Optimal Time-Adaptive Repeated Significance Tests

Eric S. Leifer and Eric V. Slud
National Institutes of Health and University of Maryland

June 20, 2002

Abstract: This paper treats the decision theoretic problem of designing optimal repeated significance tests, in the setting of a standard Wiener process with unknown drift. The loss function for testing whether the drift is zero against a specific positive alternative incorporates time-dependent costs of experimentation, alternative-dependent costs of wrong decisions concerning the direction of drift, and opportunity costs for correct, but late, decisions. Type I and II error constraints are imposed and the data are observed at discrete time-instants, as would be the case for a group-sequential, Phase III clinical trial. The procedures are time-adaptive since at each observation time, the time to the next look, or the terminal decision to accept or reject the null hypothesis, can depend on all observed data.

We prove the existence of optimal behavioral (randomized) procedures for a more general action space, one which also includes continuous monitoring procedures. For discrete-look procedures, we prove via backward induction the existence of an optimal procedure which is nonrandomized. Finally, for a hypothetical two-look clinical trial, we demonstrate how a nonrandomized optimal procedure may be feasibly computed by reformulating the two-look problem in terms of Lagrange multipliers and solving the reformulated problem to meet the desired size and power constraints. In particular, we give the resulting stopping and acceptance/rejection boundaries.

Key words and phrases: attainable strength (size and power) of an optimal test; Bayesian decision theory; behavioral decision rule; interval continuation region; group-sequential test.

1 Introduction

This paper concerns the general decision theoretic problem of optimally selecting the observation times of a Wiener process with unknown drift. Throughout, let $W_0(t)$ for $t \geq 0$ denote a standard Wiener process on a probability space $(\Omega, \mathcal{F}, P)$ and $W(t) = W_\theta(t) = W_0(t) + \theta\, t$ (for an unknown real parameter θ) a process (e.g., a repeatedly calculated statistic) which can be observed at a finite set of discrete instants $\tau_1, \tau_2, \ldots, \tau_\nu$ to be chosen adaptively by the statistician, with a view toward inference about the sign of θ. Such inference will be based on adaptively defined stopping, acceptance, and rejection regions for repeated significance tests of the null hypothesis $\theta = 0$ against a specific target alternative $\theta_1 > 0$, subject to type I and II error constraints.

1.1 Motivation

The problem setting described above is a natural abstraction of the group sequential (Phase III) two-sample clinical trial (Pocock 1977, Whitehead 1978, Slud and Wei 1982, Lan and DeMets 1983, Fleming, Harrington, and O'Brien 1984), based on a repeatedly calculated two-sample censored-data rank statistic such as the logrank (Tsiatis 1982, Slud 1984). The Wiener process approximation arises naturally from the large sample asymptotics under information time (Sellke and Siegmund 1983, Slud 1984, Gu and Lai 1992, Tsiatis 1998) for normalized weighted logrank statistics under contiguous alternatives to the null hypothesis of no treatment effect. Several other applications of the Wiener process structure described above are given in the recent book by Jennison and Turnbull (2000).

Our specific motivation for studying decision theoretically a clinical trial inference problem under very general loss structures stems from the unpublished Fisher Lecture of Herbert Robbins (delivered at the 1993 Joint Statistical Meeetings). Robbins followed the earlier formulations by Anscombe (1963) and Colton (1963) of clinical trial inference as a sequential decision problem and found a large mismatch between optimal decision procedures of the Anscombe and Colton formulations and (group) sequential clinical trials in actual practice. Motivated by these authors, we find it very reasonable to view clinical trials as an enterprise in which correct, but late, decisions incur high costs, to be weighed against costs of a wrong decision. A medical paper urging greater generality in loss structures is Ciampi and Till (1980).

In the fully sequential setting with independent and identically distributed observations from a general class of distributions including those with Monotone Likelihood Ratio, Sobel (1953), Brown, Cohen, and Strawderman (1979, 1980), and Brown and Greenshtein (1992) have considered loss functions with alternative-dependent costs of wrong decision. In the Phase II setting with binomial data, Stallard (1998) considered a loss function with costs due to a correct, but late, decision to proceed to a Phase III trial. To our knowledge, no authors have combined costs of wrong decisions and of correct, but late, decision costs in hypothesis testing.

In the normal data setting, several authors have made important computational contributions to decision problems similar to ours. Hald (1975) computed two-look Bayes procedures, subject to type I and II error constraints, that minimized expected trial duration with respect to a prior distribution on alternatives (drift). However, his procedures were restricted so that neither the second look-time nor the rejection critical value would depend on the observed data at the first look. With the restriction that look-times be equally spaced and fixed, and critical values not depend on earlier observed data, Jennison (1987) used a grid search to compute optimal decision procedures in the multi-look setting. Other papers in the group sequential literature (Eales and Jennison 1992, Pampallona and Tsiatis 1994, Chang 1996) optimize expected trial duration over parametric classes of group sequential boundaries. Therneau, Wieand, and Chang (1990) performed analogous computations in the setting where data between looks are independent and binomial.

Our computational implementation, like that of Hald (1975) and Therneau et al. (1990), depends upon the Lagrange multiplier reformulation of constrained decision problems to find optimal procedures satisfying type I and II error constraints. Such an approach is well-described in the books of Ferguson (1967) and Berger (1985), who formulate general problems of sequential decision theory. This approach is also used in the general decision theoretic formulation of a multi-look adaptive design given by Slud (1994).

1.2 Problem Formulation

Let $\tau_1 \geq 0$ be nonrandom, and 0 only if a decision is made in the absence of data; but τ_j for $j \geq 2$ may be random, and for all $j \geq 1$,

$$\tau_j \leq \tau_{j+1} < \infty \quad a.s., \quad \text{and} \quad \tau_{j+1} \in \sigma(W(\tau_i): \; 1 \leq i \leq j). \tag{1}$$

The (random) number ν of looks at the data is defined in terms of $\{\tau_j\}_{j\geq 1}$ by $\nu = \inf\{k \geq 1 : \tau_k = \tau_{k+1}\}$, and is required to be finite.

The statistician is to choose, without knowledge of the parameter θ, $\{\tau_j\}_{j\geq 1}$ together with a measurable function h mapping the observed data $(\{\tau_j\}, \{W(\tau_j)\})$ to $\{0, 1\}$, with the interpretation

$$h(\{\tau_j\}, \{W(\tau_j)\}) = \begin{cases} 1 & \text{means} \quad \text{rejection} \\ 0 & \text{means} \quad \text{acceptance} \end{cases}$$

of the null hypothesis $H_0 : \theta \leq 0$ versus the alternative $H_1 : \theta > 0$.

Thus the *action space* is the set of choices

$$\mathcal{A} = \{ (\{t_j\}_{j\geq 1}, z) : 0 \leq t_j \leq t_{j+1}, \ t_{j+1} - t_j > 0 \text{ finitely often}, \ z = 0, 1\}$$

and the set $\mathcal{D}$ of (nonrandomized) decision procedures is comprised of elements

$$\delta = \big(\{\tau_j\}_{j\geq 1}, \ h(\{\tau_j\}, \{W(\tau_j)\})\big)$$

regarded as measurable functions (subject to the restrictions (1) above) on $(\Omega, \mathcal{F}, P)$. The statistician may, instead of choosing an element $\delta \in \mathcal{D}$, choose a *randomized decision procedure* either by specifying a probability law δ^* in the set $\mathcal{D}^*$ of Borel probability measures on $\mathcal{D}$, or by specifying a *behavioral decision procedure*

$$\tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi) \, .$$

The random elements $\tilde{\delta}$, the set of which will be denoted by $\tilde{\mathcal{D}}$, are defined on the product probability space

$$(\Omega', \mathcal{F}', P') \equiv (\Omega, \mathcal{F}, P) \times ([0,1], \mathcal{B}, \lambda)$$

where $\mathcal{B}$, λ are respectively the Borel sets and Lebesgue measure on the unit interval. As was assumed for the nonrandomized variables above,

$$0 \leq \eta_j \leq \eta_{j+1} < \infty \quad a.s., \quad \chi \in \{0, 1\} \, .$$

We define the number of looks at the data in terms of $\{\eta_j\}_{j\geq 1}$ by

$$\rho = \inf\{k \geq 1 : \ \eta_{k+1} = \eta_k\}. \tag{2}$$

For $k \geq 1$, it follows from (1) that

$$[\rho \geq k],\ \eta_k \in \sigma\left(\mathcal{B}, \{W(\eta_j) :\ 1 \leq j < k\}\right)\ ,\quad [\chi = 1] \in \sigma\left(\mathcal{B}, \rho, \{W(\eta_j) :\ j \leq \rho\}\right).$$

In the foregoing definitions, the expression $W(\eta_j)$ evaluated at $\omega' = (\omega, v) \in \Omega'$ is $W(\eta_j(\omega'), \omega)$, where $W(t) \equiv W(t, \omega)$. The distinction between the variables $(\{\tau_j\}_{j\geq 1}, h(\{\tau_j\}, \{W(\tau_j)\}))$ of $\delta \in \mathcal{D}$ and $(\{\eta_j\}_{j\geq 1}, \chi)$ of $\tilde{\delta} \in \tilde{\mathcal{D}}$ is that the latter component variables can depend not only on the observation history of the process $W(\cdot)$, but also on the variable v reflecting 'auxiliary randomizations' beyond the random generating mechanism of the data. Generally, denote expectations with respect to P' as E'. Also, let

$$S_j \equiv \sqrt{\eta_j - \eta_{j-1}}, \qquad X_j \equiv I_{[S_j > 0]} \frac{W(\eta_j) - W(\eta_{j-1})}{S_j}, \quad U_j \equiv I_{[\eta_1 > 0]} \frac{W(\eta_j)}{\sqrt{\eta_j}}.$$

Therefore, in a natural way a nonrandomized procedure can be regarded as the special case of a behavioral procedure in which η_1 is deterministic; for $k \geq 2$, η_k is a Borel-measurable function of $W(\eta_1), \ldots, W(\eta_{k-1})$ alone; and χ is of the form $h(\{\eta_j\}, \{W(\eta_j)\})$ for a Borel measurable function h. There is a standard construction (Ferguson 1967) which says, for complete separable data spaces, specifying randomized decision procedures as laws δ^* or as behavioral procedures $\tilde{\delta}$ is completely equivalent. For these reasons, we define risk functions below in terms of expectations only for behavioral decision procedures $\tilde{\delta}$.

Let π denote a *prior* probability measure on $\mathbb{R}$, and let $\vartheta : \mathbb{R} \to \mathbb{R}$ be a random variable with distribution π. Define the product probability space

$$(\Omega^\pi,\ \mathcal{F}^\pi,\ P^\pi)\ \equiv\ (\Omega',\ \mathcal{F}',\ P')\ \times\ (\mathbb{R},\ \mathcal{B}(\mathbb{R}),\ \pi)$$

where $\mathcal{B}(\mathbb{R})$ is the collection of Borel sets in $\mathbb{R}$. Generally, denote expectations with respect to P^π as E^π, and conditional probabilities and expectations calculated under a specific fixed value $\vartheta = \theta$ respectively by P_θ and E_θ.

Let $\theta_1 > 0$ be a fixed positive drift which we particularly desire to discriminate from $\theta = 0$. Define the type I and type II error probabilities associated with a behavioral decision procedure $\tilde{\delta} = (\{\eta_j\}_{j\geq 1},\ \chi)$ respectively by

$$\alpha(\tilde{\delta})\ =\ P_0(\chi = 1)\quad ,\qquad \beta(\tilde{\delta})\ =\ P_{\theta_1}(\chi = 0). \tag{3}$$

We refer to the pair $(\alpha(\tilde{\delta}),\ \beta(\tilde{\delta}))$ as the *strength* of the procedure $\tilde{\delta}$.

The final ingredient of a properly specified decision problem is a loss function measuring costs of selecting a particular decision procedure. The losses we consider depend only upon the triple (t, z, θ) consisting of the terminal observation time $t = \eta_\rho$, the terminal decision $z = \chi$, and the unknown parameter θ. The form of the loss function $L(t, z, \theta)$ is assumed to be as follows:

$$L(t, z, \theta) \;=\; \begin{cases} c_1(t, \theta) + z\; c_2(\theta) + (1-z)\; c_3(t, \theta), & \text{if} \quad \theta \le 0; \\ c_1(t, \theta) + (1-z)\; c_2(\theta) + z\; c_3(t, \theta), & \text{if} \quad \theta > 0. \end{cases} \tag{4}$$

The functions c_1, c_2, and c_3 represent, respectively, the costs of trial duration; incorrect terminal decision; and correct, but late, terminal decision. They are assumed to satisfy:

Assumption 1.1. For each $t \ge 0$, the functions $c_j(t, \cdot)$ are π-integrable, for $j = 1, 2, 3$, and there exists a partition of $\mathbb{R}$ into finitely many intervals or singletons $A_1, \ldots, A_k$ such that each c_j is jointly continuous in $(t,\, \theta)$ on each set $[0, \infty) \times A_i$, $i = 1, \ldots, k$.

Assumption 1.2. For each $\theta \in \mathbb{R}$, $c_1(\cdot, \theta)$ and $c_3(\cdot, \theta)$ are nondecreasing functions with $c_1(0, \theta) = c_3(0, \theta) = 0$.

Assumption 1.3. Let $R_0 = \big\{\theta \in \{0,\, \theta_1\} \,:\, \pi(\{\theta\}) = 0\big\}$, and assume

(1) $\liminf_{t\to\infty} \inf_{\theta \notin R_0} \inf_{s \ge 0,\, u \ge s+t} \big(c_1(u, \theta) - c_1(s, \theta) - c_2(\theta)\big) \;=\; \infty;$

(2) for all $\theta_0 \in R_0$, $\exists$ a positive measure ν_{θ_0} supported on $\{\theta \,:\, 0 < |\theta - \theta_0| < \theta_1/4\}$ such that $\int (\theta - \theta_0) \nu_{\theta_0}(d\theta) = 0$ and

$$\forall \text{ Borel sets } G \;, \quad \pi(G) \ge \nu_{\theta_0}(G).$$

Assumption 1.4. $c_3(t, \theta) < \; c_2(\theta)$ for all (t, θ).

The Assumptions are sufficiently general to apply to realistic clinical trial scenarios. In particular, Assumption 1.3 requires costs of long trials to grow without bound for π-almost all θ, and requires π to assign positive mass to neighborhoods of each of $\theta = 0$ and $\theta = \theta_1$.

Define the *risk* for a behavioral procedure $\tilde{\delta}$ (which may, as indicated above, be a nonrandomized procedure) with respect to a particular drift parameter θ as

$$R(\tilde{\delta}, L, \theta) \equiv E_\theta \left(L(\eta_\rho, \chi, \vartheta) \right).$$

The problem addressed in this paper is

> **Problem P.** Fix positive α_0, β_0 with $\alpha_0 + \beta_0 \leq 1$. Find a behavioral $\tilde{\delta}$ or preferably a nonrandomized decision procedure δ to minimize
>
> $$r(\tilde{\delta}, L) \equiv \int_{\mathbb{R}} R(\tilde{\delta}, L, \theta)\, \pi(d\theta) \tag{5}$$
>
> over $\tilde{\mathcal{D}}$, subject to the constraints
>
> $$\alpha(\tilde{\delta}) \leq \alpha_0\,, \quad \beta(\tilde{\delta}) \leq \beta_0. \tag{6}$$

Such a decision procedure will be referred to as *optimal*. The notation of (5) suppresses the dependence of r on π, but not on L, because r is later evaluated at another, related loss function L_{λ_0,λ_1}.

