Av}^{(j)}).$$

The following theorem proves that if the dth test is the most difficult in order, then FWER I and FWER II are controlled asymptotically at levels α_d and β_d, respectively. The proof of this theorem is given in the Appendix.

Theorem 3.1. *If $\epsilon_{dv} = o(\epsilon_{jv})$ for all $j = 1, \ldots, d-1$, as $v \to \infty$, then*

$$\alpha_{T_v^{\max}}^{(j)} = o(1) \qquad \textit{and} \quad \beta_{T_v^{\max}}^{(j)} = o(1) \qquad \textit{for } j = 1, \ldots, d-1,$$

$$\alpha_{T_v^{\max}}^{(d)} = \alpha_d + o(1) \textit{ and } \quad \beta_{T_v^{\max}}^{(d)} = \beta_d + o(1).$$

According to Theorem 3.1, all error probabilities of less difficult tests converge to zero, whereas for the most difficult test, they converge to positive constants α_d and β_d. Hence, the family-wise error rates are controlled asymptotically at levels α_d and β_d instead of α and β. For this reason, *equal allocation of error probabilities is not optimal*. Indeed, if equal errors α/d and β/d are allocated to all d tests, FWER I and II are controlled at levels lower than required, which has to be at the cost of larger sample size. The following theorem establishes that an *optimal Type I and Type II error probability allocation* distributes the error probabilities mostly to the most difficult test. Such allocation of error probabilities leads to asymptotically efficient sequential tests that yield asymptotically optimal rate of the expected sample size.

Theorem 3.2. *Let $\epsilon_{dv} \ll \epsilon_{jv}$ for $j < d$, and $\alpha_v \downarrow 0$, $\beta_v \downarrow 0$ are chosen so that*

$$|\ln w_j \alpha_v| = o\left(\frac{\epsilon_{jv}}{\epsilon_{dv}}\right)^2 \quad \textit{and} \quad |\ln w_j \beta_v| = o\left(\frac{\epsilon_{jv}}{\epsilon_{dv}}\right)^2$$

for $w_j > 0$, $\sum_1^{d-1} w_j = 1$, and $j = 1, \ldots, d-1$. Then, error probabilities

$$\alpha_{dv} = \alpha - \alpha_v, \quad \beta_{dv} = \beta - \beta_v, \quad \textit{and}$$
$$\alpha_{jv} = w_j \alpha_v, \quad \beta_{jv} = w_j \beta_v \ \textit{ for } j < d,$$

are asymptotically optimal as $E_w(T_v^{\max})$ attains the asymptotically lower bound

$$E_w(T_v^{\max}) = \frac{k_d}{\epsilon_{dv}^2} + o\left(\frac{1}{\epsilon_{dv}^2}\right),$$

for a constant $k_d > 0$.

Proof. The error constraints are satisfied,

$$\sum_{j=1}^{d} \alpha_{jv} = \alpha \quad \text{and} \quad \sum_{j=1}^{d} \beta_{jv} = \beta.$$

Using (3.9), we have for all $j = 1, \ldots, d-1$,

$$\epsilon_{dv}^2 E_{H_{0v}^{(j)}}\left(T_{jv}\right) = \left(\frac{\epsilon_{dv}}{\epsilon_{jv}}\right)^2 \frac{\alpha_{jv} \ln\left(\frac{1-\beta_{jv}}{\alpha_{jv}}\right) + \left(1-\alpha_{jv}\right) \ln\left(\frac{\beta_{jv}}{1-\alpha_{jv}}\right)}{-I(\theta_0^{(j)})/2 + O\left(\epsilon_{jv}\right)} \to 0,$$

as $v \to \infty$, because $\left(\frac{\epsilon_{dv}}{\epsilon_{jv}}\right)^2 \ln \beta_{jv} \to 0$ from the choice of β_{jv}. Similarly,

$$\epsilon_{dv}^2 E_{H_{Av}^{(j)}}\left(T_{jv}\right) \to 0,$$

as $v \to \infty$ for $j = 1, \ldots, d-1$. Thus

$$E_w T_{jv} = \pi_j E_{H_{0v}^{(j)}}\left(T_{jv}\right) + (1-\pi_j) E_{H_{Av}^{(j)}}\left(T_{jv}\right) = o\left(\frac{1}{\epsilon_{dv}^2}\right) \quad \text{for } j = 1, \ldots, d-1.$$

Using (3.9) and (3.10) for testing $H_0^{(d)}$ vs $H_A^{(d)}$, we have

$$E_w T_{dv} = \pi_d E_{H_{0v}^{(d)}}\left(T_{dv}\right) + (1-\pi_d) E_{H_{Av}^{(d)}}\left(T_{dv}\right) = \frac{k_d}{\epsilon_{dv}^2} + o\left(\frac{1}{\epsilon_{dv}^2}\right),$$

where

$$k_d = \pi_d \frac{\alpha \ln\left(\frac{1-\beta}{\alpha}\right) + (1-\alpha) \ln\left(\frac{\beta}{1-\alpha}\right)}{-I(\theta_0^{(d)})/2} + (1-\pi_d) \frac{(1-\beta) \ln\left(\frac{1-\beta}{\alpha}\right) + \beta \ln\left(\frac{\beta}{1-\alpha}\right)}{I(\theta_0^{(d)})/2}.$$

