EFFICIENT APPROXIMATE SOLUTIONS TO THE KIEFER-WEISS PROBLEM

Thesis by

Michael David Huffman

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1980

(Submitted August 28, 1979)

Acknowledgments

I am grateful to the California Institute of Technology and its Mathematics Department for the fellowships and research and teaching assistantships which supported me financially and provided valuable teaching experience.

To all my friends and relatives in Albuquerque, Pasadena and around the country, I give many thanks for genuine concern and interest in both my work and my well-being.

My parents and my brother Cary deserve special appreciation. The love and encouragement they have given me for 25 years has certainly provided an important foundation for the completion of my graduate studies.

I wish to express heartfelt gratitude to my thesis advisor, Dr. Gary Lorden, whose incredibly patient guidance and careful tutoring enabled me to do this research.

My deepest appreciation goes to my wife, Margo, who was never-failing in her support and belief in me. No matter how frustrated I became, she was always ready with cheerful and encouraging words. I also thank her for the skillful typing and editing which she provided. With much love I dedicate this thesis to her.

Abstract

The problem is to decide on the basis of repeated independent observations whether θ_0 or θ_1 is the true value of the parameter θ of a Koopman-Darmois family of densities, where $\underline{\theta} < \theta < \overline{\theta}$. The probability of falsely rejecting θ_0 is to be at most α_0, and that of falsely rejecting θ_1, at most α_1. Procedures are studied from the point of view of minimizing the maximum (over θ) expected number of observations required when θ is the true value of the parameter.

Two types of tests are considered. The first, based on the well-known sequential probability ratio test (SPRT), dictates after each observation whether to stop and make a decision, or whether to continue sampling. An explicit method is derived for determining a combination of one-sided SPRT's, known as a 2-SPRT, which minimizes the maximum expected number of observations to within $o((n(\alpha_0,\alpha_1))^{1/2})$ as α_0 and α_1 go to 0, where $n(\alpha_0,\alpha_1)$ is the minimum of the maximum expected sample size, taken over all procedures with error probabilities at most α_0 and α_1. The second test uses several stages of observations, deciding whether to stop or continue only at the end of each stage. A procedure designed to "do what a sequential test would do", while using at most three stages, is defined and shown to minimize the maximum expected number of observations to within $O((n(\alpha_0,\alpha_1))^{1/4}(\log n(\alpha_0,\alpha_1))^{3/2})$ as α_0 and α_1 go to 0.

Finally, using backward induction, optimal procedures were developed on the computer for the case where the mean of an exponential density is tested. Then extensive computer calculations comparing

the proposed 2-SPRT with these optimal procedures show that the 2-SPRT comes within 1% of minimizing the maximum expected sample size over a broad range of error probability and parameter values.

Table of Contents

1. Background and Definition of the Problem

Hypothesis testing has long been one of the most studied statistical areas. As the name implies, two (or perhaps more) hypotheses, H_0 and H_1, are formulated, and then one of these must be chosen as correct. This decision will be based upon the observed values of random variables X_1, X_2, ... which are assumed to be independent and identically distributed with density $f_\theta(x)$ for some value of θ.

In this and the following chapters, the testing problem under consideration will be to determine which probability density, $f_{\theta_0}(x)$ or $f_{\theta_1}(x)$, is true, where $\theta_0 < \theta_1$.

When performing a test there are two errors which can be committed: rejecting θ_0 when it is true (called a type I error), or accepting θ_0 when the alternative θ_1 is true (called a type II error). The probabilities of these errors will be denoted by α_0 and α_1, respectively, so that

$$\alpha_0 = P_{\theta_0}(\text{reject } \theta_0)$$

and

$$\alpha_1 = P_{\theta_1}(\text{reject } \theta_1).$$

Of course, the most desirable test is one which will keep to a minimum the probability of these two types of error. Unfortunately, when the number of observations is given, both probabilities of error cannot be controlled simultaneously. It is customary to assign a bound to the probability of incorrectly rejecting θ_0 when it is true, and to attempt to minimize the other error probability subject to this condition.

Neyman and Pearson, who were the first to introduce the distinction between the two types of error, proposed the following "likelihood ratio test", and established a fundamental lemma bearing their names [13]. For $k = 1,2,\ldots$, the likelihood function of the observations $x_1, x_2, \ldots$ is defined by

$$f_{\theta k} = f_\theta(x_1)\ldots f_\theta(x_k).$$

Assume that a fixed number n of observations will be taken, and choose a constant $c > 0$. If

$$\frac{f_{\theta_1 n}}{f_{\theta_0 n}} > c$$

then θ_0 is rejected; otherwise θ_0 is accepted. Once n has been established, the value of c controls the value of the type I error of the test. The Neyman-Pearson lemma states that among all tests using n observations and satisfying $P_{\theta_0}(\text{reject } \theta_0) \leq \alpha_0$, the likelihood ratio test (with error probability α_0) minimizes α_1.

Hence if the sample size is fixed, the likelihood ratio test provides the best procedure. Significant improvement in the likelihood ratio test is possible, however, if the number of observations is not fixed in advance but is allowed to depend on the observations themselves. Procedures which take samples one at a time until enough information has been accumulated to make a decision are called sequential tests.

These tests based on sequential sampling will be written as the pair $T = (N,D)$. N is called a stopping rule -- it states when the sampling should end, based on the observations taken up to that point. D is the decision rule, indicating which of the hypotheses should be accepted once sampling has stopped.

The earliest of these tests, called a sequential probability ratio test (SPRT), is due to Abraham Wald [15]. Given A_0 and A_1 in (0,1) the stopping rule N is the first n (or ∞ if there is no n) such that the inequality

$$A_1 < \frac{f_{\theta_1 n}}{f_{\theta_0 n}} < \frac{1}{A_0} \tag{1.1}$$

does not hold. The decision rule D rejects θ_0 if the inequality is violated to the right, and rejects θ_1 if it fails to hold on the left.

N is a random variable which depends on the observations, that is, on the true distribution of the X_i's. If the true value of θ is less than θ_0, then the test will quickly violate the left inequality, resulting in a small N. Similarly, if the true value of θ is greater than θ_1, the right-hand inequality will fail after relatively few observations. Should θ lie between θ_0 and θ_1, the SPRT will require more observations in order to make a choice.

$E_\theta N$ will denote the expectation of N when θ is the true parameter value, and represents the average or expected number of observations which will be needed to complete the SPRT if θ is true.

The performance of any sequential test is judged on the basis of its error probabilities and its expected sample sizes. Wald and Wolfowitz [17] established the remarkable property that among all tests--sequential or not--with equal or smaller error probabilities, the SPRT minimizes both $E_{\theta_0}N$ and $E_{\theta_1}N$.

In practice the values of A_0 and A_1 in (1.1) must be chosen so that the resulting SPRT has prescribed error probabilities α_0 and α_1. Wald [15] showed that if α_0' and α_1' are the actual error probabilities of the SPRT using A_0 and A_1, then

$$\alpha_0' \leq A_0(1 - \alpha_1') \quad \text{and} \quad \alpha_1' \leq A_1(1 - \alpha_0') \tag{1.2}$$

Hence choosing $A_0 = \alpha_0$ and $A_1 = \alpha_1$ in (1.1) guarantees that the SPRT will have error probabilities at most α_0 and α_1.

Wald also provided approximations for $E_{\theta_0}N$ and $E_{\theta_1}N$. Define for $n = 1,2,\ldots$

$$Z_n = \log \frac{f_{\theta_1}(X_n)}{f_{\theta_0}(X_n)}$$

and $Y_n = Z_1 + \ldots + Z_n$. By taking logs, (1.1) is equivalent to the inequality

$$\log A_1 < Y_n < \log A_0^{-1}.$$

Since N is a stopping rule,

$$E_{\theta_1}Y_N = (E_{\theta_1}N)E_{\theta_1}Z_1 = (E_{\theta_1}N)I(\theta_1,\theta_0) \tag{1.3}$$

where the first equality is known as Wald's equation and $I(\theta_1,\theta_0)$ is the Kullback-Leibler information number defined by

$$I(\theta,\theta') = E_\theta \log \frac{f_\theta(X_1)}{f_{\theta'}(X_1)} .$$

If the SPRT rejects θ_0, then Y_N is approximately $\log A_0^{-1}$, while if it rejects θ_1, Y_N is near $\log A_1$. Hence

$$E_{\theta_1} Y_N \approx P_{\theta_1}(\text{reject } \theta_1) \log A_1 + P_{\theta_1}(\text{reject } \theta_0) \log A_0^{-1}, \quad (1.4)$$

where $\approx$ indicates that the terms $Y_N - \log A_1$ and $Y_N - \log A_0^{-1}$, called excess over the boundary, have been neglected. If A_0 and A_1 are chosen equal to α_0 and α_1, then (1.3) and (1.4) combine to yield

$$E_{\theta_1} N \approx \frac{\log \alpha_0^{-1}}{I(\theta_1, \theta_0)}$$

for small values of α_0 and α_1. In fact, if $\text{Var}_{\theta_1} Z_1$, the variance of Z_1 under θ_1, is finite, then $E_{\theta_1} N$ is within $O(1)$ of the right-hand side as α_0 and α_1 tend to 0, with a similar expression holding for θ_0.

Even though the SPRT minimizes $E_{\theta_0} N$ and $E_{\theta_1} N$ among all tests with prescribed error probabilities, its performance is unsatisfactory for values of θ between θ_0 and θ_1. In some cases $E_\theta N$ is larger than the number of observations required by a fixed sample size test with the same error probabilities. Much of the development of sequential analysis has been directed toward finding procedures which improve the performance of the SPRT for these parameter values.

Let $\mathcal{T}(\alpha_0,\alpha_1)$ denote the class of all tests (N,D) which have error probabilities at most α_0 and α_1, and define

$$n(\alpha_0,\alpha_1) = \inf\{\sup_\theta E_\theta N \mid (N,D) \in \mathcal{T}(\alpha_0,\alpha_1)\}.$$

The problem of finding a procedure (N', D') which minimizes the maximum expected sample size subject to the error probability constraints α_0 and α_1 -- that is, so that $\sup_\theta E_\theta N' = n(\alpha_0,\alpha_1)$ -- is known as the Kiefer-Weiss problem. No optimal results (in the sense of the optimality property of the SPRT) have been found for this problem.

Kiefer and Weiss [5] proved structure theorems about tests which minimize $E_{\theta_2} N$ for a fixed $\theta = \theta_2$ (this is called the modified Kiefer-Weiss problem). Weiss [18] showed that the Kiefer-Weiss problem reduces to the modified problem in symmetric cases involving normal and binomial distributions. Lai [7] investigated the Wiener process case.

A test (N,D) is customarily judged by its efficiency, which in this context is

$$\frac{n(\alpha_0,\alpha_1)}{\sup_\theta E_\theta N} . \tag{1.5}$$

A procedure is said to be asymptotically efficient if (1.5) tends to 1 as α_0 and α_1 go to 0, and for such tests the rate of approach of (1.5) to 1 is of interest. Thus, finding fairly simple procedures which are not only asymptotically efficient, but have efficiencies close to 1 for practical values of α_0 and α_1, is important.