We note that the assumptions $\alpha_0, \beta_0 > 0$ and $\alpha_0 + \beta_0 \leq 1$ in Problem P (and in the sequel) are made without loss of generality. Indeed, $\alpha_0 < 0$ or $\beta_0 < 0$ has no meaning, while $\alpha_0 = 0$ implies $\beta_0 = 1$ in which case the optimal procedure is the one that accepts H_0 without taking any data, making Problem P vacuous. However, it is not immediate that $\alpha_0 + \beta_0 \leq 1$ without loss of generality; this will be shown in Remark 4.2.

1.3 Organization of the paper

The remainder of the paper is organized as follows. In Section 2, we give a Lagrange multiplier reformulation of Problem P, and establish several general lemmas regarding its solution. Section 3, specifically Theorem 3.1, proves the existence of an optimal behavioral procedure for a quite general action space, which accomodates the discrete-look procedures studied in this paper as well as continuous-monitoring procedures. When the number of looks is bounded, Section 4 proves that a nonrandomized optimal procedure can be constructed using backward induction. In Section 5, we give an example of an optimal two-look procedure for a hypothetical clinical trial, and graphically show the stopping and acceptance/rejection boundaries. Next, in Section 6, we discuss conditions under which the continuation region at each look can be proven to be an interval, while in Section 7 we give necessary and sufficient conditions for the type I and II error constraints to be met with equality. Finally, in Section 8, we exhibit features of conditional risk functions arising in numerical calculations of optimal two-look procedures, and discuss their implications.

2 Lagrange multiplier reformulation

This section contains a standard reformulation of the constrained Problem P, by virtue of the Separating Hyperplane Theorem, to an unconstrained Problem L involving Lagrange multipliers. We begin with a reduction allowing us to assume that $\{0\}$ and $\{\theta_1\}$ have positive prior weight.

Remark 2.1. *Without loss of generality,* $\pi_0 \equiv \pi(\{0\})$ and $\pi_1 \equiv \pi(\{\theta_1\})$ are > 0. To establish this assertion, fix $\epsilon \in (0, 1/2)$ arbitrarily and recall the definition of R_0 in Assumption 1.3. Then

$$\text{for} \quad \theta \in R_0,\ i = 1, 2, 3, \quad \text{put} \quad c_i(t, \theta) \equiv 0;$$

for $\theta \notin R_0,\ i = 1, 2, 3,$ replace $c_i(t, \theta)$ with $(1+\epsilon I_{[0\in R_0]}+\epsilon I_{[\theta_1\in R_0]})c_i(t, \theta)$; finally, replace π by

$$(\pi + \epsilon\, I_{[0\in R_0]}\, \mathbf{1}_0 \;+\; \epsilon\, I_{[\theta_1\in R_0]}\, \mathbf{1}_{\theta_1})/(1 + \epsilon(I_{[0\in R_0]} + I_{[\theta_1\in R_0]}))$$

where $\mathbf{1}_u$ denotes a point mass measure at u. It is obvious that under the new definitions, for any $\tilde{\delta} \in \tilde{\mathcal{D}}$, the risk $r(\tilde{\delta}, L)$ is exactly the same as before the changes made in this Remark. Moreover, it is obvious that the changes do not alter the validity of Assumptions 1.1–1.4. □

Assuming from now on that $\pi_0,\, \pi_1 > 0$ by virtue of Remark 2.1, define for arbitrary $\lambda_0,\, \lambda_1 \geq 0$ the auxiliary loss function L_{λ_0,λ_1} by

$$L_{\lambda_0,\lambda_1}(t, z, \theta) \;=\; L(t, z, \theta) \;+\; \frac{\lambda_0}{\pi_0}\, I_{[\theta=0,\, z=1]} \;+\; \frac{\lambda_1}{\pi_1}\, I_{[\theta=\theta_1,\, z=0]}\;. \tag{7}$$

Then, for any $\tilde{\delta} \in \tilde{\mathcal{D}}$,

$$r(\tilde{\delta},\, L_{\lambda_0,\lambda_1}) \;=\; r(\tilde{\delta},\, L) \;+\; \lambda_0\, \alpha(\tilde{\delta}) \;+\; \lambda_1\, \beta(\tilde{\delta}). \tag{8}$$

Proposition 2.1. *The set of probability laws on $\mathcal{D}$ corresponding to elements of $\tilde{\mathcal{D}}$ is convex, as is the set*

$$\mathcal{R}_* \equiv \Big\{ \quad (x,y,w) \in (0,1)\times(0,1)\times(0,\infty): \quad x = \alpha(\tilde{\delta}),\; y = \beta(\tilde{\delta}), \\ w = r(\tilde{\delta}, L), \;\; \textit{for some} \quad \tilde{\delta} \in \tilde{\mathcal{D}} \Big\} .$$

Proof. If $\tilde{\delta}_1$, $\tilde{\delta}_2$ are arbitrary behavioral decision procedures in $\tilde{\mathcal{D}}$, define the procedure $\tilde{\delta}$ in terms of these and an auxiliary *Binomial*$(1,p)$ variable ξ (for any fixed $p \in (0,1)$) independent of $(\tilde{\delta}_1,\ \tilde{\delta}_2,\ W(\cdot),\ V)$, by

$$\tilde{\delta} \;=\; \begin{cases} \tilde{\delta}_1, & \text{if} \quad \xi = 0; \\ \tilde{\delta}_2, & \text{if} \quad \xi = 1. \end{cases}$$

Then $\tilde{\delta}$ is evidently also in $\tilde{\mathcal{D}}$ with probability 1. Thus the set of laws on $\mathcal{D}$ corresponding to elements of $\tilde{\mathcal{D}}$ is convex. Taking expectations over ξ first, the definitions imply that

$$(\alpha(\tilde{\delta}), \beta(\tilde{\delta}), r(\tilde{\delta}, L)) \;=\; (1-p)\,(\alpha(\tilde{\delta}_1), \beta(\tilde{\delta}_1), r(\tilde{\delta}_1, L)) + p\,(\alpha(\tilde{\delta}_2), \beta(\tilde{\delta}_2), r(\tilde{\delta}_2, L))\,.$$

Therefore $\mathcal{R}_*$ is convex. □

A point $(\alpha, \beta, r) \in \overline{\mathcal{R}_*}$ (the closure of $\mathcal{R}_*$) will be called a *lower boundary point* of $\mathcal{R}_*$ if there is no other point $(\alpha_1, \beta_1, r_1) \in \overline{\mathcal{R}_*}$ which lies below (α, β, r), i.e., which satisfies $\alpha_1 \le \alpha,\ \beta_1 \le \beta,\ r_1 \le r$. By Proposition 2.1, $\mathcal{R}_* \subset \mathbb{R}^3$ is convex, and by nonnegativity of the c_i functions, $\mathcal{R}_*$ is also bounded from below.

Remark 2.2. *For $\alpha+\beta < 1$, there is a nonrandomized procedure $\delta_{\alpha,\beta} \in \mathcal{D}$ with $\rho = \nu = 1$ of strength exactly (α, β).* Indeed, the procedure that stops at $\eta_1 = \tau_1 = (z_\alpha + z_\beta)^2/\theta_1^2$ and for which χ is the indicator of the event $[X_1 \ge z_\alpha]$ has strength (α, β). Here z_α is the upper $100\alpha^{th}$ percentile of the standard normal distribution. For $\alpha + \beta = 1$, the randomized procedure that takes no data and rejects H_0 with probability α has strength (α, β). □

Lemma 2.1. *Let (α, β, r) be a lower boundary point of $\mathcal{R}_*$. Then there exist Lagrange multipliers $\lambda_0,\ \lambda_1 \ge 0$ such that the hyperplane of points*

$(x, y, w) \in \mathbb{R}^3$ satisfying $\lambda_0(x-\alpha) + \lambda_1(y-\beta) + w = r$ separates the disjoint convex sets $\mathcal{R}_*$ and

$$\mathcal{U}(\alpha, \beta, r) \;=\; \{(x, y, w) \in \mathbb{R}^3 :\; x \le \alpha,\; y \le \beta,\; w \le r,\; x+y+w < \alpha+\beta+r\}\;.$$

That is, if $(x_1, y_1, w_1) \in \mathcal{R}_*$ and $(x_2, y_2, w_2) \in \mathcal{U}(\alpha, \beta, r)$,

$$\lambda_0\, x_1 \,+\, \lambda_1\, y_1 \,+\, w_1 \,>\, \lambda_0\, x_2 \,+\, \lambda_1\, y_2 \,+\, w_2. \tag{9}$$

Proof. By the Separating Hyperplane Theorem (Berger 1985, p.342, Theorem 13), there exists $(\overline{\lambda}_0, \overline{\lambda}_1, \overline{\lambda}_2) \neq \mathbf{0}$ such that for all $(x_1, y_1, w_1) \in \mathcal{R}_*$ and $(x_2, y_2, w_2) \in \mathcal{U}(\alpha, \beta, r)$,

$$\overline{\lambda}_0 x_1 + \overline{\lambda}_1 y_1 + \overline{\lambda}_2 w_1 > \overline{\lambda}_0 x_2 + \overline{\lambda}_1 y_2 + \overline{\lambda}_2 w_2\;. \tag{10}$$

Since $\mathcal{U}(\alpha, \beta, r)$ is unbounded below, (10) clearly implies $\overline{\lambda}_0,\, \overline{\lambda}_1,\, \overline{\lambda}_2 \ge 0$.

In fact, it must be true that $\overline{\lambda}_2 > 0$. To see this, for $0 < \epsilon < \min\{\alpha, \beta\}$, define the nonrandomized single-look procedure $\delta_1 \equiv \delta_{\alpha-\epsilon, \beta-\epsilon} \in \mathcal{D}_2$ as in Remark 2.2 to achieve strength $(\alpha-\epsilon,\, \beta-\epsilon)$. If $\overline{\lambda}_2 = 0$, then since $\overline{\lambda}_0,\, \overline{\lambda}_1 \ge 0$ with $\overline{\lambda}_0 + \overline{\lambda}_1 > 0$, the triples $(x_1, y_1, w_1) \equiv (\alpha(\delta_1), \beta(\delta_1), r(\delta_1, L)) = (\alpha-\epsilon, \beta-\epsilon, r(\delta_1, L)) \in \mathcal{R}_*$ and $(x_2, y_2, w_2) \equiv (\alpha, \beta, r-\epsilon) \in \mathcal{U}(\alpha, \beta, r)$ contradict (10).

Now (9) follows immediately from (10), with $\lambda_i = \overline{\lambda}_i / \overline{\lambda}_2$ for $i = 0, 1$. □

Corollary 2.1. *If $\tilde{\delta} \in \tilde{\mathcal{D}}$ is a solution of Problem P, then for $(\lambda_0,\, \lambda_1)$ given by Lemma 2.1 for the lower boundary point $(\alpha(\tilde{\delta}), \beta(\tilde{\delta}), r(\tilde{\delta}, L))$, $\tilde{\delta}$ is also a minimizer of $r(\cdot, L_{\lambda_0, \lambda_1})$ over $\tilde{\mathcal{D}}$.*

This Corollary justifies, as a means for finding a solution of Problem P, successively fixing pairs $(\lambda_0,\, \lambda_1)$ and searching for a solution $\tilde{\delta} \in \tilde{\mathcal{D}}$ of

Problem L. For fixed $\lambda_0,\, \lambda_1 \ge 0$, find $\tilde{\delta} \in \tilde{\mathcal{D}}$ minimizing $r(\tilde{\delta},\, L_{\lambda_0, \lambda_1})$

which has the auxiliary property of strength $\le (\alpha_0,\, \beta_0)$, i.e., property (6). Any such $\tilde{\delta}$ automatically solves Problem P. The Lagrange multipliers $\lambda_0,\, \lambda_1$ found in Corollary 2.1 can be further restricted, i.e., bounded below, under some conditions on the cost functions.

Lemma 2.2. *Suppose* $\tilde{\delta} \in \tilde{\mathcal{D}}$ *is a solution to Problem P with* $\alpha(\tilde{\delta})+\beta(\tilde{\delta}) < 1$ *for which*

$$\Delta \equiv r(\tilde{\delta}, L) - \max\Big\{ \int_{[\theta>0]} c_2(\theta)\,\pi(d\theta),\ \int_{[\theta\le 0]} c_2(\theta)\,\pi(d\theta) \Big\} > 0.$$

Then for (λ_0, λ_1) *of Lemma 2.1 for the boundary point* $(\alpha(\tilde{\delta}), \beta(\tilde{\delta}), r(\tilde{\delta}, L))$,

$$\min\{\lambda_0, \lambda_1\} \ge \Delta/(1-\alpha(\tilde{\delta})-\beta(\tilde{\delta})) \ge \Delta.$$

Proof. By Corollary 2.1, $\tilde{\delta}$ minimizes $r(\cdot, L_{\lambda_0,\lambda_1})$ over elements of $\tilde{\mathcal{D}}$. In particular, if δ_r is the nonrandomized procedure which takes no data and always rejects (i.e., has $\chi_r \equiv 1$), and if δ_a takes no data and always accepts $(\chi_a \equiv 0)$, then by Assumption 1.2,

$$r(\tilde{\delta}, L_{\lambda_0,\lambda_1}) \le \min\Big\{ r(\delta_r, L_{\lambda_0,\lambda_1}),\ r(\delta_a, L_{\lambda_0,\lambda_1}) \Big\}$$

$$= \min\Big\{ \lambda_0 + \int_{[\theta\le 0]} c_2(\theta)\,\pi(d\theta),\ \lambda_1 + \int_{[\theta>0]} c_2(\theta)\,\pi(d\theta) \Big\}$$

which implies

$$\lambda_0\alpha(\tilde{\delta}) + \lambda_1\beta(\tilde{\delta}) + \Delta \le \min\{\lambda_0, \lambda_1\}.$$

A little algebra yields the result, with $\alpha(\tilde{\delta}) + \beta(\tilde{\delta}) < 1$ by assumption. □

Remark 2.3. Lemma 2.2 applies in particular when $c_2(\theta) \equiv 0$ and $c_1(t,\theta) + c_3(t,\theta) > 0$ for all $(t,\theta) \in (0,\infty) \times \mathbb{R}$. □

3 Optimal Behavioral Procedures

The main theorem of this section, Theorem 3.1, asserts that all lower boundary points of $\mathcal{R}_*$ correspond to behavioral decision procedures, thus establishing the existence of optimal procedures. This theorem applies to a more general space of procedures than $\tilde{\mathcal{D}}$, and is possible since the loss function (4) depends only on the final look-time and the final binary decision. We begin with a lemma and its corollary, proofs of which are given in Appendix A.1, that show in the determination of a solution of Problem L there is no loss in generality in restricting attention to procedures for which the total trial time η_ρ is essentially bounded.

Lemma 3.1. *There exists a positive constant* $t_{\#} < \infty$ *depending on* (λ_0, λ_1) *such that any solution* $\tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi) \in \tilde{\mathcal{D}}$ *of Problem L satisfies* $\eta_\rho \leq t_{\#}$ *a.s.*

Remark 3.1. The proof of Lemma 3.1 shows that the same value $t_{\#}$ can be used for all (λ_0, λ_1) in a compact set. □

Corollary 3.1. *Let* (α_0, β_0, r_0) *be a lower boundary point of* $\mathcal{R}_*$. *Then there exists a positive constant* $t_{\#} < \infty$ *such that* (α_0, β_0, r_0) *is a limit point of*

$$\Big\{ \quad (x, y, w) \in (0,1) \times (0,1) \times \mathbb{R}^+ : \ x = \alpha(\tilde{\delta}), \ y = \beta(\tilde{\delta}), w = r(\tilde{\delta}, L)$$
$$\textit{for some } \ \tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi) \in \tilde{\mathcal{D}} \ \textit{ with } \ \textit{ess.sup.}\ \eta_\rho < t_{\#} \ \Big\}. \qquad (11)$$

It will be convenient to have the notation $W \equiv \big(W(t),\, 0 \leq t \leq t_{\#}\big)$.