A lower bound for $E_w\left(T_v^{\max}\right)$ is given by

$$E_w\left(T_v^{\max}\right) \geq \left\{\pi_d E_{H_{0v}^{(d)}}\left(T_{dv}\right) + (1-\pi_d) E_{H_{Av}^{(d)}}\left(T_{dv}\right)\right\} = \frac{k_d}{\epsilon_{dv}^2} + o\left(\frac{1}{\epsilon_{dv}^2}\right),$$

and an upper bound for $E_w\left(T_v^{\max}\right)$ is obtained as

$$E_w\left(T_v^{\max}\right) \leq \sum_{j=1}^{d} E_w\left(T_{jv}\right) = \frac{k_d}{\epsilon_{dv}^2} + o\left(\frac{1}{\epsilon_{dv}^2}\right).$$

Hence, $E_w\left(T_v^{\max}\right) = \frac{k_d}{\epsilon_{dv}^2} + o\left(\frac{1}{\epsilon_{dv}^2}\right)$. □

3.6. Case 2: Tests of Same Order of Difficulty

Here, we consider a situation when all tests have a common order of *difficulty*; that is, $\epsilon_{jv} = r_j \epsilon_v$ for $j = 1, \ldots, d$, where $r_1, \ldots, r_d$ are constants, and $\epsilon_v \to 0$ as $v \to \infty$. The following theorem proves that one should not allocate an error probability that is too large or too small to any particular test in this case.

Theorem 3.3. *An asymptotically optimal allocation of error probabilities* α_{jv}, β_{jv} *is such that* α_{jv} *and* β_{jv} *are bounded away from* 0 *for all* $j = 1, \ldots, d$ *and* $v = 1, 2, \ldots,$ *and*

$$\sum_{j=1}^{d} \alpha_{jv} = \alpha, \quad \sum_{j=1}^{d} \beta_{jv} = \beta.$$

Then

$$\max_{1 \le j \le d} \frac{k_j}{r_j^2} + o(1) \le \epsilon_v^2 E_w(T_v^{\max}) \le \sum_{j=1}^{d} \frac{k_j}{r_j^2} + o(1),$$

where $k_j = \pi_j k_0(T_{jv}) + (1 - \pi_j) k_A(T_{jv})$, *functions* k_0 *and* k_A *being defined in* (3.9) *and* (3.10).

Proof. An upper bound for $E_w\,(T_v^{\max})$ is obtained as

$$E_w\,(T_v^{\max}) \le \sum_{j=1}^{d} \left\{ \pi_j E_{H_{0v}^{(j)}}\left(T_{jv}\right) + \left(1 - \pi_j\right) E_{H_{Av}^{(j)}}\left(T_{jv}\right) \right\} = \sum_{j=1}^{d} \left\{ \frac{k_j}{r_j^2 \epsilon_v^2} (1 + o(1)) \right\},$$

as α_{jv}, β_{jv} are bounded away from 0. Also, from Lemma 3.3, a lower bound for $E_w\,(T_v^{\max})$ is

$$E_w(T_v^{\max}) \ge \max_{j \le d} \left\{ \frac{k_j}{r_j^2 \epsilon_v^2} (1 + o(1)) \right\}.$$

Thus,

$$\max_{1 \le j \le d} \frac{k_j}{r_j^2} + o(1) \le \epsilon_v^2 E_w(T_v^{\max}) \le \sum_{j=1}^{d} \frac{k_j}{r_j^2} + o(1). \qquad \square$$

Theorems 3.2 and 3.3 establish the asymptotically optimal error spending strategy among the d tests that attains the smallest rate of the expected sample size under the specified constraints on FWER I and II.

4. SIMULATION STUDY

Sequential procedures for testing multiple hypotheses proposed above and in the literature are compared in this section. To cover all of the interesting cases, simultaneous testing of several normal means μ_j and Bernoulli proportions p_j is considered under a strong control of FWER I and FWER II at levels $\alpha = 0.05$ and $\beta = 0.1$, respectively.

4.1. Comparison of Six Schemes with Uniform Error Spending

Table 1 compares the performance of six multiple testing procedures including the non-sequential Holm method from Holm (1979), the multistage step-down

Table 1. Comparison of six multiple testing schemes for three tests

Multiple tests	Scheme	FWER I%	FWER II%	$\widehat{ET}$	S.E. of $\widehat{ET}$
Multiple testing 1	Non-sequential Holm	4	NA	105	NA
	Bartroff–Lai	4.8	NA	104.6	NA
	Tmin	0.86	NA	14.47	0.03
	Incomplete Tmax	3.77	NA	44	0.095
	Complete Tmax	3.37	NA	44	0.095
	Intersection	2.24	NA	46.83	0.099
Multiple testing 2	Non-sequential Holm	4.4	NA	105	NA
	Bartroff–Lai	4.2	NA	98.3	NA
	Tmin	0.68	74.08	16.2	0.0314
	Incomplete Tmax	2.36	2.46	46.21	0.0951
	Complete Tmax	2.14	1.9	46.21	0.0951
	Intersection	1.33	1.01	49.34	0.1
Multiple testing 3a	Non-sequential Holm	4.3	NA	105	NA
	Bartroff–Lai	3.2	NA	92.3	NA
	Tmin	0.38	98.98	17.36	0.033
	Incomplete Tmax	1.18	4.94	48.24	0.096
	Complete Tmax	1.04	3.81	48.24	0.096
	Intersection	0.61	2.1	51.56	0.102
Multiple testing 3b	Non-sequential Holm	4.5	NA	105	NA
	Bartroff–Lai	3.6	NA	94.2	NA
	Tmin	0.03	98.78	16.16	0.025
	Incomplete Tmax	1.16	4.95	49.8	0.092
	Complete Tmax	1.14	3.96	49.8	0.092
	Intersection	0.81	2.34	53.55	0.095