Anderson [1] studied a class of easily constructable procedures for the symmetric case of testing the mean drift of a Wiener process. In a general context Lorden [8] studied a subclass of Anderson's procedures, related to SPRT's and called 2-SPRT's. Given $\theta_0 < \theta < \theta_1$, and $0 < A_0, A_1 \leq 1$, the stopping rule $M(\theta, A_0, A_1)$ is the smallest n (or ∞ if there is no n) such that

$$\frac{f_{\theta_i n}}{f_{\theta n}} \leq A_i \tag{1.6}$$

for either i = 0 or 1. The decision rule D rejects θ_0 if (1.6) does not hold for i = 1, and rejects θ_1 if it does not hold for i = 0. If (1.6) is true for both values of i, then any fixed rule can be used for deciding between θ_0 and θ_1. A useful alternate way to write the stopping rule is $M(\theta,A_0,A_1) = \min(M_0(\theta,A_0),M_1(\theta,A_1))$ where $M_i(\theta,A_i)$ is the smallest n such that (1.6) holds.

As Lorden pointed out, the method which Wald used to derive (1.2) is applicable to the 2-SPRT and yields

$$\alpha_i \leq A_i \; P_\theta(\text{reject } \theta_i), \quad i = 0,1 \tag{1.7}$$

so that setting $A_i = \alpha_i$ in (1.6) insures error probabilities of at most α_0 and α_1. The main theorem in [8] states that if α_0 and α_1 are the true error probabilities of the 2-SPRT $(M(\theta,A_0,A_1),D)$, then

$$E_\theta M(\theta,A_0,A_1) = \inf\{E_\theta N \mid (N,D) \in \mathcal{T}(\alpha_0,\alpha_1)\} + o(1)$$

as $\alpha_0,\alpha_1 \to 0$ where θ is fixed. Thus, for any fixed θ, the 2-SPRT provides an asymptotic solution to the modified Kiefer-Weiss problem. In the symmetric normal case, where θ is the mean and $\alpha_0 = \alpha_1 = \alpha$, say, the Kiefer-Weiss problem reduces to the modified problem for $\varphi = (\theta_0 + \theta_1)/2$ [18], where only procedures symmetric about φ need be considered. So in this case, the 2-SPRT gives an approximate solution to the Kiefer-Weiss problem. Lorden showed that over a wide range of values of α, θ_0 and θ_1, the 2-SPRT has an efficiency of more than 99.2%.

Hence, setting $A_i = \alpha_i$ so the 2-SPRT's are in $\mathcal{T}(\alpha_0,\alpha_1)$, if $\tilde{\theta}$ can be found so that $E_\theta M(\tilde{\theta},\alpha_0,\alpha_1)$ is nearly maximized at $\theta = \tilde{\theta}$, then the resulting 2-SPRT will be an approximate solution to the Kiefer-Weiss problem. Chapter 2 derives an explicit method for determining $\tilde{\theta}$ as a function of α_0 and α_1, in the context of the Koopman-Darmois family of densities (defined in the next paragraph). Theorem 1 shows that the efficiency of the 2-SPRT so obtained is $1 - o((\log \alpha_0^{-1})^{-1/2})$ as α_0 and α_1 go to 0, subject to the condition that $0 < C_1 < \log \alpha_0/\log \alpha_1 < C_2 < \infty$ for fixed but arbitrary constants C_1 and C_2.

In Chapters 2 and 3 it is assumed that $X_1, X_2, \ldots$ have one of the Koopman-Darmois densities given by

$$f_\theta(x) = \exp(\theta x - b(\theta)), \quad \underline{\theta} < \theta < \overline{\theta},$$

with respect to a non-degenerate σ-finite measure μ. (Common members of this family include the normal, exponential, binomial and Poisson densities.) The function $b(\theta)$ is necessarily convex and infinitely differentiable on $(\underline{\theta}, \overline{\theta})$, and its first two derivatives satisfy $b'(\theta) = E_\theta X$ and $b''(\theta) = \mathrm{Var}_\theta X = \sigma^2(\theta)$ ([6] and [15]). A simple calculation shows that

$$I(\theta,\varphi) = (\theta - \varphi)b'(\theta) - (b(\theta) - b(\varphi)).$$

For $n = 1,2,\ldots$, let $S_n = X_1 + \ldots + X_n$, and define the log-likelihood ratios

$$\ell_i(\theta,n) = \log \frac{f_{\theta n}}{f_{\theta_i n}}$$

$$= (\theta - \theta_i)S_n - n(b(\theta) - b(\theta_i)), \quad i = 0,1.$$

Then (1.6) is equivalent to $\ell_i(\theta,n) \geq \log A_i^{-1}$. The 2-SPRT can be described graphically in the plane of n and S_n. There are two converging lines given by $(\theta - \theta_i)S_n - n(b(\theta) - b(\theta_i)) = \log A_i^{-1}$. Sampling is stopped as soon as the sequence $(1,S_1), (2,S_2), \ldots$ leaves the triangular region bounded by the lines, and the decision depends on which line is crossed.

In practice it is sometimes easier to collect and use data in sets rather than one at a time. In this case it is preferable to use a testing procedure which is based on taking observations in several stages. Typically the maximum number of stages is fixed in advance, and after each set of observations is taken, a decision is made whether to stop or to continue to the next stage. The number of observations in each stage is based on the observations taken up to that point. To design such a multistage test, it is natural to try to imitate the performance of a sequential test by setting up each stage to "do what the best sequential test would do" based on the previous stages. Chapter 3 is concerned with defining a three-stage test, based on the sequential likelihood ratio test (SLRT) [9], in such a way as to achieve good asymptotic performance.

Theorem 2 shows that the procedure defined in the third chapter has an efficiency of $1 - O(\gamma^{-3/4}(\log \gamma)^{3/2})$, where $\gamma = \log \alpha_0^{-1} + \log \alpha_1^{-1}$, indicating that the performance of the three-stage test would be very good for small enough α_0 and α_1. Unfortunately, the test and theory are not refined enough to indicate what should be done to achieve high efficiencies in practical use. In particular, since there is no analog of Lorden's o(1) result for the 2-SPRT, it seems unlikely that at this level of refinement the three-stage procedure would attain efficiencies as high as those of the 2-SPRT.

Chapter 4 describes the results of computer calculations comparing the test proposed in Chapter 2 with actual Kiefer-Weiss solutions in the case of the exponential density. A method of computing the latter was developed, incorporating the well-known backward induction method for computing modified Kiefer-Weiss solutions. It is shown that the 2-SPRT comes within 1% of minimizing the maximum expected sample size over a fairly broad range of error probability and parameter values. In addition, the expected sample sizes under θ_0 and θ_1 are compared with those of the SPRT having the same error probabilities, indicating a relatively insignificant increase in the number of observations required by the 2-SPRT.

2. 2-SPRT's and the Kiefer-Weiss Problem

2.1 Introduction and Summary of Results

For the results in this chapter it suffices to choose $A_i = \alpha_i$ for i = 0, 1, which insures that all the 2-SPRT's under consideration are in $\mathcal{T}(\alpha_0,\alpha_1)$. It will also be assumed that there are fixed but arbitrary constants C_1 and C_2 such that

$$0 < C_1 < \frac{\log \alpha_0}{\log \alpha_1} < C_2 < \infty. \tag{2.1}$$

Let $S_n = X_1 + \dots + X_n$ for n = 1,2,... . In the (n,S_n) plane the boundaries of $M(\theta,\alpha_0,\alpha_1)$, defined by equality in (1.6), are lines given by

$$(\theta - \theta_i)S_n - n(b(\theta) - b(\theta_i)) = \log \alpha_i^{-1}, \quad i=0,1. \tag{2.2}$$

Defining $I_i(\theta) = I(\theta,\theta_i)$ and $a_i(\theta) = (\theta - \theta_i)/I_i(\theta)$, these lines intersect at $(n(\theta),v(\theta))$ where

$$n(\theta) = \left(\frac{\log \alpha_1^{-1}}{I_1(\theta)} a_0(\theta) - \frac{\log \alpha_0^{-1}}{I_0(\theta)} a_1(\theta)\right)(a_0(\theta) - a_1(\theta))^{-1} \tag{2.3}$$

and

$$v(\theta) = n(\theta)b'(\theta) + \left(\frac{\log \alpha_0^{-1}}{I_0(\theta)} - \frac{\log \alpha_1^{-1}}{I_1(\theta)}\right)(a_0(\theta) - a_1(\theta))^{-1}. \tag{2.4}$$

By virtue of the fact that $b(\theta)$ is convex, $I_0(\theta)$ is strictly increasing, $I_1(\theta)$ is strictly decreasing, and both are positive on (θ_0,θ_1). Thus for any θ in (θ_0,θ_1), $a_1(\theta) < 0 < a_0(\theta)$, which implies that $n(\theta)$ is positive. Hence the 2-SPRT is truncated and can take at most $[n(\theta)] + 1$ observations.

Let $\tilde{a}_i$, $\tilde{I}_i$, $\tilde{\sigma}$ and $\tilde{n}$ denote the values of $a_i(\theta)$, $I_i(\theta)$, $\sigma(\theta)$ and $n(\theta)$ for $\theta = \tilde{\theta}$. Also let $\tilde{M} = \min(\tilde{M}_0, \tilde{M}_1)$ represent $M(\tilde{\theta}, \alpha_0, \alpha_1)$. Equations (2.11), (2.14), (2.15) and (2.16) given below define $\tilde{\theta}$ and $\tilde{r}$ so that the following theorem holds.

Theorem 1. If (2.1) is satisfied then as $\alpha_0, \alpha_1 \to 0$

$$\sup_\theta E_\theta \tilde{M} = \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}) \tag{2.5}$$

and

$$n(\alpha_0, \alpha_1) = \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}), \tag{2.6}$$

where $\varphi(\cdot)$ is the standard normal density function. Thus

$$\frac{n(\alpha_0, \alpha_1)}{\sup_\theta E_\theta \tilde{M}} = 1 - o((\log \alpha_0^{-1})^{-1/2}). \tag{2.7}$$

The proof of Theorem 1 consists of establishing relations (2.8) - (2.10) below. (2.5) follows immediately from

$$E_{\tilde{\theta}} \tilde{M} \geq \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r})\tilde{n}^{1/2} - o(\tilde{n}^{1/2}) \tag{2.8}$$

and

$$\sup_\theta E_\theta \tilde{M} \leq \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}), \tag{2.9}$$

while (2.6) follows from these relations together with

$$\inf_{\mathcal{T}(\alpha_0, \alpha_1)} E_{\tilde{\theta}} N \geq E_{\tilde{\theta}} \tilde{M} - O(1). \tag{2.10}$$

(2.7) is an immediate consequence of (2.5) and (2.6).

The key to the argument is to choose $\tilde{\theta}$ so that the supremum over θ of $E_\theta \tilde{M}$ is attained at $\tilde{\theta}$, at least to within $o(\tilde{n}^{1/2})$.

To determine how to choose $\tilde{\theta}$, first define θ^* so that

$$\frac{\log \alpha_0^{-1}}{I_0(\theta^*)} = \frac{\log \alpha_1^{-1}}{I_1(\theta^*)} \tag{2.11}$$

and let n^* be the common value of the two sides (which by (2.3) equals $n(\theta^*)$). The monotonicity of the information numbers implies that θ^* is uniquely determined, and (2.1) implies that θ^* lies in a fixed closed subinterval of (θ_0, θ_1), say $[\varphi_0, \varphi_1]$. As above for $\tilde{M}$, let $M^* = M(\theta^*, \alpha_0, \alpha_1) = \min(M_0^*, M_1^*)$ and define a_i^*, I_i^* and σ^* accordingly.

