University of Alberta

On two-sample sequential testing procedures based on strong approximation principles

by

Abdulkadir Ahmed Hussein

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta

Fall 2003

Your file Votre référence
ISBN: 0-612-87993-3
Our file Notre référence
ISBN: 0-612-87993-3

University of Alberta

Library Release Form

Name of Author: Abdulkadir Ahmed Hussein

Title of Thesis: On two-sample sequential testing procedures based on strong approximation principles

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Date: July 23, 2003

UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **On two-sample sequential testing procedures based on strong approximation principles** submitted by **Abdulkadir Ahmed Hussein** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in Statistics.

Dr. Edit Gombay (Supervisor)

ces)

Dr. Richard Cook (University of Waterloo)

July 18, 2003

If all the trees on earth were pens and the ocean were ink, with seven oceans behind it to add to its supply, yet would not the knowledge of God be exhausted in the writing: for God is exalted in power, full of wisdom.

Quran, 31:27

———0———

To my father Ahmed, my mother Dahira, my wife Ayan, my son Zakaria and

to all my friends who started with me at the Somali National University in Mogadishu, but were unable to complete their degrees. Their studies, their hopes, their gifted minds and, perhaps, their lives were all shattered by the civil war.

Abstract

Originating in Abraham Wald's seminal work in the 40s, hypotheses testing in Sequential Analysis is now about 7 decades old. In spite of all the studies published in this area, little work has been done in the sequential testing of composite hypotheses. Composite hypotheses can arise when the parameter of interest is not a single point or when nuisance parameters are present along with the parameter of interest.

The objective of this thesis is to develop parametric and nonparametric fully-sequential procedures for testing two-sample hypotheses. The approach taken to reach this goal is based on the strong invariance principles (Csörgő and Révész 1981) and is inspired by the work of Gombay (1996, 1997, 2002a, 2002b, 2002c).

Specifically, we will use Rao's efficient score and Wald's statistic processes and functionals of them for the sequential testing of two-sided, two-sample, null hypotheses against two-sided alternative hypotheses with unknown nuisance parameters. The extension of these test procedures to the multi-sample case will also be discussed.

Nonparametric counterparts of the two-sample parametric tests, based on U-statistics with anti-symmetric kernels, will be developed.

Finally, Monte Carlo simulations will be carried out to compare the test procedures developed in the thesis to the fixed-sample t-test and group sequential t-tests of Pocock and O'Brien-Fleming in terms of total sample sizes, average stopping times, power, and robustness to deviations from the normality assumption.

Acknowledgements

My sincerest thanks go to my supervisor Professor Edit Gombay. Without her, this thesis would not have come the light. I thank her for teaching me the fascinating art of the Hungarian constructions and the statistical sequential analysis discipline. Thanks are also due to her for generously supporting me, from her NSERC funds for living and attending conferences. Throughout my Ph.D. programme, I always had in Edit a source of profound knowledge, warm smile and continuous encouragement.

Thanks are also due to Prof. K. C. Carriére; my being in statistics is due to her; She accepted me as her M.Sc. student when I had no clue of what statistics was all about and she passed to me some of her valuable knowledge; Profs. D. P. Wiens and T. Lau, who are the best two teachers I ever had; Profs. N. G. Prasad and Y. Wu, two oceans of intuition. If today, I have a global vision in this discipline, my vision is only due to them;

The above six icons will always remain as limsup points in my scholarly life, points that are always approachable and admirable but not reachable.

I thank Profs, R. J. Cook and A. Senthilselvan for agreeing to be external members on my exam committee and for their careful reading of my thesis. I also thank them for sacrificing some of their vacation time to accommodate my exam date.

Thanks to the Department of Mathematical and Statistical Sciences for the financial support I received during my stay. Thanks are also due to the support staff members and especially to Leona Guthrie, Dona Guelzow and Marion Benedict for their unforgettable support.

Deserved thanks go to my colleagues and friends Alex, Bhushan, Ella, Eshetu, Giseon and Roge. Each one of them is a mosaic of enthusiasm, knowledge and fun. A full description of their respective merits and the moral and material help they provided me would require an entire volume.

Needless to say, I cannot often enough express my gratitude to my mother my father and my sister Aisha, for their contribution to the completion of my degree.

I must confess that I lack the words to describe the love and the gratitude I have for my wife, Ayan, and my son, Zakaria. Although I have known them for less than three years, their support and encouragement have been instrumental in my ability to complete this thesis.

Abdulkadir

Contents

List of Figures

List of Tables

Conventions and notations:

Throughout the thesis, we will use lower-case letters to indicate random variables as well as their observed copies. Also, the following notations will be used;

Symbol	Meaning
$x_t \overset{a.s.}{=} O(\phi(t))$ or $x_t = O(\phi(t))$ $a.s.$	$\limsup_{t\to\infty} \left\lvert \frac{x_t}{\phi(t)} \right\rvert$ is almost surely finite
$x_t = O_p(\phi(t))$ ($x_t = o_p(\phi(t))$)	$\frac{x_t}{\phi(t)}$ is bounded (converges) in probability
$\xrightarrow{a.s.}$	Almost sure convergence
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\overset{\mathcal{D}}{=}$	Equality of (finite dimensional) distributions
LIL	Law of Iterated Logarithm

Chapter 1

Introduction

1.1 Sequential testing of composite hypotheses

Sequential testing of hypotheses began in 1943 when Abraham Wald developed what is known as the "Sequential Probability Ratio Test" (SPRT). Let $x_1, x_2, \ldots, x_n, \ldots$ be a sequence of independent and identically distributed random variables with a common one-parameter distribution $F(.;\theta)$ and a corresponding pdf $f(.;\theta)$. Wald's SPRT procedure for testing the simple null hypothesis $H_0 : \theta = \theta_0$ vs the simple alternative $H_A : \theta = \theta_1$ is to stop sampling and accept H_0 as soon as $L_n \geq A$ or stop sampling and accept H_A as soon as $L_n \leq B$, where $0 < B < 1 < A$ are constant stopping boundaries dictated by error probabilities $\alpha = P_{\theta_0}(\text{Reject } H_0)$ and $\beta = P_{\theta_1}(\text{Reject } H_A)$, and

$$L_n = \frac{\prod_{i=1}^{n} f(x_i;\theta_0)}{\prod_{i=1}^{n} f(x_i;\theta_1)}$$

is the likelihood ratio based on the n observations available thus far. The sample size, N, at which the boundaries are crossed, is a random variable. The mean of N is known as Average Sample Number (ASN) or average stopping

time. Wald and Wolfowitz (1948) showed that the so defined SPRT procedure is optimal in the sense of minimizing both $E_{\theta_0}\{N\}$ and $E_{\theta_1}\{N\}$ among all tests possessing a finite ASN under both H_0 and H_A and with errors α and β.

Clearly, the original SPRT did not deal with composite hypotheses, which were composite either because of the null and/or alternative parameter spaces not being single points or because of the presence of nuisance parameters. Wald attempted without much success to adopt the SPRT to the composite hypotheses case by introducing the weight functions approach. Another attempt to extend the SPRT to the case of nuisance parameters produced the so-called "Invariant SPRTs" (Ghosh 1970). This method consists of reducing the composite hypotheses to simple hypotheses through maximal invariant statistics by transforming the data as well as the hypotheses of interest and then applying Wald's SPRT procedure. According to Lai (2001), this approach has two drawbacks. Firstly, it necessitates the specification of a suitable alternative hypothesis, thus introducing some restrictions on the hypotheses to be tested under the maximal invariants. Secondly, the $\log L_n$ is no longer a random walk under the maximal invariants, so, the rich arsenal of the random walk theory, used by Wald-Wolfowitz to show optimality, is no longer applicable. A third practical drawback is the complexity of the test procedures derived from this approach, which require special tables.

In a case where the hypotheses of interest are composite because of the presence of nuisance parameters, a third approach suggested by Bartlett (1949), Cox (1963) and Breslow (1969) is based on using the likelihood ratio,

or an asymptotically equivalent form of it, under the assumption of contiguity. The method replaces nuisance parameters in the likelihood ratio by their restricted MLEs and uses Wald's SPRT procedure. This approach relies on the assumption of contiguity, i.e., on the assumption that the distance between the value of the parameter under the null hypothesis, θ_0, and its value under the alternative, θ_1, is such that $|\theta_1 - \theta_0| = O(N^{-1/2})$, where N is a sample size larger than that at which the sequential procedure reaches its decision. Alternatively, Gombay (1996, 1997) relaxed the contiguity assumption and provided some tests based on the generalized sequential likelihood ratio along with their asymptotic critical values at significance level α. Gombay (2002*c*) discussed why the Bartlett-Cox type of asymptotics fail under noncontiguous alternatives and compared the tests of Gombay (1996, 1997) to the sequential t-tests of Barnard (1947) and Rushton (1950, 1952), which are invariant SPRTs.

Whitehead (1978) adopted and improved the Bartlett-Cox approach by using closed triangular stopping boundaries. Anderson (1960) originally proposed this type of boundaries to reduce the expected sample size of the SPRT as an alternative to the open-ended Wald boundaries. Whitehead (1997) contributed much to the popularization of the triangular tests, to the extent that, recently, many clinical trials using these procedures have been conducted.

Sequential testing of hypotheses was introduced into the biomedical and clinical trials field during the 50s (Armitage 1960). As an alternative to the SPRTs, Armitage *et al.* (1969) suggested and studied the so-called "Repeated

Significance Test" (RST). Its key idea is to perform conventional fixed-sample significance testing on the cumulative data every time an observation arrives. That is, n_0 conventional fixed-sample tests will be performed if the total sample size attainable at the end of the study is n_0. The null hypothesis of interest is then rejected at the first inspection when the conventional fixed-sample test rejects it. The critical values, z_{α_i}, $i = 1, ..., n_0$, used for the intermediate testing, are obtained either by numerical integration as in Armitage *et al.* (1969) or from the approximating continuous time Wiener processes (Siegmund 1985). In any case, the RST approach did not solve the nuisance parameters problem.

Since, in double-blinded multi-centre clinical trials, frequent inspections may not be feasible, Pocock (1977) introduced a "group sequential" version of the RST. This approach performs a repeated significance testing only periodically as opposed to continuously testing after each observation. The conventional testing is performed at the pre-specified inspection times, $k = 1, ..., K$, with a fixed number of patients (group of patients) recruited between each two inspection times; that is, the number of patients, $n_k - n_{k-1}$, recruited between the $(k-1)^{th}$ and k^{th} inspection is the same for all $k = 2, ..., K$. The critical values, z_{α_k}, $k = 1, ..., K$, used for the intermediate testing, are obtained from the joint distribution of the K conventional test statistics by requiring that the overall significance level is a pre-specified α, i.e.,

$$P\{\text{Reject } H_0 \text{ at any } k \leq K\} = \alpha.$$

O'Brien and Fleming (1979) modified the constant boundary of Pocock's orig-

inal group sequential method (i.e., z_{α_k} = constant for all $k \leq K$) to a square root boundary. In 1981, the O'Brien-Fleming method was used in the famous Beta-Blocker Heart Attack Trial (BHAT 1982) . The original group sequential methods did not accommodate nuisance parameters. Jennison and Turnbull (1997) suggested group sequential t and F tests and provided recursive formulae for use in numerical computation of the boundaries. In general, obtaining exact boundaries for group sequential methods is a computationally intensive task, a disadvantage for group sequential methods if the number of interim analysis, K, is large. Jennison and Turnbull (1999) provide a detailed discussion of classical and recent developments in the group sequential methodology.

In the nonparametric field, sequential methodology for testing the equality of two or more distribution functions has undergone substantial development. However, this research has been largely concentrated on rank score statistics, developed for censored survival data. Chatterjee and Sen (1972), Majumdar and Sen (1978) and Sinha and Sen (1983) studied linear rank statistics for staggered entry survival data. Jones and Whitehead (1979, 1981) used log-rank and Gehan-Gilbert score rank tests. Tsiatis (1984), Slud and Wei (1982), Sellke and Siegmund (1983), Slud (1984), and Gu and Lai (1998) used rank score tests for censored data with random staggered entry. Murray and Tsiatis (1995) considered the sequential use of the Kaplan-Meier estimator for survival distributions. For the non-censored case, Miller and Sen (1972) derived weak invariance principles for U-statistics with symmetric kernels. Sen (1981) gives a detailed discussion of the use of such U-statistics in the sequen-

tial testing of hypotheses. Also, Gombay (2000*b*) used anti-symmetric kernels for change-point problems.

1.2 Overview of the thesis

As mentioned in the previous section, sequential testing of hypotheses in the presence of nuisance parameters has not yet received an adequate fully-sequential treatment. Most of the fully-sequential tests used in practice are based either on RST methods adapted to special cases with no nuisance parameters (c.f., Siegmund 1985) or on the Bartlett-Cox type asymptotics (c.f., Whitehead 1997) in which contiguity assumption is crucial.

In the cases where nuisance parameters are present and treatment effect differences might not be small enough (i.e., the contiguity assumption might be violated), Gombay (1996, 1997, 2002a, 2002b, 2002c) developed a class of one-sample two-sided tests using generalized sequential likelihood ratio and strong invariance principles (Csörgő and Révész 1981, Einmahl 1987, Einmahl 1989, Csörgő and Horváth 1993). The main attractive features of Gombay's approach are

a) Simple accommodation of the nuisance parameters.

b) Easy-to-compute approximate boundaries (critical values) which do not require any numerical integration.

c) A generality allowing application of the methods to a wide class of distribution families including the exponential family.

This thesis has three objectives:

1. Development of a class of two-sample sequential composite hypotheses-testing procedures and their extension to the multi-sample case. The methodology allows for random treatment allocation schemes and is based on Rao's efficient score and Wald statistics.

2. Development of a class of nonparametric procedures, counterparts of the above two-sample parametric procedures, using U-statistics with anti-symmetric kernels.

3. Empirical comparison of some of the parametric and nonparametric tests developed in the thesis to the fixed sample t-test and group sequential t-tests of Pocock and O'Brien-Fleming (Pocock 1977, O'Brien and Fleming 1979, Jennison and Turnbull 2001) in terms of power, maximum sample size (truncation point), average stopping time, and robustness to non-normality.

Specifically, in Chapter 2 we shall compare two-treatments in a clinical trial where patients arrive sequentially over time and are assigned to one of two treatments (e.g., experimental and standard) with allocation probabilities of λ and $(1-\lambda)$, $0 < \lambda < 1$. Asymptotic results for functionals of the Rao score and Wald statistics and for their weighted versions will be developed under H_0. Tests based on these asymptotic results are proposed in the same chapter, and Monte Carlo simulations are carried out in order to assess the power, Type I error and average stopping time of the tests. In Chapter 3,

the asymptotics of the functionals in Chapter 2 will be developed under the alternative hypothesis, and the consistency of the tests discussed in Chapter 2 will be proved. A methodology for obtaining the power and moments of the stopping time of these tests will be highlighted, with particular emphasis on the case of the one-dimensional parameter of interest.

In Chapter 4, we will present some straightforward extensions of the results of Chapters 2 & 3 to the multi-sample hypotheses testing and multiple comparison problems. In Chapter 5, we will construct nonparametric test procedures for the two-sample case by using U-statistics with anti-symmetric kernels. Finally, in Chapter 6, we will carry out some Monte Carlo studies to compare the two-sample parametric and nonparametric methods of Chapters 2 and 5 to the group sequential t-tests of Pocock and O'Brien-Fleming (Jennison and Turnbull 2001).

Chapter 2

Two-sample parametric tests and their asymptotics under H_0

2.1 Preliminaries

In clinical trials, the comparison of at least two treatments (e.g., active treatment and a placebo) is important and often encountered. For the two-treatment comparison, assume that patients arrive sequentially and are being assigned to either an experimental treatment (E) or a standard treatment (S) with allocation probabilities of λ and $(1-\lambda)$, respectively. At any stage, say, $k = m+n$, of such a sampling plan, we would have two independent sequences of observations, $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$, coming from patients assigned to the experimental and standard treatments, respectively. These independent observations can be continuous or discrete measurements. We assume that the distributions of the two streams of observations have densities of the same functional form but with different parameter vectors. The parameter vectors are partitioned into components of interest and nuisance components.

We assume the existence of a re-parameterisation under which the two

densities have different functional forms, $f_1(x;\boldsymbol{\theta},\boldsymbol{\eta})$ and $f_2(y;\boldsymbol{\theta},\boldsymbol{\eta})$, but share the same set of parameters, $(\boldsymbol{\theta}, \boldsymbol{\eta})$.

In this chapter, the composite null hypothesis,

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \boldsymbol{\eta} \in \Omega_2,$$

will be tested against the two-sided alternative

$$H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad \boldsymbol{\eta} \in \Omega_2, \tag{2.1}$$

where $\boldsymbol{\theta} \in \Omega_1 \subset \mathbb{R}^d$, $\boldsymbol{\eta} \in \Omega_2 \subset \mathbb{R}^p$ and thus, $\boldsymbol{\xi} = (\boldsymbol{\theta},\boldsymbol{\eta}) \in \Omega = \Omega_1 \times \Omega_2 \subset \mathbb{R}^{d+p}$.

This type of hypotheses are of interest in clinical trials where experimental and standard treatments are being compared and there is no prior knowledge of the direction of difference. It is also worth mentioning that the methods of this thesis for testing the above hypotheses do not have provision for early stopping under H_0. In fact, according to Armitage and Berry (1994) using procedures for early stopping under H_0 is not recommended because such cases have no ethical imperative, and the data collected will certainly be useful in learning more about the characteristics of the standard drug. Examples of trials in which continuing to the planned end of the trial is desired in the absence of evidence against H_0 are given, for instance, in Whitehead and Thomas (1997) and Donaldson *et al.* (2000).

Since the nuisance parameter, $\boldsymbol{\eta}$, is common to both distributions, we assume that the above null hypothesis represents equality of the densities of the two populations, i.e., $f_1(x;\boldsymbol{\theta}_0,\boldsymbol{\eta}) = f_2(y;\boldsymbol{\theta}_0,\boldsymbol{\eta}) = f(x;\boldsymbol{\theta}_0,\boldsymbol{\eta})$. In such cases, the two sequences of observations will be combined and denoted by $z_1, \ldots, z_k$, where $k = m + n$.

Before establishing the asymptotic null distribution of functionals of Rao's score and Wald statistics, we will give some necessary notations and regularity conditions. Denote by

$$I_{ij}(\boldsymbol{\xi}) = -E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f_1(x_1;\boldsymbol{\xi})\right], \tag{2.2}$$
$$J_{ij}(\boldsymbol{\xi}) = -E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f_2(y_1;\boldsymbol{\xi})\right],$$

entries of the expected Fisher information matrices of the first and second population, respectively, where $i,j = 1,2\ldots,d+p$. We shall partition these matrices based on the partition of the parameter vector $\boldsymbol{\xi} = (\boldsymbol{\theta},\boldsymbol{\eta})$. For example, we write,

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{2.3}$$

where $I_{11} = \left(-E\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_1(x_1;\boldsymbol{\xi})\right)_{d\times d}$, $I_{12} = I_{21}^t = \left(-E\frac{\partial^2}{\partial\theta_i\partial\eta_j}\log f_1(x_1;\boldsymbol{\xi})\right)_{d\times p}$, $I_{22} = \left(-E\frac{\partial^2}{\partial\eta_i\partial\eta_j}\log f_1(x_1;\boldsymbol{\xi})\right)_{p\times p}$. The inverse of I will also be partitioned and denoted by

$$I^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}. \tag{2.4}$$

Analogous notations apply to the matrix J. In general, given any partitioned matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

we shall be denoting the partitioned inverse of it by

$$A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}.$$

Also, we define

$$M(\boldsymbol{\xi}) = [\lambda I + (1-\lambda)J](\boldsymbol{\xi}),$$

which will be partitioned as above whenever needed.

Suppose that $(\boldsymbol{\theta}, \boldsymbol{\eta})$ is always a point in an open subset $\Omega \subset \mathbb{R}^{d+p}$. The following regularity conditions are required.

C1. Distribution functions $F_g(.;\boldsymbol{\theta}, \boldsymbol{\eta})$, $g = 1, 2$, are identifiable over Ω.

C2. There exists an open subset, $\Omega_0 \subset \Omega$, containing the true value of the parameter under H_0, $(\boldsymbol{\theta}_0, \boldsymbol{\eta})$, such that the partial derivatives

$$\frac{\partial}{\partial \xi_i} \log f_g(x;\boldsymbol{\xi}) \quad \frac{\partial^2}{\partial \xi_i \partial \xi_j} \log f_g(x;\boldsymbol{\xi}) \quad \frac{\partial^3}{\partial \xi_i \partial \xi_j \partial \xi_k} \log f_g(x;\boldsymbol{\xi})$$

exist and are continuous for all $x \in \mathbb{R}$, $\boldsymbol{\xi} \in \Omega_0$, and $g = 1, 2$, indicating the first and second population, respectively.

C3. For each and $m, n = 1, 2, 3, \ldots$, the score equations

$$\sum_{i=1}^{m} \nabla_{\boldsymbol{\eta}} \log f_1(x_i;\boldsymbol{\theta}_0, \boldsymbol{\eta}) + \sum_{j=1}^{n} \nabla_{\boldsymbol{\eta}} \log f_2(y_j;\boldsymbol{\theta}_0, \boldsymbol{\eta}) = 0$$

$$\sum_{i=1}^{m} \nabla_{\boldsymbol{\xi}} \log f_1(x_i;\boldsymbol{\xi}) + \sum_{j=1}^{n} \nabla_{\boldsymbol{\xi}} \log f_2(y_j;\boldsymbol{\xi}) = 0$$

have unique solutions, $\hat{\boldsymbol{\eta}}_k$, $\hat{\boldsymbol{\xi}}_k$, in Ω_0 where $k = m + n$.

C4. Under the setup of C2, there are functions $M_1(x)$, $M_2(x)$ such that

$$\int M_1(x)\nu(dx) < \infty; \quad E_{\boldsymbol{\theta}_0,\boldsymbol{\eta}}[M_2(X)] < \infty$$

with

$$\left|\frac{\partial}{\partial \xi_i} \log f_g(x;\boldsymbol{\xi})\right| \leq M_1(x), \quad \left|\frac{\partial^2}{\partial \xi_i \partial \xi_j} \log f_g(x;\boldsymbol{\xi})\right| \leq M_2(x),$$

$$\left|\frac{\partial^3}{\partial\xi_i\partial\xi_j\partial\xi_k}\log f_g(x;\boldsymbol{\xi})\right| \leq M_2(x)$$

for all $\boldsymbol{\xi} \in \Omega_0$, $1 \leq i,j,k \leq d+p$ and $g = 1,2$.

C5. $E_{\boldsymbol{\xi}}\frac{\partial}{\partial\xi_i}\log f_g(x;\boldsymbol{\xi}) = 0$, $1 \leq i \leq d+p$, $\boldsymbol{\xi} \in \Omega_0$, $g = 1,2$.

C6. $I_{ij}(\boldsymbol{\xi}) = -E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f_1(x_1;\boldsymbol{\xi})\right]$, $J_{ij}(\boldsymbol{\xi}) = -E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f_2(y_1;\boldsymbol{\xi})\right]$, $I^{-1}(\boldsymbol{\xi})$ and $J^{-1}(\boldsymbol{\xi})$ exist and are continuous for all $\boldsymbol{\xi} \in \Omega_0$ and $1 \leq i,j \leq d+p$.

C7. $\mathrm{Var}_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_i\partial\xi_j}\log f_g(x;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right] < \infty$ for $1 \leq i,j \leq d+p$, $g = 1,2$.

C8. $E_{\boldsymbol{\theta}_0,\boldsymbol{\eta}}\left[\left|\frac{\partial}{\partial\xi_i}\log f_g(x;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right|^{2+\delta}\right] < \infty$, $i = 1,2,...,d+p$, $g = 1,2$ and $\delta > 0$.

Remark 2.1 *Conditions C1 - C6 are the usual classical regularity conditions guaranteeing the existence and consistence of a sequence of MLEs (c.f., Lehmann (2001) and Serfling (1980)). The last two conditions are, respectively, required by the Law of Iterated Logarithm (Serfling 1980) and by the strong invariance principles of Komlós* et al. *(1975) and Einmahl (1987) that are used in this thesis.*

We shall state a Lemma that is useful in proving our results. Horváth (1993) provides a proof of this lemma.

Lemma 2.1 (Horváth, 1993) *Let $\{\boldsymbol{\beta}_i; 1 \leq i < \infty\}$ be a sequence of independent and identically distributed random vectors such that $E\beta_{ij} = 0$, $E\beta_{ij}^2 = 1$, $E\beta_{ij}\beta_{il} = 0$ $(j \neq l)$, $E|\beta_{ij}|^{2+\delta} < \infty$ for some $\delta > 0$, $1 \leq j,l \leq d$. Then, as $n_0 \longrightarrow \infty$,*

$$P\left[a(\log n_0)\max_{1\leq k\leq n_0}\left(k^{-1}\sum_{j=1}^{d}\left(\sum_{i=1}^{k}\beta_{ij}\right)^2\right)^{1/2} \leq x + b_d(\log n_0)\right] \longrightarrow \exp(-e^{-x}),$$

where

$$a(x) = (2\log x)^{1/2} \tag{2.5}$$

and

$$b_d(x) = 2\log x + \frac{d}{2}\log\log x - \log\Gamma(\frac{1}{2}d) \tag{2.6}$$

with $\Gamma(c) = \int_0^\infty y^{c-1}e^{-y}dy$.