procedure from Bartroff and Lai (2010), Tmin, incomplete Tmax, complete Tmax, and intersection rules for the following multiple testing scenarios.

A sequence of random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \in \mathbb{R}^3$ is observed, where $\boldsymbol{X}_i = \left(X_i^{(1)}, X_i^{(2)}, X_i^{(3)}\right)$, $X_i^{(j)} \sim N(\mu_j, 1)$ for $j = 1, 2$, and $X_i^{(3)} \sim \mathit{Bernoulli}(p)$. Three tests about parameters $\mu_{1,2}$ and p are conducted, $H^{(1)} : \mu_1 = 0$ vs. $\mu_1 = 0.5$, $H^{(2)} : \mu_2 = 0$ vs. $\mu_2 = 0.5$, and $H^{(3)} : p = 0.5$ vs. $p = 0.75$, with the uniform error spending $(\alpha_j, \beta_j) = (\alpha/d, \beta/d)$, $j = 1, 2, 3$, and stopping boundaries $(b_j, a_j) = (\ln \beta_j, -\ln \alpha_j)$ for the log-likelihood ratio statistics.

We consider the following multiple testing scenarios.

Multiple testing 1: $H_0^{(1)}, H_0^{(2)}$, and $H_0^{(3)}$ are true, and the coordinates of $\boldsymbol{X}_i$ are independent.

Multiple testing 2: $H_0^{(1)}, H_0^{(2)}$, and $H_A^{(3)}$ are true, and the coordinates of $\boldsymbol{X}_i$ are independent.

Multiple testing 3a: $H_0^{(1)}, H_A^{(2)}$, and $H_A^{(3)}$ are true, and the coordinates of $\boldsymbol{X}_i$ are independent.

Multiple testing 3b: $H_0^{(1)}, H_A^{(2)}$, and $H_A^{(3)}$ are true, and the two normal components $X_i^{(1)}$ and $X_i^{(2)}$ have correlation coefficient 0.75.

For each situation, the estimated actual FWER I and II are presented as percentage, along with the estimated expected sample size $\widehat{ET}$ and its standard error. Results are based on 55,000 simulated sequences.

4.2. Discussion of Results in Table 1

In all of the multiple testing scenarios, our proposed incomplete Tmax, complete Tmax, and intersection rule yield uniformly lower error rates than non-sequential Holm and Bartroff-Lai's multistage procedure with approximately 50% lower expected sample size. For complete Tmax rule, the decision boundary $c_j = (a_j + b_j)/2$ is used for all $j = 1, \ldots, d$. As expected, complete Tmax rule yields lower error rates than incomplete Tmax rule while using, the same sample size. Tmin rule controls FWER I conservatively and certainly saves on a sample size; however, it is at the expense of a high FWER II leading to a significant reduction of family-wise power. The estimated expected sample size and FWER I for the Bartroff–Lai procedure are adopted from Bartroff and Lai (2010).

4.3. Asymptotically Optimal Error Spending

Consider now the asymptotic setup where one or more parameters under the alternative hypotheses are close to their corresponding null values. Again, parameters μ_j and p_j of $N(\mu_j, 1)$ and $Bernoulli(p_j)$ distribution are tested with stopping boundaries $(b_j, a_j) = (\ln \beta_j, -\ln \alpha_j)$.

The following testing scenarios are compared. In 4a, 4b, 5a, and 5b, one test is more difficult than the others. Ignoring this, tests 4a and 5a continue to use the uniform error allocation, which is no longer optimal, as shown in Section 3. Conversely, tests 4b and 5b spend the error probabilities according to Theorem 3.2, allocating most of them to the most difficult test. In test 6, all four tests are equally difficult, so uniform error allocation is optimal by Theorem 3.3.

Multiple testing 4a: Test a battery of $d = 4$ hypotheses: $H^{(1)} : \mu_1 = 0$ vs. $\mu_1 = 0.1$, $H^{(2)} : \mu_2 = 0$ vs. $\mu_2 = 0.5$, and $H^{(j)} : p_j = 0.5$ vs. $p_j = 0.75$ for $j = 3, 4$ with uniform error spending $(\alpha_j, \beta_j) = (\alpha/d, \beta/d)$ for $j = 1, \ldots, 4$. The set of true null hypotheses is $J_0 = \{1, 3\}$.