In the (n, S_n) plane, relation (2.4) shows that the line determined by the points $(n, E_{\theta^*}S_n) = (n, nb'(\theta^*))$ for $n = 1,2,\ldots$ passes through the vertex $(n^*, v(\theta^*))$. So under θ^* the points (n, S_n) will tend to drift toward the vertex. In general, however, $E_\theta M^*$ is not maximized at $\theta = \theta^*$. This is because for $n < n^*$ one of the boundaries will be closer to the line $(n, nb'(\theta^*))$ so that the fluctuations in S_n will cause the 2-SPRT to end too early by going over the closer boundary.

More precisely, essentially the same argument that will be used to show (2.5) can be extended to show that for $\theta = \theta^* + c(n^*)^{-1/2}$ (where c is restricted to any bounded interval)

$$E_\theta M^* = n^* - \sigma^*(n^*)^{1/2} E(\max_{i=0,1}(a_i^*(Z + \sigma^* c))) + o((n^*)^{1/2}), \tag{2.12}$$

where the expectation on the right-hand side is with respect to the standard normal variable Z. Choosing θ to maximize $E_\theta M^*$ to within $o((n^*)^{1/2})$ is then equivalent to finding c to minimize the expectation on the right-hand side. This expectation can be written

$$\int_0^\infty P(\max_{i=0,1}(a_i^*(Z + \sigma^* c)) > t)\, dt$$

$$= \int_0^\infty \left(P(Z > \frac{t}{a_0^*} - \sigma^* c) + P(Z < \frac{t}{a_1^*} - \sigma^* c)\right) dt. \tag{2.13}$$

Using $x\Phi(x) + \varphi(x)$ as a primitive for the standard normal distribution $\Phi(x)$, straightforward integration shows that the integral equals $\sigma^* c a_1^* + (a_0^* - a_1^*)(\sigma^* c\Phi(\sigma^* c) + \varphi(\sigma^* c))$. Differentiating with respect to c shows that the minimum value of (2.13) occurs at $c = r^*/\sigma^*$ where

$$\Phi(r^*) = \frac{a_1^*}{a_1^* - a_0^*} \quad . \tag{2.14}$$

In addition, the value of (2.13) at $c = r^*/\sigma^*$ is given by $(a_0^* - a_1^*)\varphi(r^*)$.

In general, define $r(\theta)$ by the relation

$$\Phi(r(\theta)) = \frac{a_1(\theta)}{a_1(\theta) - a_0(\theta)} \quad ,$$

so that $r^* = r(\theta^*)$.

In Theorem 1 $\tilde{\theta}$ and $\tilde{r}$ are given by

$$\tilde{\theta} = \theta^* + \frac{r^*}{\sigma^*(n^*)^{1/2}} \tag{2.15}$$

and

$$\tilde{r} = r(\tilde{\theta}). \tag{2.16}$$

With this choice of $\tilde{\theta}$, it turns out that the analog of (2.12) using $\tilde{M}$ in place of M^* is extremized by the same choice of c. Thus, $E_\theta \tilde{M}$ is maximized to within $o(\tilde{n}^{1/2})$ by $\theta = \tilde{\theta}$.

It will be assumed in the remainder of the chapter that α_0 and α_1 are small enough so that θ^* and $\tilde{\theta}$ are in $[\varphi_0, \varphi_1]$. This assures that a_i^*, $\tilde{a}_i$, and the information numbers are bounded away from 0 and ∞.

Before continuing to the proofs of (2.8) - (2.10), given in the following sections, several relationships concerning the boundaries of $\widetilde{M}$ will be established.

Let $T_n = S_n - nb'(\widetilde{\theta})$ for $n = 1,2,\ldots$ and let $\widetilde{s} = v(\widetilde{\theta}) - \widetilde{n}b'(\widetilde{\theta})$. Then in the (n,T_n) plane, $(\widetilde{n},\widetilde{s})$ is the vertex of $\widetilde{M}$ and from (2.2) the boundaries are given by the equations

$$U_i(n) = \frac{-n}{\widetilde{a}_i} + \frac{\log \alpha_i^{-1}}{\widetilde{\theta} - \theta_i} \qquad \text{for } i = 0,1.$$

(Sampling is stopped as soon as either $T_n \geq U_0(n)$ or $T_n \leq U_1(n)$, the decision depending on which inequality holds.) These are lines with slope $-\widetilde{a}_i^{-1}$ passing through $(\widetilde{n},\widetilde{s})$ so the above is equivalent to

$$U_i(n) = \frac{\widetilde{n} - n}{\widetilde{a}_i} + \widetilde{s}\ . \tag{2.17}$$

Solving for $\log \alpha_i^{-1}$ using the last two equations yields

$$\log \alpha_i^{-1} = (\widetilde{n} + \widetilde{a}_i\widetilde{s})\widetilde{I}_i \tag{2.18}$$

The final relationship is given by

$$\frac{\widetilde{s}}{\widetilde{\sigma}(\widetilde{n})^{1/2}} = -\widetilde{r} + o(1). \tag{2.19}$$

To show (2.19) it suffices to establish

$$\frac{\widetilde{s}}{\sigma^*(n^*)^{1/2}} = -r^* + o(1), \tag{2.20}$$

since r^* is bounded by virtue of (2.1) and $\widetilde{\theta} - \theta^* \to 0$ ensures $\widetilde{n} \sim n^*$, $\widetilde{r} \sim r^*$ and $\widetilde{\sigma} \sim \sigma^*$. (2.4) and the definitions of $\widetilde{s}$ and n^* give

$$\tilde{s} = n^*\left(\frac{I_0(\theta^*)}{I_0(\tilde{\theta})} - \frac{I_1(\theta^*)}{I_1(\tilde{\theta})}\right)(a_0(\tilde{\theta}) - a_1(\tilde{\theta}))^{-1}$$

$$= n^*\left(\frac{I_1(\tilde{\theta}) - I_1(\theta^*)}{I_1(\tilde{\theta})} - \frac{I_0(\tilde{\theta}) - I_0(\theta^*)}{I_0(\tilde{\theta})}\right)(a_0(\tilde{\theta}) - a_1(\tilde{\theta}))^{-1} .$$

It is easily seen that $I_i'(\theta) = (\theta - \theta_i)\sigma^2(\theta)$, so that expanding $I_i(\tilde{\theta})$ in a Taylor series about θ^* yields

$$I_i(\tilde{\theta}) - I_i(\theta^*) = \frac{(\xi_i - \theta_i)\sigma^2(\xi_i)r^*}{\sigma^*(n^*)^{1/2}}$$

where the ξ_i are numbers between θ^* and $\tilde{\theta}$. Substituting into the above expression for $\tilde{s}$ and dividing by $\sigma^*(n^*)^{1/2}$ gives

$$\frac{s^*}{\sigma^*(n^*)^{1/2}} = \left\{\left(\frac{\sigma^2(\xi_1)(\xi_1 - \theta_1)}{(\sigma^*)^2 I_1(\tilde{\theta})} - \frac{\sigma^2(\xi_0)(\xi_0 - \theta_0)}{(\sigma^*)^2 I_0(\tilde{\theta})}\right)(a_0(\tilde{\theta}) - a_1(\tilde{\theta}))^{-1}\right\} r^*$$

Noting that $\sigma^2(\xi_i)(\sigma^*)^{-2} = 1 + o(1)$ and $(\xi_i - \theta_i)/I_i(\tilde{\theta}) = a_i(\tilde{\theta}) + o(1)$ completes the demonstration of (2.20).

2.2 Proof of (2.10)

As pointed out after (2.16), $\tilde{a}_i$, $\tilde{I}_i$ and $\tilde{\sigma}$ are bounded away from 0 and ∞, so that the following lemma yields (2.10).

Lemma 2.1. If $\theta \in (\theta_0, \theta_1)$, then

$$E_\theta M(\theta,\alpha_0,\alpha_1) - \inf_{\mathcal{T}(\alpha_0,\alpha_1)} E_\theta N \leq \sum_{i=0}^{1} \left(a_i^2(\theta)\sigma^2(\theta) + \frac{\log 2}{I_i(\theta)}\right). \qquad (2.21)$$

Proof. Let M and M_i denote $M(\theta,\alpha_0,\alpha_1)$ and $M_i(\theta,\alpha_i)$ respectively. Let (N,D) be any test in $\mathcal{T}(\alpha_0,\alpha_1)$, and let $\{D = i\}$ be the event that

θ_i is rejected by that test. As in the proof of Theorem 1 in [9] define

$$N_i = \min(M_i,\ N\{D = i\})$$

where $N\{D = i\} = N$ if $D = i$ and ∞ otherwise. Clearly for all θ,

$$M - N \leq \sum_{i=0}^{1} (M_i - N_i). \tag{2.22}$$

By Wald's equation

$$E_\theta M_i = \frac{\log \alpha_i^{-1} + \delta}{I_i(\theta)}$$

where $\delta = E_\theta(\ell_i(\theta, M_i) - \log \alpha_0^{-1})$ is the expected excess over the boundary $\log \alpha_i^{-1}$. By Theorem 1 of [11],

$$\delta \leq \frac{\mathrm{Var}_\theta \ell_i(\theta, 1)}{I_i(\theta)} = \frac{(\theta - \theta_i)^2 \sigma^2(\theta)}{I_i(\theta)} \tag{2.23}$$

which yields

$$E_\theta M_i \leq \frac{\log \alpha_i^{-1}}{I_i(\theta)} + a_i^2(\theta)\sigma^2(\theta). \tag{2.24}$$

To estimate $E_\theta N_i$, the inequality

$$P_{\theta_i}(N_i < \infty) \leq P_{\theta_i}(M_i < \infty) + P_{\theta_i}(D = i) \leq 2\alpha_i$$

combined with Wald's lower bound on the expected sample size of a one-sided SPRT [15] shows

$$E_\theta N_i \geq \frac{\log \alpha_i^{-1} - \log 2}{I_i(\theta)} \quad . \tag{2.25}$$

Taking expectations in (2.22) and using (2.24) and (2.25) shows that (2.21) is true, (N, D) being an arbitrary member of $\mathcal{J}(\alpha_0, \alpha_1)$.

2.3 Proof of (2.8)

Let $m = [\tilde{n} - \tilde{n}^{1/2}\log \tilde{n}]$. The derivation of (2.8) relies mainly on two facts, the first being that with overwhelming probability under $\tilde{\theta}$ the test requires at least m observations, and the second being that once the m observations are taken, the behavior of the remainder of the test is sufficiently predictable by the value of T_m. More specifically, the first claim is given by

$$P_{\tilde{\theta}}(\tilde{M} \leq m) \leq O(\tilde{n}^{-1}) \tag{2.26}$$

and is proven in Lemma 2.4 at the end of this section, while the second is given by the following lemma.