2.2 Rao's efficient score statistic

We shall use Rao's efficient score statistic to test H_0. If under H_0, $\boldsymbol{\xi} = \boldsymbol{\xi}_0$ is a completely specified parameter value, then the general form of Rao's statistic would be

$$R_k = V_k I^{-1}(\boldsymbol{\xi}_0) V_k^t,$$

where $V_k = k^{-1/2}\left[\sum_{i=1}^m \nabla_{\boldsymbol{\xi}} \log f_1(x_i;\boldsymbol{\xi}_0) + \sum_{j=1}^n \nabla_{\boldsymbol{\xi}} \log f_2(y_j;\boldsymbol{\xi}_0)\right]$, $I(\boldsymbol{\xi}_0)$ is the expected Fisher information matrix (Serfling 1980) under H_0, $\nabla_{\boldsymbol{\xi}}$ denotes the vector of partial derivatives with respect to $\boldsymbol{\xi}$, and the superscript t denotes a vector or matrix transpose. In the current problem, however, a nuisance parameter is present and has to be replaced by its restricted maximum likelihood estimator under H_0; that is, Rao's statistic becomes

$$R_k = V_k \Sigma^{-1}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) V_k^t, \tag{2.7}$$

where

$$V_k = k^{-1/2}\left[\sum_{i=1}^m \nabla_{\boldsymbol{\theta}} \log f_1(x_i;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k) + \sum_{j=1}^n \nabla_{\boldsymbol{\theta}} \log f_2(y_j;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k)\right], \tag{2.8}$$

and $\hat{\boldsymbol{\eta}}_k$ is the restricted MLE of the nuisance parameter $\boldsymbol{\eta}$, which is the solution of the equation

$$\sum_{i=1}^{m} \nabla_{\boldsymbol{\eta}} \log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + \sum_{j=1}^{n} \nabla_{\boldsymbol{\eta}} \log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta}) = \sum_{i=1}^{k} \nabla_{\boldsymbol{\eta}} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) = 0, \tag{2.9}$$

and $\Sigma(\boldsymbol{\theta}_0,\boldsymbol{\eta}) = [M^{11}(\boldsymbol{\theta}_0,\boldsymbol{\eta})]^{-1}$. This fact is rigorously stated in the following theorem which approximates R_k in (2.7) by means of Wiener processes.

Theorem 2.1 *Under H_0, if conditions C1-C8 hold, then there exist independent Wiener processes, $B_j(t)$, $j = 1, 2, \ldots, d$, such that for $\alpha \leq \frac{1}{2} - \frac{1}{2+\delta}$,*

$$\sup_{1\leq t<\infty} \left| R_{[nt]} - U(nt) \right| \stackrel{a.s.}{=} O(n^{-\alpha}(\log\log n)^{1/2}), \tag{2.10}$$

where $[nt]$ indicates the closest integer to nt,

$$\begin{aligned} R_{[nt]} &= V_{[nt]}\Sigma^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta})V_{[nt]}^t \\ U(x) &= \frac{1}{x}\sum_{j=1}^{d} B_j^2(x), \\ \Sigma(\boldsymbol{\theta}_0,\boldsymbol{\eta}) &= [\dot{M}^{11}(\boldsymbol{\theta}_0,\boldsymbol{\eta})]^{-1} = \left[(\lambda I + (1-\lambda)J)^{11}\right]^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}), \end{aligned} \tag{2.11}$$

$\delta > 0$.

Proof. Suppose for simplicity of notation, that under H_0, the true value of the nuisance parameter, common to both populations, is $\boldsymbol{\eta}$. Re-write (2.8) as

$$\begin{aligned} V_k &= k^{-1/2}\left[\sum_{i=1}^{m} \nabla_{\boldsymbol{\theta}} \log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + \sum_{j=1}^{n} \nabla_{\boldsymbol{\theta}} \log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right] \\ &+ k^{-1/2}\left[\sum_{i=1}^{m} \nabla_{\boldsymbol{\theta}} \log f_1(x_i;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k) - \sum_{i=1}^{m} \nabla_{\boldsymbol{\theta}} \log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right] \\ &+ k^{-1/2}\left[\sum_{j=1}^{n} \nabla_{\boldsymbol{\theta}} \log f_2(y_j;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k) - \sum_{j=1}^{n} \nabla_{\boldsymbol{\theta}} \log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right]. \end{aligned} \tag{2.12}$$

The last two terms in the above sum represent errors committed in estimating $\boldsymbol{\eta}$ by its MLE $\hat{\boldsymbol{\eta}}_k$. These errors can be linearized by using a three-term Taylor expansion of $\nabla_{\boldsymbol{\theta}} \log f_1(x_i;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k)$ and $\nabla_{\boldsymbol{\theta}} \log f_2(y_j;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k)$ around the true parameter value $\boldsymbol{\eta}$ so that

$$\begin{aligned} V_k &= k^{-1/2}\left[\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})+\sum_{j=1}^{n}\nabla_{\boldsymbol{\theta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right] \\ &+ k^{-1/2}(\hat{\boldsymbol{\eta}}_k-\boldsymbol{\eta})\left[\sum_{i=1}^{m}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\eta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})+\sum_{j=1}^{n}\nabla^2_{\boldsymbol{\theta}\boldsymbol{\eta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})\right] \\ &+ \epsilon(k,\boldsymbol{\eta}^*), \end{aligned} \tag{2.13}$$

where $\hat{\boldsymbol{\eta}}_k < \boldsymbol{\eta}^* < \boldsymbol{\eta}$ and $\nabla^2_{\boldsymbol{\theta}\boldsymbol{\eta}}$ denotes a $d \times p$-matrix of second order partial derivatives, first with respect to components of $\boldsymbol{\theta}$ and secondly with respect to components of $\boldsymbol{\eta}$. The term $\epsilon(k,\boldsymbol{\eta}^*)$ is a row vector whose r^{th} component has the form

$$\begin{aligned} &\frac{1}{2}k^{-1/2}\sum_{l=1}^{p}\sum_{q}^{p}k(\hat{\eta}_{kq}-\eta_q)(\hat{\eta}_{kl}-\eta_l)\left\{\frac{1}{k}\sum_{i=1}^{m}\frac{\partial^3}{\partial\eta_q\partial\eta_l\partial\theta_r}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}^*)\right\} \\ +\ &\frac{1}{2}k^{-1/2}\sum_{l=1}^{p}\sum_{q}^{p}k(\hat{\eta}_{kq}-\eta_q)(\hat{\eta}_{kl}-\eta_l)\left\{\frac{1}{k}\sum_{j=1}^{n}\frac{\partial^3}{\partial\eta_q\partial\eta_l\partial\theta_r}\log f_2(y_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}^*)\right\}. \end{aligned}$$

By the Law of Iterated Logarithm (LIL) and by C4, the terms in the curly brackets are almost surely $O(k^{-1/2}(\log\log k)^{1/2})$. On the other hand, by Lemma 2.1 of Gombay and Horváth (1994), $\frac{k}{\log\log k}(\hat{\eta}_{kq}-\eta_q)(\hat{\eta}_{kl}-\eta_l)$ is almost surely $O(1)$. Hence,

$$\epsilon(k,\boldsymbol{\eta}^*) = O\left(\frac{(\log\log k)^{3/2}}{k}\right). \tag{2.14}$$

In order to obtain an expression for $(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta})$, we shall analyse the following three-term Taylor expansion of the pooled data log-likelihood under H_0,

$$\begin{aligned}\sum_{i=1}^{k}[\nabla_{\boldsymbol{\eta}} \log f(z_i;\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k) \;-\; \nabla_{\boldsymbol{\eta}} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})] \\ = \; (\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) \sum_{i=1}^{k} \nabla^2_{\boldsymbol{\eta}^2} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + \epsilon(k,\boldsymbol{\eta}^*),\end{aligned} \tag{2.15}$$

where $\nabla^2_{\boldsymbol{\eta}^2}$ is a $p \times p$-matrix of partial derivatives with respect to the components of $\boldsymbol{\eta}$. By the same arguments leading to (2.14), the error term above is almost surely of amplitude $O\left(\frac{(\log\log k)^{3/2}}{\sqrt{k}}\right)$. On the other hand, by C4 and the LIL we have

$$\limsup_{k\to\infty} \left\| \frac{1}{k} \sum_{i=1}^{k} \nabla^2_{\boldsymbol{\eta}^2} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + I_{22}(\boldsymbol{\theta}_0,\boldsymbol{\eta}) \right\| = O\left(\frac{(\log\log k)^{1/2}}{k^{1/2}}\right) \quad \text{a.s.}$$

Thus, we have

$$\begin{aligned}-\sum_{i=1}^{k} \nabla_{\boldsymbol{\eta}} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) &= k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) \left\{ \frac{1}{k} \sum_{i=1}^{k} \nabla^2_{\boldsymbol{\eta}^2} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + I_{22} - I_{22} \right\} \\ &+ O\left(\frac{(\log\log k)^{3/2}}{\sqrt{k}}\right) \\ &= k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) \left\{ \frac{1}{k} \sum_{i=1}^{k} \nabla^2_{\boldsymbol{\eta}^2} \log f(z_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) + I_{22} \right\} \\ &- k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) I_{22} + O\left(\frac{(\log\log k)^{3/2}}{\sqrt{k}}\right) \\ &= k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) O\left(\frac{(\log\log k)^{1/2}}{k^{1/2}}\right) - k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) I_{22} \\ &+ O\left(\frac{(\log\log k)^{3/2}}{\sqrt{k}}\right) \\ &= O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right) - k(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}) I_{22} \\ &+ O\left(\frac{(\log\log k)^{3/2}}{\sqrt{k}}\right),\end{aligned}$$

where the last equality follows from the Lemma 2.1 of Gombay and Horváth (1994). Since, by C6, I_{22} is invertible, we have

$$(\hat{\eta}_k - \eta) = \left[\frac{1}{k}\sum_{i=1}^{k} \nabla_\eta \log f(z_i;\theta_0,\eta)\right] I_{22}^{-1}(\theta_0,\eta) + O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right). \quad (2.16)$$

Collecting (2.13), (2.14) and (2.16), we obtain

$$\begin{aligned} V_k &= k^{-1/2}\left[\sum_{i=1}^{m} \nabla_\theta \log f_1(x_i;\theta_0,\eta) + \sum_{j=1}^{n} \nabla_\theta \log f_2(y_j;\theta_0,\eta)\right] \\ &+ k^{-1/2}\left[\sum_{i=1}^{k} \nabla_\eta \log f(z_i;\theta_0,\eta)\right] I_{22}^{-1}(\theta_0,\eta) \\ &\times \left[\frac{1}{k}\sum_{i=1}^{m} \nabla^2_{\theta\eta} \log f_1(x_i;\theta_0,\eta) + \frac{1}{k}\sum_{j=1}^{n} \nabla^2_{\theta\eta} \log f_2(y_j;\theta_0,\eta)\right] \\ &+ k^{-1/2} O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right)\left[\sum_{i=1}^{m} \nabla^2_{\theta\eta} \log f_1(x_i;\theta_0,\eta)\right. \\ &+ \left.\sum_{j=1}^{n} \nabla^2_{\theta\eta} \log f_2(y_j;\theta_0,\eta)\right] + O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right). \end{aligned} \quad (2.17)$$

Again, by C6 and LIL,

$$\limsup_{m\to\infty} \left\| \frac{1}{m}\sum_{i=1}^{m} \nabla^2_{\theta\eta} \log f_1(x_i;\theta_0,\eta) + I_{12} \right\| = O\left(\frac{(\log\log k)^{1/2}}{k^{1/2}}\right) \quad a.s.,$$

$$\limsup_{m\to\infty} \left\| \frac{1}{n}\sum_{j=1}^{n} \nabla^2_{\theta\eta} \log f_2(y_j;\theta_0,\eta) + J_{12} \right\| = O\left(\frac{(\log\log k)^{1/2}}{k^{1/2}}\right) \quad a.s.$$

It follows, therefore, after neglecting errors that are at most $O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right)$,

$$\begin{aligned} V_k &= k^{-1/2}\left[\sum_{i=1}^{m} \nabla_\theta \log f_1(x_i;\theta_0,\eta) + \sum_{j=1}^{n} \nabla_\theta \log f_2(y_j;\theta_0,\eta)\right] \\ &+ k^{-1/2}\left[\sum_{i=1}^{k} \nabla_\eta \log f(z_i;\theta_0,\eta)\right] I_{22}^{-1}(\theta_0,\eta)\left[-\lambda I_{12} - (1-\lambda)J_{12}\right]^t \\ &+ O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right) \quad a.s. \end{aligned}$$

If we re-arrange terms, V_k takes the following final form,

$$\begin{aligned} V_k &= k^{-1/2}\sum_{i=1}^{m}\left[\nabla_{\boldsymbol{\theta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) - \nabla_{\boldsymbol{\eta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})I_{22}^{-1}\left\{\lambda I_{12}+(1-\lambda)J_{12}\right\}\right] \\ &+ k^{-1/2}\sum_{j=1}^{n}\left[\nabla_{\boldsymbol{\theta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta}) - \nabla_{\boldsymbol{\eta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})I_{22}^{-1}\left\{\lambda I_{12}+(1-\lambda)J_{12}\right\}\right] \\ &+ O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right) = C_k + D_k + O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}}\right) \quad a.s. \end{aligned} \tag{2.18}$$

Observe that V_k in (2.18) consists of the sum of the independent d-dimensional random vectors, $\mathbf{z}_i$, defined by

$$\mathbf{z}_i = \begin{cases} \nabla_{\boldsymbol{\theta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta}) - \nabla_{\boldsymbol{\eta}}\log f_1(x_i;\boldsymbol{\theta}_0,\boldsymbol{\eta})I_{22}^{-1}\left\{\lambda I_{12}+(1-\lambda)J_{12}\right\} & \text{wp } \lambda \\ \nabla_{\boldsymbol{\theta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta}) - \nabla_{\boldsymbol{\eta}}\log f_2(y_j;\boldsymbol{\theta}_0,\boldsymbol{\eta})I_{22}^{-1}\left\{\lambda I_{12}+(1-\lambda)J_{12}\right\} & \text{wp } \lambda' \end{cases} \tag{2.19}$$

where $\lambda' = 1-\lambda$, $i = 1,2,\ldots,k = m+n$. By C5, $E\mathbf{z}_i = 0$ for all i. This equation, along with our treatment allocation rule, implies that

$$\mathrm{Cov}(V_k) = \Sigma = \lambda\Sigma_1 + (1-\lambda)\Sigma_2,$$

where Σ_1 and Σ_2 are the covariance matrices of the terms in C_k and D_k respectively. Now,

$$\begin{aligned} \Sigma_1 &= I_{11} + \left\{\lambda I_{12}+(1-\lambda)J_{12}\right\} I_{22}^{-1} I_{22} I_{22}^{-1}\left\{\lambda I_{21}+(1-\lambda)J_{21}\right\} \\ &- I_{12}I_{22}^{-1}\left\{\lambda I_{21}+(1-\lambda)J_{21}\right\} - \left\{\lambda I_{12}+(1-\lambda)J_{12}\right\} I_{22}^{-1} I_{21} \\ &= I_{11} + \left\{\lambda I_{12}+(1-\lambda)J_{12}\right\} I_{22}^{-1}\left\{\lambda I_{21}+(1-\lambda)J_{21} - I_{21}\right\} \\ &- I_{12}I_{22}^{-1}\left\{\lambda I_{21}+(1-\lambda)J_{21}\right\}. \end{aligned} \tag{2.20}$$

Similarly, covariances of the terms in D_k are given by

$$\begin{aligned}
\Sigma_2 &= J_{11} + \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} I_{22} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\} \\
&- J_{12} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\} - \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} J_{21} \\
&= J_{11} + \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21} - J_{21}\} \\
&- J_{12} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\}. \qquad (2.21)
\end{aligned}$$

So,

$$\begin{aligned}
\mathrm{Cov}(V_k) &= \lambda\Sigma_1 + (1-\lambda)\Sigma_2 \\
&= \lambda I_{11} + (1-\lambda)J_{11} + \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \\
&\times \left\{\lambda^2 I_{21} + \lambda(1-\lambda)J_{21} - \lambda I_{21} + \lambda(1-\lambda)I_{21} + (1-\lambda)^2 J_{21} - (1-\lambda)J_{21}\right\} \\
&- \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\} \\
&= \lambda I_{11} + (1-\lambda)J_{11} + \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \\
&\times \{\lambda I_{21} + (1-\lambda)J_{21} - \lambda I_{21} - (1-\lambda)J_{21}\} - \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \\
&- \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\} \\
&= \lambda I_{11} + (1-\lambda)J_{11} - \{\lambda I_{12} + (1-\lambda)J_{12}\} I_{22}^{-1} \{\lambda I_{21} + (1-\lambda)J_{21}\} \\
&= \lambda I_{11} + (1-\lambda)J_{11} - \{\lambda I_{12} + (1-\lambda)J_{12}\} \{\lambda I_{22} + (1-\lambda)J_{22}\}^{-1} \\
&\times \{\lambda I_{21} + (1-\lambda)J_{21}\}, \qquad (2.22)
\end{aligned}$$

where the last equality holds since $I_{22} = \{\lambda I_{22} + (1-\lambda)J_{22}\}$ under H_0. By the inversion rules of partitioned matrices (Rao 1965, p.29), (2.22) is, in fact,

$$\mathrm{Cov}(V_k) = \Sigma = \left[(\lambda I + (1-\lambda)J)^{11}\right]^{-1}. \qquad (2.23)$$

Using C5, C8 and the fact that the above covariance matrix is symmetric and positive definite and the $\mathbf{z}_i$ in (2.19) are independent and identically

distributed random variables under H_0, we can write

$$R_k = V_k \Sigma^{-1} V_k^t = k^{-1} \sum_{j=1}^{d} \left(\sum_{i=1}^{m+n} \beta_{ij} \right)^2 + O\left(\frac{(\log\log k)^{3/2}}{k^{1/2}} \right) \quad \text{a.s.}, \qquad (2.24)$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{id})$, $1 \leq i < \infty$, are i.i.d. random vectors with $E\beta_{ij} = 0$, $E\beta_{ij}^2 = 1$, $E\beta_{ij}\beta_{il} = 0$ $(j \neq l)$, $E|\beta_{ij}|^{2+\delta} < \infty$, $\delta > 0$, $1 \leq j, l \leq d$. Now, (2.10 follows from Lemma 3.3 of Gombay (1996) using similar steps to those in the proof of Theorem 1.1 therein.

□

Remark 2.2 *From the proof of the Theorem, we can easily see that the allocation probability, λ, can be replaced by m/k since, by LIL, $|m/k - \lambda| \overset{a.s.}{=} O((\log\log k)^{1/2}/k^{1/2})$, as $k \to \infty$.*

The following Corollary gives the asymptotic distribution of the maximal functional of R_k and a weighted version of it.

Corollary 2.1 *Under the conditions of Theorem 2.1 and for any integer n_0 (truncation point),*

(i) $\lim_{n_0 \to \infty} P\left\{ a(\log n_0) \max_{1 \leq k \leq n_0} R_k^{1/2} \leq t + b_d(\log n_0) \right\} = \exp(-e^{-t})$, $t \in \mathbb{R}$ *where $R_k = V_k(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)\Sigma^{-1}(\boldsymbol{\theta}_0, \boldsymbol{\eta})V_k^t(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)$, $a(x)$ and $b_d(x)$ are defined in (2.5) and (2.6), respectively.*

(ii) $\max_{1 \leq k \leq n_0} \left(\frac{k}{n_0} R_k \right)^{1/2} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} \left(\sum_{j=1}^{d} B_j^2(t) \right)^{1/2}$ *as* $n_0 \longrightarrow \infty$.

Proof.

(i) It follows directly from (2.24) and Lemma 2.1.

(ii) From Theorem 2.1, it follows that for $\alpha = \frac{1}{2} - \frac{1}{2+\delta}$,

$$\sup_{1 \le t < \infty} \left| R_{[kt]} - U(kt) \right| \stackrel{a.s.}{=} O(k^{-\alpha} (\log \log k)^{1/2}).$$

So, for $t = 1$, we have,

$$|R_k - U(k)| \stackrel{a.s.}{=} O\left(k^{-\alpha} (\log \log k)^{-1/2}\right).$$

Taking max over $1 \le k \le n_0$ and remembering that almost sure boundedness implies boundedness in probability, we get

$$\max_{1 \le k \le n_0} \left(k^{\alpha} (\log \log k)^{-1/2}\right) |R_k - U(k)| = O_p(1).$$

Now,

$$\begin{aligned} \max_{1 \le k \le n_0} \left| \frac{k}{n_0} R_k - \frac{k}{n_0} U(k) \right| &\le \max_{1 \le k \le n_0} \frac{k}{n_0} \max_{1 \le k \le n_0} \left(k^{\alpha} (\log \log k)^{-1/2}\right)^{-1} \\ &\times \max_{1 \le k \le n_0} \left(k^{\alpha} (\log \log k)^{-1/2}\right) |R_k - U(k)| = o_p(1), \end{aligned}$$

and so, (ii) follows.

□

In order for this corollary to be useful in testing hypotheses, we need to replace the covariance matrix, Σ, by an estimated version. The following Lemma will allow us to do so.

Lemma 2.2 *If the conditions of Theorem 2.1 hold, then replacing* $\boldsymbol{\eta}$ *in* $\Sigma^{-1}(\boldsymbol{\theta}_0, \boldsymbol{\eta})$ *by its restricted MLE,* $\hat{\boldsymbol{\eta}}_k$*, does not change the limits in Cororllary 2.1(i) and (ii).*

Proof. Let $\hat{\Sigma}$ denote the covariance matrix with $\boldsymbol{\eta}$ replaced by $\hat{\boldsymbol{\eta}}_k$. From the approximation of V_k by i.i.d. terms in the proof of Theorem 2.1 we get that for any l, $\max_{1<k\leq l} ||V_k|| = O_p((\log\log l)^{1/2})$. Hence as $n_0 \to \infty$,

$$\max_{1<k\leq \log n_0} (k/n_0)^{1/2} R_k^{1/2} = O_p(n_0^{-1/2}(\log n_0)^{1/2}(\log\log\log n_0)^{1/2}).$$

By the assumptions, each component of the matrix $B = \hat{\Sigma}(\theta_0, \hat{\eta}_k) - \Sigma(\theta_0, \eta)$ is almost surely bounded, and $\max V_k$ is taken in the range $\log n_0 < k \leq n_0$, where the matrix B is almost surely $o(1)$. From these facts, we obtain the Lemma.

□

Remark 2.3 *In the preceding lemma, $\hat{\boldsymbol{\eta}}_k$ need not be the unique solution of the restricted log-likelihood. All we need is that the estimator converges weakly to the parameter $\boldsymbol{\eta}$ with a rate of at least $1/\sqrt{\log k}$. This can be attained by, say, a one-step estimator based on the Newton-Raphson iterative procedure (see Lehmann 2001, p.475). Therefore, condition C3, requiring $\hat{\boldsymbol{\eta}}_k$ and $\hat{\boldsymbol{\xi}}_k$ to be the unique MLEs can be relaxed accordingly.*

Remark 2.4 *From the theoretical point of view, we could use the unrestricted MLE, $\tilde{\boldsymbol{\eta}}_k$, instead of the restricted one, $\hat{\boldsymbol{\eta}}_k$. However, since the use of the restricted MLE is what gives Rao's score statistic its attractiveness and since it is hard to imagine any gain by using a less efficient estimator (the unrestricted MLE), we will concentrate, in this thesis, on the case where the restricted MLE is used.*

2.3 Wald's statistic

In a similar manner to that used in Section 2.2, we now consider tests based on functionals of Wald-type statistic process. All notations and conditions C1 - C8 of Section 2.1, along with the treatment allocation scheme therein, are still required for the results of this section. Furthermore, we write $\boldsymbol{\xi}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\eta})$ and $\tilde{\boldsymbol{\xi}}_k = (\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\eta}}_k)$, where $\tilde{\boldsymbol{\theta}}_k$ and $\tilde{\boldsymbol{\eta}}_k$ are the unrestricted MLEs of the parameter of interest and the nuisance parameter, respectively.

For testing H_0 of Section 2.1 in a non-sequential situation where a nuisance parameter is present, Wald's test statistic is given by

$$W = k(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[I^{11}(\tilde{\boldsymbol{\xi}})]^{-1}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t.$$

In a sequential setup, we will treat the statistic as a processes W_k and obtain its strong approximations by means of Brownian motions, after a proper normalisation.

Theorem 2.2 *Assume that conditions C1 - C8 hold under H_0, then there exist independent Wiener processes, $B_j(t)$, $j = 1, 2, \ldots, d$, such that for $\alpha \leq \frac{1}{2} - \frac{1}{2+\delta}$,*

$$\sup_{1 \leq t < \infty} \left| W_{[nt]} - U(nt) \right| = O(n^{-\alpha}(\log\log n)^{1/2}) \quad a.s., \tag{2.25}$$

where

$$\begin{aligned} U(x) &= \frac{1}{x}\sum_{j=1}^{d} B_j^2(x), \\ W_{[nt]} &= [nt](\tilde{\boldsymbol{\theta}}_{[nt]} - \boldsymbol{\theta}_0)\Sigma^{-1}(\boldsymbol{\xi}_0)(\tilde{\boldsymbol{\theta}}_{[nt]} - \boldsymbol{\theta}_0)^t \\ \Sigma(\boldsymbol{\xi}_0) &= [\lambda I + (1-\lambda)J]^{11}(\boldsymbol{\xi}_0) = M^{11}(\boldsymbol{\xi}_0), \end{aligned} \tag{2.26}$$

$\delta > 0$.