Multiple testing 4b: Tests in 4a are conducted with Type I error probabilities $(0.04, 0.004, 0.003, 0.003)$ and Type II error probabilities $(0.08, 0.008, 0.006, 0.006)$, allocating most of the errors to the most difficult test and ensuring $\sum_{j=1}^{4} \alpha_j = 0.05$ and $\sum_{j=1}^{4} \beta_j = 0.1$.

Multiple testing 5a: Same tests as in 4a, with $J_0 = \{3, 4\}$, that is, $H_0^{(3)}$ and $H_0^{(4)}$ are true.

Multiple testing 5b: Same tests as in 4b, with $J_0 = \{3, 4\}$.

Multiple testing 6: Test $H^{(j)} : \mu_j = 0$ vs. $\mu_j = 0.1$ for $j = 1, \ldots, 4$, with uniform error spending $(\alpha_j, \beta_j) = (\alpha/4, \beta/4)$, and $J_0 = \{3, 4\}$.

For each multiple testing procedure, Table 2 contains the estimates of FWER I and FWER II in percentages, expected stopping time $\widehat{ET}$, expected SPRT stopping

Table 2. Comparing expected sample size and error rates obtained from the three schemes when some of the four tests are difficult

Multiple tests	Scheme	FWER I%	FWER II%	$\widehat{ET}$	$(\widehat{ET}_1, \ldots, \widehat{ET}_4)$	$\hat{\alpha}_j\%$	$\hat{\beta}_j\%$
Multiple testing 4a	incomplete Tmax	2.2	3.65	730.7	(730.7, 36.4, 27.2, 33.7)	$(\hat{\alpha}_1, \hat{\alpha}_3) = (1.16, 1.06)$	$(\hat{\beta}_2, \hat{\beta}_4) = (1.89, 1.8)$
	Complete Tmax	1.16	0.01	730.7		$(\hat{\alpha}_1, \hat{\alpha}_3) = (1.16, 0.003)$	$(\hat{\beta}_2, \hat{\beta}_4) = (0.005, 0.005)$
	Intersection	1.16	0.002	730.8		$(\hat{\alpha}_1, \hat{\alpha}_3) = (1.16, 0)$	$(\hat{\beta}_2, \hat{\beta}_4) = (0, 0.002)$
Multiple testing 4b	Incomplete Tmax	3.76	1.1	477.4	(477.1, 46.2, 37.4, 45.5)	$(\hat{\alpha}_1, \hat{\alpha}_3) = (3.54, 0.23)$	$(\hat{\beta}_2, \hat{\beta}_4) = (0.64, 0.46)$
	Complete Tmax	3.55	0.049	477.4		$(\hat{\alpha}_1, \hat{\alpha}_3) = (3.54, 0.01)$	$(\hat{\beta}_2, \hat{\beta}_4) = (0.04, 0.009)$
	Intersection	3.54	0.0014	478.1		$(\hat{\alpha}_1, \hat{\alpha}_3) = (3.53, 0.003)$	$(\hat{\beta}_2, \hat{\beta}_4) = (0.009, 0.005)$
Multiple testing 5a	Incomplete Tmax	2.112	4.2	850.1	(850.1, 36.4, 27.2, 27.2)	$(\hat{\alpha}_3, \hat{\alpha}_4) = (1.06, 1.07)$	$(\hat{\beta}_1, \hat{\beta}_2) = (2.36, 1.88)$
	Complete Tmax	0.002	2.36	850.1		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0, 0.002)$	$(\hat{\beta}_1, \hat{\beta}_2) = (2.36, 0)$
	Intersection	0	2.36	850.1		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0, 0)$	$(\hat{\beta}_1, \hat{\beta}_2) = (2.36, 0)$
Multiple testing 5b	Incomplete Tmax	0.5	7.99	570.9	(570.9, 46.2, 37.4, 37.3)	$(\hat{\alpha}_3, \hat{\alpha}_4) = (0.23, 0.27)$	$(\hat{\beta}_1, \hat{\beta}_2) = (7.39, 0.64)$
	Complete Tmax	0.003	7.39	570.9		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0.002, 0.002)$	$(\hat{\beta}_1, \hat{\beta}_2) = (7.39, 0.007)$
	Intersection	0	7.37	571.3		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0, 0)$	$(\hat{\beta}_1, \hat{\beta}_2) = (7.37, 0)$
Multiple testing 6	Incomplete Tmax	2.25	4.43	1360.52	(848.3, 848.3, 729.3, 730.2)	$(\hat{\alpha}_3, \hat{\alpha}_4) = (1.14, 1.12)$	$(\hat{\beta}_1, \hat{\beta}_2) = (2.17, 2.3)$
	Complete Tmax	1.84	3.34	1360.52		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0.93, 0.91)$	$(\hat{\beta}_1, \hat{\beta}_2) = (1.62, 1.74)$
	Intersection	0.93	1.65	1475.12		$(\hat{\alpha}_3, \hat{\alpha}_4) = (0.48, 0.45)$	$(\hat{\beta}_1, \hat{\beta}_2) = (0.83, 0.82)$

times $\widehat{ET}_j$ for marginal tests $H^{(j)}$, $j = 1, \ldots, 4$, and marginal Type I and Type II error probabilities $\hat{\alpha}_j$ and $\hat{\beta}_j$ in percentages. Results are based on 55,000 simulated sequences.