Lemma 2.2. On the event $\{\tilde{M} > m\}$

$$E_{\tilde{\theta}}(\tilde{M}|T_m) \geq \tilde{n} - \max_{i=0,1} (\tilde{a}_i(T_m - \tilde{s})) - o(\tilde{n}^{1/2}). \tag{2.27}$$

Proof. At time m the log-likelihood ratios have values $\ell_i(\tilde{\theta},m) = (\tilde{a}_i T_m + m)\tilde{I}_i$ for $i = 0,1$. If $\tilde{M} > m$ then based on observations $Y_1, Y_2, \ldots,$ where $Y_k = X_{m+k}$, let N_i be the first n such that

$$\ell_i(\tilde{\theta},n) \geq K_i = \log \alpha_i^{-1} - (\tilde{a}_i T_m + m)\tilde{I}_i.$$

Substituting for $\log \alpha_i^{-1}$ according to (2.18) shows

$$K_i = (\tilde{n} - \tilde{a}_i(T_m - \tilde{s}) - m)\tilde{I}_i. \tag{2.28}$$

On $\{\tilde{M} > m\}$,

$$E_{\tilde{\theta}}(\tilde{M}|T_m) = m + E_{\tilde{\theta}}(\min(N_0,N_1)). \tag{2.29}$$

By Lemma 2.3 below there is a constant D such that

$$E_{\theta}(\min(N_0,N_1)) \geq \min_{i=0,1}\left(\frac{K_i}{\tilde{I}_i}\right) - D(\min_{i=0,1}\left(\frac{K_i}{\tilde{I}_i}\right))^{1/2}. \tag{2.30}$$

By (2.28),

$$\min_{i=0,1}\left(\frac{K_i}{\widetilde{I}_i}\right) = \widetilde{n} - m - \max_{i=0,1}(\widetilde{a}_i(T_m - \widetilde{s})), \tag{2.31}$$

which is at most $\widetilde{n}^{1/2}\log \widetilde{n}$. Using (2.30) and (2.31) in (2.29) yields relation (2.27).

The following lemma establishes (2.30) by giving a general lower bound on $E_\theta M(\theta, A_0, A_1)$.

Lemma 2.3. Let $D = \frac{1}{2} \max_{[\varphi_0,\varphi_1]} ((a_0(\theta) - a_1(\theta))^{1/2}\sigma(\theta))$.

For any θ in $[\varphi_0, \varphi_1]$

$$E_\theta M(\theta, A_0, A_1) \geq K - D\, K^{1/2}, \tag{2.32}$$

where $K = \min_{i=0,1}\left(\frac{\log A_i^{-1}}{I_i(\theta)}\right)$.

Proof. As in the proof of inequality (1.4) in [4], define for $n = 1,2,\ldots,$

$$Y_{i,n} = \frac{\ell_i(\theta,n)}{I_i(\theta)} \qquad i = 0,1$$

and $Y_n = Y_{0,n} - Y_{1,n}$.

Clearly $K \leq \max(Y_{0,M}, Y_{1,M})$. Therefore, writing $\max(Y_{0,M}, Y_{1,M}) = \frac{1}{2}(Y_{0,M} + Y_{1,M}) + \frac{1}{2}|Y_M|$ and noting that $E_\theta Y_{i,M} = E_\theta M$ by Wald's equation (since $E_\theta Y_{i,1} = 1$),

$$K \leq E_\theta M + \tfrac{1}{2}E_\theta |Y_M|. \tag{2.33}$$

Also,

$$E_\theta|Y_M| \leq (E_\theta Y_M^{\,2})^{1/2}. \tag{2.34}$$

Since $E_\theta Y_1 = 0$, Wald's second moment equation yields

$$E_\theta(Y_M^{\,2}) = (E_\theta M)\mathrm{Var}_\theta Y_1. \tag{2.35}$$

A simple computation shows that the variance equals $(a_0(\theta)-a_1(\theta))\sigma^2(\theta)$. Combining (2.33) - (2.35) with the definition of D leads to

$$K \leq E_\theta M + D(E_\theta M)^{1/2}$$

from which (2.32) follows easily, proving the lemma.

Lemma 2.2 and the estimate of $P_{\tilde\theta}(\tilde M \leq m)$ in (2.26) give

$$E_{\tilde\theta}\tilde M \geq \tilde n - E_{\tilde\theta} \max_{i=0,1}(\tilde a_i(T_m - \tilde s)) + o(\tilde n^{1/2}).$$

The expectation on the right-hand side can be written

$$\tilde\sigma m^{1/2} \int_0^\infty P_{\tilde\theta}\left(\max_{i=0,1}\left(\tilde a_i\left(\frac{T_m - \tilde s}{\tilde\sigma m^{1/2}}\right)\right) > t\right) dt. \tag{2.36}$$

As in the evaluation of (2.13), the integrand is the sum of the probabilities of the inequality holding for $i = 0$ and for $i = 1$. In the case $i = 0$, for example, this equals

$$P_{\tilde\theta}\left(\frac{T_m}{\tilde\sigma m^{1/2}} > \frac{t}{\tilde a_0} + \frac{\tilde s}{\tilde\sigma m^{1/2}}\right) = P\left(Z > \frac{t}{\tilde a_0} + \frac{\tilde s}{\tilde\sigma m^{1/2}}\right) + o(m^{-1/2}),$$

where Z is standard normal, by virtue of the Berry-Esseen theorem [3] and the fact that $\mathrm{Var}_{\tilde\theta}T_1$ and $E_{\tilde\theta}(|T_1|^3)$ are bounded away from 0 and ∞, respectively.

Since $\tilde s/\tilde\sigma\, \tilde n^{1/2}$--and hence $\tilde s/\tilde\sigma\, m^{1/2}$-- tends to $-\tilde r$, the first term on the right-hand side converges to $P(\tilde a_0(Z+\tilde r) > t)$. Together with a similar result for $i = 1$, this shows that the integrand in (2.36) converges pointwise to $P(\max_{i=0,1} (\tilde a_i(Z + \tilde r)) > t)$. Using Chebyshev's inequality and the boundedness of $\tilde a_i^{-1}$ and $\tilde a_i\tilde s/\tilde\sigma\, m^{1/2}$, the integrands in (2.36) are seen to be bounded above by a function that goes to 0 like t^{-2} as $t \to \infty$. Therefore, by the dominated convergence theorem

$$E_\theta \tilde{M} \geq \tilde{n} - \tilde{\sigma}\, m^{1/2} \,E \max_{i=0,1}(\tilde{a}_i(Z + \tilde{r})) + o(\tilde{n}^{1/2})$$

$$= \tilde{n} - \tilde{\sigma}\, \tilde{n}^{1/2} \,E \max_{i=0,1}(\tilde{a}_i(Z + \tilde{r})) + o(\tilde{n}^{1/2}). \qquad (2.37)$$

The evaluation of (2.13) given in section 2.1 shows that (2.37) is equivalent to (2.8).

To prove (2.8) it remains only to establish (2.26), which is contained in the following lemma (also to be used in the next section).

Lemma 2.4. Let B be a positive constant. Then

$$P_\theta(\tilde{M} \leq m) \leq O(\tilde{n}^{-1}) \qquad (2.38)$$

uniformly for $|\theta - \tilde{\theta}| \leq B\, \tilde{n}^{-1/2}$.

Proof. It suffices to show

$$P_\theta(T_k \geq U_0(k)) = O(\tilde{n}^{-2}), \quad k = 1,\ldots,m \qquad (2.39)$$

uniformly in k and $|\theta - \tilde{\theta}| \leq B\, \tilde{n}^{-1/2}$, with a similar bound for $T_k \leq U_1(k)$, since (2.38) follows by summing over $k = 1,\ldots,m$.

The boundedness of $\tilde{a}_0$ together with (2.17) and (2.19) imply there is a positive constant c such that

$$U_0(k) \geq c\, \tilde{n}^{1/2} \log \tilde{n} \equiv \gamma$$

for all $k \leq m$. For any $t > 0$,

$$P_\theta(T_k \geq \gamma) = P(\exp(t(T_k - \gamma)) \geq 1).$$

Applying Chebyshev's inequality to the right-hand side yields

$$P_\theta(T_k \geq \gamma) \leq \exp(-\gamma t)(E_\theta(\exp(tT_1)))^k$$

$$= \exp(-\gamma t + k(b(t + \theta) - b(\theta) - tb'(\tilde{\theta}))).$$

Using a Taylor expansion of $b(t + \theta)$ about θ, the continuity and boundedness of $b''(\theta)$ imply

$$P_\theta(T_k \geq \gamma) \leq \exp(-\gamma t + kt(b'(\theta) - b'(\tilde{\theta})) + qkt^2)$$

for some $q > 0$. Expanding $b'(\theta)$ about $\tilde{\theta}$ similarly yields a constant $q' > 0$ such that

$$P_\theta(T_k \geq \gamma) \leq \exp(-\gamma t + ktq'B\tilde{n}^{-1/2} + qkt^2).$$

Replacing k by $\tilde{n}$ on the right-hand side and setting $t = 2/(c\,\tilde{n}^{1/2})$ shows

$$P_\theta(T_k \geq U(k)) \leq \tilde{n}^{-2}\exp\left(\frac{4q}{c^2} + \frac{2q'B}{c}\right)$$

for any $k \leq m$, which yields (2.39).

2.4 Proof of (2.9)

The proof of (2.9) is divided into two parts, the first being to show that there is a $B > 0$ such that

$$\sup_{|\theta - \tilde{\theta}| > B\tilde{n}^{-1/2}} E_\theta \tilde{M} \leq \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}). \quad (2.40)$$

Only the case $\theta > \tilde{\theta}$ will be considered, as the case $\theta < \tilde{\theta}$ is similar.

For $B' > 0$ (to be chosen below), there is a $B > 0$ such that $\theta > \tilde{\theta} + B\,\tilde{n}^{-1/2}$ implies

$$E_\theta \ell_0(\tilde{\theta}, 1) \geq \tilde{I}_0 + B'\,\tilde{n}^{-1/2}. \quad (2.41)$$

The same analysis which established (2.23) in Lemma 2.1 shows that the expected excess of $\widetilde{M}_0$ over $\log \alpha_0^{-1}$ is at most

$$\frac{(\theta - \theta_0)^2 \sigma^2(\theta)}{E_\theta \ell_0(\widetilde{\theta}, 1)} . \qquad (2.42)$$

For $\theta > \widetilde{\theta} + B\,\widetilde{n}^{-1/2}$, (2.42) is bounded, so that by (2.41) and Wald's equation

$$E_\theta \widetilde{M} \leq E_\theta \widetilde{M}_0 \leq \frac{\log \alpha_0^{-1}}{\widetilde{I}_0 + B'\widetilde{n}^{-1/2}} + O(1).$$

Replacing $\log \alpha_0^{-1}$ according to (2.18) and using the fact that $\widetilde{s}$ is of order $\widetilde{n}^{1/2}$ yields

$$E_\theta \widetilde{M} \leq \frac{\widetilde{n} + \widetilde{a}_0 \widetilde{s}}{1 + B'\widetilde{n}^{-1/2} \widetilde{I}_0^{-1}} + O(1)$$

$$\leq \widetilde{n} - B'\widetilde{n}^{1/2} \widetilde{I}_0^{-1} + \widetilde{a}_0 \widetilde{s} + O(1).$$

Since $\widetilde{s} \sim -\widetilde{r}\,\widetilde{\sigma}\,\widetilde{n}^{1/2}$ by virtue of (2.19), this last implies

$$E_\theta \widetilde{M} \leq \widetilde{n} - (B'\widetilde{I}_0^{-1} + \widetilde{a}_0 \widetilde{r}\,\widetilde{\sigma})\widetilde{n}^{1/2} + o(\widetilde{n}^{1/2}). \qquad (2.43)$$

Choose B' sufficiently large so that (2.40) follows from (2.43).

For the remainder of this section, $\widetilde{J}(B)$ will denote the interval of θ values given by $|\theta - \widetilde{\theta}| \leq B\,\widetilde{n}^{-1/2}$.

The next lemma parallels Lemma 2.2 by establishing a bound on the conditional expectation of $\widetilde{M}$ given T_m.