Theorem 3.1. *Define* $\tilde{\mathcal{D}}_{a\#}$ *as the class of behavioral decision procedures* $\tilde{\delta} = (\tau, \chi)$, *in terms of the Uniform*$[0,1]$ *auxiliary randomization variable* V *and* $(\Omega', \mathcal{F}', P')$ *as above, where* $t_{\#} < \infty$ *is a fixed constant,* $\tau \in [0, t_{\#}]$ *is any stopping-time with respect to the filtration* $\sigma(V, \mathcal{F}^W_t)$, *and* $\chi \in \{0,\, 1\}$ *is an arbitrary binary-valued random variable on* Ω' *measurable with respect to* $\sigma(V, \mathcal{F}^W_\tau)$. *Let* π, L *be as above, and define* r, $\mathcal{R}_*$ *analogously as above, but with respect to the augmented class* $\tilde{\mathcal{D}}_{a\#}$ *of decision procedures. Then* $\mathcal{R}_*$ *contains its lower boundary points, that is, if* $(\alpha_0,\, \beta_0,\, r_0)$ *is a lower boundary point of* $\mathcal{R}_*$, *then there exists* $\tilde{\delta} = (\tau, \chi) \in \tilde{\mathcal{D}}_{a\#}$ *of strength* $(\alpha_0,\, \beta_0)$ *such that* $r(\tilde{\delta},\, L) = r_0$, *hence* $\tilde{\delta}$ *is optimal.*

Proof. Let $(\alpha_0,\, \beta_0,\, r_0)$ be a lower boundary point of $\mathcal{R}_*$. Then by hypothesis, there exists a sequence $\{\tilde{\delta}_n\}_{n=1}^\infty = \{(\tau_n,\, \chi_n)\}_{n=1}^\infty$ in $\tilde{\mathcal{D}}_{a\#}$ with $\tau_n \leq t_{\#}$ a.s., such that

$$\lim_{n\to\infty} (\alpha(\tilde{\delta}_n),\, \beta(\tilde{\delta}_n),\, r(\tilde{\delta}_n, L)) \ = \ (\alpha_0,\, \beta_0,\, r_0). \qquad (12)$$

The idea of the proof is to establish that a subsequence of the random elements defining the decision procedures $\tilde{\delta}_n$ must have a distributional limit corresponding to a decision procedure $\tilde{\delta}_0 = (\tau_0, \chi_0)$ such that

$$\tau_0 \text{ is a stopping time with respect to the filtration } \sigma(V, \mathcal{F}_t^W); \tag{13}$$

$$\chi_0 \text{ is measurable with respect to } \sigma(V, \mathcal{F}_{\tau_0}^W); \tag{14}$$

$$(\alpha(\tilde{\delta}_0), \beta(\tilde{\delta}_0), r(\tilde{\delta}_0, L)) = (\alpha_0, \beta_0, r_0). \tag{15}$$

The distributions of the random elements $(\tau_n,\, \chi_n,\, W(\tau_n))$ are *tight* under P^π. Indeed, $(\tau_n,\, \chi_n)$ are compactly supported by definition, while for nonrandom $K > 0$,

$$\begin{aligned} P^\pi(|W(\tau_n)| \geq K) \;&=\; \int P_\theta(|W(\tau_n)| \geq K)\, \pi(d\theta) \\ &\leq\; \pi\Big([-\frac{K}{2t_\#},\, \frac{K}{2t_\#}]^c\Big) \;+\; 2\,\Big(1 \,-\, \Phi\Big(\frac{K}{2\sqrt{t_\#}}\Big)\Big), \end{aligned}$$

where the last inequality is a standard estimate for the supremum of a Wiener process on an interval, based on the Reflection Principle (Billingsley 1968, p.71). Consequently, by Helly's Theorem (Billingsley 1968, Theorem 6.1), there is an infinite subsequence of integers n along which the P^π convergence in distribution holds:

$$(\vartheta, \tau_n,\, \chi_n,\, W,\, V) \;\stackrel{d}{\longrightarrow}\; (\vartheta, \tau_0,\, \chi_0,\, W,\, V)\,. \tag{16}$$

Throughout the rest of this proof, let sequence elements with subscripts n be understood to include only subscripts in the subsequence just chosen.

Properties (13) and (14) follow since (τ_n, χ_n) have the corresponding property with respect to (V, W). Indeed, since (τ_n, χ_n, W, V) converge in distribution to (τ_0, χ_0, W, V) , the Skorohod embedding theorem for separable metric spaces (van der Vaart and Wellner 1996, p. 58) implies that on some other probability space, random-element quadruples $(\tau_n', \chi_n', W', V')$, for $n \geq 1$, as well as $(\tau_0', \chi_0', W', V')$ can be defined simultaneously such that: $(\tau_n', \chi_n', W', V')$ has the same distribution as (τ_n, χ_n, W, V); $(\tau_0', \chi_0', W', V')$ has the same distribution as (τ_0, χ_0, W, V); and almost surely, $(\tau_n', \chi_n', W', V')$ $\rightarrow (\tau_0', \chi_0', W', V')$. The $\mathcal{F}_t^W$ stopping-time property (13) of τ_n implies the corresponding $\mathcal{F}_t^{W'}$ property for τ_n', and, as an a.s. limit of $\sigma(V', \mathcal{F}_t^{W'})$ stopping times, τ_0' is also a stopping time (Lipster and Shiryaev 1977, vol. 1, Lemma 1.4); note that we assume without loss of generality that all σ-algebras under discussion are complete. Thus, τ_0 satisfies (13). Similarly, χ_0 satisfies the measurability property (14).

For property (15), note first that for each set A_i from Assumption 1.1 with $\pi(A_i) > 0$, we may define a probability measure P^{A_i} on $(\Omega', \mathcal{F}', P') \times (A_i, \mathcal{B}(A_i), \pi)$, where $\mathcal{B}(A_i)$ is the collection of Borel subsets of A_i, by

$$P^{A_i}(B) \equiv \frac{P^{\pi}\big(B \cap (\Omega' \times A_i)\big)}{\pi(A_i)} \qquad \text{for } B \in \mathcal{F}' \times \mathcal{B}(A_i). \tag{17}$$

Then the same convergence in distribution in (16) (along the same subsequence of n) takes place with respect to P^{A_i}. Also, since π has atoms at 0 and θ_1, following Remark 2.1, we may similarly define measures P^0 and P^{θ_1} with $\{0\}$, respectively $\{\theta_1\}$, in place of A_i in (17), for which property (16) persists. Let E^{A_i}, E^0, and E^{θ_1} be the corresponding expectations.

Now, to see $\alpha(\tilde{\delta}_0) = \alpha_0$, note that the function z is a bounded, continuous function of $z \in \{0, 1\}$. Thus, using the first component of (12),

$$\alpha_0 = \lim_n \alpha(\tilde{\delta}_n) \equiv \lim_n E_0(\chi_n) = \lim_n E^0(\chi_n) = E^0(\chi_0) = E_0(\chi_0) \equiv \alpha(\tilde{\delta}_0).$$

We similarly deduce $\beta(\tilde{\delta}_0) = \beta_0$ using the function $1 - z$ which is bounded and continuous in $z \in \{0, 1\}$.

Finally, we show $r(\tilde{\delta}_0, L) = r_0$. Recall by the third component of (12), as well as the definitions of $r(\tilde{\delta}_n, L)$ and $r(\tilde{\delta}_0, L)$,

$$r_0 = \lim_n r(\tilde{\delta}_n, L) = \lim_n E^{\pi}\big(L(\tau_n, \chi_n, \vartheta)\big) \text{ and } r(\tilde{\delta}_0, L) = E^{\pi}\big(L(\tau_0, \chi_0, \vartheta)\big). \tag{18}$$

Moreover, since $\tau_n \le t_\#$ *a.s.*, $\{L(\tau_n, \chi_n, \vartheta)\}$ is uniformly integrable since by Assumption 1.2, $0 \le L(\tau_n, \chi_n, \theta) \le c_1(t_\#, \theta) + c_2(\theta) + c_3(t_\#, \theta)$, and $c_1(t_\#, \theta) + c_2(\theta) + c_3(t_\#, \theta)$ is π-integrable by Assumption 1.1. Since $L(t, z, \theta)$ is continuous in $(t, z, \theta) \in [0, t_\#] \times \{0, 1\} \times A_i$,

$$\begin{aligned} \lim_n E^{\pi}\big(I_{[\vartheta \in A_i]} \cdot L(\tau_n, \chi_n, \vartheta)\big) &= \pi(A_i) \lim_n E^{A_i}\big(L(\tau_n, \chi_n, \vartheta)\big) \\ &= \pi(A_i)\, E^{A_i}\big(L(\tau_0, \chi_0, \vartheta)\big) = E^{\pi}\big(I_{[\vartheta \in A_i]} \cdot L(\tau_0, \chi_0, \vartheta)\big). \end{aligned} \tag{19}$$

Summing (19) over the sets A_i, we conclude by (18) that $r(\tilde{\delta}_0, L) = r_0$. $\square$

Remark 3.2. It is clear how continuous-time behavioral monitoring procedures naturally fall within the stopping-time framework of Theorem 3.1. However, the primary interest of this paper is in $\tilde{\mathcal{D}}$, which consists of discrete-look procedures (and we assume, without loss of generality by Lemma 3.1, that those procedures have total sampling time bounded by $t_{\#} < \infty$). We now show why Theorem 3.1 remains true with $\tilde{\mathcal{D}}$ in place of $\tilde{\mathcal{D}}_{a\#}$.

Corresponding to $\tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi) \in \tilde{\mathcal{D}}$, we may write

$$\tau \equiv \eta_\rho = \sum_{j=1}^{\infty} (\eta_j - \eta_{j-1}) \text{ , where } \eta_j \in \sigma(V, W(\eta_i) : i \leq j-1). \tag{20}$$

To establish that Theorem 3.1 is true for $\tilde{\mathcal{D}}$, the only item which requires additional proof, beyond what was proved in Theorem 3.1, is that the distributional-limiting stopping-time τ_0 in (16) has the discrete-time filtration stopping-time structure (20). But if all τ_n stopping-times have the structure (20), then after a Shorokhod-embedding argument like that used in the Theorem, it suffices to observe that the a.s. limits of r.v.'s η'_{jn} which are $\sigma(V', \, (W'(\eta_i), \; i \leq j-1))$ measurable will also have the structure (20).

4 Nonrandomized Optimal Procedures

In this section, we prove that when the number of looks is bounded, say $\rho \leq M < \infty$ with M nonrandom, backward induction can be implemented to yield a nonrandomized solution to Problem L. We first show in Subsection 4.1 that the terminal-time rejection region for a Problem L solution is necessarily a half-line. Then in Subsection 4.2, we prove results for the *one-look* problem, i.e., $M = 1$, that will be needed for the multi-look case, which is discussed in Subsection 4.3.

By virtue of Lemma 3.1, we continue to assume that total trial time is bounded by $t_{\#} < \infty$. Thus, by Assumptions 1.1 and 1.2, the cost functions $c_j(t, \theta)$, $j = 1, 2, 3$, are bounded by π-integrable functions $c_j(t_{\#}, \theta)$. We will repeatedly apply this fact through the Dominated Convergence Theorem in Subsections 4.2 and 4.3.

4.1 Terminal-time rejection regions

The following lemma shows that the terminal-time rejection region for a behavioral solution of Problem L solution is a half-line.

Lemma 4.1. *Suppose that* $\tilde{\delta} = (\{\eta_j\}_{j\ge1}, \chi) \in \tilde{\mathcal{D}}$ *is a solution of Problem L. Then on the event* $[\rho = k],\ k \ge 1,$ *the rejection indicator* χ *may be expressed as a function of* $(\eta_k,\, W(\eta_k))$ *a.s. of the form* $\chi = I_{[W(\eta_k)\ge w(\eta_k)]}$ *for a real-valued function* $w : \mathbb{R} \to \mathbb{R}$.

Proof. Let $\phi(\cdot)$ denote the standard normal density and $\Phi(\cdot)$ the standard normal distribution function. The conditional expected loss on the event $[\rho = k]$ given $\{(S_j, X_j) : j \le k\}$ is equal to

$$E^\pi(L_{\lambda_0,\lambda_1}(\eta_k, \chi, \vartheta)\,|\,\{(S_j, X_j) : j \le k\}) = \Big\{ \int \prod_{j=1}^{k} \phi(X_j - \theta S_j)\pi(d\theta) \Big\}^{-1} \cdot$$

$$\int \Big\{ c_1(\eta_k, \theta) \,+\, c_3(\eta_k, \theta)\, I_{[\theta\le0]} +\, c_2(\theta)\, I_{[\theta>0]} + (\lambda_1/\pi_1)\, I_{[\theta=\theta_1]} +$$

$$E'(\chi\,|\,\{(S_j, X_j) : j \le k\}) \Big[(\lambda_0/\pi_0)\, I_{[\theta=0]} \,-\, (\lambda_1/\pi_1)\, I_{[\theta=\theta_1]} + \tag{21}$$

$$(c_2(\theta) -\, c_3(\eta_k, \theta))\,(2I_{[\theta\le0]} \,-\, 1) \Big] \Big\} \prod_{j=1}^{k} \phi(X_j - \theta S_j)\pi(d\theta). \tag{22}$$

Here $E'(\chi\,|\,\{(S_j, X_j) : j \le k\})$ corresponds to the regular conditional probability distribution of χ given $\{(S_j, X_j) : j \le k\}$, and thus, by definition of $\tilde{\mathcal{D}}$, is almost surely equal to $E^\pi(\chi|\,\vartheta,\, \{(S_j, X_j) : j \le k\})$. It follows immediately that the terminal decision χ can be replaced by one at least as good with respect to Bayes risk for loss L_{λ_0,λ_1} if $E'(\chi\,|\,\{(S_j, X_j) : j \le k\})$ does not already minimize the sum of the lines (21) and (22). That is, for a solution of Problem L, no better choice for χ exists than one which is almost surely equal to the indicator of the event that the θ integral of the square-bracketed integrand in (21)–(22) is ≤ 0, that is,

$$\chi \;=\; 1 \quad \textit{iff} \quad \lambda_0 \;\le\; \lambda_1\, e^{\theta_1 W(\eta_k) - \theta_1^2 \eta_k/2} \;-$$

$$\int \big\{ (c_2(\theta) -\, c_3(\eta_k, \theta))\,(2I_{[\theta\le0]} \,-\, 1) \big\}\; e^{\theta W(\eta_k) - \theta^2 \eta_k/2} \pi(d\theta) \tag{23}$$

and otherwise $\chi = 0$. But by Assumption 1.4, the right-hand side of the inequality in (23) is either a strictly increasing function of $W(\eta_k)$, which implies that $\chi = I_{[W(\eta_k)\ge w(\eta_k)]}$ for $w(\eta_k)$ defined as the unique value of $W(\eta_k)$ achieving equality in the inequality of (23), or is constant (when $\lambda_1 = 0$ and $c_2(\vartheta) = c_3(\eta_k, \vartheta)$ *a.s.*), and $w(\eta_k)$ can be chosen arbitrarily. $\Box$