4.4. Discussion of Results in Table 2

Proposed schemes control FWER I and FWER II at 5% and 10% levels, respectively. In all of the testing scenarios, incomplete and complete Tmax rule require the same sample size, but the complete Tmax rule yields *lower* FWER I and FWER II than the incomplete Tmax rule. This reduction in error rates is due to utilizing the full sample for each test and using a sufficient statistic based on $T^{\max}$ samples. Compared to the incomplete and complete Tmax rule, the intersection rule reduces both error rates even further but requires greater expected sample size. Equal allocation of error probabilities $(\alpha/4, \beta/4)$ on each test of multiple testing 4a yields much larger expected sample size (731) compared to the expected sample size (478) obtained in multiple testing 4b. This reduction in expected sample size is due to allocating large portions of error probabilities 0.04 and 0.08 to the first test, which is the most difficult in this case. Similarly, expected sample size reduces from 850 in multiple testing 5a to 571 in multiple testing 5b as most of the error probabilities are allocated to the most difficult test. These simulation results support Theorem 3.2.

APPENDIX

Proof of Lemma 3.1. Taylor expansion of $M_0(s, \epsilon) = E_{\theta_0}\left(e^{sZ_\epsilon}\right)$ yields

$$M_0(s, \epsilon) = M_0(s, 0) + \epsilon M_0{}'(s, 0) + \frac{\epsilon^2}{2} M_0{}''(s, 0) + R_{02,\epsilon}, \tag{A.1}$$

where $M_0(s, 0) = 1$ and $R_{02,\epsilon} = M_0{}'''(s, \tilde{\epsilon})\frac{\epsilon^3}{6} = O(\epsilon^3)$ for some $0 \le \tilde{\epsilon} \le \epsilon$ because $M_0{}'''(s, \tilde{\epsilon})$ is bounded by regularity condition R4. Differentiating $M_0(s, \epsilon)$, obtain

$$\begin{aligned}
\frac{\partial}{\partial \epsilon} E_{\theta_0}\left(e^{sZ_\epsilon}\right) &= E_{\theta_0}\left[\frac{\partial}{\partial \epsilon}\left(\frac{f(X \mid \theta_0 + \epsilon)}{f(X \mid \theta_0)}\right)^s\right] \\
&= E_{\theta_0}\left[s\left(\frac{f(X \mid \theta_0 + \epsilon)}{f(X \mid \theta_0)}\right)^{s-1} \frac{f'(X \mid \theta_0 + \epsilon)}{f(X \mid \theta_0)}\right].
\end{aligned}$$

Then, from R2, $M_0{}'(s, 0) = sE_{\theta_0}\left[\frac{f'(X \mid \theta_0)}{f(X \mid \theta_0)}\right] = 0$. Differentiating $M_0(s, \epsilon)$ again at $\epsilon = 0$,

$$\begin{aligned}
M_0{}''(s, 0) &= s(s-1)E_{\theta_0}\left(\frac{f'(X \mid \theta_0)}{f(X \mid \theta_0)}\right)^2 + sE_{\theta_0}\left(\frac{f''(X \mid \theta_0)}{f(X \mid \theta_0)}\right) \\
&= -s(1-s)I(\theta_0).
\end{aligned}$$

Therefore, (A.1) becomes

$$M_0(s, \epsilon) = 1 - s(1-s)I(\theta_0)\frac{\epsilon^2}{2} + O(\epsilon^3).$$

There exists ϵ_1 such that for all $\epsilon \le \epsilon_1$ and for fixed $s \in (0, 1)$,

$$s(1-s)I(\theta_0) + O(\epsilon) \ge \frac{s(1-s)}{2} I(\theta_0).$$

Hence, for any $\epsilon \le \epsilon_1$ and $K_0 = s(1-s)I(\theta_0)/4$,

$$\inf_{0<s<1} M_0(s, \epsilon) \le 1 - K_0\epsilon^2.$$

This proves (*i*) of Lemma 3.1. Similar to (A.1), Taylor expansion of $M_A(s,\epsilon) = E_{\theta_A}\left(e^{sZ_\epsilon}\right)$ yields

$$M_A(s,\epsilon) = M_A(s,0) + \epsilon M_A{}'(s,0) + \frac{\epsilon^2}{2} M_A{}''(s,0) + R_{A2,\epsilon}, \tag{A.2}$$

where $M_A(s,0) = 1$ and $R_{A2,\epsilon} = M_A{}'''(s,\tilde{\epsilon})\frac{\epsilon^3}{6} = O(\epsilon^3)$ for some $0 \le \tilde{\epsilon} \le \epsilon$, because $M_A{}'''(s,\tilde{\epsilon})$ is bounded (R4). Along the same steps, we derive $M_A{}'(s,0) = 0$ and $M_A{}''(s,0) = s(s+1)I(\theta_A)$. Therefore, (A.2) becomes

$$M_A(s,\epsilon) = 1 - \frac{\epsilon^2}{2}\left[(1+s)(-s)I(\theta_A) + O(\epsilon)\right].$$

There exists ϵ_2 such that for all $\epsilon \le \epsilon_2$ and for fixed $s \in (-1, 0)$,

$$(1+s)(-s)I(\theta_A) + O(\epsilon) \ge \frac{(1+s)(-s)}{2} I(\theta_A).$$

Hence, for any $\epsilon \le \epsilon_2$ and $K_A = (1+s)(-s)I(\theta_A)/4$,

$$\inf_{-1<s<0} M_A(s,\epsilon) \le 1 - K_A\epsilon^2.$$

Both (i) and (ii) of Lemma 3.1 hold for all $\epsilon \le \epsilon_0 = \min(\epsilon_1, \epsilon_2)$. This completes the proof.