Lemma 2.5. On the event $\{\widetilde{M} > m\}$

$$E_\theta(\widetilde{M} | T_m) \leq \widetilde{n} - \max_{i=0,1} (\widetilde{a}_i (T_m - \widetilde{s})) + o(\widetilde{n}^{1/2}) \qquad (2.44)$$

uniformly for θ in $\widetilde{J}(B)$.

Proof. Assume $T_m \geq \tilde{s}$, the proof for $T_m < \tilde{s}$ being similar. Define N_0 and K_0 as in the proof of Lemma 2.2. Under θ the expected excess of N_0 over K_0 is bounded above by (2.42), and is thus bounded uniformly on $\tilde{J}(B)$ (since $E_\theta \ell_0(\tilde{\theta},1)$, being continuous and positive at $\tilde{\theta}$, is eventually bounded below by a positive number for θ in $\tilde{J}(B)$).

Hence

$$E_\theta(\tilde{M}|T_m) \leq m + E_\theta N_0 \leq m + \frac{(\tilde{n} - \tilde{a}_0(T_m - \tilde{s}) - m)I_0(\tilde{\theta})}{E_\theta \ell_0(\tilde{\theta},1)} + O(1) \quad (2.45)$$

uniformly for θ in $\tilde{J}(B)$. Since the ratio of $I_0(\tilde{\theta})$ to $E_\theta \ell_0(\tilde{\theta},1)$ is $1 + O(\tilde{n}^{-1/2})$ uniformly for θ in $\tilde{J}(B)$, and $\tilde{n} - \tilde{a}_0(T_m - \tilde{s}) - m \leq \tilde{n} - m \leq O(\tilde{n}^{1/2}\log \tilde{n})$, (2.45) implies

$$E_\theta(\tilde{M}|T_m) \leq \tilde{n} - \tilde{a}_0(T_m - \tilde{s}) + O(\log \tilde{n})$$

uniformly for θ in $\tilde{J}(B)$, which yields (2.44) for the case $T_m \geq \tilde{s}$, proving the lemma.

From (2.44)

$$E_\theta \tilde{M} \leq \tilde{n} - \tilde{\sigma}\,\tilde{n}^{1/2} E_\theta\Big(\max_{i=0,1}\Big(\tilde{a}_i\Big(\frac{T_m - \tilde{s}}{\tilde{\sigma}\,\tilde{n}^{1/2}}\Big)\Big)1\{\tilde{M} > m\}\Big) + o(\tilde{n}^{1/2}),$$

where $1\{\cdot\}$ denotes the indicator function and the inequality holds uniformly for θ in $\tilde{J}(B)$.

To complete the proof of (2.9) it will suffice to show

$$\begin{aligned} &E_\theta\Big(\max_{i=0,1}\Big(\tilde{a}_i\Big(\frac{T_m - \tilde{s}}{\tilde{\sigma}\,\tilde{n}^{1/2}}\Big)\Big)\ 1\{\tilde{M} > m\}\Big) \\ &\geq \inf_r E\Big(\max_{i=0,1}\big(\tilde{a}_i(Z + r)\big)\Big) - o(1) \end{aligned} \quad (2.46)$$

uniformly for θ in $\tilde{J}(B)$, since the right-hand side is at least

$(\tilde{a}_0 - \tilde{a}_1)\varphi(\tilde{r}) + o(1)$ by the argument evaluating (2.13). To prove (2.46), note that by arguing as in the proof of Lemma 2.2, for $t > 0$

$$P_\theta\left(\max_{i=0,1} \left(\tilde{a}_i\left(\frac{T_m - \tilde{s}}{\tilde{\sigma}\,\tilde{n}^{1/2}}\right)\right) > t\right)$$

$$= P\left(\max_{i=0,1} \left(\tilde{a}_i(Z + m^{1/2}\tilde{\sigma}^{-1/2}E_\theta T_1 + \tilde{r})\right) > t\right) + o(1)$$

uniformly for θ in $\tilde{J}(B)$.

Since $P_\theta(\tilde{M} \leq m) \to 0$ uniformly by Lemma 2.4, the last relation yields for fixed $L > 0$

$$\int_0^L P_\theta\left(\max_{i=0,1} \left(\tilde{a}_i\left(\frac{T_m - \tilde{s}}{\tilde{\sigma}\,\tilde{n}^{1/2}}\right)\right) 1\{\tilde{M} > m\} > t\right) dt$$

$$= \int_0^L P\left(\max_{i=0,1} \left(\tilde{a}_i(Z + m^{1/2}\tilde{\sigma}^{-1/2}E_\theta T_1 + \tilde{r})\right) > t\right) dt + o(1) \tag{2.47}$$

uniformly for θ in $\tilde{J}(B)$.

Because $\tilde{r} + m^{1/2}\tilde{\sigma}^{-1/2}E_\theta T_1 = \tilde{r} + m^{1/2}\tilde{\sigma}^{-1/2}(b'(\theta) - b'(\tilde{\theta}))$ is bounded for θ in $\tilde{J}(B)$, there is a Q such that the integral on the right-hand side of (2.47) is at least

$$\inf_{|r|\leq Q} \int_0^L P\left(\max_{i=0,1} (\tilde{a}_i(Z + r)) > t\right) dt$$

$$\geq \inf_{|r|\leq Q} \int_0^\infty P\left(\max_{i=0,1} (\tilde{a}_i(Z + r)) > t\right) dt \; - \int_L^\infty g(t)\, dt \tag{2.48}$$

where $g(\cdot)$ is an integrable function which can be chosen to dominate the integrands (since the range of r is bounded).

(2.47) and (2.48) establish (2.46) to within the last term in (2.48), which can be made arbitrarily small by choosing L large.

Thus, (2.46) follows and the proof of (2.9) and, hence, Theorem 1 is complete.

2.5 Refinement

The actual error probabilities of the 2-SPRT $\tilde{M}$ can be evaluated asymptotically using the relations

$$P_{\theta_i}(\text{reject } \theta_i) =$$

$$P_{\tilde{\theta}}(\text{reject } \theta_i)\ A_i\ E_{\tilde{\theta}}(\exp(\log \alpha_i^{-1} - \ell_i(\tilde{\theta},\tilde{M}))|\text{reject } \theta_i),$$

$i = 0,1$. Using (2.19) and the limit distribution of T_m, $P_{\tilde{\theta}}(\text{reject } \theta_0)$ is asymptotically $P(Z > -\tilde{r}) = \tilde{a}_1/(\tilde{a}_1 - \tilde{a}_0)$. Since $\ell_0(\tilde{\theta},\tilde{M}) - \log \alpha_0^{-1}$ is the excess over the boundary when θ_0 is rejected, Theorem 5 of [10] then shows for the nonlattice case that

$$P_{\theta_0}(\text{reject } \theta_0) \sim \frac{\tilde{a}_1}{\tilde{a}_1 - \tilde{a}_0} A_0 \frac{L(\tilde{\theta},\theta_0)}{I_0(\tilde{\theta})}$$

where $L(\tilde{\theta},\theta_0)$ is defined in [5]. A similar expression holds for the other error probability.

In practice it seems advisable to use this information in defining the test, so that the actual error probabilities attained will be close to those desired. The following formulation, used for the calculations in Chapter 4, is recommended for practical use. Define

$$A_0(\theta) = \frac{a_1(\theta) - a_0(\theta)}{a_1(\theta)} \alpha_0$$

and

$$A_1(\theta) = \frac{a_0(\theta) - a_1(\theta)}{a_0(\theta)} \alpha_1 \quad .$$

Choose φ^* to satisfy

$$\frac{\log (A_0(\varphi^*))^{-1}}{I_0(\varphi^*)} = \frac{\log (A_1(\varphi^*))^{-1}}{I_1(\varphi^*)} ,$$

and let $n(\varphi^*)$ be the common value of the two sides. Let

$$\tilde{\varphi} = \varphi^* + \frac{r(\varphi^*)}{\sigma(\varphi^*)(n(\varphi^*))^{1/2}}$$

and use the 2-SPRT $\tilde{N} = M(\tilde{\varphi}, A_0(\tilde{\varphi}), A_1(\tilde{\varphi}))$.

Theorem 1 now holds for $\tilde{N}$, with α_0 and α_1 replaced by $A_0(\tilde{\varphi})$ and $A_1(\tilde{\varphi})$, and $\tilde{n}$, $\tilde{\sigma}$, $\tilde{a}_i$ and $\tilde{r}$ determined by $\tilde{\varphi}$. The proof of Theorem 1 goes through nearly unchanged, it being necessary only to modify the derivation of (2.19), using the fact that the ratios $\log A_i(\varphi^*)/\log A_i(\tilde{\varphi})$ tend to 1.

3. Three-Stage Tests and the Kiefer-Weiss Problem

In this chapter a three-stage test designed to imitate a sequential likelihood ratio test (SLRT) is defined so that it minimizes the maximum expected sample size to within $O(\gamma^{1/4} \log \gamma^{-1})$ where γ is defined as $\log \alpha_0^{-1} + \log \alpha_1^{-1}$. It is assumed that condition (2.1) holds as in Chapter 2.

For any (n, S_n), $n = 1,2,\ldots,$ define $\hat{\theta}_n$ to be the solution to $b'(\theta) = S_n/n$ if $b'(\theta_0) < S_n/n < b'(\theta_1)$. If $S_n/n \leq b'(\theta_0)$ define $\hat{\theta}_n = \theta_0$, and if $S_n/n \geq b'(\theta_1)$ define $\hat{\theta}_n = \theta_1$. $\hat{\theta}_n$ is well-defined since $b'(\theta)$ is strictly increasing, and it maximizes the likelihood function $\theta S_n - nb'(\theta)$ on $[\theta_0, \theta_1]$. Note that

$$\ell_i(\hat{\theta}_n, n) = nI_i(\hat{\theta}_n) \quad \text{if } \hat{\theta}_n \in (\theta_0, \theta_1), \tag{3.1}$$

in which case $\hat{\theta}_n$ is the maximum likelihood estimate of θ.

Given A_0 and A_1 in $(0,1)$ the SLRT consists of the stopping time $\hat{N}$ = smallest n (or ∞ if there is no n) such that

$$\ell_i(\hat{\theta}_n, n) \geq \log A_i^{-1} \quad \text{for } i = 0 \text{ or } 1, \tag{3.2}$$

and the decision rule $\hat{D}$, which rejects θ_0 if (3.2) holds only for $i = 0$, and rejects θ_1 if (3.2) holds only for $i = 1$. In the case that both inequalities hold, any fixed rule can be used to decide between θ_0 and θ_1.

SLRT's have been studied by Schwarz [14], Wong [19] and Lorden [9]. Lorden gives a concise summary of the results of Schwarz and Wong in his paper, where he extends Wong's results and shows that the SLRT with error probabilities α_0, α_1 has expected sample sizes exceeding $n(\alpha_0,\alpha_1)$ by at most M loglog $\underline{\alpha}^{-1}$ uniformly in θ, where $\underline{\alpha} = \min(\alpha_0,\alpha_1)$.

To insure that the SLRT has error probabilities at most α_0 and α_1, define $I(\theta) = \min_{i=0,1} I_i(\theta)$ and choose

$$A_i = \alpha_i / D \log \alpha_i^{-1} \quad i = 0,1 \tag{3.3}$$

where D is chosen to satisfy $D/\log D \geq 6(1 + (\min(I(\theta))^{-1})$ as in Theorem 1 of [9].