Remark 4.1. On $[\rho = k]$ and for $1 \le j < k$, χ may be written as

$$\chi = I_{[(W(\eta_k)-W(\eta_j))/\sqrt{\eta_k-\eta_j} \,\ge\, b(\eta_j, U_j, \sqrt{\eta_k-\eta_j})]}$$

where $b(\eta, u, s) \equiv (w(\eta + s^2) - \sqrt{\eta}\, u)/s$. Thus, the conditional expected loss on the event $[S_j > 0]$ given $\{(S_i, X_i) : i \le j\}$ of taking a final look at $\eta_j + s^2$ $(s^2 \ge 0)$ is a function of (η_j, U_j, s) given by

$$\begin{aligned} E^\pi[L_{\lambda_0,\lambda_1}(\eta_j + s^2, \chi, \vartheta)|\{(S_i, X_i) : i \le j\}] = \\ \frac{\prod_{i=1}^j \phi(X_i)}{\int \prod_{i=1}^j \phi(X_i - \theta S_i)\pi(d\theta)} \cdot r_2(\eta_j, U_j, s) = \\ \frac{1}{\int \exp(\theta\sqrt{\eta_j}U_j - \theta^2\eta_j/2)\pi(d\theta)} \cdot r_2(\eta_j, U_j, s). \end{aligned} \tag{24}$$

Here, $r_2 : (0, t_\#] \times \mathbb{R} \times [0, t_\#] \to [0, \infty)$ is defined by

$$r_2(\eta, u, s) \equiv \begin{cases} \tilde{c}_1(\eta, u, s) + \int \tilde{c}_2(\eta + s^2, \theta)\big[1 - \Phi(b - \theta s)\big]e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta), & \text{if } s > 0; \\ \tilde{c}_1(\eta, u, 0) + \min\big\{\int \tilde{c}_2(\eta, \theta)e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta),\, 0\big\}, & \text{if } s = 0. \end{cases} \tag{25}$$

where $b \equiv b(\eta, u, s)$,

$$\begin{aligned} \tilde{c}_1(\eta, u, s) \equiv \int \{c_1(\eta + s^2, \theta) + c_2(\theta)I_{[\theta>0]} + (\lambda_1/\pi_1)I_{[\theta=\theta_1]} + \\ c_3(\eta + s^2, \theta)I_{[\theta\le 0]}\}e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta), \end{aligned}$$

and

$$\tilde{c}_2(t, \theta) \equiv \begin{cases} c_2(\theta) + (\lambda_0/\pi_0)\, I_{[\theta=0]} - c_3(t, \theta), & \text{if } \theta \le 0, \\ c_3(t, \theta) - c_2(\theta) - (\lambda_1/\pi_1)\, I_{[\theta=\theta_1]}, & \text{if } \theta > 0. \end{cases}$$

Also note that when a final decision is made without taking *any* looks at the data, i.e., $\eta_1 = 0$,

$$\left.\begin{array}{l} \chi = 1 \\ \chi = 0 \end{array}\right\} \quad \text{if} \quad \lambda_0 - \lambda_1 + \int c_2(\theta)\,(2I_{[\theta\le 0]} - 1)\,\pi(d\theta) \;\left\{\begin{array}{l} \le 0 \\ > 0 \end{array}\right. \tag{26}$$

and the Bayes risk is

$$E^\pi\big(L_{\lambda_0,\lambda_1}(0, \chi, \vartheta)\big) = \min\Big\{\lambda_1 + \int_{[\theta>0]} c_2(\theta)\pi(d\theta),\, \lambda_0 + \int_{[\theta\le 0]} c_2(\theta)\pi(d\theta)\Big\}. \tag{27}$$

Remark 4.2. The form of χ established in Lemma 4.1 implies that a Problem L solution $\tilde{\delta} \in \tilde{\mathcal{D}}$ necessarily satisfies $\alpha(\tilde{\delta}) + \beta(\tilde{\delta}) \leq 1$, with equality if and only if $\tilde{\delta}$ observes no data. Indeed, conditionally given η_ρ, the Neyman-Pearson Lemma implies that χ defines a conditionally uniformly most powerful test. Hence, the sum of the conditional power at $\theta = \theta_1 > 0$ and the conditional size at $\theta = 0$ is less than or equal to one, with equality if and only if no data is observed at the final look (Lehmann 1986, p.76, Corollary 1). Unconditioning then justifies that without loss of generality, $\alpha_0 + \beta_0 \leq 1$ in Problem P. □

4.2 One-look case

In the one-look case, i.e., $\rho \leq 1$, the problem is to choose a single look-time $\tau \in [0, t_\#]$ which minimizes over t the Bayes risk

$$r_1(t) \equiv E^\pi\{L_{\lambda_0,\lambda_1}(t, \chi, \vartheta)\}$$

where $\chi = I_{[U_1 \geq b(0,0,\sqrt{t})]}$ and b are as defined in Remark 4.1. We see $r_1(t)$ is more explicitly written for $t > 0$ as

$$r_1(t) = \tilde{c}_1(0, 0, \sqrt{t}) + \int \tilde{c}_2(t, \theta)\{1 - \Phi(b(0, 0, \sqrt{t}) - \theta\sqrt{t})\}\pi(d\theta). \tag{28}$$

Lemma 4.2. *r_1 is a bounded, continuous function on $[0, t_\#]$ with $\lim_{t\to 0+} r_1(t) = r_1(0)$, the Bayes risk for the zero-look (no data) case (27). That is,*

$$\lim_{t\to 0+} r_1(t) = r_1(0) \equiv \begin{cases} \lambda_1 + \int c_2(\theta) I_{[\theta>0]}\pi(d\theta), & \textit{if } \zeta_0 \geq 0; \\ \lambda_0 + \int c_2(\theta) I_{[\theta\leq 0]}\pi(d\theta), & \textit{if } \zeta_0 \leq 0, \end{cases} \tag{29}$$

where $\zeta_0 \equiv \lambda_0 - \lambda_1 + \int c_2(\theta)\,(2I_{[\theta\leq 0]} - 1)\pi(d\theta)$. Thus, an optimal single look-time $\eta_1 \equiv \tau$ is given by

$$\tau \equiv \min\{\tilde{t} \in [0, t_\#] : r_1(\tilde{t}) = \min_{t\in[0,t_\#]} r_1(t)\}.$$

Proof. By Assumptions 1.1 and 1.2, r_1 is clearly bounded on $[0, t_\#]$. Next, for $t \in (0, t_\#]$, the rejection-threshold $b(0, 0, \sqrt{t})$ is determined as in Remark 4.1 by solving for b in the equation

$$F(t, b) = 0, \tag{30}$$

obtained from the equation $-\{\phi(b)\}^{-1}\frac{\partial}{\partial b}\, r_1(t) = 0$, where

$$F(t,b) \equiv \lambda_0 - \lambda_1\, e^{\theta_1 b\sqrt{t}-\theta_1^2 t/2} + \int (c_2(\theta) - c_3(t,\theta))\,(2I_{[\theta\le 0]}-1)\, e^{\theta b\sqrt{t}-\theta^2 t/2}\,\pi(d\theta). \tag{31}$$

Note that for $t \in (0, t_\#]$, $b(0,0,\sqrt{t})$ is well-defined by the Intermediate Value Theorem. This is because by Assumption 1.1, F is a continuous function, and, by Assumption 1.4, $\theta\{c_2(\theta) - c_3(t,\theta)\}\cdot(2I_{[\theta\le 0]} - 1) < 0$ for all $\theta \neq 0$, so $F(t,\cdot)$ is strictly decreasing with $F(t,-\infty) \ge \lambda_0 + \int (c_2(\theta) - c_3(t,\theta)) I_{[\theta\le 0]}\pi(d\theta) > 0$ and $F(t,\infty) = -\infty$.

We next show that r_1 is continuous on $(0, t_\#]$ which, by Assumption 1.1, only requires showing $b(0,0,\sqrt{t})$ is continuous on $(0,t_\#]$. Let $\tilde{t} \in (0, t_\#]$, $\tilde{b} = b(0,0,\sqrt{\tilde{t}}\,)$, $t_n \to \tilde{t}$, and $b_n = b(0,0,\sqrt{t_n})$. We want to conclude $b_n \to \tilde{b}$. Since $\{t_n\}$ is in a compact subset of $(0, t_\#]$, $\{b_n\}$ is a bounded sequence. By the Bolzano-Weierstrass Theorem (Apostol 1974, p.54), there exists a subsequence $\{b'_n\}$ of $\{b_n\}$ and b' such that $b'_n \to b'$. Let $\{t'_n\}$ be the subsequence of $\{t_n\}$ corresponding to $\{b'_n\}$, i.e., $b'_n = b(0,0,\sqrt{t'_n})$. Then, using the continuity of F and $F(t'_n, b'_n) = 0$,

$$F(\tilde{t},\tilde{b}) = 0 = \lim_{n\to\infty} F(t'_n, b'_n) = F(\tilde{t}, b').$$

Thus $b' = \tilde{b}$, establishing the continuity of $b(0,0,\sqrt{t})$ on $(0,t_\#]$.

It remains to establish (29). For this, observe from equation (28) that when $t_n \searrow 0$, $\tilde{c}_1(0,0,0) = \lambda_1 + \int c_2(\theta)\, I_{[\theta>0]}\,\pi(d\theta) = \lim_{n\to\infty} \tilde{c}_1(0,0,\sqrt{t_n})$ and

$$r_1(t_n) - \tilde{c}_1(0,0,0) - \big(1 - \Phi(b(0,0,\sqrt{t_n}))\big)\,\zeta_0 \longrightarrow 0.$$

Since $F(0,b) = \zeta_0$ for each b, equation (30) implies for such sequences t_n that $\sqrt{t_n}\, b(0,0,\sqrt{t_n})$ remains bounded away from 0 when $\zeta_0 \neq 0$ and has asymptotically the same sign as ζ_0. Thus $\zeta_0\,(1 - \Phi(b(0,0,\sqrt{t_n})) \to \zeta_0 I_{[\zeta_0\le 0]}$, so $r_1(0+) = \tilde{c}_1(0,0,0) + \zeta_0\, I_{[\zeta_0 \le 0]} = r_1(0)$, as was to be shown. □

4.3 Multi-look case

In this section, we assume the total number of looks $\rho \le M < \infty$. From Remark 4.1, to minimize $E^\pi[L_{\lambda_0,\lambda_1}(\eta_j + s^2, \chi, \theta)|\{(S_i, X_i) : i \le j\}]$ with respect to s, it suffices to minimize $r_2(\eta_j, U_j, s)$ in s.

Lemma 4.3. *r_2 is a bounded, continuous function of $(\eta, u, s) \in (0, t_\#] \times \mathbb{R} \times [0, t_\#]$. Thus for $(\eta_j, U_j) = (\eta, u)$, a minimizer of (24) with respect to s^2 can be defined by*

$$\sigma^2(\eta, u) \equiv \min\{\tilde{s}^2 \in [0, t_\# - \eta] : r_2(\eta, u, \tilde{s}) = \min_{s^2 \in [0, t_\# - \eta]} r_2(\eta, u, s)\}, \qquad (32)$$

and $r_2(\cdot, \cdot, \sigma(\cdot, \cdot))$ is a bounded, continuous function of $(\eta, u) \in (0, t_\#] \times \mathbb{R}$.

Proof. r_2 is seen to be a bounded, continuous function of $(s, \eta, u) \in (0, t_\#] \times \mathbb{R} \times [0, t_\#]$ in a completely analogous fashion to the proof that r_1 is bounded and continuous on $[0, t_\#]$. Here, the measure $\pi_2(d\theta|\eta, u) \equiv e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta)$ in the integrals for $r_2(\eta, u, s)$ plays the same role which $\pi(d\theta)$ did for $r_1(t)$. $\square$

We next introduce some additional notation for our backward induction, which is similar to that of Brown et al. (1979). For $0 \leq j \leq M-1$, let $\tilde{\beta}_j^M(\eta, u, s)$ be the conditional expected loss given $(\eta_j, U_j) = (\eta, u)$ and given that at least one more look is taken at time $\eta + s^2$ if $s > 0$, or that the trial stops if $s = 0$. Let $\beta_j^M(\eta, u) \equiv \inf\{\tilde{\beta}_j^M(\eta, u, s) : 0 \leq s^2 \leq t_\# - \eta\}$. Since $\tilde{\beta}_j^M(\eta, u, s)$ is bounded and nonnegative for all $0 \leq s^2 \leq t_\# - \eta$, β_j^M is well-defined, bounded, and nonnegative.

Lemma 4.4. *For $0 \leq j \leq M-2$, $(\eta_j, U_j) = (\eta, u)$, and $s > 0$,*

$$\tilde{\beta}_j^M(\eta, u, s) = \int\int \beta_{j+1}^M\Big(\eta + s^2, \frac{\sqrt{\eta}u + s x_{j+1}}{\sqrt{\eta + s^2}}\Big)\phi(x_{j+1} - \theta s)\pi(d\theta)dx_{j+1}. \quad (33)$$

Moreover, $\tilde{\beta}_j^M$ is a bounded, continuous function of $(\eta, u, s) \in (0, t_\#] \times \mathbb{R} \times [0, t_\#]$ and β_j^M is a bounded, continuous function of $(\eta, u) \in (0, t_\#] \times \mathbb{R}$. Finally, for $1 \leq j \leq M-1$ and $S_j > 0$, an optimal $(j+1)^{st}$ look-time increment S_{j+1}^2 may be defined by

$$\sigma_{j+1,M}^2(\eta_j, U_j) \equiv \min\{\tilde{s}^2 \in [0, t_\# - \eta_j] : \tilde{\beta}_j^M(\eta_j, U_j, \tilde{s}) = \min_{s^2 \in [0, t_\# - \eta_j]} \tilde{\beta}_j^M(\eta_j, U_j, s)\}. \qquad (34)$$

If instead $S_j = 0$, then $\sigma_{j+1,M}^2(\eta_j, U_j) = 0$, by definition.

Proof. Equation (33) is true by definition. To show $\tilde{\beta}_j^M$ and β_j^M are continuous, we use backward induction on j. First, let $j = M-2$. By Lemma

4.3,

$$\int \exp\big(\theta(\sqrt{\eta_{M-2}}U_{M-2}+sx_{M-1})-\theta^2(\eta_{M-2}+s^2)\big)\pi(d\theta)\cdot$$
$$\beta^M_{M-1}\Big(\eta_{M-2}+s^2,\frac{\sqrt{\eta_{M-2}}U_{M-2}+sx_{M-1}}{\sqrt{\eta_{M-2}+s^2}}\Big)$$
$$=\inf\Big\{r_2\Big(\eta_{M-2}+s^2,\frac{\sqrt{\eta_{M-2}}U_{M-2}+sx_{M-1}}{\sqrt{\eta_{M-2}+s^2}},s_M\Big):0\le s_M^2\le t_\#-\eta_{M-2}-s^2\Big\}$$

is a continuous function of (η_{M-2},U_{M-2},s), hence the same is true for $\tilde\beta^M_{M-2}$. Consequently, β^M_{M-2} is also continuous. But then the continuity of $\tilde\beta^M_j$ and β^M_j for $0\le j\le M-2$ follow by the inductive hypothesis. Equation (34) then follows immediately for $1\le j\le M-2$, and for $j=M-1$ by Lemma 4.3. □

Since $\sigma^2_{2,M},\dots,\sigma^2_{M,M}$ can be chosen in an optimal nonrandom fashion, to obtain a fully nonrandomized optimal procedure, it remains to choose a single look-time $\eta_1\equiv\tau_{1,M}\ge 0$ which minimizes

$$r_0^M(t_1)\equiv\tilde\beta^M_0(0,0,\sqrt{t_1})=\begin{cases}\int\beta^M_1(t_1,x_1)\phi(x_1-\theta\sqrt{t_1})\pi(d\theta)dx_1, & \text{if } t_1>0;\\ r_1(0), & \text{if } t_1=0.\end{cases}$$

By Lemma 4.4, r_0^M is a bounded, continuous function on $(0,t_\#]$. Moreover,

Lemma 4.5. *A nonrandom solution to Problem L is given by*

$$\big((\tau_{1,M},\sigma^2_{2,M},\dots,\sigma^2_{M,M}),\chi\big)\tag{35}$$

where $\tau_{1,M}\equiv\min\{\tilde t_1\in[0,t_\#]:r_0^M(\tilde t_1)=\min_{t_1\in[0,t_\#]}r_0^M(t_1)\}$*;* $(\sigma^2_{2,M},\dots,\sigma^2_{M,M})$ *are as defined in (34); and* χ *is defined as in Lemma 4.1.*

Proof. The proof is by induction on M. When $M=1$, the lemma is true by Lemma 4.2.