Proof of Lemma 3.2. For an arbitrary $\epsilon > 0$,

$$P\left(\frac{1}{g(\epsilon)} \le \epsilon^2 T_\epsilon^{\text{SPRT}} \le g(\epsilon)\right) = 1 - P\left(\epsilon^2 T_\epsilon^{\text{SPRT}} \le \frac{1}{g(\epsilon)}\right) - P\left(\epsilon^2 T_\epsilon^{\text{SPRT}} > g(\epsilon)\right). \tag{A.3}$$

For all $i = 1, 2, \ldots$, using (3.2) and (3.3), $|\mu_\epsilon| = |E\left(Z_{i,\epsilon}\right)| \le k\epsilon^2$ for some positive constant k. Let $A_0 = \min(A, |B|)$. Then

$$\begin{aligned}
P\left(\epsilon^2 T_\epsilon^{\text{SPRT}} \le \frac{1}{g(\epsilon)}\right) &= P\left(T_\epsilon^{\text{SPRT}} \le \frac{1}{\epsilon^2 g(\epsilon)}\right) \\
&\le P\left(\max_{1\le n\le \frac{1}{\epsilon^2 g(\epsilon)}} |\Lambda_{n,\epsilon}| \ge A_0\right) \\
&\le P\left(\max_{1\le n\le \frac{1}{\epsilon^2 g(\epsilon)}} |\Lambda_{n,\epsilon} - n\mu_\epsilon| \ge A_0 - \frac{k}{g(\epsilon)}\right) \\
&\le \frac{1}{\left(A_0 - \frac{k}{g(\epsilon)}\right)^2} \sum_{n=1}^{\left\lfloor \frac{1}{\epsilon^2 g(\epsilon)} \right\rfloor} E(Z_{n,\epsilon} - \mu_\epsilon)^2.
\end{aligned} \tag{A.4}$$

The last inequality in (A.4) follows from the Kolmogorov maximal inequality for independent random variables (e.g., Billingsley, 1995). From (3.4), $E(Z_{n,\epsilon} - \mu_\epsilon)^2 \leq c_1\epsilon^2$ for some $c_1 > 0$. Therefore, (A.5) implies

$$P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} \leq \frac{1}{g(\epsilon)}\right) \leq \frac{c_1/g(\epsilon)}{(A_0 - k/g(\epsilon))^2} = o(1) \quad \text{as } \epsilon \to 0. \tag{A.5}$$

From the Markov inequality,

$$P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} > g(\epsilon)\right) \leq \frac{\epsilon^2 E(T_\epsilon^{\mathrm{SPRT}})}{g(\epsilon)}.$$

Then, using (3.9) and (3.10), $\epsilon^2 E(T_\epsilon^{\mathrm{SPRT}}) = O(1)$ as $\epsilon \to 0$. Hence,

$$P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} \geq g(\epsilon)\right) = o(1) \quad \text{as } \epsilon \to 0. \tag{A.6}$$

The proof is completed by applying (A.5) and (A.6) in (A.3).

Proof of Corollary 3.1. Consider δ such that $\frac{c_1\delta}{(A_0-k\delta)^2} < 1$. Such a choice of δ exists because the quadratic form in δ has discriminant $c_1^2 + 4A_0kc_1 > 0$. Substituting δ for $\frac{1}{g(\epsilon)}$ in the inequality (A.5), obtain

$$P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} \leq \delta\right) \leq \frac{c_1\delta}{(A_0 - k\delta)^2},$$

and therefore,

$$P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} > \delta\right) \geq 1 - \frac{c_1\delta}{(A_0 - k\delta)^2} > 0.$$

Thus, $\epsilon^2 T_\epsilon^{\mathrm{SPRT}}$ is bounded away from 0 in probability as $\epsilon \to 0$. Next, applying the Markov inequality,

$$\lim_{\lambda\to\infty} \limsup_{\epsilon\to 0} P\left(\epsilon^2 T_\epsilon^{\mathrm{SPRT}} > \lambda\right) \leq \lim_{\lambda\to\infty} \limsup_{\epsilon\to 0} \frac{\epsilon^2 E(T_\epsilon^{\mathrm{SPRT}})}{\lambda} = 0.$$

Hence, $\epsilon^2 T_\epsilon^{\mathrm{SPRT}} = O_P(1)$ as $\epsilon \to 0$.