Proof. Using Taylor expansion and error analysis similar to that in the proof of Theorem 2.1, we have

$$\begin{aligned} -\sum_i \nabla_{\xi} \log f_1(x_i;\xi_0) \; - \; & \sum_j \nabla_{\xi} \log f_2(y_j;\xi_0) \\ = \; & k(\tilde{\xi}_k - \xi_0)\frac{1}{k}\left[\sum_i \nabla^2_{\xi} \log f_1(x_i;\xi_0) + \sum_j \nabla^2_{\xi} \log f_2(y_j;\xi_0)\right] \\ + \; & \left(O((\log\log k)^{3/2}k^{-1/2})\right) \quad a.s. \end{aligned} \tag{2.27}$$

By using C4, C6 and the Law of Iterated Logarithm,

$$\begin{aligned} \sqrt{k}(\tilde{\xi}_k - \xi_0) \; = \; & \frac{1}{\sqrt{k}}\left[\sum_i \nabla_{\xi} \log f_1(x_i;\xi_0) + \sum_j \nabla_{\xi} \log f_2(y_j;\xi_0)\right][\lambda I + (1-\lambda)J]^{-1} \\ + \; & \left(O((\log\log k)^{3/2}k^{-1/2})\right) \quad a.s. \end{aligned} \tag{2.28}$$

Hence, for θ we have

$$\begin{aligned} \sqrt{k}(\tilde{\theta}_k - \theta_0) \; = \; & \frac{1}{\sqrt{k}}\left[\sum_i \nabla_{\xi} \log f_1(x_i;\xi_0) + \sum_j \nabla_{\xi} \log f_2(y_j;\xi_0)\right]\begin{pmatrix} M^{11} \\ M^{21} \end{pmatrix} \\ + \; & \left(O((\log\log k)^{3/2}k^{-1/2})\right) \quad a.s. \end{aligned} \tag{2.29}$$

The covariance of the main terms on the right-hand side of the above equality is found to be

$$\Sigma(\xi_0) = \left[\begin{pmatrix} M^{11} \\ M^{21} \end{pmatrix}^t M \begin{pmatrix} M^{11} \\ M^{21} \end{pmatrix}\right](\xi_0) = M^{11}(\xi_0).$$

Now the rest of the proof proceeds in the same way as that of Theorem 2.1.

□

The following corollary and lemma are similar in spirit and proof to Corollary 2.1 and Lemma 2.2, so their proofs will be omitted.

Corollary 2.2 *Under the conditions of Theorem 2.2,*

(i) $\lim_{n_0\to\infty} P\left\{a(\log n_0)\max_{1\le k\le n_0} W_k^{1/2} \le t + b_d(\log n_0)\right\} = \exp(-e^{-t}),\ \ t \in$ IR *where* $W_k = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)\Sigma^{-1}(\boldsymbol{\xi}_0)(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t$, $a(x)$ *and* $b_d(x)$ *are defined in (2.5) and (2.6), respectively.*

(ii) $\max_{1\le k\le n_0}\left(\frac{k}{n_0}W_k\right)^{1/2} \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1}\left(\sum_{j=1}^d B_j^2(t)\right)^{1/2}$ *as* $n_0 \longrightarrow \infty$.

Lemma 2.3 *Replacing* $\boldsymbol{\xi}_0$ *in* $\Sigma^{-1}(\boldsymbol{\xi}_0)$ *by either* $\tilde{\boldsymbol{\xi}}_k = (\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\eta}}_k)$, $\tilde{\boldsymbol{\xi}}_{0k} = (\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_k)$ *or restricted MLE,* $\hat{\boldsymbol{\xi}}_{0k} = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)$ *and* λ *by* m/k *do not change the limits in Corollary 2.2(i) and (ii).*

2.4 Test procedures

Let $stat_k$ denote either of

$$R_k^* = V_k(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)(M^{11})(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)V_k^t(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k), \tag{2.30}$$

$$W_k^{*(1)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)(M^{11})^{-1}(\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\eta}}_k)(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t, \tag{2.31}$$

$$W_k^{*(2)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)(M^{11})^{-1}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_k)(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t \tag{2.32}$$

or

$$W_k^{*(3)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)(M^{11})^{-1}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t. \tag{2.33}$$

As a consequence of the Corollaries 2.1, 2.2 and Lemmas 2.2 and 2.3, we can identify two, α-level sequential testing procedures, truncated at n_0 observations.

Test 2.1 (Test 1) *Compute* $\sqrt{stat_k}$ *for* $k = 1, 2, ..., n_0$ *and reject* H_0 *at the first* k *when it exceeds*

$$CV_1 = CV_1(\alpha, n_0, d) = (a(\log n_0))^{-1} \left[-\log(-\log(1-\alpha)) + b_d(\log n_0) \right]. \tag{2.34}$$

Fail to reject H_0 *if it is not rejected by* $k = n_0$.

The functions $a(x)$ and $b_d(x)$ are defined in (2.5) and (2.6), respectively.

A better finite sample approximation to the critical value in Test 2.1 can be obtained by using a result in Vostrikova (1981). In fact, using methods similar to those used in proving Corollary 2.1(ii), we can show that

$$\left| \max_{1 \le k \le n_0} (stat_k)^{1/2} - \sup_{1 \le t \le n_0} U^{1/2}(t) \right| = o_p(1). \tag{2.35}$$

Also, from Itô and McKean (1965),

$$\{U^{1/2}(t); 1 \le t < \infty\} \stackrel{\mathcal{D}}{=} \{N(\log t); 1 \le t < \infty\}, \tag{2.36}$$

where $\{N(t), 0 \le t < \infty\}$ is a diffusion process with backward equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} + \frac{1}{2} \left(\frac{d-1}{x} - x \right) \frac{\partial u}{\partial x}$$

and boundary condition

$$\lim_{x \downarrow 0} x^{d-1} \frac{\partial u}{\partial x} = 0.$$

On the other hand, from Vostrikova (1981), we have for a fixed $T > 0$,

$$P\left(\sup_{0\le t\le T} N(t) > x\right) = \frac{x^d \exp(-x^2/2)}{2^{d/2}\gamma(d/2)}\left\{T - \frac{d}{x^2}T + \frac{4}{x^2} + O\left(\frac{1}{x^4}\right)\right\} \quad (2.37)$$

as $x \longrightarrow \infty$.

From (2.35), (2.36) and (2.37), we can get a better approximation, $CV_1' = CV_1'(\alpha, n_0, d)$, to the critical value of Test 1 by solving

$$\alpha = \frac{(CV_1')^d \exp(-(CV_1')^2/2)}{2^{d/2}\Gamma(d/2)}\left\{\log n_0 - \frac{d}{(CV_1')^2}\log n_0 + \frac{4}{(CV_1')^2} + O\left(\frac{1}{(CV_1')^4}\right)\right\}. \quad (2.38)$$

Test 2.2 *(Test 2) For $k = 1, 2, \ldots, n_0$, compute $\sqrt{\frac{k}{n_0}stat_k}$ and reject H_0 the first time it exceeds $CV_2 = CV_2(\alpha, d)$. Fail to reject H_0 if it is not rejected by $k = n_0$. The critical value, CV_2, is obtained by using part (ii) of Corollaries 2.1, 2.2 and approximations to the crossing probabilities of a Bessel process. One such approximation is available from Borodin and Salminen (1996).*

For the case of a one-dimensional parameter of interest, CV_2 can be obtained from the well known formula (Borodin and Salminen 1996)

$$1 - \alpha = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(\frac{\pi^2(2k+1)^2}{8CV_2^2}\right). \quad (2.39)$$

Using numerical integration, Delong (1980) tabulated the exact values of CV_2 for several dimensions and for almost all α of practical importance. Also, Betensky (1998) offered a handy, but not very accurate, analytic approximation to boundary-crossing probabilities of a Bessel process which can be used for calculating CV_2.

For the sake of completeness, we give in Table 2.1 some values of CV_1, CV_1' and CV_2 for α and n_0 used in later simulations.

Remark 2.5 *In the important case of the Adaptive Biased Coin sampling scheme of Wei (1978), the* $(k+1)^{th}$ *patient is assigned to one of the two treatments with probability* $p = (1 - D_k/k)/2$ *and to the other with probability* $q = 1 - p$, *where* $D_k = m - n$ *in our notation. In our case, we can write* $(D_k/2k) + 1/2 = m/k$. *Therefore, using Lemma 3 of Wei(1978) in which he shows* $D_k/k \to 0$ *a.s., as* $k \to \infty$, *we have* $m/k \to 1/2 = \lambda$ *a.s. as* $k \to \infty$. *Thus, in connection with Remark 2.2, our results remain valid.*

2.5 Applications

Example 2.1 Consider a sequential clinical trial involving the comparison of two treatments, E and S, (Experimental and a Standard treatment) with continuous outcomes. We assume that patients are assigned to treatment E with probability λ and to treatment S with probability $1 - \lambda$ and that the two streams of observations have normal densities; that is, we have $x_1, \ldots, x_m, \ldots$ $y_1, \ldots, y_n, \ldots$, assumed to be, respectively, from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. We are interested in testing $H_0 : \mu_1 = \mu_2$ ($\sigma > 0$, unknown) vs $H_A : \mu_1 \neq \mu_2$ ($\sigma > 0$, unknown). Under the re-parametrization

$$\theta = \frac{\mu_1 - \mu_2}{\sigma^2}, \quad \eta_1 = \frac{\mu_1 + \mu_2}{2\sigma^2}, \quad \eta_2 = \frac{1}{\sigma^2},$$

the hypothesis to be tested is $H_0 : \theta = 0$ ($\eta_1 \in (-\infty, \infty), \eta_2 \in (0, \infty)$) vs $H_A : \theta \neq 0$ ($\eta_1 \in (-\infty, \infty), \eta_2 \in (0, \infty)$), and the log-likelihood of the two

sequences up to observations n and m are

$$\begin{aligned}\log f_1(\theta,\eta_1,\eta_2;\mathbf{x}) &= \theta\left[\frac{1}{2}\sum_{i=1}^{m}x_i\right]+\eta_1\left[\sum_{i=1}^{m}x_i\right]+\eta_2\left[-\frac{1}{2}\sum_{i=1}^{m}x_i^2\right]\\ &- m\left[\frac{1}{2}\eta_2\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)^2-\frac{1}{2}\log\eta_2\right]-\frac{m}{2}\log 2\pi,\end{aligned}$$

$$\begin{aligned}\log f_2(\theta,\eta_1,\eta_2;\mathbf{y}) &= \theta\left[-\frac{1}{2}\sum_{j=1}^{n}y_j\right]+\eta_1\left[\sum_{j=1}^{n}y_j\right]+\eta_2\left[-\frac{1}{2}\sum_{j=1}^{n}y_j^2\right]\\ &- n\left[\frac{1}{2}\eta_2\left(\frac{-\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)^2-\frac{1}{2}\log\eta_2\right]-\frac{n}{2}\log 2\pi.\end{aligned}$$

Expected information matrices of the two populations under the new parameters are

$$I=\begin{pmatrix} \frac{1}{4\eta_2} & \frac{1}{2\eta_2} & -\frac{1}{2\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)\\ \frac{1}{2\eta_2} & \frac{1}{\eta_2} & -\frac{1}{\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)\\ -\frac{1}{2\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right) & -\frac{1}{\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right) & \frac{1}{\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)^2+\frac{1}{2\eta_2^2}\end{pmatrix},$$

$$J=\begin{pmatrix} \frac{1}{4\eta_2} & -\frac{1}{2\eta_2} & \frac{1}{2\eta_2}\left(\frac{-\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)\\ -\frac{1}{2\eta_2} & \frac{1}{\eta_2} & -\frac{1}{\eta_2}\left(\frac{-\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)\\ \frac{1}{2\eta_2}\left(\frac{-\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right) & -\frac{1}{\eta_2}\left(\frac{\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right) & \frac{1}{\eta_2}\left(\frac{-\theta}{2\eta_2}+\frac{\eta_1}{\eta_2}\right)^2+\frac{1}{2\eta_2^2}\end{pmatrix}.$$

Rao's efficient score of equation (2.8) can be verified to be

$$V_k=\frac{1}{\sqrt{k}}\left(\frac{ns_1-ms_2}{m+n}\right),$$

where $s_1=\sum x_i$ and $s_2=\sum y_j$. Now, using the restricted MLE,

$$\hat{\sigma}^2=\frac{1}{\hat{\eta}_2}=\frac{\sum_i x_i^2+\sum_j y_j^2}{m+n}-\left(\frac{s_1+s_2}{m+n}\right)^2,$$

and transforming back to the original parametrization under $H_0 : \mu_1 = \mu_2 = \mu$, we have

$$\hat{\Sigma}^{-1} = \frac{1}{\hat{\sigma}^2} \frac{1}{\lambda(1-\lambda)}.$$

Therefore,

$$R_k^* = V_k \hat{\Sigma}^{-1} V_k^t = \frac{1}{k} \left(\frac{n s_1 - m s_2}{m+n} \right)^2 \frac{1}{\hat{\sigma}^2} \frac{1}{\lambda(1-\lambda)}. \tag{2.40}$$

For Wald's statistic we have

$$W_k^{*(1)} = k \left[\frac{\bar{x} - \bar{y}}{\tilde{\sigma}_k^2} \right]^2 \times \frac{\tilde{\sigma}_k^2 \lambda(1-\lambda)}{(1 + 2(\frac{\bar{x}-\bar{y}}{\tilde{\sigma}_k^2})^2 \tilde{\sigma}_k^2 \lambda(1-\lambda))}, \tag{2.41}$$

$$W_k^{*(2)} = \frac{k\lambda(1-\lambda)(\bar{x}-\bar{y})^2}{\tilde{\sigma}_k^2}, \tag{2.42}$$

if we use the unrestricted MLE $\tilde{\sigma}_k^2 = (\sum(x_i - \bar{x})^2 + \sum(y_j - \bar{y})^2)/m+n$. Using the restricted MLE, $\hat{\sigma}_k^2$, and $\boldsymbol{\theta}_0 = 0$ we have

$$W_k^{*(3)} = k \left[\frac{\bar{x} - \bar{y}}{\tilde{\sigma}_k^2} \right]^2 \times \hat{\sigma}_k^2 \lambda(1-\lambda). \tag{2.43}$$

Remark 2.6 *In the above example, the unrestricted MLE, $\tilde{\sigma}^2$, can be replaced by the pooled estimator,*

$$s_p^2 = \frac{\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2}{m+n-2}.$$

This replacement would slightly improve power and does not disturb the asymptotic results developed in this chapter.

Example 2.2 As in Example 2.1, suppose that we have patients enrolling sequentially into a clinical trial and being assigned to one of two treatments, E

or S, with probability λ or $1-\lambda$, respectively, and binary outcomes being measured afterwards; that is, we obtain two streams of independent observations: $x_1, \ldots, x_i, \ldots$ i.i.d. with density $f(x;\pi_1) = \pi_1^x(1-\pi_1)^{1-x}$ and $y_1, \ldots, y_j, \ldots$ i.i.d. with density $f(y;\pi_2) = \pi_2^y(1-\pi_2)^{1-y}$, and the two Bernoulli populations are independent of each other.

Let the hypothesis of interest be $H_0 : \pi_1 = \pi_2$ vs $H_A : \pi_1 \neq \pi_2$. A natural re-parametrization is $(\theta, \eta) = (\log(\frac{\pi_1}{1-\pi_1}\frac{1-\pi_2}{\pi_2}), \log(\frac{\pi_2}{1-\pi_2}))$ (Robbins 1974, Whitehead 1978). Under this re-parametrization, the two populations will have densities of different functional forms sharing the same vector of parameters. The log-likelihoods of the two sequences are now

$$\log f_1(\theta, \eta; \mathbf{x}) = \theta s_1 + \eta s_1 - m \log(1 + e^{\theta+\eta}) \tag{2.44}$$

$$\log f_2(\theta, \eta; \mathbf{x}) = \eta s_2 - n \log(1 + e^{\eta}), \tag{2.45}$$

respectively, with $s_1 = \sum_{i=1}^{m} x_i$, $s_2 = \sum_{j=1}^{n} y_j$. It can be verified that the expected information matrices are

$$I = \begin{pmatrix} \frac{e^{\theta+\eta}}{(1+e^{\theta+\eta})^2} & \frac{e^{\theta+\eta}}{(1+e^{\theta+\eta})^2} \\ \frac{e^{\theta+\eta}}{(1+e^{\theta+\eta})^2} & \frac{e^{\theta+\eta}}{(1+e^{\theta+\eta})^2} \end{pmatrix},$$

$$J = \begin{pmatrix} 0 & 0 \\ 0 & \frac{e^{\eta}}{(1+e^{\eta})^2} \end{pmatrix}.$$

Therefore, in terms of the original parameters and under $H_0 : \theta = 0$, the estimated version of the covariance matrix in (2.22) becomes

$$\hat{\Sigma} = \hat{\pi}_k(1-\hat{\pi}_k) \times \lambda(1-\lambda),$$

where $\hat{\pi}_k = \frac{s_1+s_2}{m+n}$ is the MLE of the common Bernoulli parameter under H_0. Consequently,

$$V_k = \frac{1}{\sqrt{k}}\left[s_1 - \frac{me^{\theta+\eta}}{1+e^{\theta+\eta}}\right]_{\theta=0,\eta=\hat{\eta}_k} = \frac{1}{\sqrt{k}}\left[\frac{ns_1 - ms_2}{m+n}\right].$$

Hence,

$$R_k^* = \frac{1}{k}\left[\frac{ns_1 - ms_2}{m+n}\right]^2 \times \frac{1}{\hat{\pi}_k(1-\hat{\pi}_k)} \times \frac{1}{\lambda(1-\lambda)}. \tag{2.46}$$

Similarly, for Wald's statistic we have

$$W_k^{*(1)} = k\left[\log\left(\frac{\tilde{\pi}_{1k}}{1-\tilde{\pi}_{1k}}\frac{1-\tilde{\pi}_{2k}}{\tilde{\pi}_{2k}}\right)\right]^2 \times \frac{\lambda\tilde{\pi}_{1k}(1-\tilde{\pi}_{1k})(1-\lambda)\tilde{\pi}_{2k}(1-\tilde{\pi}_{2k})}{\lambda\tilde{\pi}_{1k}(1-\tilde{\pi}_{1k}) + (1-\lambda)\tilde{\pi}_{2k}(1-\tilde{\pi}_{2k})} \tag{2.47}$$

and

$$W_k^{*(2)} = k\left[\log\left(\frac{\tilde{\pi}_{1k}}{1-\tilde{\pi}_{1k}}\frac{1-\tilde{\pi}_{2k}}{\tilde{\pi}_{2k}}\right)\right]^2 \times \lambda(1-\lambda)\tilde{\pi}_{2k}(1-\tilde{\pi}_{2k}) \tag{2.48}$$

if we use the unrestricted MLEs, $\tilde{\pi}_{1k} = s_1/m$ and $\tilde{\pi}_{2k} = s_2/n$, for estimating the covariance matrix. In contrast, using restricted MLE and the null hypothesis value, $\boldsymbol{\theta}_0 = 0$, we obtain

$$W_k^{*(3)} = k\left[\log\left(\frac{\tilde{\pi}_{1k}}{1-\tilde{\pi}_{1k}}\frac{1-\tilde{\pi}_{2k}}{\tilde{\pi}_{2k}}\right)\right]^2 \times \lambda(1-\lambda)\hat{\pi}_k(1-\hat{\pi}_k), \tag{2.49}$$

where, $\hat{\pi}_k = \frac{s_1+s_2}{m+n}$.

2.6 Simulation study

We have carried out Monte Carlo experiments using Examples 2.1 and 2.2 to assess the performance of Test 1 and Test 2 in terms of power and average stopping times. Both normally distributed outcomes and binary outcomes had the following parameter sets in common:

1. Allocation probability: $\lambda = 0.5$

2. Truncation points: $n_0 = 50, 100, 200, 500$

3. Nominal significance levels; $\alpha = 0.01, 0.05, 0.1$

For the normal case, we kept $\mu_1 = 0$, and μ_2 was varied over the set $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The common variance of the two populations was $\sigma^2 = 1$, which allowed assessment of power and average stopping time for Test 1 & Test 2 under standardized treatment differences of magnitude μ_2.

For the binary outcomes case, π_1 was kept at 0.5 and π_2 varied over the set $\{0.1, 0.2, 0.3, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. These values correspond to odds ratios that are above as well as below 1.

Each scenario in these Monte Carlo simulations is based on 10^4 replicates. Taking values of λ other than 0.5 did not change the results of the simulations and hence, we concentrated only on $\lambda = .5$. Also, it is worth mentioning that for the Bernoulli case, values of π_1 that are more extreme resulted in lower power and higher ASN, however, this did not affect the empirical α

Generally, these simulations suggest that the test procedures are consistent against fixed alternatives, a fact justified by Theorems 3.1 & 3.2; that is, as the truncation point, n_0, grows to infinity, the power of the tests goes to unity for any fixed alternative.

Specific comments about the Monte Carlo results are provided in the following subsections.

2.6.1 Comments on Test 1

1. The results for Test 1 using normally distributed outcomes, Rao's efficient score statistic, R_k^*, and Wald's statistic, $W_k^{*(1)}$ are reported in Table 2.2, using the CV_1' of (2.38), and in Table 2.3, using the CV_1 of (2.34). The truncation points are limited to $n_0 = 50,500$ to illustrate the performance at small and large sample. We notice:

   a. Test 1 using R_k^* is extremely conservative and has much lower power than Test 2 (Tables 2.4-2.8) at small treatment differences but stops much earlier than Test 2 at large truncation points.

   b. Test 1 using $W_k^{*(1)}$ is even more conservative than Test 1 using R_k^*.

   c. In some simulations (not reported in this thesis), we observed that Test 1 using $W_k^{*(2)}$ is very liberal and has, for instance, empirical $\alpha = 0.13$ corresponding to a nominal 0.05. This feature makes it even less attractive than Test 1 with $W_k^{*(1)}$. $W_k^{*(3)}$ shows similar anti-conservative behaviour.

2. In some other simulations (not reported), less severe but similar patterns appear for Test 1 when binary outcomes are used.

2.6.2 Comments on Test 2

1. Test 2 using Rao's efficient score statistic, R_k^*, and critical values CV_2 from Equation (2.39) performs very well for both normal (Table 2.4) and binary outcomes (Table 2.5). Its empirical Type I error (rows headed by

$\mu_2 = 0.0$ or $\pi_2 = 0.5$) is slightly on the conservative side and gets better for large truncation points.

2. Test 2 using $W_k^{*(1)}$ has a much larger average stopping time and a much more conservative significance level than Test 2 using R_k^*. This pattern persists for both the cases of normal outcomes (Tables 2.6) and binary outcomes (Tables 2.7).

3. Test 2 using $W_k^{*(2)}$ has a liberal empirical significance level at small sample sizes (truncation points $n_0 = 50, 100$), a phenomenon which disappears at large truncation points ($n_0 = 200$ or larger) for both normally distributed outcomes (Tables 2.8) and binary outcomes (Tables 2.9). The average stopping time for this test is much smaller than that of Test 2 using R_k^*.

4. In other simulations, not reported, Test 2 using either $W_k^{*(1)}$ or $W_k^{*(3)}$ resulted very conservative.

2.6.3 Conclusion

From the above comments on the Monte Carlo simulations, we recommend Test 2 with Rao's efficient score for practical use for all truncation points. However, if early stopping is crucial and the truncation point (i.e., the total sample size attainable at the end of the trial) is large, it is safe and better to use Test 2 with Wald's statistic $W_k^{*(2)}$.