Next, suppose the lemma is true for $M=m$. Let

$$\big((\tau_{1,m},\sigma^2_{2,m},\dots,\sigma^2_{m,m}),\chi\big)\tag{36}$$

be an optimal m look procedure of the type described in this lemma. Suppose there is a sequence of minimizers $t_{1n}\searrow 0$ of minimizers of r_0^{m+1}. By Lemma

4.4 and the Dominated Convergence Theorem, for all $(x_1, s^2) \in \mathbb{R} \times (0, t_\#]$,

$$\begin{aligned}
\tilde{\beta}_1^{m+1}(t_{1n}, x_1, s) &= \int\int \beta_2^{m+1}\Big(t_{1n} + s^2, \frac{\sqrt{t_{1n}}x_1 + sx_2}{\sqrt{t_{1n} + s^2}}\Big)\phi(x_2 - \theta s)\pi(d\theta)dx_2 \\
&\longrightarrow \int\int \beta_2^{m+1}(s^2, x_2)\phi(x_2 - \theta s)\pi(d\theta)dx_2 \\
&= \int\int \beta_1^{m}(s^2, x_1)\phi(x_1 - \theta s)\pi(d\theta)dx_1 \\
&= r_0^m(s^2) \geq r_0^m(\tau_{1,m}).
\end{aligned}$$

Thus by the continuity of $\tilde{\beta}_1^{m+1}$,

$$\begin{aligned}
\liminf_{n\to\infty} r_0^{m+1}(t_{1n}) &= \liminf_{n\to\infty} \int\int \beta_1^{m+1}(t_{1n}, x_1)\phi(x_1 - \theta\sqrt{t_{1n}})\pi(d\theta)dx_1 \\
&\geq \int\int r_0^m(\tau_{1,m})\phi(x_1 - \theta\sqrt{t_{1n}})\pi(d\theta)dx_1 = r_0^m(\tau_{1,m}).
\end{aligned}$$

Hence, an optimal $m+1$ look procedure is given by the m look procedure (36).

Next, if such a sequence $t_{1n} \searrow 0$ does not exist, then by the continuity of r_0^{m+1} on compact subsets of $(0, t_\#]$, it follows that $\tau^*_{1,m+1} = \min\{\tilde{t}_1 \in (0, t_\#] : r_0^{m+1}(\tilde{t}_1) = \min_{t_1 \in (0,t_\#]} r_0^{m+1}(t_1)\} > 0$. If the Bayes risk for the zero-look (no-data) procedure (given in (26)) $r_1(0) \leq r_0^{m+1}(\tau^*_{1,m+1})$, then an optimal first look-time is $\tau_{1,m+1} = 0$, which is equivalent to saying that an optimal $m+1$ look procedure is the zero-look procedure and (36) is again optimal. Otherwise, take $\tau_{1,m+1} = \tau^*_{1,m+1}$ in (35). □

5 Example of an Optimal Two-Look Procedure

We now give an example of a nonrandomized, optimal, two-look (i.e., $\rho = \nu \leq 2$) procedure in a hypothetical two-armed (control vs. new treatment) clinical trial with a time-to-event endpoint using the logrank statistic. In particular, we give the resulting stopping and acceptance/rejection boundaries.

We wish to test, at one-sided significance level $\alpha_0 = 0.025$, and power $1 - \beta_0 = 0.9$ versus $\theta_1 = \log(1.5)$,

$$H_0 : \theta \leq 0 \quad \text{vs.} \quad H_1 : \theta > 0 \tag{37}$$

where θ is the log-hazard ratio of the control treatment to the new treatment. Based on this parametrization for θ, the 'usual' single-look test of (37) with the same $(\alpha_0, \beta_0) = (0.025, 0.1)$ (i.e., with single critical value 1.96) has trial time $\tau_{\text{fix}} = 63.9$. For the logrank statistic, this corresponds to information time $4 \cdot \tau_{\text{fix}} = 4 \cdot (63.9) \approx 256$ events.

The discrete prior distribution and loss function with respect to which we determine the Problem P solution are given in Table 1. The discrete prior masses $\pi(\{\theta\})$ are most heavily concentrated around the point null hypothesis (hazard ratio $e^\theta = 1.0$) of no treatment difference between the two treatment arms. This is quite appropriate for many chronic disease trials (Freedman, Spiegelhalter, and Parmar 1994). The $c_1(t, \theta)$ trial time costs are equal costs of patient accrual. The $c_2(\theta)$ costs due to wrong terminal decision are largest under the most distant alternatives. Here we take $c_3(t, \theta) \equiv 0$.

Table 1: Prior distribution and elements of the loss function $L(t, z, \theta)$.

e^θ = hazard ratio	0.9	1.0	1.25	1.5	1.75
$1.51 \cdot \pi(\{\theta\})$	0.2	1.0	0.2	0.1	0.01
$c_1(t, \theta)$	t	t	t	t	t
$c_2(\theta)$	200	100	50	250	500

The description of the nonrandomized optimal procedure δ is completely given in Figures 1 and 2 below. Figure 1 is the graph of the total trial time as a function of the first-look statistic $U_1 = W(\tau_1)/\sqrt{\tau_1}$. The first look time is $\tau_1 = 0.42\,\tau_{\text{fix}}$. If at this time $U_1 \leq 0.56$, the trial is terminated and H_0 is accepted; similarly, if $U_1 \geq 2.56$, the trial is also terminated with rejection of H_0. Otherwise, on the continuation region $0.56 < U_1 < 2.56$, the trial continues to a second look at time given by the ordinate value for U_1 in Figure 1. The corresponding critical value at the second look for the normalized statistic $W(\tau_2)/\sqrt{\tau_2}$ is the ordinate value in Figure 2. For instance, the maximum trial time occurs when $U_1 = 1.78$, and when this happens the second look is taken at $\tau_2 = 1.31\,\tau_{\text{fix}}$. The corresponding second-look critical value for $W(\tau_2)/\sqrt{\tau_2}$ is 2.01. As a basis for comparison with Figure 2, recall that the usual single-look test of (37) with the same strength $(\alpha_0, \beta_0) = (0.025, 0.1)$ has critical value 1.96.

Figures 1 and 2 here; they're currently at the end.

6 Interval Continuation Regions

It is important for the common-sense appeal of a Problem L solution $\tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi) \in \tilde{\mathcal{D}}$ that it have interval continuation regions. That is, on the event $[S_j > 0]$ (i.e., a j^{th}-look has been taken) and given η_j, the continuation region $\{u \in \mathbb{R} : U_j = u \text{ and } S_{j+1} > 0\}$ should be an interval. As in the Example from Section 5, this has been the case in our numerical experience with nonrandomized two-look designs. However, because of the generality of the loss function $L(t, z, \theta)$, a theoretical result valid for all α, β in Problem P is difficult to obtain.

In the fully sequential setting with independent and identically distributed observations from a general class of distributions including those with Monotone Likelihood Ratio, Sobel (1953) and Brown et al. (1979) showed that Bayes sequential tests have interval continuation regions. However, their loss functions had the same, linear sampling time costs across all alternatives (i.e., $c_1(t, \theta) = t$), and costs of wrong terminal decision (i.e., $c_2(\theta)$) which were increasing in $|\theta|$ on each of $(-\infty, 0]$ and $(0, \infty)$. Those authors did not consider $c_3(t, \theta)$ costs. An important difference between their work and this paper, and a key barrier to generalizing their results, is that their Bayes procedures did not satisfy prespecified constraints on type I and II error probabilities, while we solve the constrained optimization problem Problem P. Since we introduce Lagrange multipliers to solve Problem P, the loss function $L_{\lambda_0,\lambda_1}(t, z, \theta)$ has a component, $c_2(\theta) + I_{[\theta=0]}\lambda_0/\pi_0 + I_{[\theta=\theta_1]}\lambda_1/\pi_1$, which is often *not* increasing in $|\theta|$ on each of $(-\infty, 0]$ and $(0, \infty)$. When this component is increasing in $|\theta|$ on each of $(-\infty, 0]$ and $(0, \infty)$ (as could be the case when π has discrete support), and when $c_1(t, \theta) = \gamma_1(t)$ does not depend on θ and $c_3(t, \theta) = 0$, their proof can be adapted to our set-up to show that the continuation regions are intervals. For the general loss function (4), it is straightforward to adapt the results from Section 4 of Kiefer and Weiss (1957) to show that the continuation regions are intervals when $\text{support}(\pi) \subset [0, \theta_1]$. For the more interesting case when $\text{support}(\pi) \not\subseteq [0, \theta_1]$, we supply a lemma applicable to the loss functions considered by Hald (1975) and Jennison (1987), i.e., $L(t, z, \theta) = \gamma_1(t)c_0(\theta)$.

In preparation for the lemma, define the functions

$$g_0(\eta, u) \equiv \int c_0(\theta) e^{\theta\sqrt{\eta}u - \theta^2\eta/2} \pi(d\theta) \tag{38}$$

and

$$g_1(\eta, u) \equiv \int c_0(\theta) e^{(\theta-\theta_1)\sqrt{\eta}u - (\theta^2-\theta_1^2)\eta/2} \pi(d\theta). \tag{39}$$

Since $g_0(\eta, \cdot)$ is a convex function, it has a unique local, hence global, minimum $u_0(\eta)$ if support$(\pi) \not\subseteq [0, \infty)$; otherwise, it is strictly increasing for all u. Similarly, if support$(\pi) \not\subseteq (-\infty, \theta_1]$, $g_1(\eta, \cdot)$ has a unique local, hence global, minimum $u_1(\eta)$; otherwise, it is strictly decreasing for all u.

Lemma 6.1. *Suppose $L(t, z, \theta) = \gamma_1(t)c_0(\theta)$ and let $\tilde{\delta} = (\{\eta_j\}_{j\geq 1}, \chi)$ be a solution of Problem L. On the event $[S_j > 0]$ (i.e., a j^{th} look has been taken) and given η_j, the j^{th}-look continuation region is an interval if*

$$\textit{support}(\pi) \subset [0, \infty) \quad \textit{or} \quad u_0(\eta_j) \leq \frac{w(\eta_j)}{\sqrt{\eta_j}} \tag{40}$$

and

$$\textit{support}(\pi) \subset (-\infty, \theta_1] \quad \textit{or} \quad \frac{w(\eta_j)}{\sqrt{\eta_j}} \leq u_1(\eta_j), \tag{41}$$

where $w(\eta_j)$ is as defined in Lemma 4.1.

Proof. First recall from Remark 4.1 that on $[S_j > 0]$, the conditional expected loss of taking a final look at $\eta_j + s^2$ given $\{(S_i, X_i) : i \leq j\}$ depends only on (η_j, U_j, s).

Now for $(u, s) \in \mathbb{R} \times (0, \infty)$, define

$$\begin{aligned} f_0(\eta_j, u, s) &\equiv \frac{\int \phi(u - \theta\sqrt{\eta_j})\pi(d\theta)}{\phi(u)} E^{\pi}\{L(\eta_j + s^2, \chi, \vartheta) - L(\eta_j, 1, \vartheta)\} | \eta_j, U_j = u\} \\ &= -\lambda_0\Phi(b) + \lambda_1\Phi(b - \theta_1 s)e^{\theta_1\sqrt{\eta_j}u - \theta_1^2\eta_j/2} + (\gamma_1(\eta_j + s^2) - \gamma_1(\eta_j))g_0(\eta_j, u), \end{aligned}$$

where $b = b(\eta_j, u, s)$ is as defined in Remark 4.1. Since $g_0(\eta_j, \cdot)$ strictly increases on $\mathbb{R}$ if support$(\pi) \subset [0, \infty)$, or on $[u_0(\eta_j), \infty)$ if support$(\pi) \not\subseteq [0, \infty)$, the j^{th} look rejection region is 'to the right' of the continuation region, i.e., $f_0(\eta_j, u, s) > 0$ implies $f_0(\eta_j, u', s) > 0$ for $u' \geq u \geq w(\eta_j)/\sqrt{\eta_j}$. Similarly, for $(u, s) \in \mathbb{R} \times (0, \infty)$, define

$$f_1(\eta_j, u, s) \equiv \frac{\int \phi(u - \theta\sqrt{\eta_j})\pi(d\theta)}{\phi(u - \theta_1\sqrt{\eta_j})} E^\pi\{L(\eta_j + s^2, \chi, \vartheta) - L(\eta_j, 0, \vartheta)\}|\eta_j, U_j = u\}$$
$$= \lambda_0(1 - \Phi(b))\, e^{-\theta_1\sqrt{\eta_j}u + \theta_1^2\eta_j/2} + \lambda_1\big(\Phi(b - \theta_1 s) - 1\big) + (\gamma_1(\eta_j + s^2) - \gamma_1(\eta_j))g_1(\eta_j, u).$$

Since $g_1(\eta_j, \cdot)$ strictly decreases on $\mathbb{R}$ if support$(\pi) \subset (-\infty, \theta_1]$, or on $(-\infty, u_1(\eta_j)]$ if support$(\pi) \nsubseteq (-\infty, \theta_1]$, the acceptance region at the j^{th} look is 'to the left' of the continuation region, i.e., $f_1(\eta_j, u, s) < 0$ implies $f_0(\eta_j, u', s) < 0$ for $u \leq u' \leq w(\eta_j)/\sqrt{\eta_j}$. Thus the continuation region at the j^{th} look must be an interval. $\square$

For the general loss function (4) with $c_1(t, \theta) = \gamma_1(t)c_0(\theta)$, it can similarly be shown that (40) implies the j^{th} look rejection region is to the right of the continuation region . However, (41) does not imply the acceptance region is to the left of the continuation region , because $\theta = \theta_1$ is in the interior of the alternative hypothesis parameter set.

7 Attainable Strengths via Lagrange Multipliers

Because of the generality of loss functions (4), the strength $\big((\alpha(\tilde{\delta}), \beta(\tilde{\delta})\big)$ for any $\tilde{\delta} \in \tilde{\mathcal{D}}$ solving Problem P may have one or both components less than the corresponding values α_0, β_0. In fact, it is possible that an optimal *unconstrained* procedure, that is, a procedure $\tilde{\delta}$ which minimizes $r(\cdot, L)$ over all procedures in $\tilde{\mathcal{D}}$, has both components less than the corresponding α_0, β_0. This phenomenon occurs, for example, if all the $c_2(\theta)$ quantities in Table 1 are increased by a multiple of 10. In such an instance, the magnitudes of the wrong decision costs $c_2(\theta)$ are sufficiently larger than the trial time costs $c_1(t, \theta)$ so as to force the unconstrained optimal procedure to have strength (0.011, 0.03). Consequently, in this example it would not be desirable (from a decision theoretic standpoint) to use a procedure of strength $(0.025, 0.1)$.

We refer to (α_0, β_0) in Problem P as an *attainable strength* if there exists a minimizer of $r(\cdot, L)$ over all $\tilde{\delta} \in \tilde{\mathcal{D}}$ satisfying $\alpha(\tilde{\delta}) = \alpha_0$ and $\beta(\tilde{\delta}) = \beta_0$. The following lemma, proved in Appendix A.3, provides necessary and sufficient conditions for (α_0, β_0) to be an attainable strength.