Proof of Theorem 3.1. Let $c_d = A_d = \ln\left(\frac{1-\beta_d}{\alpha_d}\right)$, and $B_d = \ln\left(\frac{\beta_d}{1-\alpha_d}\right)$. Consider

$$\begin{aligned}
\alpha_{T_v^{\max}}^{(d)} &= P_{H_{0v}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} \geq A_d, T_v^{\max} = T_{dv}\right) + P_{H_{0v}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} \geq A_d, T_v^{\max} \neq T_{dv}\right) \\
&= P_{H_{0v}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} \geq A_d\right) + o(1) = \alpha_d + o(1),
\end{aligned}$$

because $P\left(T_v^{\max} \neq T_{dv}\right) = o(1)$. The last equality is subject to Wald's approximation ignoring the overshoot for the dth test. Similarly,

$$\begin{aligned}
\beta_{T_v^{\max}}^{(d)} &= P_{H_{Av}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} < A_d, T_v^{\max} = T_{dv}\right) + P_{H_{Av}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} < A_d, T_v^{\max} \neq T_{dv}\right) \\
&= P_{H_{Av}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} < A_d\right) + o(1) \\
&= P_{H_{Av}^{(d)}}\left(\Lambda_{T_v^{\max},v}^{(d)} \leq B_d\right) + o(1) = \beta_d + o(1).
\end{aligned}$$

Then, for any sequence of positive integers $g_v \to \infty$ as $v \to \infty$ such that $g_v = \mathrm{o}\left(\epsilon_{dv}^{-2}\right)$ and $\max_{j<d}\left(\epsilon_{jv}^{-2}\right) = \mathrm{o}\left(g_v\right)$,

$$
\begin{aligned}
P\left(T_v^{\max} < g_v\right) &\le P\left(T_{dv} < g_v\right) \\
&\le P\left(\max_{1\le m\le g_v} |\Lambda_{m,v}^{(d)}| \ge A_{0d}\right) \quad \text{where } A_{0d} = \min\left(A_d, |B_d|\right) \\
&\le \sum_{m=1}^{\lfloor g_v \rfloor} \frac{E\left({Z_{m,v}^{(d)}}^2\right)}{\left(A_{0d} - C\epsilon_{dv}^2 g_v\right)^2} \quad \text{using (A.4)} \\
&\le \frac{K_d \epsilon_{dv}^2 g_v}{\left(A_{0d} - C\epsilon_{dv}^2 g_v\right)^2} = \mathrm{o}\left(1\right),
\end{aligned}
$$

where the last inequality follows from (3.4). Here, C and K_d are some positive constants. For all $j = 1, 2, \ldots, d-1$,

$$
\begin{aligned}
\alpha_{T_v^{\max}}^{(j)} &\le P_{H_{0v}^{(j)}}\left(\Lambda_{T_v^{\max},v}^{(j)} \ge c_j,\, T_v^{\max} \ge g_v\right) + P_{H_{0v}^{(j)}}\left(T_v^{\max} < g_v\right) \\
&\le P_{H_{0v}^{(j)}}\left(\max_{m\ge g_v} \sum_{i=1}^{m} Z_{i,v}^{(j)} \ge c_j\right) + o(1) \\
&= P_{H_{0v}^{(j)}}\left(\max_{m\ge g_v} e^{s\sum_{i=1}^{m} Z_{i,v}^{(j)}} \ge e^{sc_j}\right) + o(1)
\end{aligned}
$$

for any $0 \le s \le 1$. Note that $\left\{e^{s\sum_{i=1}^{m} Z_{i,v}^{(j)}}\right\}_{m=\eta}^{\infty}$ is a nonnegative supermartingale. Applying Theorem 1 in Lake (2000), which is analogous to the Doob's maximal inequality for nonnegative submartingales,

$$
P_{H_{0v}^{(j)}}\left(\max_{m\ge g_v} e^{s\sum_{i=1}^{m} Z_{i,v}^{(j)}} \ge e^{sc_j}\right) \le e^{-sc_j} E_{H_{0v}^{(j)}}\left(e^{s\sum_{i=1}^{g_v} Z_{i,v}^{(j)}}\right)
$$

An upper bound of $\alpha_{T_v^{\max}}^{(j)}$ is obtained by taking infimum over all s,

$$
\begin{aligned}
\alpha_{T_v^{\max}}^{(j)} &\le \inf_{0\le s\le 1} e^{-sc_j} E_{H_{0v}^{(j)}}\left(e^{s\sum_{i=1}^{g_v} Z_{i,v}^{(j)}}\right) + \mathrm{o}\left(1\right) \\
&\le \inf_{0\le s\le 1} E_{H_{0v}^{(j)}}\left(e^{s\sum_{i=1}^{g_v} Z_{i,v}^{(j)}}\right) + \mathrm{o}\left(1\right) \\
&= \left\{\inf_{0\le s\le 1} E_{H_{0v}^{(j)}}\left(e^{sZ_{i,v}^{(j)}}\right)\right\}^{g_v} + \mathrm{o}\left(1\right) \\
&\le \left\{1 - L_j \epsilon_{jv}^2\right\}^{g_v} + \mathrm{o}\left(1\right) \quad \text{for some large } v \\
&= \mathrm{o}\left(1\right).
\end{aligned}
$$

The last inequality is obtained by using Lemma 3.1. A similar approach proves that for all $j < d$, $\beta_{T_v^{\max}}^{(j)} \to 0$ as $v \to \infty$.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grant DMS 1007775. Research of the second author was partially supported by the National Security Agency grant H98230-11-1-0147. We thank the associate editor and Professor Mukhopadhyay for their help and attention to our work.