Table 2.1: Critical values for $d = 1$ and α, n_0 of interest in later simulations.

n_0	α	$CV_1(n_0, \alpha, d)$	$CV_1'(n_0, \alpha, d)$	$CV_2(\alpha, d)$
50	.10	2.76	2.74	1.96
	.05	3.20	3.02	2.24
	.01	4.18	3.56	2.80
100	.10	2.83	2.80	1.96
	.05	3.24	3.07	2.24
	.01	4.17	3.60	2.80
200	.10	2.89	2.85	2.80
	.05	3.28	3.12	2.24
	.01	4.17	3.64	2.80
500	.10	2.95	2.91	1.96
	.05	3.32	3.17	2.24
	.01	4.18	3.69	2.80

Table 2.2: Simulated power (P), and average stopping time (EN) for Test 1 using normally distributed outcomes (Example 2.2) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2. Rao's score statistic R_k^* and Wald's $W_k^{*(1)}$ were used with critical value CV_1' obtained from Vostrikova's approximation, Equation (2.38). Treatment allocation probability was $\lambda = 0.5$.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	P	EN	P	EN	P	EN
		Rao's efficient score statistic					
50							
	0.0	0.000	50.00	0.008	49.82	0.030	49.30
	0.1	0.001	49.99	0.014	49.72	0.033	49.21
	0.2	0.002	49.97	0.025	49.55	0.057	48.73
	0.3	0.008	49.92	0.042	49.28	0.086	48.29
	0.4	0.014	49.86	0.074	48.79	0.147	47.13
	0.5	0.034	49.62	0.133	47.94	0.235	45.66
	0.6	0.059	49.34	0.208	46.74	0.329	43.78
	0.7	0.104	48.84	0.303	45.17	0.442	41.60
	0.8	0.174	48.02	0.426	43.01	0.569	38.73
	0.9	0.270	46.96	0.560	40.33	0.690	35.74
500							
	0.0	0.004	498.92	0.019	493.73	0.045	483.64
	0.1	0.016	496.40	0.065	483.54	0.117	467.60
	0.2	0.126	478.23	0.293	436.66	0.399	403.30
	0.3	0.469	412.50	0.687	339.64	0.783	295.09
	0.4	0.828	312.58	0.939	234.82	0.963	195.04
	0.5	0.979	218.80	0.995	159.42	0.998	131.89
	0.6	0.999	158.36	1.000	115.27	1.000	96.83
	0.7	1.000	120.00	1.000	87.61	1.000	73.45
	0.8	1.000	96.00	1.000	69.89	1.000	58.50
	0.9	1.000	79.70	1.000	57.22	1.000	48.44
		Wald's statistic					
50							
	0.0	0.000	50.00	0.002	49.98	0.006	49.94
	0.1	0.000	50.00	0.002	49.99	0.009	49.89
	0.2	0.000	50.00	0.004	49.96	0.019	49.75
	0.3	0.000	50.00	0.011	49.91	0.039	49.55
	0.4	0.001	50.00	0.028	49.74	0.070	49.20
	0.5	0.004	49.98	0.046	49.60	0.126	48.55
	0.6	0.008	49.97	0.092	49.19	0.202	47.54
	0.7	0.015	49.93	0.151	48.64	0.304	46.21
	0.8	0.034	49.85	0.242	47.77	0.425	44.56
	0.9	0.069	49.67	0.355	46.46	0.544	42.65
500							
	0.0	0.001	499.71	0.010	496.97	0.029	491.12
	0.1	0.014	497.70	0.051	488.81	0.099	475.78
	0.2	0.113	482.48	0.269	446.88	0.366	419.46
	0.3	0.435	426.67	0.660	355.19	0.766	311.88
	0.4	0.821	327.22	0.925	252.06	0.964	213.33
	0.5	0.978	237.79	0.996	173.57	0.998	147.28
	0.6	0.999	175.68	1.000	129.82	1.000	109.19
	0.7	1.000	137.32	1.000	100.82	1.000	85.49
	0.8	1.000	112.12	1.000	83.03	1.000	70.47
	0.9	1.000	95.07	1.000	70.11	1.000	59.21

Table 2.3: Simulated power (P), and average stopping time (EN) for Test 1 using normally distributed outcomes (Example 2.2) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2. Rao's score statistic R_k^* and Wald's $W_k^{*(1)}$ were used with critical value CV_1 obtained from (2.34) and treatment allocation probability $\lambda = 0.5$.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	P	EN	P	EN	P	EN
		Rao's efficient score statistic					
50							
	0.0	0.000	50.00	0.004	49.92	0.026	49.35
	0.1	0.000	50.00	0.005	49.92	0.031	49.27
	0.2	0.000	50.00	0.012	49.82	0.052	48.92
	0.3	0.000	50.00	0.026	49.60	0.083	48.36
	0.4	0.001	50.00	0.051	49.25	0.134	47.45
	0.5	0.002	49.98	0.089	48.69	0.221	45.88
	0.6	0.006	49.95	0.147	47.98	0.324	44.05
	0.7	0.015	49.89	0.226	46.78	0.436	41.70
	0.8	0.033	49.77	0.341	45.02	0.560	39.16
	0.9	0.063	49.55	0.460	43.13	0.676	36.21
500							
	0.0	0.000	499.96	0.013	495.43	0.041	485.18
	0.1	0.003	499.46	0.044	489.04	0.110	470.68
	0.2	0.045	493.12	0.230	452.50	0.378	408.36
	0.3	0.272	459.32	0.616	365.19	0.768	301.94
	0.4	0.673	379.44	0.915	257.47	0.962	201.78
	0.5	0.937	279.39	0.993	175.57	0.998	136.70
	0.6	0.996	204.83	1.000	126.35	1.000	98.44
	0.7	1.000	155.15	1.000	97.67	1.000	74.80
	0.8	1.000	123.67	1.000	77.08	1.000	60.70
	0.9	1.000	102.29	1.000	63.67	1.000	49.88
		Wald's statistic					
50							
	0.0	0.000	50.00	0.001	49.99	0.006	49.92
	0.1	0.000	50.00	0.001	49.99	0.007	49.92
	0.2	0.000	50.00	0.002	49.99	0.017	49.80
	0.3	0.000	50.00	0.005	49.97	0.035	49.59
	0.4	0.000	50.00	0.012	49.92	0.067	49.27
	0.5	0.000	50.00	0.024	49.83	0.118	48.64
	0.6	0.000	50.00	0.049	49.66	0.195	47.74
	0.7	0.000	50.00	0.087	49.38	0.299	46.29
	0.8	0.000	50.00	0.148	48.88	0.410	44.82
	0.9	0.000	50.00	0.240	48.01	0.535	42.85
500							
	0.0	0.000	499.96	0.008	498.00	0.025	492.52
	0.1	0.003	499.54	0.036	492.34	0.089	479.67
	0.2	0.035	495.30	0.218	460.28	0.355	423.59
	0.3	0.237	468.20	0.600	378.04	0.744	320.37
	0.4	0.641	395.34	0.904	275.24	0.959	217.68
	0.5	0.928	300.48	0.991	192.30	0.998	150.39
	0.6	0.994	226.06	1.000	141.80	1.000	110.88
	0.7	1.000	176.18	1.000	111.07	1.000	87.17
	0.8	1.000	142.47	1.000	91.34	1.000	71.93
	0.9	1.000	121.49	1.000	77.36	1.000	60.89

Table 2.4: Simulated power (P), and average stopping time (EN) for Test 2 using Rao's statistic, (2.40), normally distributed outcomes (Example 2.2) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2. Critical value CV_2 is obtained from (2.39), and treatment allocation probability $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	P	EN	P	EN	P	EN
50							
	0.0	0.006	49.96	0.041	49.62	0.086	49.02
	0.1	0.010	49.94	0.051	49.53	0.103	48.83
	0.2	0.021	49.87	0.087	49.19	0.156	48.20
	0.3	0.042	49.73	0.148	48.56	0.239	47.18
	0.4	0.081	49.45	0.238	47.58	0.347	45.71
	0.5	0.143	49.00	0.351	46.23	0.476	43.74
	0.6	0.228	48.30	0.478	44.55	0.608	41.45
	0.7	0.342	47.29	0.612	42.43	0.729	38.90
	0.8	0.466	45.98	0.730	40.19	0.828	36.35
	0.9	0.597	44.38	0.828	37.85	0.898	33.90
100							
	0.0	0.008	99.88	0.043	99.13	0.089	97.87
	0.1	0.016	99.78	0.067	98.62	0.124	97.02
	0.2	0.044	99.35	0.145	96.91	0.228	94.24
	0.3	0.110	98.26	0.278	93.69	0.394	89.40
	0.4	0.227	96.14	0.456	88.79	0.581	82.80
	0.5	0.394	92.56	0.645	82.25	0.754	75.08
	0.6	0.585	87.51	0.802	74.92	0.878	67.23
	0.7	0.757	81.43	0.909	67.61	0.951	59.90
	0.8	0.879	75.11	0.966	61.14	0.984	53.76
	0.9	0.951	69.10	0.990	55.62	0.996	48.87
200							
	0.0	0.008	199.75	0.045	198.12	0.092	195.38
	0.1	0.026	199.19	0.097	195.77	0.165	191.57
	0.2	0.101	196.61	0.258	187.93	0.370	179.45
	0.3	0.280	189.49	0.519	172.50	0.635	159.75
	0.4	0.541	175.98	0.763	151.90	0.850	136.45
	0.5	0.789	157.58	0.922	129.89	0.957	115.01
	0.6	0.931	138.28	0.982	111.63	0.992	97.96
	0.7	0.985	121.54	0.997	97.38	0.999	85.41
	0.8	0.998	108.28	1.000	86.74	1.000	76.01
	0.9	1.000	97.98	1.000	78.31	1.000	68.69
500							
	0.0	0.009	499.27	0.047	494.92	0.096	487.87
	0.1	0.063	494.61	0.182	479.06	0.279	461.92
	0.2	0.334	466.01	0.570	419.41	0.684	385.91
	0.3	0.748	399.34	0.897	331.65	0.942	293.69
	0.4	0.961	320.91	0.991	257.57	0.996	225.75
	0.5	0.998	260.94	1.000	208.55	1.000	183.07
	0.6	1.000	220.37	1.000	176.12	1.000	154.52
	0.7	1.000	191.76	1.000	152.96	1.000	134.25
	0.8	1.000	170.51	1.000	136.36	1.000	119.31
	0.9	1.000	154.44	1.000	123.18	1.000	107.95

Table 2.5: Simulated power (P), and average stopping time (EN) for Test 2 using Rao's statistic, (2.46), binary outcomes (Example 2.2) with $\pi_1 = 0.5$ and various π_2. Critical value CV_2 is obtained from (2.39), and allocation probability $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	π_2	P	EN	P	EN	P	EN
50							
	0.1	0.673	42.67	0.871	36.07	0.926	32.23
	0.2	0.317	47.25	0.551	42.93	0.682	39.49
	0.3	0.102	49.20	0.274	46.95	0.380	44.81
	0.4	0.025	49.80	0.098	48.99	0.166	47.95
	0.5	0.009	49.94	0.045	49.57	0.098	48.86
	0.6	0.026	49.81	0.098	48.97	0.171	47.88
	0.7	0.105	49.18	0.271	46.92	0.377	44.90
	0.8	0.314	47.26	0.560	42.78	0.684	39.41
	0.9	0.666	42.88	0.864	36.15	0.926	32.30
100							
	0.1	0.971	65.24	0.994	52.64	0.998	45.92
	0.2	0.686	83.13	0.866	69.86	0.921	62.14
	0.3	0.258	95.29	0.494	87.14	0.615	80.67
	0.4	0.055	99.08	0.150	96.65	0.251	93.48
	0.5	0.009	99.87	0.042	99.14	0.088	97.88
	0.6	0.047	99.23	0.156	96.44	0.240	93.67
	0.7	0.265	95.03	0.498	86.97	0.608	80.94
	0.8	0.686	83.32	0.869	69.64	0.929	61.69
	0.9	0.972	65.31	0.996	52.33	0.997	46.05
200							
	0.1	1.000	92.15	1.000	73.54	1.000	64.92
	0.2	0.967	127.39	0.993	102.23	0.998	89.58
	0.3	0.594	171.71	0.802	146.52	0.874	131.80
	0.4	0.112	195.97	0.268	187.07	0.378	178.63
	0.5	0.009	199.73	0.049	197.87	0.096	195.04
	0.6	0.106	196.36	0.271	186.81	0.388	178.10
	0.7	0.588	172.28	0.806	146.31	0.869	132.12
	0.8	0.967	127.37	0.993	101.97	0.997	89.16
	0.9	1.000	92.22	1.000	73.75	1.000	64.51
500							
	0.1	1.000	144.79	1.000	115.97	1.000	101.57
	0.2	1.000	200.89	1.000	161.05	1.000	140.36
	0.3	0.972	309.36	0.993	248.51	0.997	216.69
	0.4	0.353	462.64	0.582	416.32	0.688	384.08
	0.5	0.011	499.21	0.048	495.19	0.095	487.61
	0.6	0.352	463.34	0.573	416.89	0.700	381.70
	0.7	0.970	308.43	0.994	248.40	0.997	218.28
	0.8	1.000	201.43	1.000	161.01	1.000	141.07
	0.9	1.000	144.62	1.000	115.88	1.000	101.36

Table 2.6: Simulated power (P), and average stopping time (EN) for Test 2 using Wald's statistic, (2.41), for normally distributed outcomes (Example 2.2) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2. Critical value CV_2 is obtained from (2.39), and allocation probability of $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	P	EN	P	EN	P	EN
50							
	0.0	0.002	49.99	0.029	49.80	0.064	49.44
	0.1	0.004	49.99	0.033	49.77	0.079	49.29
	0.2	0.009	49.96	0.062	49.56	0.125	48.91
	0.3	0.020	49.93	0.110	49.22	0.201	48.11
	0.4	0.042	49.84	0.187	48.56	0.308	46.97
	0.5	0.080	49.64	0.297	47.61	0.427	45.46
	0.6	0.144	49.33	0.412	46.35	0.565	43.50
	0.7	0.226	48.88	0.542	44.78	0.688	41.49
	0.8	0.335	48.18	0.668	43.10	0.796	39.16
	0.9	0.457	47.26	0.779	41.04	0.880	36.96
100							
	0.0	0.006	99.94	0.036	99.39	0.079	98.33
	0.1	0.013	99.84	0.059	98.95	0.108	97.67
	0.2	0.032	99.59	0.126	97.71	0.209	95.27
	0.3	0.083	98.90	0.256	94.95	0.374	90.92
	0.4	0.179	97.56	0.431	90.61	0.563	84.85
	0.5	0.336	94.88	0.620	85.13	0.735	77.74
	0.6	0.527	90.92	0.781	78.38	0.870	70.74
	0.7	0.712	85.71	0.896	71.99	0.947	63.76
	0.8	0.844	80.48	0.960	65.55	0.981	57.67
	0.9	0.934	74.94	0.987	60.75	0.994	53.37
200							
	0.0	0.006	199.85	0.043	198.36	0.083	196.26
	0.1	0.019	199.46	0.089	196.50	0.156	192.43
	0.2	0.093	197.11	0.246	189.63	0.354	181.84
	0.3	0.256	191.32	0.503	175.28	0.622	162.66
	0.4	0.513	179.26	0.753	155.24	0.837	140.21
	0.5	0.763	162.45	0.913	134.92	0.957	119.62
	0.6	0.921	143.77	0.980	117.15	0.991	102.91
	0.7	0.981	129.21	0.997	103.44	0.999	90.78
	0.8	0.998	116.11	1.000	92.86	1.000	81.51
	0.9	1.000	106.53	1.000	84.95	1.000	74.61
500							
	0.0	0.009	499.33	0.043	495.60	0.095	487.77
	0.1	0.060	495.45	0.183	480.22	0.271	463.81
	0.2	0.329	468.24	0.567	422.43	0.684	388.58
	0.3	0.735	403.41	0.894	337.61	0.940	298.75
	0.4	0.958	329.59	0.991	262.59	0.996	232.51
	0.5	0.998	268.98	1.000	216.72	1.000	189.23
	0.6	1.000	230.31	1.000	184.73	1.000	161.88
	0.7	1.000	202.57	1.000	162.04	1.000	141.99
	0.8	1.000	182.54	1.000	145.86	1.000	128.61
	0.9	1.000	166.87	1.000	134.02	1.000	117.60

Table 2.7: Simulated power (P), and average stopping time (EN) for Test 2 using Wald's statistic, (2.47), for binary outcomes (Example 2.2) with $\pi_1 = 0.5$ and various π_2. Critical value CV_2 is obtained from (2.39), and allocation probability $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	π_2	P	EN	P	EN	P	EN
50							
	0.1	0.421	47.70	0.744	42.31	0.829	38.86
	0.2	0.223	48.76	0.508	45.19	0.649	42.13
	0.3	0.072	49.63	0.233	47.95	0.349	46.18
	0.4	0.015	49.92	0.082	49.34	0.143	48.56
	0.5	0.006	49.97	0.034	49.70	0.076	49.25
	0.6	0.016	49.92	0.079	49.34	0.147	48.49
	0.7	0.071	49.62	0.235	47.93	0.351	46.02
	0.8	0.211	48.84	0.507	45.18	0.655	42.10
	0.9	0.421	47.64	0.739	42.42	0.831	38.94
100							
	0.1	0.956	74.19	0.988	60.47	0.992	54.12
	0.2	0.659	86.25	0.857	73.28	0.921	65.37
	0.3	0.243	95.93	0.479	88.30	0.605	82.30
	0.4	0.041	99.42	0.145	97.03	0.230	94.34
	0.5	0.007	99.91	0.041	99.17	0.088	97.98
	0.6	0.043	99.41	0.152	96.81	0.229	94.28
	0.7	0.251	95.97	0.476	88.37	0.606	82.53
	0.8	0.650	86.63	0.851	73.49	0.918	65.56
	0.9	0.954	74.24	0.988	60.77	0.992	53.86
200							
	0.1	1.000	103.75	1.000	83.66	1.000	74.00
	0.2	0.964	132.19	0.992	106.49	0.997	93.67
	0.3	0.576	173.96	0.791	149.31	0.863	134.85
	0.4	0.113	196.16	0.259	187.83	0.378	179.01
	0.5	0.009	199.76	0.045	198.14	0.095	195.40
	0.6	0.103	196.53	0.268	187.35	0.369	179.26
	0.7	0.570	174.19	0.792	148.86	0.868	133.74
	0.8	0.966	131.51	0.992	106.26	0.997	93.83
	0.9	1.000	103.53	1.000	83.41	1.000	74.04
500							
	0.1	1.000	161.14	1.000	129.21	1.000	113.97
	0.2	1.000	208.14	1.000	167.07	1.000	146.12
	0.3	0.974	311.54	0.994	250.34	0.998	219.07
	0.4	0.336	465.29	0.573	419.06	0.687	385.48
	0.5	0.009	499.29	0.049	494.74	0.098	487.23
	0.6	0.333	465.72	0.584	417.27	0.688	385.27
	0.7	0.972	312.15	0.995	249.04	0.997	219.49
	0.8	1.000	208.15	1.000	166.44	1.000	145.70
	0.9	1.000	161.17	1.000	129.46	1.000	113.64

Table 2.8: Simulated power (P), and average stopping time (EN) for Test 2 using Wald's statistic, (2.42), for normally distributed outcomes (Example 2.2) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2. Critical value CV_2 is obtained from (2.39), and allocation probability $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	P	EN	P	EN	P	EN
50							
	0.0	0.022	49.62	0.072	48.63	0.134	47.46
	0.1	0.030	49.53	0.092	48.41	0.153	47.17
	0.2	0.048	49.31	0.131	47.83	0.206	46.24
	0.3	0.086	48.84	0.206	46.77	0.296	44.88
	0.4	0.140	48.18	0.297	45.53	0.405	42.97
	0.5	0.226	47.20	0.420	43.44	0.542	40.42
	0.6	0.332	45.77	0.538	41.25	0.660	37.72
	0.7	0.457	43.89	0.671	38.33	0.762	34.72
	0.8	0.576	41.83	0.783	35.41	0.863	31.59
	0.9	0.700	39.23	0.867	32.46	0.917	28.96
100							
	0.0	0.012	99.71	0.055	98.38	0.112	96.37
	0.1	0.023	99.50	0.088	97.55	0.149	95.44
	0.2	0.067	98.61	0.163	95.46	0.255	92.11
	0.3	0.141	97.08	0.308	91.67	0.427	86.72
	0.4	0.267	94.04	0.490	85.59	0.610	79.03
	0.5	0.441	89.44	0.678	78.01	0.772	70.91
	0.6	0.631	82.93	0.821	70.15	0.891	61.79
	0.7	0.798	75.81	0.920	62.30	0.955	54.49
	0.8	0.901	68.67	0.971	55.00	0.985	48.27
	0.9	0.962	61.72	0.991	48.70	0.996	42.41
200							
	0.0	0.012	199.54	0.051	197.48	0.101	194.30
	0.1	0.033	198.73	0.107	194.92	0.174	190.33
	0.2	0.119	195.43	0.278	185.79	0.383	176.94
	0.3	0.305	187.32	0.527	169.57	0.651	156.70
	0.4	0.581	171.35	0.774	147.69	0.852	132.02
	0.5	0.801	152.18	0.926	124.77	0.960	109.57
	0.6	0.939	131.08	0.986	105.53	0.993	92.23
	0.7	0.988	112.85	0.998	89.93	0.999	78.99
	0.8	0.998	99.13	1.000	78.64	1.000	68.97
	0.9	1.000	87.96	1.000	70.09	1.000	61.21
500							
	0.0	0.008	499.28	0.049	494.83	0.098	487.00
	0.1	0.067	494.12	0.189	476.99	0.287	460.10
	0.2	0.342	463.31	0.575	417.27	0.678	383.31
	0.3	0.750	394.03	0.903	326.53	0.944	288.31
	0.4	0.962	313.62	0.993	251.91	0.997	218.50
	0.5	0.998	251.83	1.000	200.78	1.000	175.40
	0.6	1.000	209.07	1.000	167.52	1.000	146.98
	0.7	1.000	179.22	1.000	143.38	1.000	125.56
	0.8	1.000	157.21	1.000	125.16	1.000	109.54
	0.9	1.000	140.10	1.000	111.00	1.000	97.54

Table 2.9: Simulated power (P), and average stopping time (EN) for Test 2 using Wald's statistic, (2.48), for binary outcomes (Example 2.2) with $\pi_1 = 0.5$ and various π_2. Critical value CV_2 is obtained from (2.39), and allocation probability $\lambda = 0.5$ is used.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	π_2	P	EN	P	EN	P	EN
50							
	0.1	0.718	40.93	0.830	36.55	0.866	33.89
	0.2	0.414	45.15	0.620	40.76	0.717	37.81
	0.3	0.158	48.36	0.320	45.63	0.416	43.60
	0.4	0.044	49.52	0.125	48.35	0.197	47.07
	0.5	0.020	49.78	0.067	49.08	0.117	48.28
	0.6	0.045	49.50	0.125	48.35	0.197	47.07
	0.7	0.160	48.25	0.322	45.57	0.429	43.35
	0.8	0.412	45.21	0.620	40.88	0.714	37.82
	0.9	0.713	40.99	0.831	36.45	0.868	34.09
100							
	0.1	0.975	58.19	0.990	48.80	0.994	44.46
	0.2	0.731	77.89	0.883	65.36	0.930	58.32
	0.3	0.305	93.18	0.523	84.24	0.633	78.12
	0.4	0.064	98.66	0.165	95.69	0.262	92.07
	0.5	0.012	99.75	0.057	98.61	0.106	96.90
	0.6	0.062	98.70	0.169	95.69	0.256	92.28
	0.7	0.305	92.84	0.521	84.43	0.628	78.25
	0.8	0.736	77.84	0.882	65.01	0.932	57.95
	0.9	0.973	58.49	0.991	49.00	0.993	44.46
200							
	0.1	1.000	79.15	1.000	64.86	1.000	57.71
	0.2	0.972	118.81	0.995	94.92	0.997	83.29
	0.3	0.602	168.90	0.805	143.66	0.882	127.45
	0.4	0.122	195.46	0.278	186.16	0.391	176.15
	0.5	0.011	199.57	0.050	197.65	0.100	194.46
	0.6	0.117	195.42	0.278	185.55	0.378	177.37
	0.7	0.614	168.30	0.814	142.51	0.874	127.97
	0.8	0.973	118.55	0.994	94.93	0.998	82.99
	0.9	1.000	79.60	1.000	64.82	1.000	58.15
500							
	0.1	1.000	123.02	1.000	98.61	1.000	87.00
	0.2	1.000	189.39	1.000	150.97	1.000	131.67
	0.3	0.976	302.35	0.993	241.70	0.998	211.71
	0.4	0.335	464.20	0.580	413.11	0.700	379.31
	0.5	0.010	499.14	0.050	494.54	0.099	487.12
	0.6	0.346	463.24	0.592	412.83	0.695	379.62
	0.7	0.971	302.65	0.995	241.51	0.998	211.85
	0.8	1.000	190.35	1.000	151.14	1.000	132.37
	0.9	1.000	123.04	1.000	98.51	1.000	86.45

Chapter 3

Asymptotic results under H_A

3.1 Preliminaries

In this chapter, we will derive asymptotics of the test statistics in Chapter 2 under the alternative hypothesis, H_A. In Section 3.2, we will discuss the distribution of R_k^* under H_A, whereas Sections 3.3 deals with W_k^* under H_A. In Section 3.4, we highlight some approximate methods for obtaining power and moments of the stopping time when the parameter of interest has dimension one.

To render the discussion and conditions more tractable, we assume throughout this chapter that the observations come from the exponential family of distributions (Serfling 1980). To this end, let the canonical form of the log-likelihood be

$$\log f_i(x;\boldsymbol{\theta},\boldsymbol{\eta}) = T_1^{(i)}(x)\boldsymbol{\theta}^t + T_2^{(i)}\boldsymbol{\eta}^t + S^{(i)}(x) - A^{(i)}(\boldsymbol{\theta},\boldsymbol{\eta}), \quad i = 1,2, \qquad (3.1)$$

where i indicates population, f_i is the density function of the i^{th} population after eventual parameter transformation, $T_j^{(i)}$ are vector valued functions of the data and $A^{(i)}(\boldsymbol{\theta},\boldsymbol{\eta})$ are functions of the parameters only. Under H_0, i.e.,

when $f_1 = f_2 = f$, we may drop the superscript denoting the population so that $T_2 = T_2^{(1)} = T_2^{(2)}$, and $A^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = A^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = A(\boldsymbol{\theta}_0, \boldsymbol{\eta})$; hence,

$$\nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = \nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = \nabla_{\boldsymbol{\eta}} A(\boldsymbol{\theta}_0, \boldsymbol{\eta}).$$

Under H_A, we assume that there exists λ^*, $0 < \lambda^* < 1$, such that $m/k \to \lambda^*$ *a.s.* We let $\boldsymbol{\eta}^*$ denote the true value of parameter $\boldsymbol{\eta}$ and Γ^* an open neighborhood of the line $\lambda^*(\boldsymbol{\theta}_0, \boldsymbol{\eta}^*) + (1 - \lambda^*)(\boldsymbol{\theta}, \boldsymbol{\eta}^*)$. The following regularity conditions will be needed for the results of the coming sections.

C9. All first- and second-order partial derivatives of $A^{(i)}$, $i = 1, 2$, exist; the functions $\nabla_{\boldsymbol{\eta}} A^{(i)}(\boldsymbol{\theta}, \boldsymbol{\eta})$, $i = 1, 2$, are continuous, and $\nabla_{\boldsymbol{\eta}} A(\boldsymbol{\theta}_0, \boldsymbol{\eta})$ has a unique inverse in Γ^* that is Lipschitz continuous of order one in each component of $\boldsymbol{\eta}$.

C10. Matrices $\nabla^2_{\boldsymbol{\eta}^2} A^{(i)}(\boldsymbol{\theta}, \boldsymbol{\eta})$, $\nabla^2_{\boldsymbol{\theta}^2} A^{(i)}(\boldsymbol{\theta}, \boldsymbol{\eta})$, $i = 1, 2$, are positive definite, Lipschitz continuous of order one in each variable in Γ^*. Furthermore, their inverses and the inverse of matrix $\lambda^* \nabla^2_{\boldsymbol{\theta}^2} A^{(1)} + (1 - \lambda^*) \nabla^2_{\boldsymbol{\theta}^2} A^{(2)}$ exist in Γ^*.

C11. The third derivatives of $A(\boldsymbol{\theta}, \boldsymbol{\eta})$ are bounded.

C12. $E_{\boldsymbol{\theta}} \| T_j^{(i)} \|^{2+\delta} < \infty$, $i, j = 1, 2$, for some $\delta > 0$.

3.2 Consistency for R_k^*

As a consequence of the following theorem, we will see that Test 1 and Test 2 are consistent for Rao's score statistic process.