Lemma 7.1. *Suppose* (α_0, β_0) *in Problem P satisfies* $\alpha_0 + \beta_0 < 1$. *Then* (α_0, β_0) *is an attainable strength if and only if there exists* $\tilde{\delta} \in \tilde{\mathcal{D}}$ *satisfying*

$$\alpha(\tilde{\delta}) \le \alpha_0 \qquad \text{and} \qquad \beta(\tilde{\delta}) \le \beta_0 \tag{42}$$

along with the two conditions

$$\alpha \ge \alpha_0 \quad \forall\, (\alpha, \beta, r) \in \overline{\mathcal{R}_*} \,:\, r = r_{\beta_0} \equiv \inf\{r(\tilde{\delta}, L) :\, \tilde{\delta} \in \tilde{\mathcal{D}},\, \beta(\tilde{\delta}) = \beta_0\} \tag{43}$$

$$\beta \ge \beta_0 \quad \forall\, (\alpha, \beta, r) \in \overline{\mathcal{R}_*} \,:\, r = r_{\alpha_0} \equiv \inf\{r(\tilde{\delta}, L) :\, \tilde{\delta} \in \tilde{\mathcal{D}},\, \alpha(\tilde{\delta}) = \alpha_0\} \tag{44}$$

Remark 7.1. The sets of allowed $\tilde{\delta}$ in (43) and (44) are nonempty by Remark 2.2. Also, (43) and (44) hold automatically when $c_2(\theta) \equiv 0$. To see this, note by Assumption 1.2 that the no-data procedure $\tilde{\delta}_0 \in \tilde{\mathcal{D}}$ with acceptance probability β_0 (i.e., $\beta(\tilde{\delta}_0) = \beta_0$) satisfies

$$\tilde{\delta}_0 = \arg\min \left\{ r(\tilde{\delta}, L) :\, \tilde{\delta} \in \tilde{\mathcal{D}},\, \beta(\tilde{\delta}) = \beta_0 \right\}.$$

Then $\alpha(\tilde{\delta}_0) = 1 - \beta(\tilde{\delta}_0) = 1 - \beta_0 > \alpha_0$, so (43) holds. Similarly, (44) holds. $\square$

8 Computation of Optimal Two-Look Procedures and Concluding Remarks

The computation of an optimal two-look procedure of strength (α_0, β_0) proceeds in two stages. In the first stage, we determine for given $\tau_1 > 0$ a minimizer $\delta_{\tau_1,\alpha_0,\beta_0} \in \mathcal{D}$ of $r(\cdot, L)$ among those two-look procedures with first look-time τ_1 and strength (α_0, β_0). This first stage is theoretically validated by successively fixing pairs (λ_0, λ_1) and applying Lemma 4.3, until the desired strength (α_0, β_0) is attained. Then, in the second stage, we compute the optimal first look-time $\arg\min_{\tau_1 > 0}\ r(\delta_{\tau_1,\alpha_0,\beta_0}, L)$.

When the loss function is of the form $L(t, z, \theta) = c_1(t, \theta)$ where $c_1(t, \theta) = \gamma_1(t) c_0(\theta)$ and $\lambda_0, \lambda_1 > 0$, the qualitative behavior of $r_2(\tau_1, u_1, s)$ (as a function of s) which determines $\sigma(\tau_1, u_1)$ is established rigorously in Appendix A.2. For more interesting loss functions such as the one in Table 1, such qualitative behavior has been intractable to prove. However, we can describe our numerical experience with two-look procedures in terms of the quantities

$$c_-(t) \equiv \frac{\int_{(-\infty,0]} c_2(\theta)\pi(d\theta)}{\int_{(-\infty,0]} \big[c_1(t,\theta) + c_3(t,\theta)\big]\pi(d\theta)}\,, \quad c_+(t) \equiv \frac{\int_{(0,\infty)} c_2(\theta)\pi(d\theta)}{\int_{(0,\infty)} \big[c_1(t,\theta) + c_3(t,\theta)\big]\pi(d\theta)},$$

and of the notation $\mathcal{O}_2$ denoting 'order of magnitude' in the sense that for two positive quantities a and b, and j an integer,

$$\mathcal{O}_2\Big(\frac{a}{b}\Big)\equiv 10^j \quad \text{if} \quad \frac{1}{2}\cdot 10^j < \frac{a}{b} \leq 2\cdot 10^j.$$

When

$$\mathcal{O}_2\Big(\frac{c_+(t)}{c_-(t)}\Big)= 10^j, \quad j\in\{-1,0,1\} \tag{45}$$

and

$$\mathcal{O}_2(t\,c_+(t)),\ \mathcal{O}_2(t\,c_-(t)) = 10^k, \quad k\in\{1,2,3\}, \tag{46}$$

our numerical experience has been that for (λ_0,λ_1) corresponding to $(\alpha,\beta)\in(0.01,0.1]\times(0.05,0.2]$:

1. The function $r_2(\tau_1,u_1,\cdot)$ either has a single local, or is bounded below by $r_2(\tau_1,u_1,0)$; see Figure 3. The first case corresponds to $\nu=2$, the second to $\nu=1$. Under some additional assumptions on the loss function, which are satisfied by the example in Table 1, it can be shown that $r_2(\tau_1,u_1,\cdot)$ is strictly increasing in a neighborhood of 0; see Appendix A.4.

2. The function $r(\delta_{\tau_1,\alpha_0,\beta_0},L)$ has a unique local minimum in $\tau_1\in(0,\infty)$ which is also a global minimum; see Figure 4.

3. $\mathcal{O}_2\big(\lambda_0 t/c_-(t)\big),\ \mathcal{O}_2\big(\lambda_1 t/c_+(t)\big)= 10^k, \quad k\in\{-1,0,1\}$.

Moreover, when $j\notin\{-1,0,1\}$ in (45), or $k>3$ in (46), $(\alpha,\beta)\in(0.01,0.1]\times(0.05,0.2]$ is not an attainable strength.

Figures 3 and 4 here; they're currently at the end.

The first paragraph of this section described the algorithm used for computing an optimal time-adaptive two-look procedure. Such an algorithm is needed, as opposed to a two-dimensional grid search, since both the second look-time and corresponding critical value can depend on the first-look statistic $W(\tau_1)/\sqrt{\tau_1}$ in a completely unrestricted manner. It is a consequence of Items 1 and 2 above that this algorithm may be feasibly implemented. Moreover, these Items suggest time-adaptive optimal procedures with more than two looks may be computed, a topic we are currently investigating.

A Appendix

A.1 Proofs of Lemma 3.1 and Corollary 3.1

Proof of Lemma 3.1. By Assumption 1.3 and Remark 2.1, a nonrandom constant $t_\# < \infty$ can be found so large that for $\theta_1\sqrt{t_\#} \ge 4$, all nonnegative $s,\ u$ such that $u - s \ge t_\#/4$, and all $\theta \notin R_0$:

$$c_1(u,\theta) - c_1(s,\theta) - c_2(\theta) \ \ge\ K_1/\gamma, \tag{47}$$

where

$$K_1 \ \equiv 2 \cdot \left\{ c_2(0) + c_2(\theta_1) \ +\ \frac{\lambda_0}{\pi_0} \ +\ \frac{\lambda_1}{\pi_1} \ +1 \right\}$$

and

$$\gamma \ \equiv \begin{cases} \inf\limits_{\theta_0 \in R_0} \nu_{\theta_0}\big((\theta_0 - \frac{\theta_1}{4},\, \theta_0 + \frac{\theta_1}{4})\big), & \text{if } R_0 \neq \emptyset; \\ 1, & \text{otherwise.} \end{cases}$$

Now let $\tilde\delta \in \tilde{\mathcal{D}}$ be arbitrary. We define a modified element $\tilde\delta_{\rm alt} = (\{\eta_{j,{\rm alt}}\}_{j\ge 1}, \chi_{\rm alt}) \in \tilde{\mathcal{D}}$ for which $\eta_{\rho_{\rm alt}} \le t_\#/2$ a.s., as follows: on the event $[\eta_\rho > t_\#/2]$,

$$\eta_{j,{\rm alt}} \equiv \min\{\eta_j, t_\#/2\}\ , \qquad \chi_{\rm alt} \equiv I_{[U_{\rho_{\rm alt}} \ge \theta_1\sqrt{t_\#}/4]};$$

on the event $[\eta_\rho \le t_\#/2]$, $\tilde\delta_{\rm alt} \equiv \tilde\delta$. For notational simplicity, define $B \equiv [\eta_\rho > t_\#]$ and note that $\eta_{\rho_{\rm alt}} \le t_\#/2$. We will prove under the hypotheses of the Lemma that the modified procedure $\tilde\delta_{\rm alt}$, with total trial time bounded in this way, has Bayes risk with respect to L_{λ_0,λ_1} strictly smaller than $\tilde\delta$ unless $P_\theta(B) = 0$ π-a.e., in which case $\tilde\delta_{\rm alt} = \tilde\delta$ a.s. This will complete the proof.

Since $\eta_{\rho_{\rm alt}} \equiv t_\#/2$ on the event B,

$$P_0(U_{\rho_{\rm alt}} \ge \theta_1\sqrt{t_\#}/4 \ \big|\ B) \le 1 - \Phi(\theta_1\sqrt{t_\#}/4) \le \exp(-\frac{1}{32}\theta_1^2 t_\#),$$

and similarly

$$P_{\theta_1}(U_{\rho_{\rm alt}} < \theta_1\sqrt{t_\#}/4 \ \big|\ B) \le 1 - \Phi(\theta_1\sqrt{t_\#}/4) \le \exp(-\frac{1}{32}\theta_1^2 t_\#);$$

the last two displays both rely on the standard inequality (Feller 1968, p.175)

$$\text{for} \quad z \geq 1 \ , \qquad 1 - \Phi(z) \ \leq \ \frac{e^{-z^2/2}}{z\sqrt{2\pi}} \ \leq \ e^{-z^2/2}.$$

Now, by Assumption 1.4 and the nonnegativity of c_3, we bound $c_2 - c_3$ costs below by 0 under $\tilde{\delta}$, and c_3 costs below by 0 under $\tilde{\delta}_{\rm alt}$, to find that

$$r(\tilde{\delta},\, L_{\lambda_0,\lambda_1}) - r(\tilde{\delta}_{\rm alt},\, L_{\lambda_0,\lambda_1}) \ \geq \ \int_{R_0^c} E_\theta\Big(I_B\big(c_1(\eta_1,\theta) - c_1(t_\#/2,\theta) - c_2(\theta)\big)\Big)\pi(d\theta)$$

$$- \ \int_{R_0} E_\theta\Big(I_B \cdot \Big[I_{[\theta=\theta_1]}\Big\{c_2(\theta) + \frac{\lambda_1}{\pi_1}\Big\}(1-\chi_{\rm alt}) + I_{[\theta=0]}\Big\{c_2(\theta) + \frac{\lambda_0}{\pi_0}\Big\}\chi_{\rm alt}\Big]\Big)\ \pi(d\theta).$$

By (47) and the estimates given above for conditional error probabilities,

$$r(\tilde{\delta},\, L_{\lambda_0,\lambda_1}) - r(\tilde{\delta}_{\rm alt},\, L_{\lambda_0,\lambda_1}) \ \geq \ \frac{K_1}{\gamma}\int_{R_0^c} P_\theta(B)\,\pi(d\theta) \ -$$

$$\exp(-\frac{1}{32}\theta_1^2 t_\#) \cdot \Big(\big[\pi_1 c_2(\theta_1) + \lambda_1\big]\, P_{\theta_1}(B)\, I_{[\theta_1 \in R_0]} + \big[\pi_0 c_2(0) + \lambda_0\big]\, P_0(B)\, I_{[0\in R_0]}\Big). \tag{48}$$

We require one further idea to complete the proof. The Radon-Nikodym derivative of the law P_θ on Ω' with respect to P_{θ_0} when both are restricted to $\mathcal{F}_{t_\#} \equiv \sigma\big((W_s,\ s \leq t_\#), V\big)$ is (Liptser and Shiryayev 1977)

$$\frac{dP_\theta}{dP_{\theta_0}}\Big|_{\mathcal{F}_{t_\#}}(w) \ = \ \exp\Big((\theta-\theta_0)(w - t_\#\theta_0) \ - \ \frac{1}{2}\,(\theta-\theta_0)^2\, t_\#\Big). \tag{49}$$

We apply this result on the event $B \in \mathcal{F}_{t_\#}$, to obtain for each $\theta_0 \in R_0$, via Assumption 1.3(2)

$$\int_{0<|\theta-\theta_0|<\theta_1/4} P_\theta(B)\pi(d\theta) \ \geq \ \int_{\mathbb{R}}\int_B \frac{dP_\theta}{dP_{\theta_0}}\Big|_{\mathcal{F}_{t_\#}}(w)\ P_{\theta_0}(dw)\ \nu_{\theta_0}(d\theta)$$

which by (49) and the Fubini theorem is

$$\geq \ \int_B \int_{\mathbb{R}} \exp\Big((\theta-\theta_0)(w - \ t_\#\,\theta_0) - \frac{\theta_1^2 t_\#}{8}\Big)\ \nu_{\theta_0}(d\theta)\ P_{\theta_0}(dw).$$

Finally, by Jensen's inequality applied to the inner integral with its convex integrand as a function of θ, and using Assumption 1.3(2) to give $\int(\theta - \theta_0)\nu_{\theta_0}(d\theta) = 0$,

$$\int_{0<|\theta-\theta_0|<\theta_1/4} P_\theta(B)\pi(d\theta) \;\geq\; e^{-\theta_1^2 t_\#/32} \int_B \nu_{\theta_0}\big((\theta_0 - \theta_1/4, \theta_0 + \theta_1/4)\big)\, P_{\theta_0}(dw).$$

Since Assumption 1.3(2) and (48) also say that $\nu_{\vartheta_0}\big((\vartheta_0 - \vartheta_1/4, \vartheta_0 + \vartheta_1/4)\big) \geq \gamma > 0$, we have shown

$$\int_{R_0^c} P_\theta(B)\pi(d\theta) \;\geq\; \gamma\, \exp(-\frac{1}{32}\,\theta_1^2 t_\#) \sum_{\theta\in R_0} P_\theta(B).$$

Substitute this bound into (48) and recall the definition of K_1 in (47) to find

$$\begin{aligned} r(\tilde\delta,\, L_{\lambda_0,\lambda_1}) - r(\tilde\delta_{\text{alt}}, L_{\lambda_0,\lambda_1}) &\geq K_1\Big\{\frac{1}{\gamma}\int_{R_0^c} P_\theta(B)\,\pi(d\theta) - \frac{1}{2}e^{-\theta_1^2 t_\#/32}\sum_{\theta\in R_0} P_\theta(B)\Big\} \\ &\geq \frac{K_1}{2\gamma}\int_{R_0^c} P_\theta(B)\,\pi(d\theta) \;\geq\; 0\,, \qquad (50) \end{aligned}$$

with the last inequality strict, unless $P_\theta(B) = 0$ π-a.e. □

Proof of Corollary 3.1. For $(\alpha_0,\, \beta_0,\, r_0)$ a lower boundary point of $\mathcal{R}_*$, Lemma 2.1 implies that there exist $\lambda_0,\, \lambda_1 \geq 0$ such that $(\alpha_0,\, \beta_0,\, r_0)$ lies on a hyperplane $\lambda_0 x + \lambda_1 y + w = \lambda_0\alpha_0 + \lambda_1\beta_0 + r_0$ which separates $\mathcal{R}_*$ and $\mathcal{U}(\alpha_0,\, \beta_0,\, r_0)$.