REFERENCES

Armitage, P. (1950). Sequential Analysis with More Than Two Alternative Hypotheses, and Its Relation to Discriminant Function Analysis, *Journal of Royal Statistical Society, Series B* 12: 137–144.

Baillie, D. H. (1987). Multivariate Acceptance Sampling—Some Applications to Defence Procurement, *Statistician* 36: 465–478.

Bartroff, J. and Lai, T. L. (2010). Multistage Tests of Multiple Hypotheses, *Communications in Statistics—Theory and Methods* 39: 1597–1607.

Baum, C. W. and Veeravalli, V. V. (1994). A Sequential Procedure for Multihypotheses Testing, *IEEE Transactions on Information Theory* 40: 1994–2007.

Benjamini, Y., Bertz, F., and Sarkar, S. (2004). *Recent Developments in Multiple Comparison Procedures, IMS Lecture Notes – Monograph Series*, Beachwood: Institute of Mathematical Statistics.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of Royal Statistical Society, Series B* 57: 289–300.

Berger, J. O. (1985). *Statistical Decision Theory*, New York: Springer.

Betensky, R. A. (1996). An O'Brien-Fleming Sequential Trial for Comparing Three Treatments, *Annals of Statistics* 24: 1765–1791.

Billingsley, P. (1995). *Probability and Measure*, New York: Wiley.

Borovkov, A. A. (1997). *Matematicheskaya Statistika*, Novosibirsk: Nauka.

Darkhovsky, B. (2011). Optimal Sequential Tests for Testing Two Composite and Multiple Simple Hypotheses, *Sequential Analysis* 30: 479–496.

De, S. K. and Baron, M. (2012). Step-Up and Step-Down Methods for Testing Multiple Hypotheses in Sequential Experiments, *Journal of Statistical Planning and Inference*, in press.

Edwards, D. (1987). Extended-Paulson Sequential Selection, *Annals of Statistics* 15: 449–455.

Edwards, D. G. and Hsu, J. C. (1983). Multiple Comparisons with the Best Treatment, *Journal of American Statistical Association* 78: 965–971.

Glaspy, J., Jadeja, J. S., Justice, G., Kessler, J., Richards, D., Schwartzberg, L., Rigas, J., Kuter, D., Harmon, D., Prow, D., Demetri, G., Gordon, D., Arseneau, J., Saven, A., Hynes, H., Boccia, R., O'Byrne, J., and Colowick, A. B. (2001). A Dose-Finding and Safety Study of Novel Erythropoiesis Stimulating Protein (NESP) for the Treatment of Anemia in Patients Receiving Multicycle Chemotherapy, *British Journal of Cancer* 84: 17–23.

Govindarajulu, Z. (1987). *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*, Columbus: American Sciences Press.

Hamilton, D. C. and Lesperance, M. L. (1991). A Consulting Problem Involving Bivariate Acceptance Sampling by Variables, *Canadian Journal of Statistics* 19: 109–117.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics* 6: 65–70.

Hughes, M. D. (1993). Stopping Guidelines for Clinical Trials with Multiple Treatments, *Statistics in Medicine* 12: 901–915.

Jennison, C. and Turnbull, B. W. (1993). Group Sequential Tests for Bivariate Response: Interim Analyses of Clinical Trials with Both Efficacy and Safety Endpoints, *Biometrics* 49: 741–752.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, Boca Raton: Chapman & Hall.

Lake, D. E. (2000). Minimum Chernoff Entropy and Exponential Bounds for Locating Changes, *IEEE Transactions on Information Theory* 46: 1168–1170.

Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L., and Greenhouse, J. (1994). *Case Studies in Biometry*, New York: Wiley.

Lee, M.-L. T. (2004). *Analysis of Microarray Gene Expression Data*, Boston: Kluwer.

Novikov, A. (2009). Optimal Sequential Multiple Hypothesis Testing, *Kybernetika* 45: 309–330.

O'Brien, P. C. (1984). Procedures for Comparing Samples with Multiple Endpoints, *Biometrics* 40: 1079–1087.

Pocock, S. J., Geller, N. L., and Tsiatis, A. A. (1987). The Analysis of Multiple Endpoints in Clinical Trials, *Biometrics* 43: 487–498.

Sarkar, S. K. (2007). Step-up Procedures Controlling Generalized FWER and Generalized FDR, *Annals of Statistics* 35: 2405–2420.

Sidak, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions, *Journal of American Statistical Association* 62: 626–633.

Sobel, M. and Wald, A. (1949). A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution, *Annals of Mathematical Statistics* 20: 502–522.

Tang, D.-I. and Geller, N. L. (1999). Closed Testing Procedures for Group Sequential Clinical Trials with Multiple Endpoints, *Biometrics* 55: 1188–1192.

Tartakovsky, A. G., Li, X. R., and Yaralov, G. (2003). Sequential Detection of Targets in Multichannel Systems, *IEEE Transactions on Information Theory* 49: 425–445.

Tartakovsky, A. G. and Veeravalli, V. V. (2004). Change-Point Detection in Multichannel and Distributed Systems with Applications, in *Applications of Sequential Methodologies*, N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, eds., pp. 339–370, New York: Dekker.