Theorem 3.1 *Under H_A, if* C9 - C11 *hold, then*

$$\begin{aligned}\limsup_{k\to\infty} \; & |R_k^*(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)/k - [\lambda^* c_1 + (1-\lambda^*)c_2] \\ & \times\big(\lambda^* I(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) + (1-\lambda^*)J(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big)^{11}[\lambda^* c_1 + (1-\lambda^*)c_2]^t| \\ & = O\big(k^{-1/2}(\log\log k)^{1/2}\big) \;\; a.s.,\end{aligned}$$

where

$$c_1 = \nabla_{\boldsymbol{\theta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) - \nabla_{\boldsymbol{\theta}} A^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A), \tag{3.2}$$

$$c_2 = \nabla_{\boldsymbol{\theta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) - \nabla_{\boldsymbol{\theta}} A^{(2)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A), \tag{3.3}$$

and $\boldsymbol{\eta}_0^A$ is the solution of the equation

$$\lambda^* \nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) + (1-\lambda^*)\nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) = \nabla_{\boldsymbol{\eta}} A(\boldsymbol{\theta}_0,\boldsymbol{\eta}). \tag{3.4}$$

Before proving the theorem, we need a technical Lemma showing that under H_A, the restricted MLE, $\hat{\boldsymbol{\eta}}_k$, almost surely converges to the point $\boldsymbol{\eta}_0^A$.

Lemma 3.1 *Under the conditions of Theorem 3.1, we have*

$$\limsup_{k\to\infty} \|\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_0^A\| \overset{a.s.}{=} O\big(k^{-1/2}(\log\log k)^{1/2}\big). \tag{3.5}$$

Proof Lemma 3.1. Since $\widehat{\boldsymbol{\eta}}_k$ is the MLE of $\boldsymbol{\eta}$ under H_0,

$$\sum_{i=1}^m T_2(x_i) + \sum_{j=1}^n T_2(y_j) = m\nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k) + n\nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k),$$

or equivalently,

$$\frac{m}{k}\frac{1}{m}\sum_{i=1}^m T_2(x_i) + \frac{n}{k}\frac{1}{n}\sum_{j=1}^n T_2(y_j) = \nabla_{\boldsymbol{\eta}} A(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k). \tag{3.6}$$

Subtracting $\lambda^*\nabla_{\boldsymbol{\eta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) + (1-\lambda^*)\nabla_{\boldsymbol{\eta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)$ from both sides of (3.6) and taking into account (3.4), we have

$$\begin{aligned}\frac{m}{k}\frac{1}{m}\sum_{i=1}^{m}T_2(x_i) - \lambda^*\nabla_{\boldsymbol{\eta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) \;+\; & \frac{n}{k}\frac{1}{n}\sum_{j=1}^{n}T_2(y_j) - (1-\lambda^*)\nabla_{\boldsymbol{\eta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*) \\ = \;& \nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k) - \nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\boldsymbol{\eta}_k^A).\end{aligned} \tag{3.7}$$

Noticing that by the law of iterated logarithm,

$$\limsup_{m\to\infty}\left\|\frac{1}{m}\sum_{i=1}^{m}T_2(x_i) - \nabla_{\boldsymbol{\eta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right\| = O\big(m^{-1/2}(\log\log m)^{1/2}\big)\;\; a.s.,$$

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{j=1}^{n}T_2(y_j) - \nabla_{\boldsymbol{\eta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right\| = O\big(n^{-1/2}(\log\log n)^{1/2}\big)\;\; a.s.,$$

$$\limsup_{k\to\infty}\left|\frac{m}{k} - \lambda^*\right| = O\big(k^{-1/2}(\log\log k)^{1/2}\big)\;\; a.s.,$$

and

$$\limsup_{k\to\infty}\left|\frac{n}{k} - (1-\lambda)^*\right| = O\big(k^{-1/2}(\log\log k)^{1/2}\big)\;\; a.s.,$$

we obtain

$$\limsup_{k\to\infty}\left\|\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k) - \nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\right\| = O\big(k^{-1/2}(\log\log k)^{1/2}\big)\;\; a.s.$$

This result combined with the fact that, by C9, $\nabla_{\boldsymbol{\eta}}A$ has a unique inverse which is Lipschitz continuous of order one, will complete the proof of the lemma.

□

Proof of Theorem 3.1. Re-write the score vector, V_k, as

$$\begin{aligned} V_k &= k^{-1/2}\sum_{i=1}^{m}\Big(T_1^{(1)}(x_i)-\nabla_{\boldsymbol{\theta}}A^{(1)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)\Big)+k^{-1/2}\sum_{j=1}^{n}\Big(T_1^{(2)}(y_j)-\nabla_{\boldsymbol{\theta}}A^{(2)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)\Big) \\ &= k^{-1/2}\sum_{i=1}^{m}\Big(T_1^{(1)}(x_i)-\nabla_{\boldsymbol{\theta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\Big)+k^{-1/2}\sum_{j=1}^{n}\Big(T_1^{(2)}(y_j)-\nabla_{\boldsymbol{\theta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\Big) \\ &+ k^{-1/2}mc_1+k^{-1/2}nc_2+k^{-1/2}m\big(\nabla_{\boldsymbol{\theta}}A^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)-\nabla_{\boldsymbol{\theta}}A^{(1)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)\big) \\ &+ k^{-1/2}n\big(\nabla_{\boldsymbol{\theta}}A^{(2)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)-\nabla_{\boldsymbol{\theta}}A^{(2)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)\big). \end{aligned} \tag{3.8}$$

By C10, C11 and (3.5),

$$\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)-\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)=(\widehat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_0^A)\nabla^2_{\boldsymbol{\eta}^2}A(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)+\Big(O\big(k^{-1/2}(\log\log k)^{1/2}\big)\Big)\ \ a.s.,$$

and hence, by C11, the relation (3.4), and (3.6),

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_0^A &= -\big(\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)-\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big)I_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)+\Big(O\big(k^{-1/2}(\log\log k)^{1/2}\big)\Big) \\ &= -\frac{m}{k}\frac{1}{m}\sum_{i=1}^{n}\big(T_2(x_i)-\nabla_{\boldsymbol{\eta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big)I_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\ &- \frac{n}{k}\frac{1}{n}\sum_{j=1}^{n}\big(T_2(y_j)-\nabla_{\boldsymbol{\eta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big)J_{22}^{-1}(\boldsymbol{\theta}_0\boldsymbol{\eta}_0^A) \\ &+ \Big(O\big(k^{-1/2}(\log\log k)^{1/2}\big)\Big)\ \ a.s. \end{aligned} \tag{3.9}$$

Similarly, by using C10 and (3.5),

$$\nabla_{\boldsymbol{\theta}}A^{(i)}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\eta}}_k)-\nabla_{\boldsymbol{\theta}}A^{(i)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)=(\widehat{\boldsymbol{\eta}}_k-\boldsymbol{\eta}_0^A)\nabla^2_{\boldsymbol{\theta}\boldsymbol{\eta}^2}A^{(i)}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)+\Big(O\big(k^{-1/2}(\log\log k)^{1/2}\big)\Big)\ \ a.s. \tag{3.10}$$

Now, by inserting (3.9) and (3.10) in (3.8), the efficient score vector V_k can be written as

$$\begin{aligned}
V_k &= k^{1/2}\lambda^* c_1 + k^{1/2}(1-\lambda^*)c_2 + k^{-1/2}\sum_{i=1}^{m}\left(T_1^{(1)}(x_i) - \nabla_{\boldsymbol{\theta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) \\
&+ k^{-1/2}\sum_{j=1}^{n}\left(T_1^{(2)}(y_j) - \nabla_{\boldsymbol{\theta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) + k^{-1/2} m(\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_0^A) I_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\
&+ k^{-1/2} n(\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_0^A) J_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) + \left(O(k^{-1/2}(\log\log k)^{1/2})\right) \\
&= k^{1/2}\lambda^* c_1 + k^{1/2}(1-\lambda^*)c_2 + k^{-1/2}\sum_{i=1}^{m}\left(T_1^{(1)}(x_i) - \nabla_{\boldsymbol{\theta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) \\
&+ k^{-1/2}\sum_{j=1}^{n}\left(T_1^{(2)}(y_j) - \nabla_{\boldsymbol{\theta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) \\
&- k^{-1/2}\frac{m}{k}\Big[\sum_{i=1}^{m}\left(T_2(x_i) - \nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) I_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\
&+ \sum_{j=1}^{n}\left(T_2(y_j) - \nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) J_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\Big] I_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\
&- k^{-1/2}\frac{n}{k}\Big[\sum_{i=1}^{m}\left(T_2(x_i) - \nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) I_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\
&+ \sum_{j=1}^{n}\left(T_2(y_j) - \nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right) J_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\Big] \times J_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) \\
&+ \left(O(k^{-1/2}(\log\log k)^{1/2})\right) \\
&= k^{1/2}\lambda^* c_1 + k^{1/2}(1-\lambda^*)c_2 + C_k + D_k + \left(O(k^{-1/2}\log\log k)\right) \quad a.s.
\end{aligned} \tag{3.11}$$

where, after re-arranging terms,

$$C_k = k^{-1/2}\sum_{i=1}^{m}\Big[\big(T_1^{(1)}(x_i) - \nabla_{\boldsymbol{\theta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big)$$

$$-\big(T_2(x_i) - \nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big) I_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big(\lambda^* I_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) + (1-\lambda^*) J_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big)\Big],$$

$$D_k = k^{-1/2}\sum_{j=1}^{n}\Big[\big(T_1^{(2)}(y_j) - \nabla_{\boldsymbol{\theta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big)$$

$$-\big(T_2(y_j) - \nabla_{\boldsymbol{\eta}} A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\big) J_{22}^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big(\lambda^* I_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A) + (1-\lambda^*) J_{21}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)\big)\Big].$$

The independent terms in C_k and D_k have mean zero. Their covariance structure can be calculated, but it gives a lengthy formula not needed for the proof of this theorem. Instead, by the law of iterated logarithm,

$$\limsup_{k\to\infty} \|k^{-1/2} C_k\| = O\big(k^{-1/2}(\log\log k)^{1/2}\big) \ \ a.s.,$$

$$\limsup_{k\to\infty} \|k^{-1/2} D_k\| = O\big(k^{-1/2}(\log\log k)^{1/2}\big) \ \ a.s.,$$

and since $\widehat{\boldsymbol{\eta}}_k \to \boldsymbol{\eta}_0^A$ $a.s.$ implies that we can replace $\Sigma^{-1}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\eta}}_k)$ by $\Sigma^{-1}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)$ without disturbing the error magnitude, the theorem follows.

□

Remark 3.1 *From the proof of Theorem 3.1, one can see that the efficient score process* $\{V_k\}$ *behaves, in the limit, as a zero-mean Gaussian process with finite covariance structure and a drift of order* $k^{1/2}$. *As* $R_k^*/k \xrightarrow{a.s.}$ *constant* > 0, *by Theorem 3.1, the consistency of Tests 1 and 2, based on* R_k^*, *follows.*

3.3 Consistency for W_k^*

Using arguments similar to those in Remark 3.1, and the following theorem, one can conclude consistency for the procedures based on the test statistics W_k^*.

Theorem 3.2 *Under H_A, if C9-C10 hold, then*

$$\limsup_{k\to\infty}\left|W_k^{*(1)}/k-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta},\boldsymbol{\eta}^*)]^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t\right| \overset{a.s.}{=} \left(O((\log\log k)^{1/2}k^{-1/2})\right), \tag{3.12}$$

$$\limsup_{k\to\infty}\left|W_k^{*(2)}/k-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}_0,\boldsymbol{\eta}^*)]^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t\right| \overset{a.s.}{=} \left(O((\log\log k)^{1/2}k^{-1/2})\right) \tag{3.13}$$

and

$$\limsup_{k\to\infty}\left|W_k^{*(3)}/k-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0^A)]^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t\right| \overset{a.s.}{=} \left(O((\log\log k)^{1/2}k^{-1/2})\right), \tag{3.14}$$

where $\boldsymbol{\eta}_0^A$ is the solution of

$$\nabla_{\boldsymbol{\eta}}A(\boldsymbol{\theta}_0,\boldsymbol{\eta})=\frac{1}{k}[mE(T_2^{(1)}(x_i))+nE(T_2^{(2)}(y_j))]=\frac{m}{k}\nabla_{\boldsymbol{\eta}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)+\frac{n}{k}\nabla_{\boldsymbol{\eta}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*). \tag{3.15}$$

Proof. Using arguments similar to those leading to (2.29), we can write

$$\begin{aligned}\sqrt{k}(\tilde{\boldsymbol{\theta}}_k-\boldsymbol{\theta}_0) &= \sqrt{k}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\sqrt{k}(\tilde{\boldsymbol{\theta}}_k-\boldsymbol{\theta})\\ &\overset{a.s.}{=} \sqrt{k}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+C_k+D_k+\left(O((\log\log k)^{3/2}k^{-1/2})\right),\end{aligned} \tag{3.16}$$

where

$$C_k=\frac{1}{\sqrt{k}}\left[\sum_i T_1^{(1)}(x_i)+T_2^{(1)}(x_i)-\nabla_{\boldsymbol{\xi}}A^{(1)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right]\begin{pmatrix}M^{11}\\ M^{21}\end{pmatrix}(\boldsymbol{\theta}_0,\boldsymbol{\eta}^*) \tag{3.17}$$

and

$$D_k=\left[\sum_j T_1^{(2)}(y_j)+T_2^{(2)}(y_j)-\nabla_{\boldsymbol{\xi}}A^{(2)}(\boldsymbol{\theta},\boldsymbol{\eta}^*)\right]\begin{pmatrix}M^{11}\\ M^{21}\end{pmatrix}(\boldsymbol{\theta}_0,\boldsymbol{\eta}^*). \tag{3.18}$$

Since the independent terms in C_k and D_k have mean zero, by the law of iterated logarithm it can be shown that

$$\limsup_{k\to\infty} \|k^{-1/2}C_k\| = O\big(k^{-1/2}(\log\log k)^{1/2}\big) \ \ a.s.$$

and

$$\limsup_{k\to\infty} \|k^{-1/2}D_k\| = O\big(k^{-1/2}(\log\log k)^{1/2}\big) \ \ a.s.$$

Hence, almost surely, $(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \big(O((\log\log k)^{1/2}k^{-1/2})\big)$. Now, if $W_k^{*(1)}$ is used, since Lemma 2.1 of Gombay and Horváth (1994) guarantees enough closeness of the MLE $\tilde{\boldsymbol{\xi}}$ to $\boldsymbol{\xi}$, by C10 we have

$$W_k^{*(1)}/k = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}, \boldsymbol{\eta}^*)]^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t + \big(O((\log\log k)^{1/2}k^{-1/2})\big)$$

and

$$W_k^{*(2)}/k = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}_0, \boldsymbol{\eta}^*)]^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t + \big(O((\log\log k)^{1/2}k^{-1/2})\big),$$

so that (3.12) and (3.13) follow. Analogous arguments would lead to (3.14) for $W_k^{*(3)}$ if we realize that, almost surely, $||\boldsymbol{\eta}_0^A - \hat{\boldsymbol{\eta}}_k|| = \big(O((\log\log k)^{1/2}k^{-1/2})\big)$. This result is a consequence of C9, Law of Iterated Logarithm and the fact that, under H_0, $\hat{\boldsymbol{\eta}}_k$ is the solution of

$$k\nabla_{\boldsymbol{\eta}} A(\boldsymbol{\theta}_0, \boldsymbol{\eta}) = \sum_i T_2^{(1)}(x_i) + \sum_j T_2^{(2)}(y_j),$$

thus completing the proof.

□

3.4 Examples

Example 3.1 (Example 2.1 continued): For Example 2.1, it can be easily seen that under the alternative hypothesis, H_A,

$$\frac{1}{(\eta_2)_0^A} = (\sigma_0^A)^2 = \sigma^2 + \lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2,$$

and

$$R_k/k \xrightarrow{a.s.} (\mu_1-\mu_2)^2/\{\sigma^2 + \lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2\},$$

where μ_1, μ_2, σ^2 denote the values of the parameters under H_A. Similarly, for the Wald statistics we have

$$W_k^{*(1)}/k \xrightarrow{a.s.} \frac{\lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2}{\sigma^2 + \lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2}$$

and

$$W_k^{*(2)}/k \xrightarrow{a.s.} \frac{\lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2((\sigma^2 + \lambda^*(1-\lambda^*)(\mu_1-\mu_2)^2)}{\sigma^4}.$$

Example 3.2 (Example 2.2 continued): For the problem of Example 2.2, by solving

$$\lambda^*\nabla_{\boldsymbol{\eta}} A^{(1)}(\boldsymbol{\theta}, \boldsymbol{\eta}^*) + (1-\lambda^*)\nabla_{\boldsymbol{\eta}} A^{(21)}(\boldsymbol{\theta}, \boldsymbol{\eta}^*) = \nabla_{\eta} A(\boldsymbol{\theta}_0, \boldsymbol{\eta})$$

for $\boldsymbol{\eta}$, we can see that, under the alternative hypothesis,

$$(\pi_2)_0^A = \lambda^*\pi_1 + (1-\lambda^*)\pi_2,$$

where π_1, π_2 denote the true values of the parameters. Notice that if $\lambda^* = 1/2$, then $(\pi_2)_0^A = \frac{\pi_1+\pi_2}{2}$; that is, the point to which the restricted MLE converges is actually the midpoint between H_0 and H_A.

Calculating the drift for R_k^*, we see that

$$\frac{R_k^*}{k} \to \frac{(1-\lambda^*)^2(\pi_1+\pi_2)^2}{\big(\lambda^*\pi_1+(1-\lambda^*)\pi_2\big)\big(1-\lambda^*\pi_1-(1-\lambda^*)\pi_2\big)} \quad a.s.$$

For Wald statistics we have

$$W_k^{*(1)}/k \longrightarrow \left[\log\left(\frac{\pi_1}{1-\pi_1}\frac{1-\pi_2}{\pi_{2k}}\right)\right]^2 \times \frac{\lambda^*\pi_1(1-\pi_1)(1-\lambda^*)\pi_2(1-\pi_2)}{\lambda^*\pi_1(1-\pi_1)+(1-\lambda^*)\pi_2(1-\pi_2)},$$

$$W_k^{*(2)}/k \longrightarrow \left[\log\left(\frac{\pi_1}{1-\pi_1}\frac{1-\pi_2}{\pi_{2k}}\right)\right]^2 \times \lambda^*(1-\lambda^*)(\pi_2)_0^A(1-(\pi_2)_0^A).$$

3.5 Approximations for Power and sample size distribution

In order to design a sequential trial, having a method for at least approximating the power, the expected sample size, and the probability that sample size exceeds a specified threshold, is quite important. For example, the probability that the sample size for the sequential procedure exceeds that of a fixed sample procedure with the same power and significance level as the sequential procedure is often needed.

In this section, we shall discuss these issues with particular emphasis on the case of a one-dimensional parameter of interest.

The following two Lemmas give the covariance structure of V_k and $\sqrt{k}(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)$, respectively, under H_A. For ease of notation, we adopt the convention that superscript * for a matrix means evaluation at $(\boldsymbol{\theta}, \boldsymbol{\eta}^*)$, the superscript 0 means evaluation at $(\boldsymbol{\theta}_0, \boldsymbol{\eta}^*)$, and no superscript means evaluation at $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0^A)$. In particular, this convention means that for the matrix M,

$$M = \lambda^* I(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0^A) + (1-\lambda^*)J(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0^A) \tag{3.19}$$

$$M^* = \lambda^* I(\boldsymbol{\theta}, \boldsymbol{\eta}^*) + (1-\lambda^*) J(\boldsymbol{\theta}, \boldsymbol{\eta}^*), \tag{3.20}$$

$$M^0 = \lambda^* I(\boldsymbol{\theta}_0, \boldsymbol{\eta}^*) + (1-\lambda^*) J(\boldsymbol{\theta}_0, \boldsymbol{\eta}^*). \tag{3.21}$$

Lemma 3.2 *Under H_A and C9-C12, $V_{[kt]} - \sqrt{[kt]}d$ is approximately a mean-zero Gaussian process with the covariance structure*

$$\begin{aligned} \Sigma_A = M_{11}^* &+ M_{12}\left\{\lambda^* I_{22}^{-1} I_{22}^* I_{22}^{-1} + (1-\lambda^*) J_{22}^{-1} J_{22}^* J_{22}^{-1}\right\} M_{21} \\ &- \left\{\lambda^* I_{12}^* I_{22}^{-1} + (1-\lambda^*) J_{12}^* J_{22}^{-1}\right\} M_{21} \\ &- M_{12}\left\{\lambda^* I_{22}^{-1} I_{21}^* + (1-\lambda^*) J_{22}^{-1} J_{21}^*\right\}, \end{aligned} \tag{3.22}$$

where $d = \lambda^* c_1 + (1-\lambda^*) c_2$, and c_1, c_2 are defined in equations (3.2), (3.3).

Proof. From (3.11), it is clear that V_k is approximately a d-dimensional Gaussian processes with covariance

$$\Sigma_A = Cov(C_k + D_k) = \lambda^* \Sigma_1^A + (1-\lambda^*)\Sigma_2^A,$$

where Σ_1^A, Σ_2^A are, respectively, covariances of terms in C_k and D_k. It can be easily seen that

$$\Sigma_1^A = I_{11}^* + M_{12} I_{22}^{-1} I_{22}^* I_{22}^{-1} M_{21} - I_{12}^* I_{22}^{-1} M_{21} - M_{12} I_{22}^{-1} I_{21}^*,$$

$$\Sigma_2^A = J_{11}^* + M_{12} J_{22}^{-1} J_{22}^* J_{22}^{-1} M_{21} - J_{12}^* J_{22}^{-1} M_{21} - M_{12} J_{22}^{-1} J_{21}^*.$$

Hence, the desired result follows by simple algebraic manipulations.

□

Lemma 3.3 *Under H_A C9-C12, $\sqrt{k}(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0) - \sqrt{k}d$ is approximately a mean-zero Gaussian process with the covariance structure*

$$\Sigma_A = \begin{pmatrix} M^{11} \\ M^{21} \end{pmatrix}^t (\boldsymbol{\xi}_0) \left[\lambda^* I(\boldsymbol{\theta}, \boldsymbol{\eta}^*) + (1-\lambda^*) J(\boldsymbol{\theta}, \boldsymbol{\eta}^*)\right] \begin{pmatrix} M^{11} \\ M^{21} \end{pmatrix} (\boldsymbol{\xi}_0), \quad (3.23)$$

where $d = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$.

Proof. The proof uses equations (3.16)-(3.18) and is similar to that of the previous Lemma. Hence, the proof is omitted.

□

Example 3.3 For Example 2.1 of the two normally distributed treatment outcomes, the Σ_A for Rao's efficient score depends on μ_1 and μ_2 only through θ, which measures the standardized minimum clinically significant betterment achieved by the experimental treatment. The formula for Σ_A is given by

$$\begin{aligned}
\Sigma_A &= \{ \frac{1}{2}\lambda^*(2\sigma + 16\lambda^{*2}\sigma^3\theta^2 - 12\lambda^*\sigma^3\theta^2 - \lambda^*\theta^2 + 4\sigma^6\theta^4\lambda^* - 4\lambda^{*2}\sigma^4\theta^2 + 2\sigma^4\theta^2\lambda^* \\
&+ \theta^2 - 2\lambda^*\sigma - 8\lambda^{*3}\sigma^3\theta^2 + 12\sigma^5\theta^4\lambda^{*2} + 2\sigma^5\theta^4 - 8\sigma^5\theta^4\lambda^* + 4\sigma^3\theta^2 \\
&- 16\lambda^{*2}\sigma^6\theta^4 + 28\lambda^{*3}\sigma^6\theta^4 - 24\sigma^6\lambda^{*4}\theta^4 + 6\sigma^8\theta^6\lambda^{*5} - 18\sigma^8\theta^6\lambda^{*4} + 20\sigma^8\theta^6\lambda^{*3} \\
&- 10\sigma^8\theta^6\lambda^{*2} + 2\sigma^8\theta^6\lambda^* + 8\sigma^6\lambda^{*5}\theta^4 - 6\lambda^{*3}\sigma^5\theta^4 + 2\sigma^4\lambda^{*3}\theta^2) \} \\
&\div \{\sigma^2 (-\lambda^*\sigma^3\theta^2 + \lambda^{*2}\sigma^3\theta^2 - 1)^2)\}
\end{aligned}$$

For Wald's statistic, however, the covariance under H_A is free of the parameter of interest θ and is actually the same as the covariance under H_0,

$$\Sigma_A = \frac{1}{\sigma^2\lambda^*(1-\lambda^*)}.$$

Suppose that for the one-dimensional θ, by using the above lemmas and results of Theorem 3.1, one has been able to calculate the variances of V_k, under both H_0 and H_A, and denoted them by σ_0^2 and σ_A^2, respectively.

Let

$$\begin{aligned}\tau_1 &= \inf\left\{1 \le k \le n_0 : (R_k^*)^{1/2} \ge CV_1\right\} \\ &= \inf\left\{1 \le k \le n_0 : \left|\sigma_A^{-1}V_k\right| \ge \frac{\sigma_0}{\sigma_A}CV_1\right\}.\end{aligned}$$

Since by Lemma 3.1, $\sigma_A^{-1}V_{[kt]} - \sigma_A^{-1}\sqrt{[kt]}d$ is approximately a Wiener process, one can approximate τ_1 by

$$\inf\left\{t \le n_0 : \left|W(t) + \sqrt{t}\sigma_A^{-1}d\right| \ge \frac{\sigma_0}{\sigma_A}CV_1\right\}. \tag{3.24}$$

Hence, the stopping time for Test 1 using Rao's efficient score statistics can be approximated by $\min\{n_0, \tau_1\}$. Similarly, the stopping time for Test 2 can be approximated by $\min\{n_0, \tau_2\}$ with

$$\tau_2 = \inf\left\{t \le n_0 : \left|\sqrt{t}W(t) + t\sqrt{\frac{n_0}{\sigma_A}}\sigma_A^{-1}d\right| \ge \sqrt{\frac{n_0\sigma_0}{\sigma_A}}CV_2\right\}. \tag{3.25}$$

Therefore, the problem is reduced to computing the moments of the first passage time of a Wiener process through a non-linear barrier. In such cases, no closed form exists for the moments or for the density of the first passage time; however, much literature has provided approximate methods requiring numerical integration. We believe that the most tractable one is that of Pötzelberger and Wang (2001). Other related references include Scheike (1992), Novikov *et al.* (1999), Daniels (1996) and Ragimov (1993).