By the definition of lower boundary points and the convexity of $\mathcal{R}_*$, there exists a sequence $\{\tilde\delta_n\}_{n=1}^\infty = \{(\{\eta_{j,n}\}_{j\geq 1}, \chi_n)\}_{n=1}^\infty \subset \tilde{\mathcal{D}}$ such that the triplets $(\alpha(\tilde\delta_n),\, \beta(\tilde\delta_n),\, r(\tilde\delta_n, L)) \in \mathcal{R}_*$ converge as $n \to \infty$ to $(\alpha_0,\, \beta_0,\, r_0)$, and there does not exist any other element $(\alpha_1, \beta_1, r_1) \in \mathcal{R}_*$ which lies below $(\alpha_0,\, \beta_0,\, r_0)$. Arguing as in the proof of Lemma 3.1, there exists a positive constant $t_\# < \infty$ and a sequence $\{\tilde\delta_{n,\text{alt}}\}_{n=1}^\infty \subset \tilde{\mathcal{D}}$ such that $\tilde\delta_{n,\text{alt}} = \tilde\delta_n$ on the complementary event B_n^c where $B_n \equiv [\eta_{\rho_n,n} > t_\#/2]$ and such that $r(\tilde\delta_{n,\text{alt}}, L_{\lambda_0,\lambda_1}) \leq r(\tilde\delta_n, L_{\lambda_0,\lambda_1})$.

Since $\lim_{n\to\infty} r(\tilde\delta_n, L_{\lambda_0,\lambda_1}) = \lambda_0\alpha_0 + \lambda_1\beta_0 + r_0$,

$$\limsup_{n\to\infty} \Big(\lambda_0\alpha(\tilde\delta_{n,\text{alt}}) \;+\; \lambda_1\beta(\tilde\delta_{n,\text{alt}}) \;+\; r(\tilde\delta_{n,\text{alt}}, L)\Big) \;\leq\; \lambda_0\alpha_0 + \lambda_1\beta_0 + r_0.$$

Also, since (α_0, β_0, r_0) lies on the separating hyperplane defined in the first paragraph of the proof,

$$r(\tilde{\delta}_{n,\mathrm{alt}}, L_{\lambda_0,\lambda_1}) \geq \lambda_0\alpha_0 + \lambda_1\beta_0 + r_0 \;\text{ for all }\; n. \tag{51}$$

Consequently,

$$\lim_{n\to\infty} r(\tilde{\delta}_n, L_{\lambda_0,\lambda_1}) = \lim_{n\to\infty} r(\tilde{\delta}_{n,\mathrm{alt}}, L_{\lambda_0,\lambda_1}) = \lambda_0\alpha_0 + \lambda_1\beta_0 + r_0.$$

But then the lower bound in (50) in the proof of Lemma 3.1 implies

$$0 = \limsup_{n\to\infty} \left[r(\tilde{\delta}_n, L_{\lambda_0,\lambda_1}) - r(\tilde{\delta}_{n,\mathrm{alt}}, L_{\lambda_0,\lambda_1})\right] \geq \limsup_{n\to\infty} \frac{K_1}{2} \int_{R_0^c} P_\theta(B_n)\,\pi(d\theta) \;\geq\; 0.$$

In particular, as $n \to \infty$, $P_0(B_n)$, $P_{\theta_1}(B_n) \;\longrightarrow\; 0$, and so

$$\int P_\theta(\tilde{\delta}_{n,\mathrm{alt}} \neq \tilde{\delta}_n)\,\pi(d\theta) \;\leq\; \int P_\theta(B_n)\,\pi(d\theta) \;\longrightarrow\; 0\;.$$

Thus, $\alpha(\tilde{\delta}_{n,\mathrm{alt}}) - \alpha(\tilde{\delta}_n) \to 0$, $\beta(\tilde{\delta}_{n,\mathrm{alt}}) - \beta(\tilde{\delta}_n) \to 0$, and $r(\tilde{\delta}_{n,\mathrm{alt}}, L) - r(\tilde{\delta}_n, L) \to 0$ as $n \to \infty$. Hence

$$(\alpha(\tilde{\delta}_{n,\mathrm{alt}}), \beta(\tilde{\delta}_{n,\mathrm{alt}}), r(\tilde{\delta}_{n,\mathrm{alt}}, L)) \;\longrightarrow\; (\alpha_0,\, \beta_0,\, r_0),$$

and the assertion is proved. □

A.2 Loss functions with only trial time costs

In this section, we assume the loss function $L(t, z, \theta) = c_1(t, \theta)$ and $\lambda_0, \lambda_1 > 0$ (recall Remark 2.3).

Lemma A.1. *Suppose the loss function* $L(t, z, \theta) = c_1(t, \theta)$ *and* $\tilde{\delta} = (\{\eta_j\}_{j\geq 1},\, \chi) \in \tilde{\mathcal{D}}$ *is a solution of Problem L with* $\lambda_0, \lambda_1 > 0$*. Then*

$$b(\eta, u, s) \;=\; \frac{\log[\lambda_0/(\lambda_1\, e^{\theta_1\sqrt{\eta}u - \theta_1^2\eta/2})]}{\theta_1 s} \;+\; \frac{\theta_1 s}{2}. \tag{52}$$

Proof. By Remark 4.1 with $b = b(\eta, u, s)$,

$$r_2(\eta, u, s) \;=\; \left\{\int \phi(u - \theta\sqrt{\eta})\,\pi(d\theta)\right\}^{-1} \cdot \Big\{\lambda_0\,(1 - \Phi(b))\,\phi(u) \,+$$
$$\lambda_1\,\Phi(b - \theta_1 s)\,\phi(u - \theta_1\sqrt{\eta}) \;+\; \int c_1(\eta + s^2, \theta)\,\phi(u - \theta\sqrt{\eta})\,\pi(d\theta)\Big\}. \tag{53}$$

As a function of b, $r_2(\eta, u, s)$ has the same minimum as the function

$$\psi(\eta, u, s, b) \;\equiv\; \frac{\int \phi(u - \theta\sqrt{\eta})\,\pi(d\theta)}{\phi(u)}\; r_2(\eta, u, s) \quad =$$

$$\lambda_0\,[1 - \Phi(b)] \;+\; \lambda_1\,\Phi(b - \theta_1 s)\,e^{\theta_1\sqrt{\eta}u - \theta_1^2\eta/2} + \int c_1(\eta + s^2, \theta) e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\,\pi(d\theta).$$

The equation

$$0 \;=\; -\frac{\partial\psi}{\partial b}(b, \eta, u, s) \;=\; \lambda_0\phi(b) - \lambda_1\phi(b - \theta_1 s) e^{\theta_1\sqrt{\eta}u - \theta_1^2\eta/2}$$

has a unique solution in b given by (52). $\square$

The explicit form of $b = b(\eta, u, s)$ in (52) allows (53) to be written as an explicit function of (η, u, s):

$$r_2(\eta, u, s) \;=\; \frac{1}{2}\,q\,\cdot\,(\eta + s^2) \;+\; C\,e^{\theta_1\sqrt{\eta}u}\,\Phi\Big(-\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)$$

$$+\;\lambda\,\Big[1 \;-\; \Phi\Big(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)\Big]$$

where

$$q = q(\eta, u) \equiv \frac{2\int c_1(\eta, \theta) e^{\theta\sqrt{\eta}u\, - \theta^2\eta/2}\pi(d\theta)}{\int e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta)}$$

$$\kappa = \kappa(\eta, u) \equiv \log\frac{\lambda_1 e^{\theta_1\sqrt{\eta}u - \theta_1^2\eta/2}}{\lambda_0}$$

$$C = C(\eta, u) \equiv \frac{\lambda_1 e^{-\theta_1^2\eta/2}}{\int e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta)}$$

$$\lambda = \lambda(\eta, u) \equiv \frac{\lambda_0}{\int e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta)}.$$

This leads to the following result (where we use notation from Lemmas 4.1 and 4.3).

Lemma A.2. *Let* $L(t, z, \theta) = c_0(\theta)t$, *and* $\lambda_0, \lambda_1 > 0$. *Suppose* $S_k > 0$ *and* $(\eta_k, U_k) = (\eta, u)$. *Then for* $u \neq w(\eta)/\sqrt{\eta}$, $\sigma(\eta, u) > 0$ *only if* $\dfrac{\partial^2}{\partial s^2} r_2(s|\eta, u)\Big|_{s = s(\eta,u)} < 0$, *where* $s(\eta, u) \equiv \{2(\sqrt{1 + \kappa^2(\eta, u)} - 1)\}^{1/2}/\theta_1$. *In* *this case,* $\sigma(\eta,u)$ *is the unique root* $\tilde{s}$ *of* $\dfrac{\partial}{\partial s} r_2(s|\eta,u)$ *in* $(s(\eta,u),\infty)$ *only if* $r_2(\eta,u,0) > r_2(\eta,u,\tilde{s})$. *Otherwise,* $\sigma(\eta,u) = 0$, *and* H_0 *is accepted (respectively, rejected) if and only if* $u < w(\eta)/\sqrt{\eta}$ *(respectively,* $u > w(\eta)/\sqrt{\eta}$*).*

Proof. By equation (23) and since $c_2 = c_3 \equiv 0$, $u \neq w(\eta)/\sqrt{\eta}$ if and only if $\kappa = \kappa(\eta,u) \neq 0$. Now

$$\begin{aligned}
\frac{\partial}{\partial s} r_2(s|\eta,u) &= qs + Ce^{\theta_1\sqrt{\eta}u}\phi\Big(-\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)\Big(-\frac{\theta_1}{2} + \frac{\kappa}{\theta_1 s^2}\Big) \\
&\qquad - \lambda\phi\Big(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)\Big(\frac{\theta_1}{2} + \frac{\kappa}{\theta_1 s^2}\Big) \\
&= qs - \lambda\phi\Big(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)\Big[\frac{\theta_1}{2} + \frac{\kappa}{\theta_1 s^2} - \frac{C}{\lambda}e^{\theta_1 u\sqrt{\eta}-\kappa}\Big(-\frac{\theta_1}{2} + \frac{\kappa}{\theta_1 s^2}\Big)\Big] \\
&= qs - \lambda\theta_1\phi\Big(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big) \qquad (54)
\end{aligned}$$

which is zero only when

$$qs = \lambda\theta_1\phi(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}). \qquad (55)$$

At any point s satisfying $\dfrac{\partial}{\partial s} r_2(s|\eta,u) = 0$, by substitution from (55),

$$\begin{aligned}
\frac{\partial^2}{\partial s^2} r_2(s|\eta,u) &= q + qs\Big(\frac{\theta_1 s}{2} - \frac{\kappa}{\theta_1 s}\Big)\Big(\frac{\theta_1}{2} + \frac{\kappa}{\theta_1 s^2}\Big) \\
&= q\Big[1 + \Big(\frac{\theta_1 s}{2}\Big)^2 - \Big(\frac{\kappa}{\theta_1 s}\Big)^2\Big].
\end{aligned}$$

Now for all $u \neq w(\eta)/\sqrt{\eta}$, $q(\eta,u)\Big[1 + \Big(\frac{\theta_1 s}{2}\Big)^2 - \Big(\frac{\kappa(\eta,u)}{\theta_1 s}\Big)^2\Big]$ is strictly increasing in s on $(0,\infty)$ and has unique root

$$s(\eta,u) = \frac{1}{\theta_1}\{2(\sqrt{1+\kappa^2(\eta,u)} - 1)\}^{1/2}.$$

Since $r_2(\eta,u,s) \longrightarrow \infty$ as $s \to \infty$, $r_2(\eta,u,\cdot)$ has a unique local minimum in $(0,\infty)$ if and only if $\dfrac{\partial}{\partial s} r_2(\eta,u,s)\Big|_{s=s(\eta,u)} < 0$, in which case the local minimum is the unique root $\tilde{s}$ of $\dfrac{\partial}{\partial s} r_2(\eta,u,s)$ in $(s(\eta,u),\infty)$. Otherwise, if

$\frac{\partial}{\partial s} r_2(\eta, u, s)\Big|_{s=s(\eta,u)} \geq 0$, so $r_2(s|\eta, u)$ is nondecreasing in s for all $s > 0$, or if $r_2(\eta, u, 0) \leq r_2(\eta, u, \tilde{s})$, $\sigma(\eta, u) = 0$, with H_0 accepted (respectively, rejected) if and only if $u < w(\eta)/\sqrt{\eta}$ (respectively, $u > w(\eta)/\sqrt{\eta}$). □

Remark A.1. Since U_k has P_θ conditional distribution $\mathcal{N}(\theta\sqrt{\eta_k}, 1)$ given η_k, $P_\theta\{U_k = w(\eta_k)/\sqrt{\eta_k} \,\big|\, \eta_k\} = 0$. Thus, on $[S_k > 0]$, $\sigma(\eta_k, w(\eta_k)/\sqrt{\eta_k})$ may be defined arbitrarily without affecting $r(\cdot, L_{\lambda_0,\lambda_1})$. □

A.3 Proof of Lemma 7.1

Proof of Lemma 7.1. The necessity of (43) and (44) are clear. For sufficiency, suppose (43) and (44) are true. It suffices to show that if $\tilde{\delta}_1 \in \tilde{\mathcal{D}}$ has strength (γ, μ) with

$$0 < \gamma \leq \alpha_0 < 1 - \beta_0 \leq 1 - \mu < 1 \qquad \text{and} \qquad \gamma + \mu < \alpha_0 + \beta_0, \tag{56}$$

then there exists $\tilde{\delta}_2 \in \tilde{\mathcal{D}}$ of strength (α_0, β_0) such that

$$r(\tilde{\delta}_2, L) \quad < \quad r(\tilde{\delta}_1, L).$$

For $\tilde{\delta}_1$ as above, if $\alpha(\tilde{\delta}_1) = \alpha_0$, then also $r(\tilde{\delta}_1, L) > r_{\alpha_0}$, while if $\beta(\tilde{\delta}_1) = \beta_0$, then $r(\tilde{\delta}_1, L) > r_{\beta_0}$. By definition of the infima and Remark 2.2, there exists a procedure $\delta_1^* \in \tilde{\mathcal{D}}$ for which either

$$\beta(\delta_1^*) = \beta_0,\ \alpha(\delta_1^*) \ \geq\ \alpha_0 \ >\ \alpha(\tilde{\delta}_1), \quad \text{and} \quad r(\delta_1^*, L) \ <\ r(\tilde{\delta}_1, L), \tag{57}$$

or

$$\alpha(\delta_1^*) = \alpha_0,\ \beta(\delta_1^*) \ \geq\ \beta_0 \ >\ \beta(\tilde{\delta}_1), \quad \text{and} \quad r(\delta_1^*, L) \ <\ r(\tilde{\delta}_1, L), \tag{58}$$

or both. If both hold, then we are done. Otherwise, define a $Binom(1, \lambda)$ random variable ξ_1, independent of all other random variables defined so far, with

$$\lambda \ = \ \begin{cases} \frac{\alpha(\delta_1^*) - \alpha_0}{\alpha(\delta_1^*) - \alpha(\tilde{\delta}_1)} & \text{if} \quad \text{(57) holds, but not (58);} \\ \frac{\beta(\delta_1^*) - \beta_0}{\beta(\delta_1^*) - \beta(\tilde{\delta}_1)} & \text{if} \quad \text{(58) holds, but not (57).} \end{cases} \tag{59}$$

Then define a procedure $\tilde{\delta}_1' \in \tilde{\mathcal{D}}$ by

$$\tilde{\delta}_1' \ \equiv\ \tilde{\delta}_1 \quad \text{if } \xi_1 = 1 \ , \qquad \text{else} \quad \tilde{\delta}_1' \ =\ \delta_1^*.$$