In this chapter θ^* is defined by the relation

$$\frac{\log A_0^{-1}}{I_0(\theta^*)} = \frac{\log A_1^{-1}}{I_1(\theta^*)} \tag{3.4}$$

and n^* is the common value of the two sides. Let $m^* = [n^*] + 1$. Then the SLRT can take at most m^* observations. To see this, note that if, say, $\hat{\theta}_{m^*} \geq \theta^*$, then

$$\begin{aligned} \ell_0(\hat{\theta}_{m^*}, m^*) &= (\hat{\theta}_{m^*} - \theta_0)S_{m^*} - m^*(b(\hat{\theta}_{m^*}) - b(\theta_0)) \\ &\geq m^* \, I_0(\hat{\theta}_{m^*}) \\ &\geq n^* \, I_0(\theta^*) \\ &= \log A_0^{-1}, \end{aligned}$$

and the case $\hat{\theta}_{m^*} < \theta^*$ is similar. Finally, under condition (2.1), the ratio γ/n^* is bounded away from 0 and ∞.

The three-stage test $(N(A_0,A_1), \hat{D})$ is defined as follows:

(i) Take

$$m = [n^* - (n^*)^{1/2}\log n^*] \tag{3.5}$$

observations and stop if (3.2) holds, deciding as $\hat{D}$ prescribes.

(ii) Otherwise, continue to time

$$\tau = \min_{i=0,1}\left(m^*,\ 1 + \left[\frac{\log A_i^{-1}}{I_i(\hat{\theta}_m)} + (n^*)^{1/4}(\log n^*)^{3/2}\right]\right). \tag{3.6}$$

If (3.2) holds, stop and use $\hat{D}$.

(iii) Otherwise, continue to time m^* and follow $\hat{D}$.

Note that case (iii) arises only if $\tau < m^*$.

By (3.4) and the monotonicity of the information numbers, if $\hat{\theta}_m \geq \theta^*$ then the term inside the greatest integer function in (3.6) is minimized by $i = 0$, and if $\hat{\theta}_m < \theta^*$ it is minimized by $i = 1$. Assume $\theta_1 > \hat{\theta}_m \geq \theta^*$. If $\hat{\theta}_n$ stays close to $\hat{\theta}_m$ for $n = m+1, m+2, \ldots$, then

$$\ell_0(\hat{\theta}_n, n) \approx nI_0(\hat{\theta}_m) \approx n\frac{\log A_0^{-1}}{\tau},$$

so when n reaches τ (3.2) should be satisfied. The extra term of $(n^*)^{1/4}(\log n^*)^{3/2}$ in the definition of τ is designed to insure that the three-stage test will end with high probability at time τ.

Let

$$J(A) = \{\theta \mid |\theta - \theta^*| \leq A\ (n^*)^{-1/2}\log n^*\}.$$

A routine calculation shows that there exists an A small enough so that

$$mI_i(\theta) < n^*I_i(\theta^*)$$

for all θ in $J(A)$, $i = 0,1$. Fixing this A, note that if $\hat{\theta}_m \in J(A) \cap (\theta_0,\theta_1)$ then (3.1) and (3.4) imply

$$\ell_i(\hat{\theta}_m,m) = mI_i(\hat{\theta}_m) < \log A_i^{-1}, \quad i = 0,1,$$

so that $N(A_0,A_1) > m$.

Theorem 2. For the A_i given by (3.3), $(N(A_0,A_1), \hat{D})$ is in $\mathcal{T}(\alpha_0,\alpha_1)$. In addition, if (2.1) is satisfied, then

$$\sup_\theta E_\theta N(A_0,A_1) = n(\alpha_0,\alpha_1) + O(\gamma^{1/4}(\log \gamma)^{3/2}) \qquad (3.7)$$

as $\alpha_0,\alpha_1 \to 0$. Hence

$$\frac{n(\alpha_0,\alpha_1)}{\sup_\theta E_\theta N(A_0,A_1)} = 1 - O(\gamma^{-3/4}(\log \gamma)^{3/2}). \qquad (3.8)$$

Proof. To show that the three-stage procedure is in $\mathcal{T}(\alpha_0,\alpha_1)$, note that in the proof of Theorem 1 of [9], Lorden shows that for the SLRT,

$$P_{\theta_i}(\text{reject } \theta_i) \leq P_{\theta_i}(\ell_i(\hat{\theta}_n,n) \geq \log A_i^{-1} \text{ for some } n \leq m^*)$$

$$\leq \alpha_i \quad \text{for } i = 0,1.$$

Since $(N(A_0,A_1), \hat{D})$ can reject θ_i only if $\ell_i(\hat{\theta}_n,n) \geq \log A_i^{-1}$ for some $n \leq m^*$, its error probabilities are likewise at most α_0 and α_1.

Using Theorem 1, (3.8) follows immediately from (3.7). The proof of (3.7) consists of establishing the following three relations.

$$E_\theta M(\theta,A_0,A_1) - \inf_{\mathcal{T}(\alpha_0,\alpha_1)} E_\theta N \leq O(\log \gamma) \qquad (3.9)$$

uniformly for $\theta \in J(A/2)$.

$$E_\theta N(A_0,A_1) - E_\theta M(\theta,A_0,A_1) \leq O(\gamma^{1/4}(\log \gamma)^{3/2}) \qquad (3.10)$$

uniformly for $\theta \in J(A/2)$.

$$\sup_{\theta \notin J(A/2)} E_\theta N(A_0,A_1) \leq E_{\theta*} N(A_0,A_1) + O(1). \qquad (3.11)$$

(3.9) and (3.10) show that

$$E_\theta N(A_0,A_1) \leq n(\alpha_0,\alpha_1) + O(\gamma^{1/4}(\log \gamma)^{3/2})$$

uniformly for $\theta \in J(A/2)$, and this combines with (3.11) to give (3.7).

To verify (3.9), note that the method of proof in Lemma 2.1 still applies when $M(\theta,\alpha_0,\alpha_1)$ is replaced by $M(\theta,A_0,A_1)$, so in fact the left-hand side of the inequality in (3.9) is at most

$$\sum_{i=0}^{1} (a_i^2(\theta)\sigma^2(\theta) + \frac{\log (1 + D_i \log \alpha_i^{-1})}{I_i(\theta)}) ,$$

which establishes (3.9).

For the proofs of (3.10) and (3.11), first note that the proof of Lemma 2.4 shows that for any $\theta \in (\theta_0,\theta_1)$ and for all $k \leq n^*$,

$$P_\theta(|S_k - kb'(\theta)| \geq c(n^*)^{1/2}\log n^*) \leq (n^*)^{-2} \exp\left(\frac{4q}{c^2}\right), \qquad (3.12)$$

where q is a constant independent of θ (since it depends only on $b''(\theta)$ being continuous and bounded on (θ_0,θ_1)). The same proof also establishes for all k with $m \leq k \leq m^*$ that

$$\begin{aligned} P_\theta(|S_k - S_m - (k - m)b'(\theta)| &\geq c(n^* - m)^{1/2}\log n^*) \\ &= O((n^*)^{-2}) \end{aligned} \qquad (3.13)$$

uniformly in θ on (θ_0,θ_1).

Let

$$E_1 = \{\max_{k \le m} |S_k - kb'(\theta^*)| \le Q_1 (n^*)^{1/2} \log n^*\}$$

and

$$E_2 = \{\max_{m \le k \le m^*} |\hat{\theta}_k - \hat{\theta}_m| \le Q_2 (n^*)^{-3/4} (\log n^*)^{3/2}\}$$

where Q_1 and Q_2 are constants to be determined.

It will next be established that for sufficiently small Q_2,

$$N(A_0, A_1) \le \tau \quad \text{on } E_2. \tag{3.14}$$

Also,

$$1 - P_\theta(E_2) = O((n^*)^{-1}) \tag{3.15}$$

uniformly in θ. It follows that

$$P_\theta(N(A_0, A_1) > \tau) = O((n^*)^{-1}) \tag{3.16}$$

uniformly in θ.

To see (3.14) note that if $N(A_0, A_1) > m$ and $\hat{\theta}_m \ge \theta^*$, then either $\tau = m^*$ (so that (3.14) is trivial) or else

$$\tau \ge \frac{\log A_0^{-1}}{I_0(\hat{\theta}_m)} + (n^*)^{1/4} (\log n^*)^{3/2} \ge \frac{\log A_0^{-1}}{I_0(\hat{\theta}_\tau)}, \tag{3.17}$$

the last inequality holding provided Q_2 is chosen sufficiently small. Arguing as in the verification following (3.4) that the SLRT takes at most m^* observations, (3.17) implies that $\ell_0(\hat{\theta}_\tau, \tau) \ge \log A_0^{-1}$, which yields (3.14). The case where $\hat{\theta}_m < \theta^*$ is similar.

To establish (3.15), first assume $\theta \in [\theta_0', \theta_1']$ where $\underline{\theta} < \theta_0' < \theta_0$, and $\theta_1 < \theta_1' < \bar{\theta}$ are fixed, and note that for $m \le n \le m^*$

$$\frac{S_n}{n} - \frac{S_m}{m} = n^{-1}(S_n - S_m - (n - m)b'(\theta)) + \frac{m - n}{mn}(S_m - mb'(\theta)).$$

Since $b''(\theta)$ is bounded on $[\theta_0', \theta_1']$, (3.13) and (3.12) extend to show that therefore

$$\max_{m \le n \le m^*} \left|\frac{S_n}{n} - \frac{S_m}{m}\right| = O((n^*)^{-3/4}(\log n^*)^{3/2}) + O((n^*)^{-1}(\log n^*)^2)$$
$$= O((n^*)^{-3/4}(\log n^*)^{3/2}) \qquad (3.18)$$

with probability at least $1 - O((n^*)^{-1})$ uniformly for θ in $[\theta_0', \theta_1']$. Now, $\hat{\theta}_n$, which is a function of S_n/n, has bounded derivative on (θ_0, θ_1) since $b''(\theta)$ is bounded away from 0, and is constant for $S_n/n \le \theta_0$ or $S_n/n \ge \theta_1$. Hence (3.18) implies that (3.15) holds uniformly for $\theta \in [\theta_0', \theta_1']$. To show (3.15) holds uniformly for all θ, note that if, say, $\theta > \theta_1'$ then

$$P_\theta(\hat{\theta}_m = \dots = \hat{\theta}_{m^*} = \theta_1)$$
$$\ge P_{\theta_1'}(S_n - nb'(\theta_1') \ge \rho n \quad \text{for } n = m, \dots, m^*)$$

where $\rho = b'(\theta_1') - b'(\theta_1) > 0$. By (3.12) this last probability is at least $1 - O((n^*)^{-1})$. A similar argument for $\theta < \theta_0'$ shows that (3.15) is uniform for all θ.

The derivation of (3.11) can now be completed. The basic idea is that the second stage when θ^* is true will with high probability be at least as large as the second stage when any θ outside of $J(A/2)$ is true.

Let $\theta^-(c)$ and $\theta^+(c)$ denote the left-hand and right-hand endpoints of $J(c)$, respectively. Choose B small enough so that

$$\frac{\log A_0^{-1}}{I_0(\theta^*(A/2))} \leq \min\left(\frac{\log A_0^{-1}}{I_0(\theta^+(B))}, \frac{\log A_1^{-1}}{I_1(\theta^-(B))}\right) \tag{3.19}$$

Suppose $\theta \geq \theta^+(A/2)$. Then by (3.12), the last part of the argument for (3.15), and (3.16), the probability is at least $1 - O((n^*)^{-1})$ that both $\hat{\theta}_m > \theta^+(A/4)$ and $N(A_0,A_1) \leq \tau$. In this case,

$$N(A_0,A_1) \leq \min(m^*, \text{LHS} + (n^*)^{1/4}(\log n^*)^{3/2}), \tag{3.20}$$

where LHS stands for the left-hand side of (3.19). Now if θ^* is true, (3.12) implies that with probability at least $1 - O((n^*)^{-2})$, $\hat{\theta}_m \in J(B)$, in which case

$$N(A_0,A_1) \geq \min(m^*, \text{RHS} + (n^*)^{1/4}(\log n^*)^{3/2}), \tag{3.21}$$

where RHS stands for the right-hand side of (3.19). (3.11) follows from (3.19) - (3.21) and a similar argument for $\theta \leq \theta^-(A/2)$.