Remark 3.2 *In Tests 2.1 and 2.2, when the process crosses the boundaries,*

we have $stat_{\tau_1} \geq CV_1^2$ *and* $\tau_2 stat_{\tau_2}/n_0 \geq CV_2^2$. *On the other hand, since Theorems 3.1 and 3.2, imply* $stat_k/k \xrightarrow{a.s.} d$ *(positive drift), we can, heuristically, approximate* $stat_{\tau_i}/\tau_i$ *by the drift. Therefore, we can conclude that*

$$n_0 \geq E\tau_1 \geq CV_1^2/d$$

and

$$n_0 \geq E\tau_2 \geq \sqrt{\frac{n_0 CV_2^2}{d}}.$$

Using numerical values from Tables 2.2-2.9, we have seen that this lower bound is a good approximation of the average stopping time for tests based on Wald statistics.

Chapter 4

Generalization to g treatment groups

4.1 Preliminaries

Large multi-armed clinical trials in which several treatments have to be compared with each other or against a standard one are becoming quite frequent. However, until now, few studies have been published on this issue. For instance, by using a two-stage hypothesis-testing procedure, Siegmund (1993) and Betensky (1996) considered the comparison of three treatments with outcomes that are normally distributed with known common variance and observations coming in triplets. This procedure starts with a sequential test of the null hypothesis of no difference among the three treatments (henceforth, the "global null hypothesis") and once this hypothesis is rejected, the apparently inferior treatment will be removed and another sequential testing starts to compare the remaining two treatments. Both authors based their procedure on RST methods, and their strategy is similar to the Fisher's protected LSD in the fixed sample literature (Kuehl 2000). Gombay (2002a) gives an example

of a sequential procedure comparing g normally distributed populations with known, common, variance by using generalized likelihood ratio and assuming that observations arrive in g-tuples. Also, Examples 6 and 7 in Gombay (2002c) provide sequential change-point procedures for testing the ANOVA global hypotheses.

In the group sequential literature, Jennison & Turnbull (1999, [16.1]) addressed the problem of testing a global hypothesis of no difference among g treatment groups with outcomes that are normally distributed. The authors provided O'Brien-Fleming & Pocock-type critical values for use in the case of balanced designs (i.e., designs in which groups of equal sizes are recruited in each treatment arm at each interim analysis) with known common standard deviation across treatments. In the case of unbalanced designs with unknown variances, Jennison and Turnbull (1999) suggested the use of the critical values for the balanced case with known variances as good approximations, whereas Follmann *et al.* (1994) suggested obtaining critical values by simulations.

In this chapter, we shall give some straightforward extensions of the results of Chapter 2 to the case of global hypotheses for g treatment groups. This extension will be discussed in Section 4.2. Application to the case of the one-way ANOVA for g treatment groups along with a small Monte Carlo simulation assessing power and ASN for $g = 3$ will be given in Section 4.3.

4.2 Extension to g treatment groups

As in Chapter 2, we assume that patients arriving sequentially are assigned to one of $i = 1, 2, ..., g$ treatments with probabilities λ_i such that $\sum_{i=1}^{g} \lambda_i = 1$. At each stage of the trial, $\{x_{ij}; i = 1, ..., g, \ \ j = 1, ..., n_i\}$ independent observations are available, with the total number of observations being $k = \sum_{i=1}^{g} n_i$. Let $f_i(x; \boldsymbol{\theta}, \boldsymbol{\eta})$, $i = 1, ..., g$ denote the population densities with different functional forms sharing the same d-dimensional parameter of interest, $\boldsymbol{\theta}$, and the same p-dimensional nuisance parameter, $\boldsymbol{\eta}$. These densities are obtained by appropriately transforming the parameters of the original problem so that the hypotheses of interest become $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and $\boldsymbol{\eta}$ unknown.

Conditions similar to C1-C8 of Chapter 2 are still required in order to extend Theorems 2.1, 2.2 and their corollaries to the case of g groups. The extension of the conditions to the general case is quite simple. For example, C3 requires that, for each $\boldsymbol{\xi} \in \Omega_0$ and $n_i = 1, 2, 3, ...$, the score equations

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \nabla_{\boldsymbol{\eta}} \log f_i(x_{ij}; \boldsymbol{\theta}_0, \boldsymbol{\eta}) = 0$$

and

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \nabla_{\boldsymbol{\xi}} \log f_i(x_{ij}; \boldsymbol{\xi}) = 0,$$

have unique solutions denoted by $\hat{\boldsymbol{\eta}}_k$ and $\hat{\boldsymbol{\xi}}_k$, where $k = \sum_{i=1}^{g} n_i$. However, for the error terms to be bounded, the number of treatment groups must be small compared to the total sample size at any given stage of the trial.

For the i^{th} treatment group, let I_i be the population Fisher information matrix. Partitioned matrices will follow the notation of equations (2.3) and

(2.4).

Under the above setup, Theorems 2.1, 2.2, Corollaries 2.1, 2.2, Lemmas 2.2, 2.3 and their proofs will, apart from notational complications, extend to the general case of g treatments in a straightforward manner. The resulting Rao's test statistic process is

$$R_k^* = V_k(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k) M^{11}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k) V_k^t(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}}_k),$$

with

$$V_k(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}}_k) = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \nabla_{\boldsymbol{\theta}} \log f_i(x_{ij}; \boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k),$$

$$M^{11} = \left(\sum_{i=1}^{g} \frac{n_i}{k} I_i \right)^{11}.$$

Similarly, for Wald test statistics,

$$W_k^{*(1)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)[M^{11}(\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\eta}}_k)]^{-1}(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t,$$

$$W_k^{*(2)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\eta}}_k)]^{-1}(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t,$$

and

$$W_k^{*(3)} = (\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)[M^{11}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)]^{-1}(\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^t,$$

where $\tilde{\boldsymbol{\theta}}_k$, $\tilde{\boldsymbol{\eta}}_k$ are unrestricted MLEs, and $\hat{\boldsymbol{\eta}}_k$ is a restricted MLE. Furthermore, Tests 2.1 and 2.2 will straightforwardly apply in the general case.

Remark 4.1 *By using the above global hypotheses testing procedures, one can adopt a strategy similar to that of Siegmund (1993) and Betensky (1996). In that case, one would start with all the g treatments and use the tests of this chapter to eventually reject the null hypothesis and then discard the, apparently,*

most inferior treatment and continue with the remaining ones until only one best treatment is left or the terminal planned sample size is reached. The overall α can be controlled by using a methods such as Bonferroni.

4.3 Application and simulations

As an extension of Example 2.1, suppose the measured outcomes, $x_{ij} \sim N(\mu_i, \sigma^2)$, are coming from treatment $i = 1, ..., g$ with probability λ_i. Consider the global hypotheses of interest $H_0 : \mu_1 = \mu_2 = ... = \mu_g = \mu$ vs $H_A : \mu_i \neq \mu$ for at least one i and that the variance, σ^2, is an unknown nuisance parameter. Without loss of generality, let $\mu_1 = \mu$. A transformation such as

$$\theta_{i-1} = \frac{\mu_i - \mu}{\sigma^2} \quad i = 2, ..., g; \quad \theta_g = \eta_1 = \frac{\mu_g + \mu}{2\sigma^2}; \quad \theta_{g+1} = \eta_2 = \frac{1}{\sigma^2}$$

leads to the equivalent null hypotheses $H_0 : \theta_1 = \theta_2 = ... = \theta_{g-1} = 0$ with nuisance parameters η_1, η_2, and the log-likelihood of the data is given by

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\eta}; \boldsymbol{x}) &= \sum_j \log(\frac{1}{\sqrt{2\pi}}) + \log \eta_2 - \frac{1}{2}\eta_2 x_{1j}^2 + \eta_2 x_{1j}\left(\frac{\theta_{g-1}}{2\eta_2} + \frac{\eta_1}{\eta_2}\right) \\ &- \frac{1}{2}\eta_2\left(\frac{\theta_{g-1}}{2\eta_2} + \frac{\eta_1}{\eta_2}\right)^2 + \sum_{i=2}^{g}\sum_{j=1}^{n_i}\left\{\log(\frac{1}{\sqrt{2\pi}}) + \log \eta_2 - \frac{1}{2}\eta_2 x_{ij}^2\right. \\ &+ \left.\eta_2 x_{ij}\left(\frac{-\theta_{i-1}}{\eta_2} + \frac{\theta_{g-1}}{2\eta_2} + \frac{\eta_1}{\eta_2}\right) - \frac{1}{2}\eta_2\left(\frac{-\theta_{i-1}}{\eta_2} + \frac{\theta_{g-1}}{2\eta_2} + \frac{\eta_1}{\eta_2}\right)^2\right\}. \end{aligned}$$

Example 4.1 By using the above described ANOVA setup, it can be shown that for $g = 3$,

$$V_k = \frac{1}{\sqrt{k}}\left(x_{2.} - n_2 x_{..}, x_{(3).} - n_3 x_{..}\right)^t,$$

$$M^{11}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k) = \frac{1}{\hat{\sigma}_k^2 \hat{\lambda}_1 \hat{\lambda}_2 \hat{\lambda}_3}\begin{pmatrix} \hat{\lambda}_3(1-\hat{\lambda}_3) & \hat{\lambda}_2\hat{\lambda}_3 \\ \hat{\lambda}_2\hat{\lambda}_3 & \hat{\lambda}_2(1-\hat{\lambda}_2) \end{pmatrix},$$

and

$$[M^{11}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}_k)]^{-1} = \hat{\sigma}_k^2 \begin{pmatrix} \hat{\lambda}_3(1-\hat{\lambda}_3) & \hat{\lambda}_2\hat{\lambda}_3 \\ \hat{\lambda}_2\hat{\lambda}_3 & \hat{\lambda}_2(1-\hat{\lambda}_2) \end{pmatrix},$$

where $x_{i.} = \sum_{j=1}^{n_i} x_{ij}$, $x_{..} = \sum_{i=1}^{g}\sum_{j=1}^{n_i} x_{ij}$, $\hat{\lambda}_i = n_i/k$ and

$$\hat{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{3}(x_{ij} - x_{..})^2$$

is the restricted MLE of σ^2 under H_0. Also, the unrestricted MLE of σ^2 is

$$\tilde{\sigma}_k^2 = \frac{1}{k}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2.$$

Hence,

$$R_k^* = \frac{1}{k\hat{\sigma}_k^2\hat{\lambda}_1\hat{\lambda}_2\hat{\lambda}_3}\Big\{(x_{2.} - n_2x_{..})^2\hat{\lambda}_3(1-\hat{\lambda}_3) + 2(x_{2.} - n_2x_{..})(x_{3.} - n_3x_{..})\hat{\lambda}_2\hat{\lambda}_3 + (x_{3.} - n_3x_{..})^2\hat{\lambda}_2(1-\hat{\lambda}_2)\Big\}, \tag{4.1}$$

$$W_k^{*(2)} = \tilde{\sigma}_k\{\,\hat{\lambda}_2(1-\hat{\lambda}_2)(\bar{x}_{2.} - \bar{x}_{1.})^2 + 2\hat{\lambda}_2\hat{\lambda}_3(\bar{x}_{2.} - \bar{x}_{1.})(\bar{x}_{3.} - \bar{x}_{1.}) + \hat{\lambda}_3(1-\hat{\lambda}_3)(\bar{x}_{3.} - \bar{x}_{1.})^2\,\} \tag{4.2}$$

and

$$W_k^{*(3)} = \hat{\sigma}_k\{\,\hat{\lambda}_2(1-\hat{\lambda}_2)(\bar{x}_{2.} - \bar{x}_{1.})^2 + 2\hat{\lambda}_2\hat{\lambda}_3(\bar{x}_{2.} - \bar{x}_{1.})(\bar{x}_{3.} - \bar{x}_{1.}) + \hat{\lambda}_3(1-\hat{\lambda}_3)(\bar{x}_{3.} - \bar{x}_{1.})^2\,\}. \tag{4.3}$$

The matrix $[M^{11}(\tilde{\boldsymbol{\theta}}_k, \tilde{\boldsymbol{\eta}}_k)]^{-1}$ and hence $W_k^{*(1)}$ have lengthy expressions and will not be reported here.

Using the above example, we have carried out a small Monte Carlo simulation for Test 2 with R_k^* by using allocation probabilities $\lambda_i = 1/3$ for

$i = 1, 2, 3$ and common variance $\sigma^2 = 1$. The results of the simulation, in Table 4.1, indicate that our conclusions for Test 2 with R_k^*, given in Section 2.6, are still valid for $g = 3$.

Table 4.1: Simulated power (P), and average stopping time (EN) for Test 2 using R_k^* of equation (4.1) with normally distributed outcomes (Example 4.1) with $\mu_1 = 0$, $\sigma = 1$ and various μ_2, μ_3. The critical value, CV_2, obtained from Delong (1980) and treatment allocation probabilities $\lambda_i = 1/3$ for $i = 1, 2, 3$ were used.

			$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
n_0	μ_2	μ_3	P	EN	P	EN	P	EN
50								
	0.0	0.0	0.006	49.97	0.038	49.72	0.084	49.24
	0.0	0.5	0.072	49.61	0.227	48.12	0.347	46.47
	0.0	0.9	0.385	47.37	0.664	42.81	0.782	39.52
	0.5	0.0	0.072	49.61	0.228	48.13	0.347	46.47
	0.5	0.5	0.073	49.60	0.228	48.11	0.348	46.46
	0.5	0.9	0.263	48.36	0.533	44.73	0.663	41.90
	0.9	0.0	0.388	47.35	0.664	42.81	0.777	39.57
	0.9	0.5	0.260	48.37	0.531	44.77	0.663	41.84
	0.9	0.9	0.386	47.36	0.664	42.83	0.779	39.54
100								
	0.0	0.0	0.008	99.91	0.043	99.28	0.087	98.27
	0.0	0.5	0.239	96.43	0.477	89.75	0.608	84.29
	0.0	0.9	0.849	79.41	0.953	66.96	0.977	60.13
	0.5	0.0	0.241	96.43	0.476	89.75	0.604	84.29
	0.5	0.5	0.241	96.40	0.477	89.75	0.608	84.25
	0.5	0.9	0.696	85.64	0.878	73.72	0.931	66.67
	0.9	0.0	0.850	79.39	0.954	66.85	0.978	60.11
	0.9	0.5	0.697	85.59	0.879	73.64	0.929	66.68
	0.9	0.9	0.851	79.33	0.954	66.96	0.977	60.13
200								
	0.0	0.0	0.008	199.79	0.045	198.43	0.091	196.24
	0.0	0.5	0.614	174.31	0.817	152.02	0.887	138.20
	0.0	0.9	0.997	116.00	1.000	96.17	1.000	86.20
	0.5	0.0	0.613	174.40	0.817	151.97	0.889	138.07
	0.5	0.5	0.615	174.37	0.816	152.05	0.887	137.98
	0.5	0.9	0.981	130.19	0.996	108.28	0.999	96.82
	0.9	0.0	0.998	116.05	1.000	96.25	1.000	86.21
	0.9	0.5	0.980	130.31	0.996	108.26	0.998	96.88
	0.9	0.9	0.998	116.02	1.000	96.17	1.000	86.20
500								
	0.0	0.0	0.009	499.39	0.046	495.74	0.093	490.09
	0.0	0.5	0.986	308.87	0.997	255.76	0.999	229.36
	0.0	0.9	1.000	184.23	1.000	152.97	1.000	137.11
	0.5	0.0	0.986	308.83	0.997	255.79	0.999	229.26
	0.5	0.5	0.987	309.02	0.998	255.86	0.999	228.96
	0.5	0.9	1.000	207.78	1.000	172.28	1.000	154.54
	0.9	0.0	1.000	184.22	1.000	152.89	1.000	137.13
	0.9	0.5	1.000	207.92	1.000	172.33	1.000	154.49
	0.9	0.9	1.000	184.19	1.000	153.23	1.000	137.11

Chapter 5

Two-sample nonparametric tests using U-statistics with anti-symmetric kernels

5.1 Preliminaries

Consider the sequential comparison of two populations (treatments) in which, as in Chapter 2, observations come from the first population with probability λ and from the second with probability $1-\lambda$; that is, at each stage of the sampling, the investigator has a total cumulative sample of size $k = m + n$, $x_1, \ldots, x_m$ independent each from distribution F and $y_1, \ldots, y_n$ independent from distribution G. One would like to test

$$\begin{aligned} H_0 : &\quad F(x) = G(x) \quad \text{for all } x \in \mathbb{R} \\ H_A : &\quad F(x_0) \neq G(x_0) \quad \text{for some } x_0 \in \mathbb{R} \end{aligned} \tag{5.1}$$

based on the sample at hand at every stage and come to a decision as soon as possible. Early decision-making is dictated, especially in trials involving human lives, by ethical and economical reasons.

Many studies have been published on nonparametric methods for the

sequential testing of (5.1). These studies concentrate mainly on rank score tests when observations are survival times with certain censoring mechanisms.

In this chapter, we will focus on non-censored observations and, unlike Sen (1981), use U-statistic based procedures with anti-symmetric kernels, i.e., kernels $h(x,y)$ such that $h(x,y) = -h(y,x)$ for $x, y \in \mathbb{R}$, to test (5.1). In Chapter 6, we will compare the sequential methods so constructed with the two-sample t-tests of the Pocock and O'Brien-Fleming types studied in Jennison and Turnbull (2001).

Originally, U-statistics were invented for unbiased estimation of a parameter θ. For that purpose, people usually considered symmetric kernels without loss of generality, as all other kernels useful for that purpose can be symmetrized. For anti-symmetric kernels,

$$E[h(x,y)] = 0 \text{ if } x \stackrel{\mathcal{D}}{=} y.$$

Therefore, anti-symmetric kernels are suitable for testing hypotheses of the form (5.1). We point out that symmetrizing an anti-symmetric kernel will result in an identically zero U-statistic which is, obviously, not useful. For this reason, the weak invariance principles of Miller and Sen (1972) are not applicable. A detailed discussion of the general theory of U-statistics is given in Koroljuk and Borovskich (1994) and Lee (1990).

The chapter is organized as follows: in section 5.2, we will give a definition of the U-statistic and its kernel, and the regularity conditions for the kernel, along with the main asymptotic results for the maximal functionals of the U-statistic. Based on these asymptotic results, we define in Section 5.3

two testing procedures for (5.1), similar to Test 1 & 2 of Chapter 2, and give an illustrative application using sign and Gehan-Gilbert kernels. In Section 5.5, the consistency of the test statistics is proved.

5.2 Asymptotic results under H_0

A kernel function, $h(x, y)$, is called "anti-symmetric" if $h(x, y) = -h(y, x)$. For a given anti-symmetric kernel, $h(x, y)$, define

$$h_1(u) = E\big(h(x, y)|x = u\big). \tag{5.2}$$

We require the following regularity conditions.

C1. For all $|t| \leq t_0$ with $t_0 > 0$, the moment generating function of h_1 is finite; that is, $E\{\exp(t\, h_1(x))\} < \infty$.

C2. The kernel, h, is non-degenerate; that is,

$$\sigma^2 = \text{Var}\,\big(h_1(x)\big) > 0.$$

To compare the two populations at stage $k = m + n$ of the sequential sampling when there are m observations from population 1 and n observations from population 2, we use the U-statistic

$$U_k = \sum_{i=1}^{m} \sum_{j=1}^{n} h\big(x_i, y_j\big).$$

Let $S_j^{(1)} = \sum_{i=1}^{j} h_1(x_i)$ and $S_j^{(2)} = \sum_{i=1}^{j} h_1(y_i)$ with h_1 defined in (5.2).

Theorem 5.1 *Under H_0, C1 and C2,*

$$\limsup_{k\to\infty} \left| k^{-1}U_k - (1-\lambda)S_{\lambda k}^{(1)} + \lambda S_{(1-\lambda)k}^{(2)} \right| \overset{\text{a.s.}}{=} O(\beta_k^{-1}),$$

where

$$\beta_k = k^{-1/4}(\log\log k)^{-1/4}\big(\log(k/\log\log k)^{1/2} - \log\log k\big)^{-1/2}. \tag{5.3}$$

For the proof of Theorem 5.1, we need the following Lemma:

Lemma 5.1 *Assuming the conditions of Theorem 5.1,*

$$\limsup_{k\to\infty} \left| S_m^{(i)} - S_{\lambda k}^{(i)} \right| \overset{\text{a.s.}}{=} O(\beta_k^{-1}),$$

where β_k is defined in (5.3).

Proof of Lemma 5.1 We have $m = \sum_{i=1}^{k} I_i$, where $I_i = 1$ if the i^{th} observation is from population 1 and zero otherwise. Since observations come from population 1 with probability λ, it is clear that $E(m) = \lambda k$, and so the law of iterated logarithm gives

$$\limsup_{k\to\infty} \left| m - \lambda k \right| \overset{\text{a.s.}}{=} O(a_k), \tag{5.4}$$

where $a_k = (k\log\log k)^{1/2}$. From Theorem 3.1.1 of Csörgő and Révész (1981) on the increments of partial sums of independent identically distributed random variables, we get

$$\limsup_{k\to\infty} \max_{1\le l\le k-a_k} \max_{1\le j\le a_k} \beta_k \left| S_{l+j}^{(i)} - S_l^{(i)} \right| \overset{\text{a.s.}}{=} 1,$$

and from this the Lemma follows.

□

Proof of Theorem 5.1 We write,

$$\begin{aligned} U_k &= n\sum_{i=1}^{m} h_1(x_i) - m\sum_{j=1}^{n} h_1(y_j) \\ &+ \sum_{i=1}^{m}\sum_{j=1}^{n}\left[h(x_i,y_j) - h_1(x_i) + h_1(y_j)\right] \\ &= A_k - B_k + C_k. \end{aligned}$$

Defining a new kernel $h^*(u,v) = h(u,v) - h_1(u) + h_1(v)$, we see that $C_k = \sum\sum h^*(x_i,y_j)$. From the anti-symmetry property of h, it is easy to verify that the kernel h^* is also anti-symmetric. By using properties of conditional expectation and Jensen's inequality,

$$\begin{aligned} E|h_1(x)|^\nu &= E\left\{|E[h(x,y)|x]|^\nu\right\} \\ &\leq E\left\{E[|h(x,y)|^\nu|x]\right\} \\ &= E[|h(x,y)|^\nu] < \infty. \end{aligned}$$

Therefore, it is easy to see that $E|h^*| < \infty$ for some $\nu > 2$. Also, since under H_0, $E[h_1(x)] = E[h_1(y)]$, it can be seen that

$$\begin{aligned} Var[h_1^*(x)] &= Var\left\{E[h^*(x,y)|y]\right\} \\ &= Var\left\{E[h(x,y)|x] - E[h_1(x)] + E[h_1(y)]\right\} = Var\{0\} = 0, \end{aligned}$$

and hence, the kernel h^* is degenerate.

Now denoting by z_j elements of $(x_1, \ldots, x_m, y_1, \ldots, y_n)$, we have by Theorem 2 of Dehling *et al.* (1986)

$$\begin{aligned} C_k &= \sum_{1 \le i_1, i_2 \le k} h^*(z_{i_1}, z_{i_2}) \\ &\quad - \sum_{1 \le i_1, i_2 \le m} h^*(x_{i_1}, x_{i_2}) - \sum_{1 \le i_1, i_2 \le n} h^*(y_{i_1}, y_{i_2}) \\ &= O(k \log\log k) \end{aligned} \tag{5.5}$$

By the law of iterated logarithm as in (5.4), we get

$$\left| \frac{n}{k} - (1-\lambda) \right| \overset{\text{a.s.}}{=} O\big(k^{-1/2} (\log\log k)^{1/2}\big).$$

Thus,

$$\begin{aligned} k^{-1} A_k &= n/k \sum_{i=1}^{m} h_1(x_i) \\ &= (1-\lambda) S^{(1)}_{\lambda k} + (n/k)\big(S^{(1)}_m - S^{(1)}_{\lambda k}\big) + \big(n/k - (1-\lambda)\big) S^{(1)}_{\lambda k} \\ &= (1-\lambda) S^{(1)}_{\lambda k} + A_{1k} + A_{2k}. \end{aligned}$$

By Lemma 5.1, (5.4), and the law of iterated logarithm,

$$|A_{1k}| \overset{\text{a.s.}}{=} O(\beta_k^{-1}) \tag{5.6}$$

$$|A_{2k}| \overset{\text{a.s.}}{=} O(\log\log k), \tag{5.7}$$

and hence,

$$k^{-1/2} A_k \overset{a.s}{=} (1-\lambda) S^1_{\lambda k} + O(\beta_k^{-1}). \tag{5.8}$$

Similarly, we have

$$k^{-1} B_k \overset{\text{a.s.}}{=} \lambda S^{(2)}_{(1-\lambda)k} + O(\beta_k^{-1}). \tag{5.9}$$

From (5.5), (5.6), (5.7), and (5.9), the Theorem follows.

□

Based on Theorem 5.1, we can derive the following corollary giving weak convergence results for two functionals of U_k.

Corollary 5.1 *Under the conditions of Theorem 5.1,*

$$\text{(i)} \lim_{n_0\to\infty} P\Big\{a(T)\max_{1\le k\le n_0} k^{-3/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k| \le t + d(T)\Big\} = \exp(-2e^{-t}),$$
$$\lim_{n_0\to\infty} P\Big\{a(T)\max_{1\le k\le n_0} k^{-3/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}U_k \le t + d(T)\Big\} = \exp(-e^{-t}),$$

where $T = \log(n_0)$, $a(T) = (2\log T)^{1/2}$,

$$d(T) = 2\log T + (1/2)\log\log T - (1/2)\log\pi.$$

$$\text{(ii)} \lim_{n_0\to\infty} P\Big\{\max_{1<k\le n_0} k^{-1}n_0^{-1/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k| \le t\Big\}$$
$$= P\Big\{\sup_{0<t<1}|W(t)| \le t\Big\} = \tfrac{4}{\pi}\sum_{m=0}^{\infty}\tfrac{(-1)^m}{2m+1}\exp\big(-\pi^2(2m+1)^2/(8t^2)\big),$$

where $W(t)$ *is a standard Brownian motion,*

$$\lim_{n_0\to\infty} P\Big\{\max_{1<k\le n_0} k^{-1}n_0^{-1/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}U_k \le t\Big\}$$
$$= P\Big\{\sup_{0<t<1} W(t) \le t\Big\} = 2\Phi(t) - 1,$$

where $\Phi(\cdot)$ *is the standard normal distribution function.*

Proof of Corollary 5.1 Theorem 5.1 implies that $U_k/k\sigma\sqrt{\lambda(1-\lambda)}$ is well approximated by

$$M_k = \frac{(1-\lambda)S^{(1)}_{\lambda k} - \lambda S^{(2)}_{(1-\lambda)k}}{\sigma\sqrt{\lambda(1-\lambda)}}.$$

But, since M_k is the partial sum of i.i.d. random variables with mean zero and standard deviation one (we can see this by calculating the variance of

the terms in the sum defining M_k), the Darling-Erdös and KMT invariance principles will apply to $U_k/k\sigma\sqrt{\lambda(1-\lambda)}$ to give (i) and (ii), respectively (c.f., Csörgö and Révész 1981).

□

The following Lemma shows that the asymptotic results in Corollary 5.1(ii) remain valid if the unknown variance is replaced by a consistent estimator. In the case of Corollary 5.1(i), we need the convergence rate of the estimator replacing the variance to be of the order $O\big((\log k)^{-1/2}\big)$.

Lemma 5.2 *Under H_0 and conditions of Theorem 5.1,*

(i) *if $\widehat{\sigma}_k^2 \xrightarrow{\text{a.s.}} \sigma^2$, then*

$$\max_{1<k\leq n_0} |(\hat{\sigma}_k^{-1} - \sigma^{-1})\, k^{-1} n_0^{-1/2} \lambda^{-1/2}(1-\lambda)^{-1/2} U_k| = o_p(1).$$

(ii) *if $|\widehat{\sigma}_k - \sigma| = o_p((\log k)^{-1/2})$, then*

$$a(T) \max_{1<k\leq n_0} |(\hat{\sigma}_k^{-1} - \sigma^{-1})\, k^{-3/2} \lambda^{-1/2}(1-\lambda)^{-1/2} U_k| = o_p(1).$$

Proof. (i) By Corollary 5.1(ii)

$$\max_{1<k\leq \log n_0} k^{-1} n_0^{-1/2} |U_k| = O_p\big(n_0^{-1/2}(\log n_0)^{1/2}\big),$$

and

$$\max_{1<k\leq n_0} k^{-1} n_0^{-1/2} |U_k| = O_p(1).$$

As

$$\max_{1<k\leq \log n_0} |\sigma^{-1} - \widehat{\sigma}_k^{-1}| \overset{\text{a.s}}{=} O(1), \tag{5.10}$$

$$\begin{aligned}
&\max_{1<k\le \log n_0} |k^{-1}n_0^{-1/2}\sigma^{-1}U_k - k^{-1}n_0^{-1/2}\widehat{\sigma}_k^{-1}U_k| \\
&\qquad \le \max_{1<k\le \log n_0} |\sigma^{-1} - \widehat{\sigma}_k^{-1}| \max_{1<k\le \log n_0} k^{-1}n_0^{-1/2}|U_k| \\
&\qquad = O_p\big(n_0^{-1/2}(\log n_0)^{1/2}\big).
\end{aligned}$$

Thus it is enough to examine the process in the range $\log n_0 < k \le n_0$.

$$\begin{aligned}
&\max_{\log n_0<k\le n_0} k^{-1}n_0^{-1/2}|\widehat{\sigma}_k^{-1}U_k - \sigma^{-1}U_k| \\
&\qquad \le \max_{\log n_0<k\le n_0} |\widehat{\sigma}_k^{-1} - \sigma^{-1}| \max_{\log n_0<k\le n_0} k^{-1}n_0^{-1/2}|U_k| \\
&\qquad = o_p(1),
\end{aligned}$$

as $\max_{\log n_0<k\le n_0} |\widehat{\sigma}_k^{-1} - \sigma^{-1}| \overset{\text{a.s.}}{=} o(1)$.

(ii) By Corollary 5.1(i),

$$\max_{1<k\le \log n_0} k^{-3/2}|U_k| = O_p\big((\log\log\log n_0)^{1/2}\big)$$

and

$$\max_{1<k\le n_0} k^{-3/2}|U_k| = O_p\big((\log\log n_0)^{1/2}\big). \tag{5.11}$$

By (5.10),

$$\begin{aligned}
\max_{1<k\le \log n_0} |k^{-3/2}\sigma^{-1}U_k - k^{-3/2}\widehat{\sigma}_k^{-1}U_k| &\le \max_{1<k\le \log n_0} |\sigma^{-1} - \widehat{\sigma}_k^{-1}| \max_{1<k\le \log n_0} k^{-3/2}|U_k| \\
&= O_p\big((\log\log\log n_0)^{1/2}\big).
\end{aligned}$$

For the range $\log n_0 < k \le n_0$, we have by the assumption

$$\begin{aligned}
\max_{\log n_0<k\le n_0} k^{-3/2}|U_k|\,|\widehat{\sigma}_k^{-1} - \sigma^{-1}| &= \max_{\log n_0<k\le n_0} k^{-3/2}|U_k| o_p((\log k)^{-1/2}) \\
&\le o_p\big((\log\log n_0)^{-1/2}\big) \max_{\log n_0<k\le n_0} k^{-3/2}|U_k| \\
&\le o_p\big((\log\log n_0)^{-1/2}\big) O_p\big((\log\log n_0)^{1/2}\big) \\
&= o_p(1).
\end{aligned}$$

By (5.11), the proof of part (i) of the Lemma is now complete.

□

5.3 Test procedures and examples

Let α denote the level of significance, $T = log(n_0)$, $a(T) = (2\,log\,T)^{1/2}$, $d(T) = 2\,log\,T + (1/2)\,log\,log\,T - (1/2)log\,\pi$ and $\hat{\sigma}_k$ be estimator of σ with appropriate convergence rates. Corollary 5.1 and Lemma 5.2 allow us to define the following sequential tests, truncated after n_0 observations.

Test 5.1 (Test 1) *For $k = 2, 3, \ldots, n_0$ calculate*

$$k^{-3/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k|.$$

Stop and reject H_0 the first time it exceeds the critical value $CV_1(\alpha, n_0)$, where

$$CV_1(\alpha, n_0) = a^{-1}(T)\big[-\log\big(-1/2\,\log(1-\alpha)\big) + d(T)\big];$$

otherwise do not reject H_0.

If a one-sided version is desired, we replace $|U_k|$ by U_k, and $CV_1(\alpha, n_0)$ by $CV_1^*(\alpha, n_0) = a^{-1}(T)\big[-\log\big(-\log(1-\alpha)\big) + d(T)\big]$.

As we have mentioned in Chapter 2, critical values CV_1 and CV_1^* give conservative tests. However, we have seen how, by using Vostrikova's (1981) results, they can be improved.

Test 5.2 (Test 2) *For $k = 2, 3, \ldots, n_0$ calculate*

$$k^{-1}n_0^{-1/2}\sigma^{-1}\lambda^{-1/2}(1-\lambda)^{-1/2}|U_k|.$$

Stop and reject H_0 the first time it exceeds critical value $CV_2(\alpha)$.

$CV_2(\alpha)$ is obtained from the distribution in Corollary 5.1(ii), and some of its values were already given in Table 2.1. For the one-sided version using U_k instead of $|U_k|$, we replace $CV_2(\alpha)$ by $CV_2^*(\alpha) = \Phi^{-1}(1-\alpha/2)$, where Φ is the standard normal distribution function.

5.4 Examples

Example 5.1 The sign kernel defined by

$$h(x,y) = \begin{cases} +1 \text{ if } x > y \\ 0 \text{ if } x = y \\ -1 \text{ if } x < y \end{cases}$$

is useful for comparing the locations of two populations. In this case, $\sigma^2 = 1/3$ under the null hypothesis (Lee 1990).

Example 5.2 Consider a sequential trial comparing two treatments with a survival endpoint. Let $xi = \{z_i, \delta_{z_i}\}$ be the observation from the i^{th} subject under treatment 1, where z_i is time from entry to study until either death or random right censoring occurs, and δ_{z_i} is an indicator of death or censoring. Similarly, let $y_j = \{v_j, \delta_{v_j}\}$ be an observation of a subject under treatment 2. Assume the censoring mechanism is the same in the two groups, and independent from the survival time, and that the data is analysed each time an event (death or censoring) occurs. In this situation, the Gehan-Gilbert kernel (Kalbfleisch and Prentice 2002),

$$h(x,y) = I(z > v, \delta_v = 1) - I(z < v, \delta_z = 1),$$

where I is an indicator function and $x = (z, \delta_z), y = (v, \delta_v)$, can be used to test (5.1) sequentially.

5.5 Consistency results

Now we will analyse the process under the alternative hypothesis H_A. Define

$$\mu = Eh(x_1, y_1) \tag{5.12}$$

$$h_2(x_i) = E\big(h(x_i, y_1)|x_i\big) \tag{5.13}$$

$$\sigma_2^2 = \mathrm{Var}\left(h_2(x_1)\right) \tag{5.14}$$

$$\sigma_3^2 = E\big(h(x_1, y_2)h(x_2, y_1)\big) - \mu^2. \tag{5.15}$$

Under H_A, we need the following regularity conditions:

C1. $0 < \sigma_2^2 < \infty$

C2. $E\,|h(y_1, y_2)|^\nu < \infty$ with some $\nu > 2$

C3. $0 < E\,|E\{h[(x_1, y_1)h(x_2, y_2)]\,|\,x_1, y_2\}| < \infty$

C4. $Eh^2(x_1, y_1) < \infty$

In the next theorem, we will show that the process $k^{-3/2}U_k$ converges weakly to the linear combination of two independent Ornstein-Uhlenbeck processes.

Theorem 5.2 *Under H_A, if $\mu \neq 0$ and C1-C4 are satisfied, then*

$$\begin{aligned} |\,k^{-3/2}U_k &- (1-\lambda)\lambda^{1/2}k^{1/2}\mu \\ &- \lambda(1-\lambda)^{1/2}\sigma_3\Gamma_1\big(\log[(1-\lambda)k]\big) \\ &- (1-\lambda)\lambda^{1/2}\sigma_2\Gamma_2(\log[\lambda k])\big| = o_p(1), \end{aligned}$$

where Γ_1 and Γ_2 are independent Ornstein-Uhlenbeck processes.

Proof. Define $U_k^* = \sum_{i=1}^{[\lambda k]} \sum_{j=1}^{[(1-\lambda)k]} h(x_i, y_j)$.

Since, $m < [\lambda k]$ implies $n > [(1-\lambda)k]$ and conversely, $m > [\lambda k]$ implies $n < [(1-\lambda)k]$ we can see that

$$|U_k - U_k^*| \leq \begin{cases} 0 & m = [\lambda k], \\ \left|\sum_{i=m}^{[\lambda k]} \sum_{j=1}^{[(1-\lambda)k]} h(x_i, y_j)\right| + \left|\sum_{i=1}^{m} \sum_{j=[(1-\lambda)k]}^{n} h(x_i, y_j)\right| & m < [\lambda k], \\ \left|\sum_{i=[\lambda k]}^{m} \sum_{j=1}^{n} h(x_i, y_j)\right| + \left|\sum_{i=[\lambda k]}^{m} \sum_{j=n}^{[(1-\lambda)k]} h(x_i, y_j)\right| & m > [\lambda k], \end{cases}$$

where [.] indicates the integer part of its argument. The terms in the right-hand side of the above inequality are almost surely of order $O((\log\log k)^{1/2})$. Therefore, the asymptotic distributions of $U_k/k^{3/2}$ and $U_k^*/k^{3/2}$ are the same, and we will concentrate on U_k^*.

For simplicity, from now on we will omit the integer part notation $[\cdot]$. To derive the asymptotic distribution of U_k^*, we follow some of the ideas in the proof of Theorem 3 of Gombay (2000*b*). Re-write U_k^* as

$$\begin{aligned} U_k^* &= \sum_{i=1}^{\lambda k} \sum_{j=1}^{(1-\lambda)k} [h(x_i, y_j) - h_2(x_i)] \\ &\quad + (1-\lambda)k \sum_{i=1}^{\lambda k} h_2(x_i) \\ &= A_k + B_k, \end{aligned}$$

and let $\mathcal{F}^{(1)}$ be the sigma-algebra generated by the sequence $x_1, x_2, \ldots$. Note the orthogonality of A_k to $\mathcal{F}^{(1)}$, i.e., $E[A_k|\mathcal{F}^{(1)}] = 0$.

Define the conditionally standardized, independent, identically distributed and centered random variables

$$z_j = \Big(\sum_{i=1}^{\lambda k} [h(x_i, y_j) - h_2(x_i)] \Big) / \sigma_*,$$

where

$$\sigma_*^2 = E\Big\{ \Big(\sum_{i=1}^{\lambda k} [h(x_i, y_j) - h_2(x_i)] \Big)^2 \Big| \mathcal{F}^{(1)} \Big\},$$

so that

$$A_k/\sigma_* = \sum_{j=1}^{(1-\lambda)k} z_j.$$

Given $\mathcal{F}^{(1)}$, there exists an Ornstein-Uhlenbeck process Γ_1 such that, as $k \to \infty$,

$$\Big| \sigma_*^{-1} \big((1-\lambda)k \big)^{-1/2} \sum_{j=1}^{(1-\lambda)k} z_j - \Gamma_1\big(\log[(1-\lambda)k] \big) \Big| \stackrel{\text{a.s}}{=} o(1),$$

using strong invariance principles for independent and identically distributed random variables (Csörgő and Révész 1981). To analyse the large sample behavior of σ_*^2, we use properties of conditional expectations, and re-write it

as,

$$\begin{aligned}
\sigma_*^2 &= E\left\{\left(\sum_{i=1}^{\lambda k}[h(x_i,y_j)-h_2(x_i)]\right)^2 \mid \mathcal{F}^{(1)}\right\} \\
&= E\left\{\left(\sum_{i=1}^{\lambda k}[h(x_i,y_j\right)^2 \mid \mathcal{F}^{(1)}\right\} + E\left\{\left(\sum_{i=1}^{\lambda k} h_2(x_i)\right)^2 \mid \mathcal{F}^{(1)}\right\} \\
&- E\left\{\sum_{i=1}^{\lambda k} 2[h(x_i,y_j)][h_2(x_i)] \mid \mathcal{F}^{(1)}\right\} \\
&= E\left\{\left(\sum_{i=1}^{\lambda k}[h(x_i,y_j\right)^2 \mid \mathcal{F}^{(1)}\right\} - \left(\sum_{i=1}^{\lambda k} h_2(x_i)\right)^2 \\
&= \sum_{i=1}^{\lambda k} E[h^2(x_i,y_j) \mid \mathcal{F}^{(1)}] + 2\sum_{1\le i<l\le \lambda k} E\left\{h(x_i,y_j)h(x_l,y_j) \mid \mathcal{F}^{(1)}\right\} \\
&- \left(\sum_{i=1}^{\lambda k} h_2(x_i)\right)^2 = C_{\lambda k} + 2D_{\lambda k} - H_{\lambda k}^2
\end{aligned}$$

As $k \longrightarrow \infty$, by the strong law of large numbers and C2 and C4, we get,

$$\frac{C_{\lambda k}}{\lambda k} \xrightarrow{a.s} Eh^2(x_1,y_1) \tag{5.16}$$

and

$$\frac{H_{\lambda k}}{\lambda k} \xrightarrow{a.s} Eh^2(x_1,y_1). \tag{5.17}$$

Noticing that $\binom{\lambda k}{2} D_{\lambda k}$ is a U-statistic with kernel

$$\phi(x_i,y_l) = E\left\{h(x_i,x_l)h(x_l,y_j) \mid \mathcal{F}^{(1)}\right\},$$

by the strong law of large numbers for U-statistics (Koroljuk and Borovskich 1994), we get

$$\frac{2}{(\lambda k)(\lambda k-1)} D_{\lambda k} \xrightarrow{a.s.} E[\phi(x_1,x_2)] = E[h(x_1,y_1)h(x_2,y_1) \mid \mathcal{F}^{(1)}]. \tag{5.18}$$

Now from (5.16)- (5.18) and the fact that, $\frac{1}{\lambda k} \stackrel{a.s}{\to} o$ and $\frac{\lambda k-1}{\lambda k} \stackrel{a.s}{\to} 1$ as $k \longrightarrow \infty$, we have

$$\sigma_*^2/(\lambda k)^2 \stackrel{\text{a.s.}}{\to} E[h(x_1, y_1)h(x_2, y_1)] - \mu^2 = \sigma^3.$$

Thus, we get

$$\left|A_k/[\sigma_3(\lambda k)\big((1-\lambda)k\big)^{1/2}] - \Gamma_1\big(\log[(1-\lambda)k]\big)\right| \stackrel{\text{a.s}}{=} o(1). \tag{5.19}$$

Finally, the term B_k is the sum of independent, identically distributed random variables, so by strong invariance principles (Csörgő and Révész 1981), there exists an Ornstein-Uhlenbeck process Γ_2, such that

$$\left|\big((1-\lambda)k(\lambda k)^{1/2}\big)^{-1}[B_k - (\lambda k)^{1/2}\mu] - \sigma_2\Gamma_2\big(\log(\lambda k)\big)\right| \stackrel{\text{a.s}}{=} o(1). \tag{5.20}$$

Putting (5.16), (5.19) and (5.20) together, we have

$$\left|k^{-3/2}U_k^* - (1-\lambda)\lambda^{1/2}k^{1/2}\mu - \lambda(1-\lambda)^{1/2}\sigma_3\Gamma_1\big(\log[(1-\lambda)k]\big)\right.$$

$$\left.-(1-\lambda)\lambda^{1/2}\sigma_2\Gamma_2\big(\log(\lambda k)\big)\right| = o_p(1).$$

□

Theorem 5.2 derives the distribution of the test statistic under the alternative hypothesis H_A. From it, we see that under H_A and for large k, $k^{-3/2}U_k$ is approximately normal with a bounded variance and an expected value in the order of $k^{1/2}$. From here, the consistency of the test procedures described in section 5.3 will follow.

Chapter 6

Empirical comparisons and concluding remarks

6.1 Preliminaries

From a practical point of view, determining how the test procedures proposed in this thesis compare, at least empirically, to the group sequential methods is very important.

In Section 6.2, we will examine the performance of Test 2 (Test 2.2) by using both Rao's efficient score and Wald's statistics, and Test 2 (Test 5.2) by using the sign kernel of Example 5.1. These tests will be compared to the fixed-sample t-test and to the group sequential t-tests proposed by Pocock (1977) and Jennison and Turnbull (2001).

In Section 6.3, we will provide concluding remarks for the thesis and directions for further research.

6.2 Monte Carlo simulations

Throughout this section, we will call the Test 2 (Test 2.2) procedure using Rao's efficient score process, defined in (2.30), the " sequential Rao test", the

Test 2 (Test 2.2) procedure using Wald's statistic process, defined in (2.32), will be called the "sequential Wald test". Similarly, the Test 2 (Test 5.2) procedure using the sign kernel of Example 5.1 will be called the "sequential sign test".

The design and results of the Monte Carlo simulations comparing these test procedures to the fixed-sample t-test and to the group sequential t-tests of Pocock and O'Brien-Fleming will be discussed in the next sections.

6.2.1 Parametric case

This section compares the sequential Rao and Wald tests to the fixed-sample t-test and to the group sequential t-tests of Pocock and O'Brien-Fleming (OBF) given in Jennison and Turnbull (2001). For this purpose, we use the setup of Example 2.1.

To briefly describe the Pocock and OBF group sequential t-tests with K planned interim analyses, let n_k denote the cumulative number of observations from both treatment groups at the k^{th} interim analysis, $k = 4, 2, ..., K-1$. At the terminal analysis, we have $n_K = n_0$, which is equivalent to our truncation point. Jennison and Turnbull (2001) suggested the following t-test procedure: *Reject H_0 at the first interim analysis when $|t_k| \geq t_{f,\alpha_k}$ where, t_k is the usual two-sample t-statistic for testing H_0 based on the n_k observations available at the k^{th} interim analysis, $f = n_k - 2$, $\alpha_k = 1 - \Phi(c_k)$, and Φ denotes the cdf of the standard normal.*

For Pocock, $c_k = C_P$, and for OBF, $C_k = c_{OBF}\sqrt{K/k}$. The two constants, C_P and C_{OBF}, are obtained by numerical integration assuming the known

variance, $\sigma^2 = 1$.

For the current simulations, we have chosen K as small as possible for the Pocock test and as many as $K = 20$ for OBF test wherever possible. In fact, Jennison and Turnbull (1999) suggest that choices of $K = 10 - 20$ for the OBF and $K = 2 - 5$ for the Pocock test are optimal for the two methods.

Each entry in Tables 6.1 and 6.2 and each point in Figures 6.1 and 6.2 is based on 10^4 Monte Carlo replicates using pseudorandom number generators of the IMSL fortran library. The simulations used the normally distributed outcomes of Example 2.1.

An initial guess of the truncation points, n_0, needed for the OBF and Pocock methods to attain a given fixed power at a fixed α level, is calculated by using S+SeqTrial (2001) under the assumption of known standard deviation, $\sigma = 1$. Consequently, this value is fine-tuned by simulations until the correct n_0 for the unknown σ case is found. For the sequential Rao and Wald tests, truncation points corresponding to fixed power and α levels are obtained by simulation search.

In general, when comparing sequential tests, several performance measures have to be considered. These are the truncation point, average stopping time, number of interim analyses, and power. Figures 6.1 and 6.2 summarise simulation results when the truncation point, n_0, is fixed and therefore, comparison is made in terms of power and average stopping time for some optimal number of interim analyses, K. The meaning of the optimality here is regorously defined in Barber and Jennison (2002) and Barber and Jennison (1999).

Figure 6.1 shows that for a fixed, moderately large truncation point ($n_0 = 240$), the sequential Rao test is practically identical in terms of average stopping time and operating characteristics to the O'Brien-Fleming procedure, and that both procedures have higher power than that of Pocock. The sequential Wald test is as powerful as the OBF and sequential Rao tests, but has a smaller average stopping time than all other tests. Pocock's test stops earlier than other tests, but it has the least power. Jennison and Turnbull (1999) reported this behaviour when comparing the Pocock and OBF tests to the fixed-sample test. The situation remains almost the same for small truncation points (See Figure 6.2), except that the sequential Rao and Wald tests are less powerful. This phenomenon was expected since the tests in this thesis were based on large sample theory.

On the other hand, if we fix power, then the comparison must be done in terms of n_0 and average stopping times. Let n_f be the sample size of the fixed, two-sample t-test attaining the same power as the sequential test under the same Type I error and treatment difference $|\mu_1 - \mu_2| = \theta$. Also, let $E_\theta N$ denote the average stopping time when the true absolute difference is θ, and let n_0 denote the truncation point of the sequential test. We use $100(\frac{E_\theta N}{n_f})\%$ and $100(\frac{n_0}{n_f})\%$ as measures of the performance of the sequential tests.

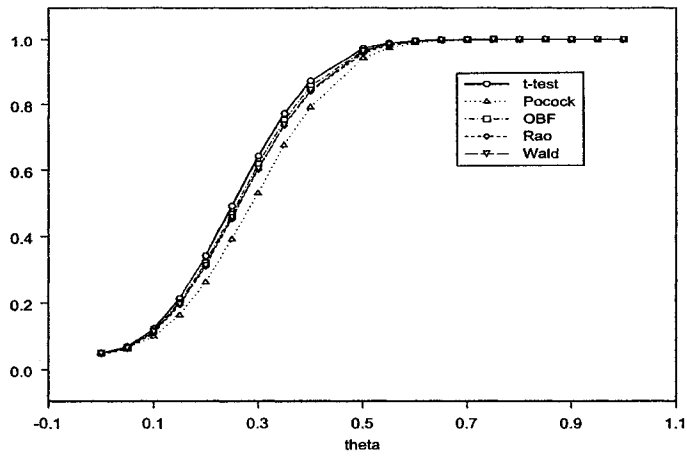
Tables 6.1 and 6.2 show the simulation results in terms of these measures of performance for $1 - \beta = .8, .9$; $\alpha = .05$; and $\theta = .1, .2, .5, .9$. It is clear from Table 6.1 that, at low power (80%) and small treatment difference, the sequential Rao and OBF tests have a similar performance with respect to

both measures. The sequential Wald test saves, on average, more sample than do both the OBF and sequential Rao test while having the same truncation points. For low power and large treatment difference, the sequential Rao and Wald tests are inferior to OBF.