To complete the proof of Theorem 2 it remains only to show (3.10). For $\theta \in J(A/2)$ and $k \leq n^*$, $|S_k - kb'(\theta^*)| \leq |S_k - kb'(\theta)| + O((n^*)^{1/2} \log n^*)$, so that by (3.12), $P_\theta(E_1) \geq 1 - O((n^*)^{-1})$ uniformly for $\theta \in J(A/2)$. With (3.15) this shows

$$P_\theta(E_1 \cap E_2) \geq 1 - O((n^*)^{-1} \tag{3.22}$$

uniformly for $\theta \in J(A/2)$. It will be shown that for a Q_1 chosen sufficiently small,

$$M(\theta,A_0,A_1) \geq N(A_0,A_1) - O(\gamma^{1/4}(\log \gamma)^{3/2}) \tag{3.23}$$

on $E_1 \cap E_2$, uniformly for $\theta \in J(A/2)$. Since $N(A_0,A_1) \leq m^*$, relations (3.22) and (3.23) yield (3.10).

It will first be established that for $\theta \in J(A/2)$

$$M(\theta,A_0,A_1) > m \quad \text{and} \quad N(A_0,A_1) > m \quad \text{on } E_1 \cap E_2.$$

If $\theta \in J(A/2)$, then for sufficiently small α_0,α_1, θ is sufficiently close to θ^* so that in the relation

$$\ell_0(\theta,k) = (\theta-\theta_0)(S_k - kb'(\theta^*)) + k\{(\theta-\theta_0)b'(\theta^*) - (b(\theta) - b(\theta_0))\}$$

the bracketed expression is positive. Hence on E_1, for $k \leq m$

$$\ell_0(\theta,k) \leq (\theta-\theta_0)Q_1(n^*)^{1/2}\log n^* + m\{(\theta-\theta_0)b'(\theta^*) - (b(\theta) - b(\theta_0))\}. \tag{3.24}$$

The right-hand side of (3.24) is the value of $\ell_0(\theta,m)$ when $S_m = mb'(\theta^*) + Q_1(n^*)^{1/2}\log n^*$. By the argument following the definition of $J(A)$, it follows that for sufficiently small Q_1 this $\ell_0(\theta,m)$ is less than $\log A_0^{-1}$. Thus on $E_1 \cap E_2$ and for $\theta \in J(A/2)$, $M(\theta,A_0,A_1)$ does not reject θ_0 by time m and, similarly, does not reject θ_1 by time m. In addition, for sufficiently small Q_1, $|S_m - mb'(\theta^*)| \leq Q_1(n^*)^{1/2}\log n^*$ implies $\hat{\theta}_m \in J(A)$, so that $N(A_0,A_1) > m$ also.

Let M denote $M(\theta,A_0,A_1)$. If $M > m^*$ then (3.23) holds trivially. Otherwise, for sufficiently small α_0,α_1, on the event $E_1 \cap E_2$ $\hat{\theta}_M \in (\theta_0,\theta_1)$ and hence either $MI_0(\hat{\theta}_M) \geq \log A_0^{-1}$ or $MI_1(\hat{\theta}_M) \geq \log A_1^{-1}$. Therefore, for all $\theta \in J(A/2)$, on $E_1 \cap E_2$

$$M \geq \min_{i=0,1}\left(\frac{\log A_i^{-1}}{I_i(\hat{\theta}_M)}\right) \geq \min_{i=0,1}\left(\frac{\log A_i^{-1}}{I_i(\hat{\theta}_m)}\right) - O((n^*)^{1/4}(\log n^*)^{3/2}).$$

Using the definition of τ and the fact that $N(A_0,A_1) \leq \tau$ on E_2 gives

$$M \geq \tau - O(\gamma^{1/4}(\log \gamma)^{3/2}) \geq N(A_0,A_1) - O(\gamma^{1/4}(\log \gamma)^{3/2}).$$

Since the lower bounds on M do not depend on θ, this proves (3.23) and concludes the proof of Theorem 2.

4. Comparison of 2-SPRT's with Kiefer-Weiss Solutions

4.1 Summary of Results

Calculations were carried out comparing the 2-SPRT as described in section 2.5 with Kiefer-Weiss solutions, in the case of testing the parameter θ of the exponential density.

$$f_\theta(x) = \theta \exp(-\theta x),\ \theta > 0,\ x \geq 0.$$

In testing $\theta = \theta_0$ against $\theta = \theta_1$, it can be assumed that $\theta_0 = 1$, since that can always be achieved by scaling the X's.

Desired values of α_0 and α_1 were used to define the 2-SPRT $\tilde{N}$ and $\tilde{\omega}$. $E_{\tilde{\omega}}\tilde{N}$, $\sup_\theta E_\theta \tilde{N}$ and the actual error probabilities α_0' and α_1' of the 2-SPRT were computed. Then, as described in section 4.2, the boundaries of the Kiefer-Weiss solution with error probabilities α_0' and α_1' were calculated, along with its operating characteristics. This provided the values of $n(\alpha_0',\alpha_1')$ used to compute the efficiency $n(\alpha_0',\alpha_1')/\sup_\theta E_\theta \tilde{N}$ of the 2-SPRT.

In Figure 1 are the boundaries attained by this process for testing $\theta_0 = 1$ against $\theta_1 = 1.5$ with desired error probabilities of $\alpha_0 = \alpha_1 = .05$. The straight line boundaries are those of the 2-SPRT, which had actual error probabilities of $\alpha_0' = .045$ and $\alpha_1' = .044$. The curved boundaries are those of the corresponding Kiefer-Weiss solution. For convenience these were drawn in the (n, T_n) plane where $T_n = S_n - .8n$, $S_n = X_1 + \ldots + X_n$. To conduct the 2-SPRT, observations are taken and the points $(1,T_1)$, $(2,T_2)$, ... plotted until one of the 2-SPRT boundaries is crossed. The Kiefer-Weiss test is conducted

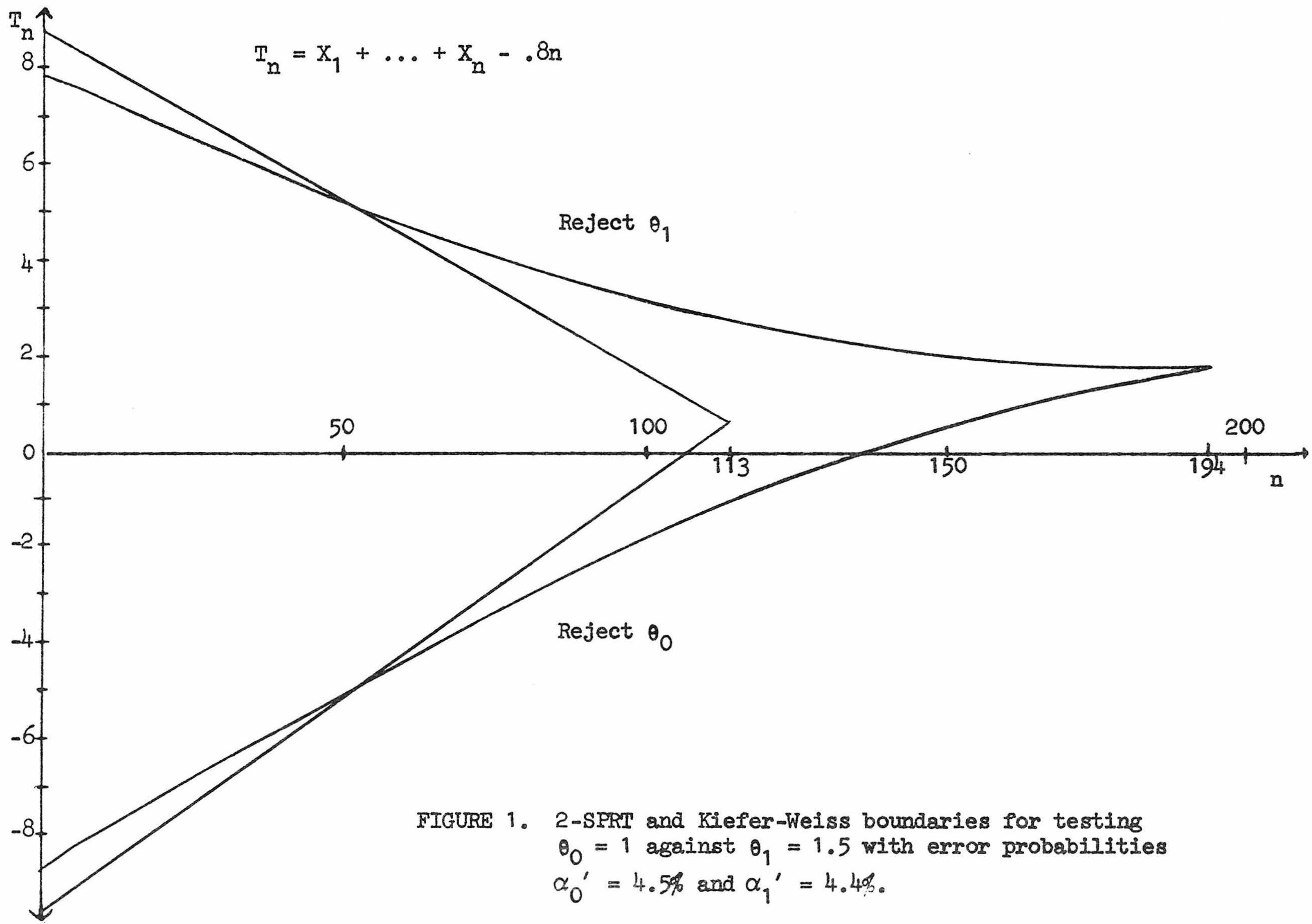

FIGURE 1. 2-SPRT and Kiefer-Weiss boundaries for testing $\theta_0 = 1$ against $\theta_1 = 1.5$ with error probabilities $\alpha_0' = 4.5\%$ and $\alpha_1' = 4.4\%$.

similarly. In both cases, if the top boundary is crossed, then $\theta_1 = 1.5$ is rejected, while if the lower boundary is crossed, $\theta_0 = 1$ is rejected. For this test $\sup_\theta E_\theta \tilde{N} = 51.72$, whereas $n(\alpha_0', \alpha_1') = 51.39$, resulting in an efficiency of 99.3%. A typical feature is that the maximum possible number of observations with the 2-SPRT is much smaller than that of the Kiefer-Weiss solution. The truncation point of the 2-SPRT is at 113 observations, while that of the Kiefer-Weiss solution is at 194 observations.

The most extensive calculations were carried out for the case $\theta_1 = 2$ and are recorded in Table 1. The 2-SPRT is seen to have an efficiency of over 99% over a broad range of desired error probabilities, with both the efficiency and the closeness of the actual error probabilities to the desired ones decreasing as the ratio of α_0 to α_1 becomes extreme. The last column records the values of $E_{\tilde{\varphi}} \tilde{N}$, which are in general within ½% of $\sup_\theta E_\theta \tilde{N}$, indicating that $\tilde{\varphi}$ indeed nearly maximizes $E_\theta \tilde{N}$.