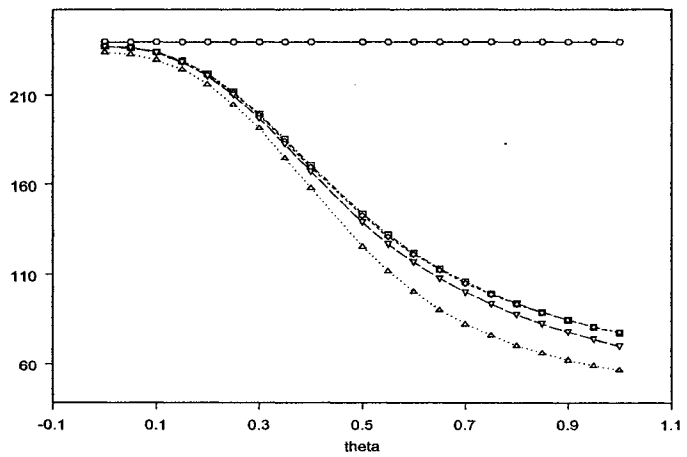
For high power (90%), the sequential Wald test is the most economical, followed at all treatment differences by the Rao and then by the OBF tests. All three methods have the same truncation points. Although, at high power, Pocock's method may save slightly more sample than the sequential Rao, sequential Wald and OBF tests, it requires a substantially much larger total sample than the others, and hence, it may not be preferable.

In terms of the empirical significance level, the first column of Tables 6.1 and 6.2 shows that, for large total sample size, all tests maintain the nominal level. For small total sample sizes, the Pocock and OBF tests are statistically significantly anti-conservative (as Jennison and Turnbull (2001) also noticed), whereas the sequential Rao and Wald tests are statistically significantly conservative.

Other simulation results, not reported, for the case when $\lambda \neq .5$ showed the same behaviour as above.

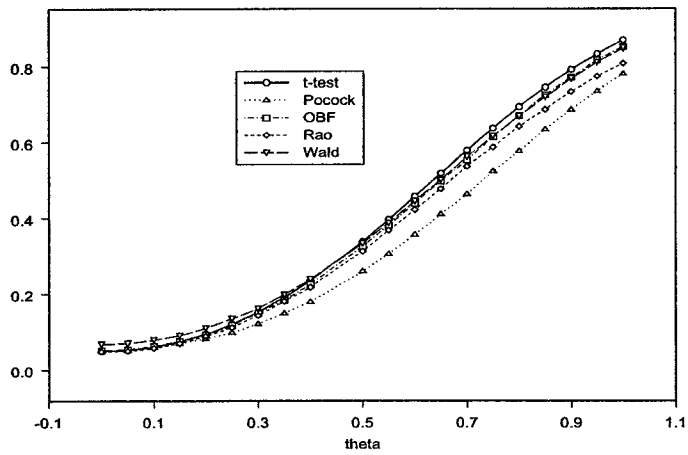


(a) Empirical power

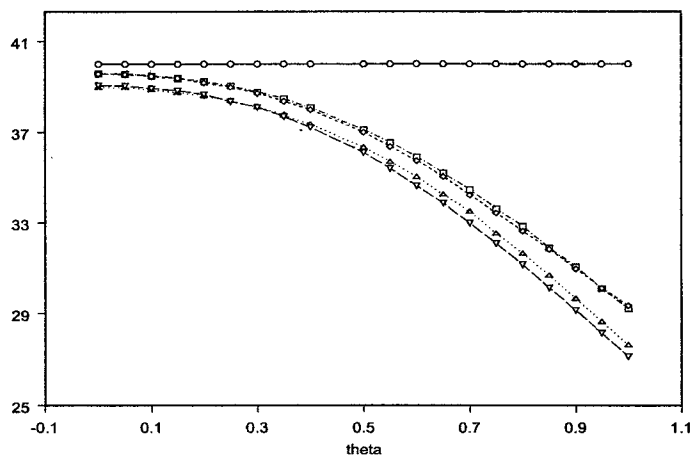


(b) Average stopping time

Figure 6.1: Plots for normally distributed outcomes with $\theta = \mu_1 - \mu_2$, common $\sigma = 1$, treatment allocation rate $\lambda = .5$ and total sample $n_0 = 240$. For the OBF and Pocock test, the number of interim analyses are $K = 20, 5$, respectively.



(a) Empirical power



(b) Average stopping time

Figure 6.2: Plots for normally distributed outcomes with $\theta = \mu_1 - \mu_2$, common $\sigma = 1$, treatment allocation rate $\lambda = .5$ and total sample $n_0 = 40$. For the OBF and Pocock tests, the number of interim analyses are $K = 20, 5$, respectively.

Table 6.1: Results of simulations based on Example 2.1 Sequential Rao and Wald are compared to the results Pocock and OBF t-tests. Nominal $\alpha = 0.05$, $1 - \beta = .8$, $\theta = |\mu_1 - \mu_2|$, and allocation rule $\lambda = .5$ are used. $E_\theta N$ is the average stopping time for sequential trials when the true treatment difference is θ, $n_0 = n_K$ is the truncation point, K is the number of interim analyses, n_f is the sample size of the fixed, two-sample t-test attaining the same power at the same α-level as the sequential tests. $\hat{\alpha}$ is the empirical size at the given total sample size, n_0.

$\hat{\alpha}$	K	θ	$E_\theta N$	$100(\frac{E_\theta N}{n_f})\%$	n_0	$100(\frac{n_0}{n_f})\%$
O'Brien-Fleming						
0.0498	20	0.1	2420.25	77.08	3240	103.18
0.0502	20	0.2	622.70	79.22	840	106.87
0.0506	15	0.5	101.73	79.48	135	105.47
0.0527	11	0.9	34.04	81.05	44	104.76
Pocock						
0.0493	5	0.1	2506.10	79.81	3860	122.93
0.0506	5	0.2	632.20	80.43	970	123.41
0.0510	5	0.5	104.80	81.88	160	125.00
0.0520	3	0.9	35.06	83.48	48	114.29
Sequential Rao						
0.0500		0.1	2392.49	76.19	3260	103.82
0.0480		0.2	619.97	78.88	850	108.14
0.0450		0.5	104.35	81.52	140	109.38
0.0410		0.9	35.83	85.31	46	109.52
Sequential Wald						
0.0490		0.1	2377.36	75.71	3240	103.18
0.0490		0.2	610.89	77.72	840	106.87
0.0510		0.5	98.40	76.88	135	105.47
0.0640		0.9	31.11	74.07	44	104.76

Table 6.2: Results of simulations based on Example 2.1 Sequential Rao and Wald are compared to the results from Pocock and OBF t-tests. Nominal $\alpha = 0.05$, $1 - \beta = .9$, $\theta = |\mu_1 - \mu_2|$, and allocation rule $\lambda = .5$ are used. $E_\theta N$ is the average stopping time for sequential trials when the true treatment difference is θ, $n_0 = n_K$ is the truncation point, K is the number of interim analyses, n_f is the sample size of the fixed, two-sample t-test attaining the same power at the same α-level as the sequential tests. $\hat{\alpha}$ is the empirical size at the given total sample size, n_0.

$\hat{\alpha}$	K	θ	$E_\theta N$	$100(\frac{E_\theta N}{n_f})\%$	n_0	$100(\frac{n_0}{n_f})\%$
O'Brien-Fleming						
0.0499	20	0.1	2946.60	70.09	4380	104.19
0.0504	20	0.2	741.30	70.47	1110	105.51
0.0503	15	0.5	122.90	71.45	180	104.65
0.0512	4	0.9	43.20	80.00	56	103.70
Pocock						
0.0495	5	0.1	2890.30	68.75	5100	121.31
0.0490	5	0.2	723.60	68.78	1280	121.67
0.0515	5	0.5	118.60	68.95	205	119.19
0.0510	3	0.9	40.90	75.74	63	116.67
Sequential Rao						
0.0490	n_0	0.1	2894.90	68.86	4400	104.66
0.0480	n_0	0.2	734.50	69.82	1110	105.51
0.0450	n_0	0.5	122.20	71.05	180	104.65
0.0410	n_0	0.9	40.60	75.19	56	103.70
Sequential Wald						
0.0490		0.1	2885.54	68.64	4380	104.19
0.0490		0.2	730.64	69.45	1110	105.51
0.0510		0.5	118.05	68.63	180	104.65
0.0610		0.9	36.45	67.50	56	103.70

6.2.2 Nonparametric case

In this section, we compare the sequential sign test (Test 5.2, using the sign kernel of Example 5.1) to the two-sample group sequential t-tests of Pocock and O'Brien-Fleming defined in Jennison and Turnbull (2001). We employ a class of location and scale-contaminated normal distributions used by Afifi and Kim (1972). With probability λ, we sample from the mixture population

$$f(x) = (1-\gamma)\phi(x;0,1) + \gamma\phi(x;a,c^2) \tag{6.1}$$

and with probability $(1-\lambda)$ from

$$g(x) = (1-\gamma)\phi(x;\delta\tau,1) + \gamma\phi(x;a+\delta\tau,c^2), \tag{6.2}$$

where $\phi(x;\mu,\sigma^2)$ indicates density of a normal distribution with mean μ, and variance σ^2, $0 \leq \gamma < 1$ is the mixing proportion, and a, c, δ are fixed parameters useful for tuning the skewness and kurtosis of the populations. A nice feature of these mixtures is that the two populations have all moments identical. Hence, the hypotheses $H_0 : F \equiv G$ vs $H_A : F \not\equiv G$ are equivalent to $H_0 : \delta = 0$ vs $H_A : \delta \neq 0$. The common variance of the two populations is given by $\tau^2 = 1 - \gamma + \gamma c^2 + \gamma(1-\gamma)a^2$. (Afifi and Kim 1972) provide formulae for the skewness and kurtosis.

We have considered populations generated from all combinations of the parameters: $c = 1, 2, 3$; $a = 0, .51, 2.56$; $\gamma = 0, .1, .3$ and $\delta = 0, .2, .4$. To aid interpretation, we reported in Table 6.3 the skewness (first row) and kurtosis (second row) of the populations considered. The nominal $\alpha = .05$ and allocation probabilities $\lambda = .5$ and .7 were used. Since our methods hinge

on large sample results, we used truncation points $n_0 = 150, 300$. The group sequential tests had $K = 10$ and $K = 3$ interim analyses, respectively, for the O'Brien-Fleming and Pocock tests. Beyond these numbers of interim analyses, no substantial gain in sample size saving occurred. Indeed, a substantial loss of power occurred in the Pocock method. Each scenario was replicated 10^5 times by using Fortran IMSL random number generators.

Since the results for $n_0 = 300$ were similar, we reported in Table 6.4 results for $n_0 = 150$ only. We can summarise our findings in the following two points:

1. All methods maintained their nominal $\alpha = .05$ very well under all combinations of a, c, γ; that is, non-normality did not matter under H_0 for all total samples considered.

2. Under the alternatives, the sequential sign test is more powerful and saves, on average, more sample than the two group sequential t-tests do when populations are non-normal and remains comparable whenever the populations are perfectly normal (e.g., $c = 1$, $a = 0$ and any γ). To illustrate this phenomenon, we reported in Table 6.4 results for the case where $\delta = .4$, $\lambda = .5$, $n_0 = 150$, with all combinations of c, a, γ given in Table 6.3.

Finally, our simulation results are consistent with those of Afifi and Kim (1972), who compared Wilcoxon's rank-sum test to the two-sample t-test for the fixed-sample design.

Table 6.3: Skewness (first row) and kurtosis (second row) for several combinations of mixture parameters for (6.1) and (6.2).

	c=1		c=2		c=1	
a	$\gamma = .1$	$\gamma = .3$	$\gamma = .1$	$\gamma = .3$	$\gamma = .1$	$\gamma = .3$
0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3.00	3.00	4.44	4.57	8.33	6.49
0.51	0.00	0.00	0.27	0.35	0.45	0.40
	3.00	3.00	4.58	4.59	8.47	6.47
2.56	0.06	0.12	0.84	0.89	1.53	1.28
	3.70	2.58	6.56	4.23	10.30	5.82

Table 6.4: Power (first row) and average stopping time (second row) of the sequential testing procedures using (6.1) and (6.2). Test 2 uses the sign kernel of Example 5.1. Truncation point $n_0 = 150$ and $\lambda = .5$ were used.

	c=1		c=2		c=1	
a	$\gamma = .1$	$\gamma = .3$	$\gamma = .1$	$\gamma = .3$	$\gamma = .1$	$\gamma = .3$
			Sign test			
0.00	0.621	0.625	0.674	0.708	0.779	0.839
	123.5	123.3	120.2	117.9	112.7	107.0
0.51	0.622	0.625	0.678	0.713	0.782	0.841
	123.4	123.3	120.1	117.5	112.3	106.8
2.56	0.688	0.638	0.782	0.784	0.864	0.891
	119.3	122.5	112.4	112.2	104.7	101.2
			O'Brien-Fleming			
0.00	0.660	0.661	0.662	0.665	0.674	0.668
	124.6	124.5	124.1	123.9	122.9	123.3
0.51	0.661	0.662	0.666	0.668	0.673	0.671
	124.6	124.4	124.0	123.9	122.8	123.3
2.56	0.662	0.660	0.671	0.663	0.681	0.670
	124.3	124.5	123.4	124.1	122.0	123.4
			Pocok			
0.00	0.600	0.600	0.607	0.608	0.619	0.616
	120.9	121.0	120.1	120.1	118.6	119.0
0.51	0.602	0.600	0.608	0.607	0.618	0.611
	120.8	120.9	120.2	120.1	118.5	119.3
2.56	0.606	0.599	0.614	0.606	0.628	0.613
	120.4	121.0	119.0	120.3	117.2	119.3

6.2.3 Discussion

The Test 2 procedure of Section 2.2 using Rao's efficient score and Wald's statistic of Example 2.1 are compared, via simulations, to the almost exact group sequential t-tests of Pocock and O'Brien-Fleming. Test 2 using Rao's score showed a very good performance comparable to that of O'Brien-Fleming even at small total sample sizes, and much better performance than Pocock's method. Test 2 using Wald statistic, $W_k^{*(2)}$ showed better performance than all other tests for moderately large truncation points.

We have also seen that the Test 2 procedure of Section 5.2, using the sign kernel of Example 5.1, outperformed both O'Brien-Fleming's and Pocock's two-sample t-tests in terms of power and average stopping times even for moderately large sample sizes under distributions with non-normal shapes.

Moreover, no numerical integration is required in order to obtain the boundaries of this thesis' procedures. Given these advantages, these procedures might be more suitable than the group sequential procedures in many situations including those where continuous monitoring is feasible. However, the user has to be warned from using these methods if the final sample size (truncation point) is small.

6.3 Concluding remarks

In this thesis, the following results were obtained:

R1. A class of parametric sequential testing procedures, based on Rao and Wald type statistics, were derived. These procedures were designed for

testing a two-sample, simple null hypothesis of the form $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the composite two-sided alternative hypothesis $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ when there is a vector of unknown nuisance parameters, $\boldsymbol{\eta}$. The procedures have simple decision boundaries that do not require heavy computations. Values of the boundaries can be found in the literature for many Type I errors and several parameter-of-interest dimensions of practical importance. The test statistics were examined under H_A and their consistency was demonstrated.

R2. The simple extension of the above procedures to the case of multi-sample global hypotheses was indicated. An application of this extension to the one-way ANOVA for comparing three treatments was provided.

R3. Nonparametric two-sample procedures using U-statistics with anti-symmetric kernels were derived. These procedures dealt with the null hypothesis of equality of two distribution functions against a composite alternative. as in R1, decision boundaries and consistency results for the nonparametric test statistics were obtained.

R4. An important Monte Carlo simulation was carried out to compare the most recommendable of our test procedures, namely Test 2 using Rao's efficient score and Wald's statistic, to the popular group sequential t-tests of Pocock (1977) and O'Brien and Fleming (1979) (see also Jennison and Turnbull 2001). The simulation revealed that Test 2 performs better than O'Brien-Fleming's and Pocock's t-tests in terms of average

stopping time, power, and total sample required. Another set of Monte Carlo simulations compared the two group sequential t-tests against our nonparametric Test 2 based on U-statistics with observations from non-normal distribution shapes. These simulations showed that the non-parametric test had substantially higher power and lower stopping time compared to the power and stopping of the group sequential t-tests under deviations from the normality.

The results of the thesis are an incentive and a ground for further research in the following directions:

D1. The methodology can be further extended to the generalized linear models (GLM) adjusting treatment effect for prognostic factors (McCullagh and Nelder 1989). Usually, the GLMs have a vector of parameters, $\boldsymbol{\beta}$, measuring effects of a set of covariates, $\boldsymbol{x}_i$, characterizing observations, on the response. The response comes from exponential family whose mean is linked to $\boldsymbol{x}_i'\boldsymbol{\beta}$ through some function, g. For example, in clinical trials comparing two treatments, one would be interested in the hypothesis of the form $H_0 : \beta_1 = 0$ where, β_1 is the coefficient of the dummy covariate, x_{i1}, representing treatments. In this setup, the rest of the β parameters are nuisance parameters in addition to the scale parameter, σ. Therefore, the two-sample sequential methods of Chapters 2-3 can be reformulated in terms of GLM methods. The latter approach's disadvantage is that it does not incorporate the random allocation mechanism allowed in Chapters 2-3, which may lead to some loss in efficiency.

However, GLM approach would allow us to easily perform multiple comparisons by using the Bonferroni-type adjustment for α.

D2. Further research should look into applying the sequential analysis approach of this thesis to correlated data. This application can be done in two directions: in testing hypotheses about the parameters of the Generalized Estimating Equation (GEE) models for longitudinal data and in testing hypotheses about time series coefficients.

Bibliography

Afifi, A. A. and P. J. Kim (1972). Comparison of some two-sample tests for non-normal alternatives. *Journal of the Royal Statistical Society* **B34**, 448–455.

Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Annals of Mathematical Statistics* **31**, 165–197.

Armitage, P. (1960). *Sequential Medical Trials.* Blackwell.

Armitage, P. and G. Berry (1994). *Statistical Methods in Medical Research.* Blackwell Scientific Publications Ltd (Oxford).

Armitage, P., C. K. McPherson and B. C. Rowe (1969). Repeated significance test on accumulating data. *Journal of the Royal Statistical Society* **A132**, 235–244.

Barber, Stuart and Christopher Jennison (1999). Symmetric tests and confidence intervals for survival probabilities and quantiles of censored survival data. *Biometrics* **55**, 430–436.

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

Barnard, G. A. (1947). Review of A. Wald's sequential analysis. *Journal of the American Statistical Association* **42**, 658–664.

Bartlett, M. S. (1949). The large sample theory of sequential tests. *Proc. Camb. Phil. Soc.* **42**, 239–244.

Betensky, R. A. (1996). An O'Brien-Fleming sequential trial for comparing three treatments. *Annals of Statistics* **24**(4), 1765–1791.

Betensky, R. A. (1998). A boundary crossing probability for the Bessel process. *Advances in Applied Probability* **30**, 807–830.

BHAT (1982). A randomized trial of propranolol in patients with acute myocardial infarction. *Journal of the American Medical Association* **147**, 1707–1714.

Borodin, A. N. and P. Salminen (1996). *Handbook of Brownian Motion - Facts and Formulae.* Birkhauser.

Breslow, N. (1969). On large sample sequential analysis with applications to survivorship data. *Journal of Applied Probability* **6**, 261–274.

Chatterjee, S. K. and P. K. Sen (1972). Nonparametric testing under progressive censoring. *Calcutta Statistical Association Bulletin* **22**, 13–50.

Cox, D. R. (1963). Large sample sequential tests for composite hypotheses. *Sankhya* **A25**, 5–12.

Csörgő, M. and L. Horváth (1993). *Weighted Approximations in Probability and Statistics.* Wiley.

Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics.* Academic Press.

Daniels, H. E. (1996). Approximating the first crossing-time density for a curved boundary. *Bernoulli* **2**, 133–143.

Dehling, H., M. Denker and W. Philipp (1986). A bounded law of iterated logarithm for Hilbert space valued martingales and its application to U-statistics. *Probability Theory and Related Fields* **72**, 111–131.

Delong, David M. (1980). Some asymptotic properties of a progressively censored nonparametric test for multiple regression. *Journal of Multivariate Analysis* **10**, 363–370.

Donaldson, N., R. O. Dillman, J. Wallace and A. Ortiz-Hurtado (2000). Sequential re-analysis of Phase-III clinical trial in non-small cell lung cancer. *European Respiratory Journal* **15**, 821–827.

Einmahl, U. (1987). Strong invariance principles for partial sums of independent random vectors. *Annals of Probability* **15**, 1419–1440.

Einmahl, U. (1989). Extensions of results of Komlós and Tusnády to the multivariate case. *Journal of Multivariate Analysis* **28**, 20–68.

Follmann, D. A., M. A. Prochan and N. L. Geller (1994). Monitoring pairwise comparisons in multi-armed clinical trials. *Biometrics* **50**, 325–336.

Ghosh, B. K. (1970). *Sequential Tests of Hypotheses.* Addison-Wesley.

Gombay, E. (1996). The weighted sequential likelihood ratio. *Canadian Journal of Statistics* **24**, 229–239.

Gombay, E. (1997). The likelihood ratio under noncontiguous alternatives. *Canadian Journal of Statistics* **25**, 417–423.

Gombay, E. (2000*a*). Sequential change-detection with likelihood ratios. *Statistics and Probability Letters* **49**, 195–204.

Gombay, E. (2000*b*). U-statistics for sequential change detection. *Metrika* **54**, 133–145.

Gombay, E. (2002*a*). Parametric sequential tests in the presence of nuisance parameters. *Theory of Stochastic Processes* **8**(24), 107–118.

Gombay, E. (2002*b*). Sequential change-point detection and estimation. *under review.*

Gombay, E. (2002*c*). Sequential testing of composite hypothesis. *Limit Theorems in Probability and Statistics II (I. Berkes, E. Csáki, M. Csörgő, eds.)* pp. 107–125.

Gombay, E. and L. Horváth (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications* **50**, 161–171.

Gu, M. and T. L. Lai (1998). Repeated significance testing with censored rank statistics in interim analysis of clinical trials. *Statistica Sinica* **8**, 411–428.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics* **21**, 671–680.

Itô, K. and H. P. McKean (1965). *Diffusion proccesses and their sample paths.* Academic Press.

Jennison, C. and B. W. Turnbull (1997). Distribution theory of group sequential t, χ^2 and F tests for general linear models. *Sequential Analysis* **16**, 295–317.

Jennison, C. and B. W. Turnbull (1999). *Group Sequential Methods with Applications to Clinical Trials.* Chapman & Hall.

Jennison, C. and B. W. Turnbull (2001). On group sequential tests for data in unequally sized groups and with unkown variance. *Journal of Statistical Planning and Inference* **96**, 263–288.

Jones, D. and J. Whitehead (1979). Sequential forms of the log-rank and modified Wilcoxon tests for censored data (corr: V68 p576). *Biometrika* **66**, 105–114.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons.

Komlós, J., P. Major and G. Tusnády (1975). An approximation of partial sums of independent R.V.'s and the sample DF.1. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **34**, 33–58.

Koroljuk, V. S. and Yu. V. Borovskich (1994). *Theory of U-Statistics.* Kluwer Academic Publishers Group (Dordrecht; Norwell, MA).

Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis.* Duxbury Press.

Lai, T. Z. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica* **11**, 303–408.

Lee, A. J. (1990). *U-Statistics.* Marcel Dekker.

Lehmann, E. L. (2001). *Elements of Large-Sample Theory.* Springer-Verlag.

Majumdar, H. and P. K. Sen (1978). Nonparametric testing for simple regression under progressive censoring with staggered entry and random withdrawal. *Communications in Statistics-Theory and Methods* **A7**, 349–371.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models.* Chapman and Hall.

Miller, R. J. Jr. and P. K. Sen (1972). Weak convergences of U-statistics and Von Mises' differentiable statistical functions. *Annals of Mathematical Statistics* **43**, 31–41.

Murray, S. and A. A. Tsiatis (1995). Sequential methods for comparing years of life saved in the two-sample censored data problem. *Biometrics* **55**, 1085–1092.

Novikov, A., V. Frishling and N. Kordzakhia (1999). Approximations of boundary crossing probabilities for a Brownian motion. *Journal of Applied Probability* **36**, 1019–1030.

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

Pötzelberger, K. and L. Wang (2001). Boundary crossing probability for Brownian motion. *Journal of Applied Probability* **38**(1), 152–164.

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

Ragimov, F. G. (1993). Asymptotic expansion for the distribution of nonlinear boundary crossing time. *Theory of Probability and its Applications* **37**, 560–564.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* John Wiley & Sons.

Robbins, H. (1974). A sequential test for two binomial populations. *Proc. Nat. Acad. Sci. U.S.A* **71**, 4435–4436.

Rushton, S. (1950). On a sequential t-test. *Biometrika* **37**, 326–333.

Rushton, S. (1952). On a two-sided sequential t-test. *Biometrika* **39**, 302–308.

Scheike, T. H. (1992). A boundary-crossing result for Brownian motion. *Journal of Applied Probability* **29**, 448–453.

Sellke, T. and D. Siegmund (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70**, 315–326.

Sen, P. K. (1981). *Sequential nonparametrics: Invariance principles and statistical inferences.* John Wiley.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

Siegmund, D. (1985). *Sequential Analysis: Tests and confidence intervals.* Springer-Verlag.

Siegmund, D. (1993). A sequential clinical trial for comparing three treatments. *Annals of Statistics* **21**, 464–483.

Sinha, A. N. and P. K. Sen (1983). Tests based on empirical processes for progressive censoring schemes with staggered entry and random withdrawal. *Sankhya* **B44**, 1–18.

Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics* **12**, 551–571.

Slud, E. V. and L. J. Wei (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *Journal of the American Statistical Association* **77**, 862–868.

S+SeqTrial (2001). *Software for Design and Analysis of Sequential Trials.* Insightful Corp.

Tsiatis, A. A. (1984). Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association* **77**, 855–861.

Vostrikova, L. J. (1981). Detection of a 'disorder' in a Wiener process. *Theory of Probability and its Applications* **26**, 356–362.

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

Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *Annals of Statistics* **6**, 92–100.

Whitehead, J. (1978). Large sample sequential methods with application to the analysis of 2×2 contingency tables. *Biometrika* **65**, 351–356.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials.* Wiley & Sons.

Whitehead, J. and P. Thomas (1997). A sequential trial of pain killer in arthritis: Issues of multiple comparisons with control and interval-censored data. *Journal of Biopharmaceutical Statistics* **7**, 333–353.