Lorden indicates in [8] that in the symmetric normal case the observed efficiencies depended on the desired error probabilities, but that over a broad range they depended hardly at all on the parameter values. To confirm this for the exponential density, two cases were computed for $\theta_1 = 1.5$. As stated earlier, the $\alpha_0 = \alpha_1 = .05$ case resulted in 99.3% efficiency. The case $\alpha_0 = .1$, $\alpha_1 = .05$ obtained $\alpha_0' = .11$ and $\alpha_1' = .035$ with an efficiency of 99.2%. Both of these efficiencies agree exactly with the corresponding cases for $\theta_1 = 2$.

In addition to the characteristics already mentioned, $E_{\theta_0} \tilde{N}$ and $E_{\theta_1} \tilde{N}$ were computed. For the exponential case Lorden and Eisenberger [12]

TABLE 1

Error probabilities and efficiencies, $\theta_0 = 1$, $\theta_1 = 2$.

α_0	α_1	α_0'	α_1'	$n(\alpha_0',\alpha_1')$	$\sup_\theta E_\theta\tilde{N}$	% efficiency	$E_{\tilde{\varphi}}\tilde{N}$
10	5	9.5	3.3	14.84	14.95	99.2	14.88
5	5	4.1	4.1	18.96	19.08	99.3	19.00
5	1	5.1	.6	25.65	25.83	99.3	25.76
1	5	.7	5.8	26.98	27.24	99.1	27.11
.1	5	.06	8.2	37.02	37.60	98.4	37.35

α_i = desired error probabilities (in %).

α_i' = actual error probabilities attained by 2-SPRT (in %).

$n(\alpha_0',\alpha_1') = \inf\{\sup_\theta E_\theta N\}$ where the inf is taken over all tests with error probabilities at most α_0',α_1'.

efficiency $= n(\alpha_0',\alpha_1')/\sup_\theta E_\theta\tilde{N}$.

$\tilde{N}$, $\tilde{\varphi}$ = 2-SPRT and $\tilde{\varphi}$ as defined in section 2.5.

give an accurate approximation to the expected sample sizes of the SPRT with error probabilities α_0' and α_1'. Typical results are recorded in Table 2, and indicate that the loss in performance of the 2-SPRT is fairly mild.

4.2 Method for Computing Kiefer-Weiss Solutions

Given α_0', α_1', it is desired to find a procedure which attains $n(\alpha_0', \alpha_1')$, i.e. to solve the Kiefer-Weiss problem, which is related to the modified Kiefer-Weiss problem in the following way. If (N,D) solves the modified problem for α_0', α_1' and $\theta = \theta_2$, say, and if in addition $\sup_\theta E_\theta N = E_{\theta_2} N$, then (N,D) solves the Kiefer-Weiss problem. To see this, note that for any (N′,D′) in $\mathcal{T}(\alpha_0', \alpha_1')$,

$$\begin{aligned} \sup_\theta E_\theta N' &\geq E_{\theta_2} N' \\ &\geq E_{\theta_2} N = \sup_\theta E_\theta N. \end{aligned}$$

For any test T = (N,D), let $\alpha_0(T)$ and $\alpha_1(T)$ denote the error probabilities of T. Kiefer and Weiss [5] showed that finding all solutions to the modified problems of minimizing $E_{\theta_2} N$ is equivalent to finding all procedures which, for some positive constants ρ_0, ρ_1, ρ_2 summing to 1, minimize

$$\rho_0 \alpha_0(T) + \rho_1 \alpha_1(T) + \rho_2 E_{\theta_2} N$$

over all tests T. This is known as a Bayes problem and $\rho = (\rho_0, \rho_1, \rho_2)$ is called the prior distribution of θ.

If n observations have been taken and $S_n = s$, the vector $\rho(n,s) = (\rho_0(n,s), \rho_1(n,s), \rho_2(n,s))$ where

TABLE 2

Comparison of expected sample sizes of 2-SPRT and SPRT N, $\theta_0 = 1$, $\theta_1 = 2$.

α_0'	α_1'	$E_{\theta_0}\tilde{N}$	$E_{\theta_0}N$	$E_{\theta_1}\tilde{N}$	$E_{\theta_1}N$
9.5	3.3	10.60	9.93	12.33	11.15
4.1	4.1	11.47	10.36	16.27	15.08
5.1	.6	17.30	16.08	18.21	15.22
.7	5.8	12.33	10.07	24.38	23.13

$$\rho_i(n,s) = \frac{\rho_i\theta_i^{\,n}\exp(-\theta_i s)}{\sum_{j=0}^{2}\rho_j\theta_j^{\,n}\exp(-\theta_j s)}$$

is called the posterior distribution of θ. The Bayes problem is characterized by the following "stationarity" property [16]: Given that n observations have been taken and $S_n = s$, the optimal test proceeds so as to minimize

$$\rho_0(n,s)\alpha_0(T) + \rho_1(n,s)\alpha_1(T) + \rho_2(n,s)E_{\theta_2}N', \qquad (4.1)$$

where N' denotes the number of observations taken from now on. Another way to state this property is that at each (n,s), the optimal procedure either stops or it takes one observation x and then proceeds from $(n + 1, s + x)$ according to the optimal procedure from that point.

Let $R(n,s)$ be the infimum of (4.1) over all tests, $R_1(n,s)$ the infimum of (4.1) over all tests which take at least one observation, and define $R_0(n,s) = \min(\rho_0(n,s),\rho_1(n,s))$. R is known as the Bayes risk, R_1 is called the continuation risk, and R_0 is the stopping risk. As a result of the stationarity property the following relations hold [16]:

$$R(n,s) = \min(R_0(n,s),R_1(n,s)) \qquad (4.2)$$

and

$$R_1(n,s) = \rho_2(n,s) + \sum_{i=0}^{2}\rho_i(n,s)\int_0^\infty R(n+1,s+x)f_{\theta_i}(x)\,dx. \qquad (4.3)$$

The second relation is particularly important. If it is assumed that at least one observation will be taken from (n,s), then

$$\sum_{i=0}^{2}\rho_i(n,s)\,f_{\theta_i}(x) \qquad (4.4)$$

is called the posterior distribution of the next observation X_{n+1}. (4.3) says the continuation risk is the risk (or "cost") incurred by taking the next observation, plus the expectation of the Bayes risk $R(n + 1, s + x)$ under the posterior distribution (4.4).

(4.3) also forms the basis for calculating solutions by "backward" induction [2]. For a given ρ_0, ρ_1, ρ_2, there is an n_0 such that $\rho_2(n_0, s) > \min(\rho_0(n_0, s), \rho_1(n_0, s))$ for all s, so that $R_1(n_0, s) > R_0(n_0, s)$ for all s. Hence the solutions to the Bayes problem can take at most n_0 observations. $R(n_0, s) = R_0(n_0, s)$ is easily calculable for all s, so that numerical integration and (4.3) can be used to compute $R_1(n_0 - 1, s)$ for any s, and (4.2) provides the value of $R(n_0 - 1, s)$. After calculating $R(n_0 - 1, s)$ on a suitable grid of s-values, the induction proceeds backwards to $n_0 - 2$, and so on. For each n, the upper boundary of the test is calculated by finding the s where $R_1(n, s) - \rho_1(n, s)$ changes sign, and the lower boundary by finding the s where $R_1(n, s) - \rho_0(n, s)$ changes sign. Once these boundaries have been located, further numerical integration yields the operating characteristics.

In fact, every set of operating characteristics so obtained provides the error probabilities and expected sample size of some solution to the modified Kiefer-Weiss problem. To see this, assume that from (n,s) the calculations yield α_0^*, α_1^* and m^* for the values of the error probabilities and the expected sample size. These values minimize (4.1), so that for any $T = (N, D)$,

$$\rho_0(n,s)\alpha_0(T) + \rho_1(n,s)\alpha_1(T) + \rho_2(n,s)E_{\theta_2}N$$
$$\geq \rho_0(n,s)\alpha_0{*} + \rho_1(n,s)\alpha_1(T) + \rho_2(n,s)m{*}.$$

Hence if $\alpha_0(T) \leq \alpha_0{*}$ and $\alpha_1(T) \leq \alpha_1{*}$, then $m{*} \leq E_{\theta_2}N$.

For each n in the range of the computations, a Kiefer-Weiss solution is found by obtaining an s_1 between the lower and upper boundaries so that the optimal test starting from (n, s_1) has its maximum expected sample size at $\theta = \theta_2$. If $\alpha_0{*}$ and $\alpha_1{*}$ are the computed error probabilities, then the expected sample size equals $n(\alpha_0{*},\alpha_1{*})$. (It seems reasonable that such an s_1 can be found, because one expects that for s near one boundary, the maximizing θ should be smaller than θ_2, while for s near the other boundary, it should be larger.)

Unfortunately, this routine requires large amounts of computer time and, because the relationship between ρ_0, ρ_1, ρ_2 and the attained error probabilities is unknown, the procedure must be iterated in order to obtain the solution with the desired error probabilities α_0' and α_1'. Not only is the program laborious to write, but it is dependent upon the family of distributions being tested. Thus, calculation of Kiefer-Weiss solutions by this method proves impractical for standard use. The much simpler 2-SPRT appears to provide a more convenient, yet efficient procedure.

References

[1] Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. Ann. Math. Statist. 31, 165-197.

[2] Chow, Y. S., Robbins, H. and Siegmund, D. (1971). Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin, Boston.

[3] Feller, W. (1971). An Introduction to Probability Theory and its Applications, Vol. 2. Wiley, New York.

[4] Hoeffding, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. Ann. Math. Statist. 31, 352-368.

[5] Kiefer, J. and Weiss, L. (1957). Some properties of generalized sequential probability ratio tests. Ann. Math. Statist. 28, 57-75.

[6] Koopman, B. O. (1936). On distributions admitting a sufficient statistic. Trans. Amer. Math. Soc. 39, 399-409.

[7] Lai, T. L. (1973). Optimal stopping and sequential tests which minimize the maximum expected sample size. Ann. Statist. 1, 659-673.

[8] Lorden, G. (1976). 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size. Ann. Statist. 4, 281-291.

[9] Lorden, G. (1972). Likelihood ratio tests for sequential k-decision problems. Ann. Math. Statist. 43, 1412-1427.

[10] Lorden, G. (1977). Nearly-optimal sequential tests for finitely many parameter values. Ann. Statist. 5, 1-21.

[11] Lorden, G. (1970). On excess over the boundary. Ann. Math. Statist. 41, 520-527.

[12] Lorden, G. and Eisenberger, I. (1973). Detection of failure rate increases. Technometrics 15, 167-175.

[13] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. Roy. Soc., Ser. A, 231, 289-337.

[14] Schwarz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. Ann. Math. Statist. 33, 224-236.

[15] Wald, A. (1947). Sequential Analysis. Wiley, New York.

[16] Wald, A. and Wolfowitz, J. (1950). Bayes solutions of sequential problems. Ann. Math. Statist. 21, 82-99.

[17] Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. Ann. Math. Statist. 19, 326-339.

[18] Weiss, L. (1962). On sequential tests which minimize the maximum expected sample size. J. Amer. Statist. Assoc. 57, 551-566.

[19] Wong, S. P. (1968). Asymptotically optimum properties of certain sequential tests. Ann. Math. Statist. 39, 1244-1263.