As a consequence, the identities

$$\alpha(\tilde{\delta}_1') \;=\; \lambda\,\alpha(\tilde{\delta}_1) + (1-\lambda)\,\alpha(\delta_1^*) \;\le\; \alpha_0 \;\;, \quad \beta(\tilde{\delta}_1') \;=\; \lambda\,\beta(\tilde{\delta}_1) + (1-\lambda)\,\beta(\delta_1^*) \;\le\; \beta_0$$

result in

$$\alpha(\tilde{\delta}_1') = \alpha_0 \quad \text{or} \quad \beta(\tilde{\delta}_1') = \beta_0 \quad , \qquad r(\tilde{\delta}_1',\, L) \;<\; r(\tilde{\delta}_1,\, L). \tag{60}$$

If both equalities hold in (60), then putting $\tilde{\delta}_2 = \tilde{\delta}_1'$ completes the argument. If only one of the equalities holds, then repeating the argument with $\tilde{\delta}_1'$ in place of $\tilde{\delta}_1$ results in a new procedure $\tilde{\delta}_2 \equiv \tilde{\delta}_1''$ of strength $(\alpha_0,\, \beta_0)$ such that $r(\tilde{\delta}_2,\, L) \;<\; r(\tilde{\delta}_1,\, L)$. □

A.4 Behavior of the Bayes risk for small sampling times

In this section, we prove under some additional assumptions on the loss function that $r_1(\cdot)$ is increasing in a neighborhood of 0. The same reasoning will apply to $r_2(\eta, u, \cdot)$, since the measure $\pi_2(d\theta|\eta, u) \equiv e^{\theta\sqrt{\eta}u - \theta^2\eta/2}\pi(d\theta)$ in the integrals for $r_2(\eta, u, s)$ plays the same role which $\pi(d\theta)$ does for $r_1(t)$. In addition to Assumptions 1.1–1.4, we assume the following for the remainder of this section:

Assumption A.1. $\liminf_{t\to 0+} \; t^{-1}\int c_1(t,\theta)\pi(d\theta) > 0.$

Assumption A.2. $\int \theta^4 c_2(\theta)\pi(d\theta) < \infty.$

Assumption A.3. $\limsup_{t\to 0+} \; t^{-1}\int c_3(t,\theta)\pi(d\theta) < \infty.$

Lemma A.3. *Let*

$$\xi_0 \equiv \frac{\lambda_0 - \lambda_1 + \int c_2(\theta)\,(2I_{[\theta\le 0]} - 1)\pi(d\theta)}{\theta_1\lambda_1 - \int \theta c_2(\theta)(2I_{[\theta\le 0]} - 1)\pi(d\theta)}. \tag{61}$$

Then as $t \to 0+$,

$$b(0,0,\sqrt{t}) = \frac{\xi_0}{\sqrt{t}} + O(\sqrt{t}). \tag{62}$$

Proof. First we note that $-\theta(2I_{[\theta\le 0]}-1)\ge 0$ for all θ implies the denominator of (61) is positive. Using Taylor's formula with remainder (Apostol 1974, p.113) and Assumptions A.2 and A.3, there exist $t_1, t_2, t_{1,\theta}, t_{2,\theta}\in[0, t/2]$ such that as $t\to 0+$, (30) may be rewritten as, for $b=b(0,0,\sqrt{t})$,

$$\begin{aligned}\lambda_0 =&\lambda_1\{1+\theta_1 b\sqrt{t}+(\theta_1 b)^2 t_1\}(1-\theta_1^2 t_2)-\\ &\int\{c_2(\theta)-O(t)\}(2I_{[\theta\le 0]}-1)\{1+\theta b\sqrt{t}+(\theta b)^2 t_{1,\theta}\}(1-\theta^2 t_{2,\theta})\pi(d\theta).\end{aligned}$$

Thus

$$\begin{aligned}\lambda_0 =&\lambda_1+\int c_2(\theta)(2I_{[\theta\le 0]}-1)\pi(d\theta)+\\ &b\sqrt{t}\Big\{\lambda_1\theta_1-\int\theta c_2(\theta)(2I_{[\theta\le 0]}-1)\pi(d\theta)\Big\}+O(t),\end{aligned}$$

establishing (62). □

Lemma A.4. *There exists $\varepsilon>0$ such that r_1 is increasing on $[0,\varepsilon]$.*

Proof. Assume $\xi_0=0$ so

$$r_1(0)=1/2\,\Big\{\lambda_1+\int c_2(\theta)I_{[\theta>0]}\pi(d\theta)+\lambda_0+\int c_2(\theta)I_{[\theta\le 0]}\pi(d\theta)\Big\}.$$

(The case when $\xi_0\neq 0$ is even more straightforward than the following argument.) Now, for t near 0,

$$\begin{aligned}\frac{r_1(t)-r_1(0)}{t}\ge\frac{1}{t}\Big[&\int\{c_1(t,\theta)+c_2(\theta)I_{[\theta>0]}+\lambda_1/\pi_1 I_{[\theta=\theta_1]}\}\pi(d\theta)+\\ &\int I_{[\theta\le 0]}\big\{c_2(\theta)+\lambda_0/\pi_0\, I_{[\theta=\theta_0]}\big\}\big\{1-\Phi(\xi_0/\sqrt{t}-\theta\sqrt{t}+O(\sqrt{t}))\big\}\pi(d\theta)-\\ &\int I_{[\theta>0]}\big\{c_2(\theta)+\lambda_1/\pi_1\, I_{[\theta=\theta_1]}\big\}\big\{1-\Phi(\xi_0/\sqrt{t}-\theta\sqrt{t}+O(\sqrt{t}))\big\}\pi(d\theta)-\\ &1/2\Big[\int c_2(\theta)\pi(d\theta)+\lambda_0+\lambda_1\Big].\end{aligned}$$

Thus by Assumption A.1,

$$\liminf_{t\to 0+}\frac{r_1(t)-r_1(0)}{t}\ge\liminf_{t\to 0+}\frac{1}{t}\int c_1(t,\theta)\pi(d\theta)>0. \tag{63}$$

□

Acknowledgments. This paper incorporates and extends parts of the first author's Ph.D. thesis at the University of Maryland, College Park. The authors would like to thank Dean Follmann and Michael Proschan for their careful reading of an earlier version of the paper.

References

[1] Anscombe, F. J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58**, 365–383.

[2] Apostol, T. M. (1974). *Mathematical Analysis, 2nd ed.* Reading, Massachusetts: Addison-Wesley.

[3] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd ed.* New York: Springer-Verlag.

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley & Sons.

[5] Brown, L. D., Cohen, A., Strawderman, W. E. (1979). Monotonicity of Bayes sequential tests. *Ann. Statist.* **7**, 1222–1230.

[6] Brown, L. D., Cohen, A., Strawderman, W. E. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8**, 377–398.

[7] Brown, L. D., Greenshtein, E. (1992). Two-sided sequential tests. *Ann. Statist.* **20**, 555–561.

[8] Chang, M. N. (1996). Optimal designs for group sequential clinical trials. *Commun. Statist. A.* **25**, 361–379.

[9] Ciampi, A., Till, J. (1980) Null results in clinical trials: the need for a decision theory approach. *British Journal of Cancer* **41**, 618–629.

[10] Colton, T. (1963). A model for selecting one of two medical treatments. *J. Amer. Statist. Assoc.* **58**, 388–401.

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

[12] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: John Wiley & Sons.

[13] Ferguson, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* San Francisco: Academic Press.

[14] Fleming, T. R., Harrington, D. P., O'Brien, P. C. (1984). Designs for group sequential tests. *Controlled Clin. Trials* **5**, 348–361.

[15] Freedman, L. S., Spiegelhalter, D. J., Parmar, M. K. B. (1994). The what, why and how of Bayesian clinical trial monitoring. *Statist. Med.* **13**, 1371–1383.

[16] Gu, M. G., Lai, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Ann. Statist.* **19**, 1403–1433.

[17] Hald, A. (1975). Optimum double sampling tests of given strength I. The normal distribution. *J. Amer. Statist. Assoc.* **70**, 451–456.

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

[19] Jennison, C., Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* New York: Chapman & Hall/CRC.

[20] Kiefer, J., Weiss, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* **28**, 57–74.

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

[22] Lehmann, E. (1986). *Testing Statistical Hypotheses, 2nd ed.* New York: John Wiley & Sons.

[23] Lipster, R. S., Shiryayev, A. N. (1977). *Statistics of Random Processes, Vol. 1.* New York: Springer-Verlag.

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

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

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

[27] Slud, E. V., Wei, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Amer. Statist. Assoc.* **77**, 862-868.

[28] Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *Ann. Statist.* **12**, 551–571.

[29] Slud, E. (1994). Adaptive repeated significance tests. Tech. Report 94-12, Department of Mathematics, University of Maryland, College Park.

[30] Sobel, M. (1953). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Statist.* **23**, 319–337.

[31] Stallard, N. (1998). Sample size determination for Phase II clinical trials based on Bayesian decision theory. *Biometrics* **54**, 279–294.

[32] Therneau, T. M., Wieand, H. S., Chang, M. (1990). Optimal designs for a grouped sequential binomial trial. *Biometrics* **46**, 771–781.

[33] Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *J. Amer. Statist. Assoc.* **77**, 855–861.

[34] Tsiatis, A. S. (1998). Interim analysis of censored data. In *Encyclopedia of Biostatistics.* New York: John Wiley & Sons.

[35] van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

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

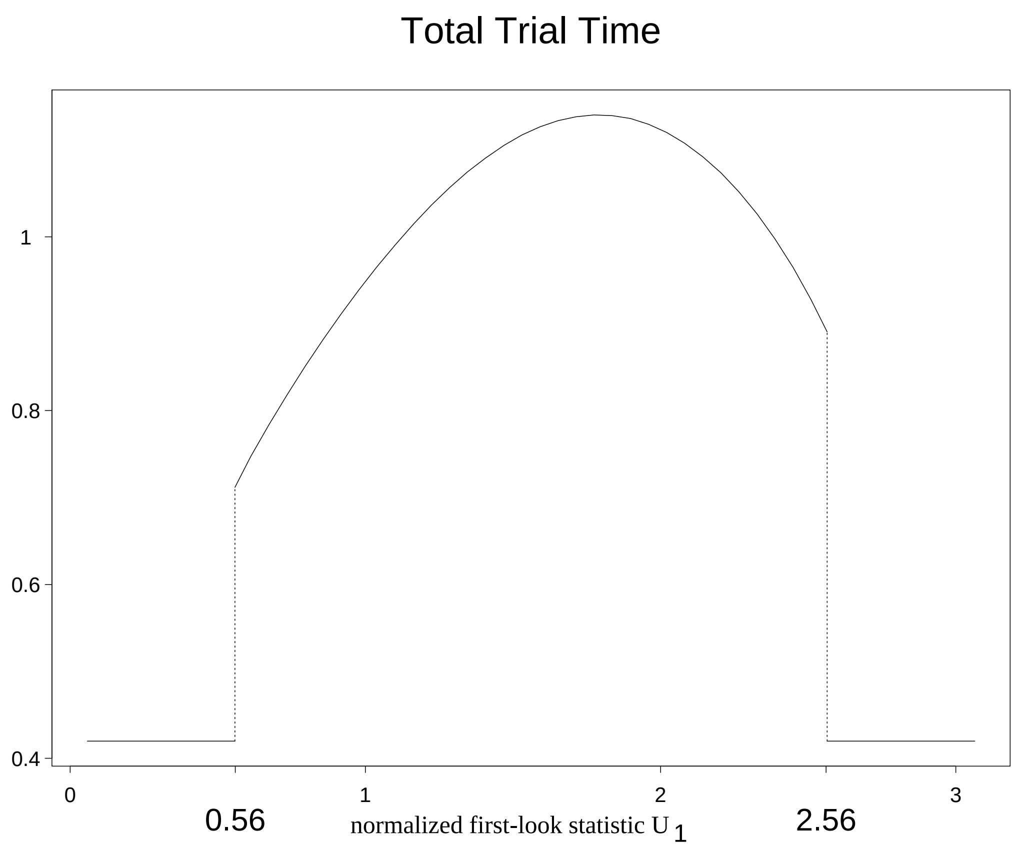

Figure 1: Total trial time for the optimal procedure of strength $(0.025, 0.1)$ with respect to the prior distribution and loss function in Table 1. The ordinate value is the total trial time graphed as a function of the first-look statistic $U_1 = W(\tau_1)/\sqrt{\tau_1}$. Time units have been scaled so that $\tau_{\text{fix}} = 1.0$. The continuation region is $0.56 < U_1 < 2.56$.

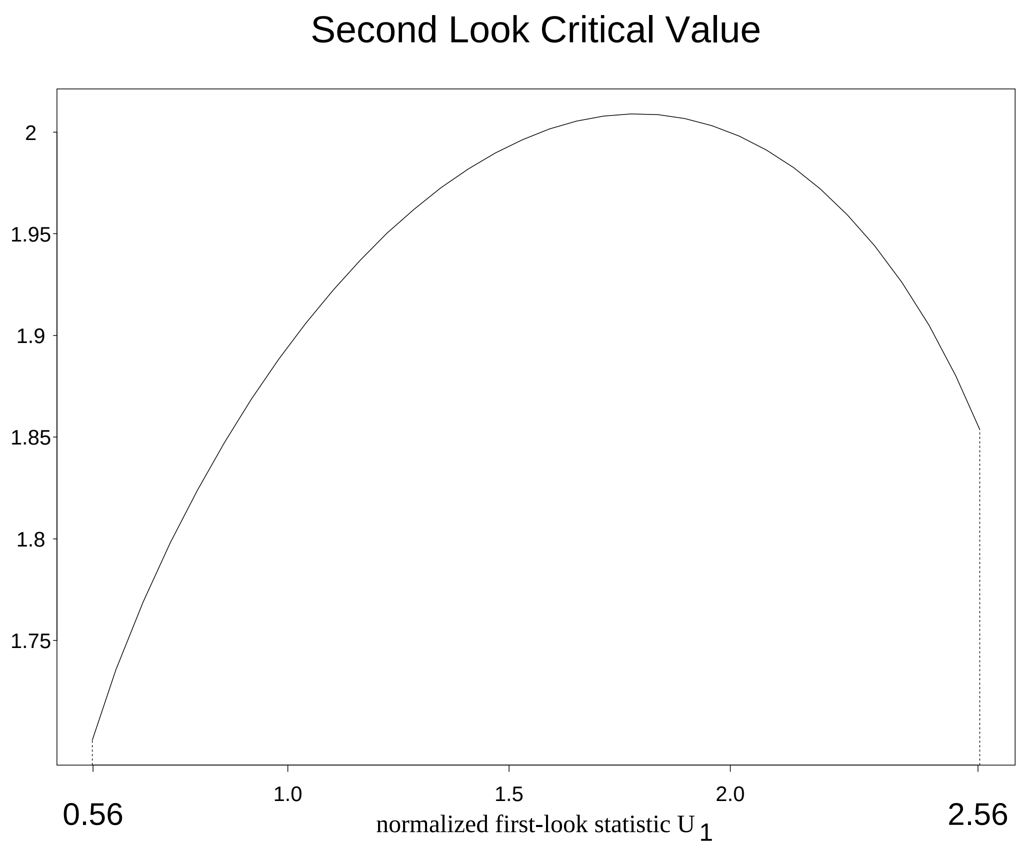

Figure 2: Terminal-time second-look critical value for the optimal procedure of strength $(0.025, 0.1)$ with respect to the prior distribution and loss function in Table 1. The ordinate value is the second-look critical value for $W(\tau_2)/\sqrt{\tau_2}$, graphed as a function of the first-look statistic $U_1 = W(\tau_1)/\sqrt{\tau_1}$.

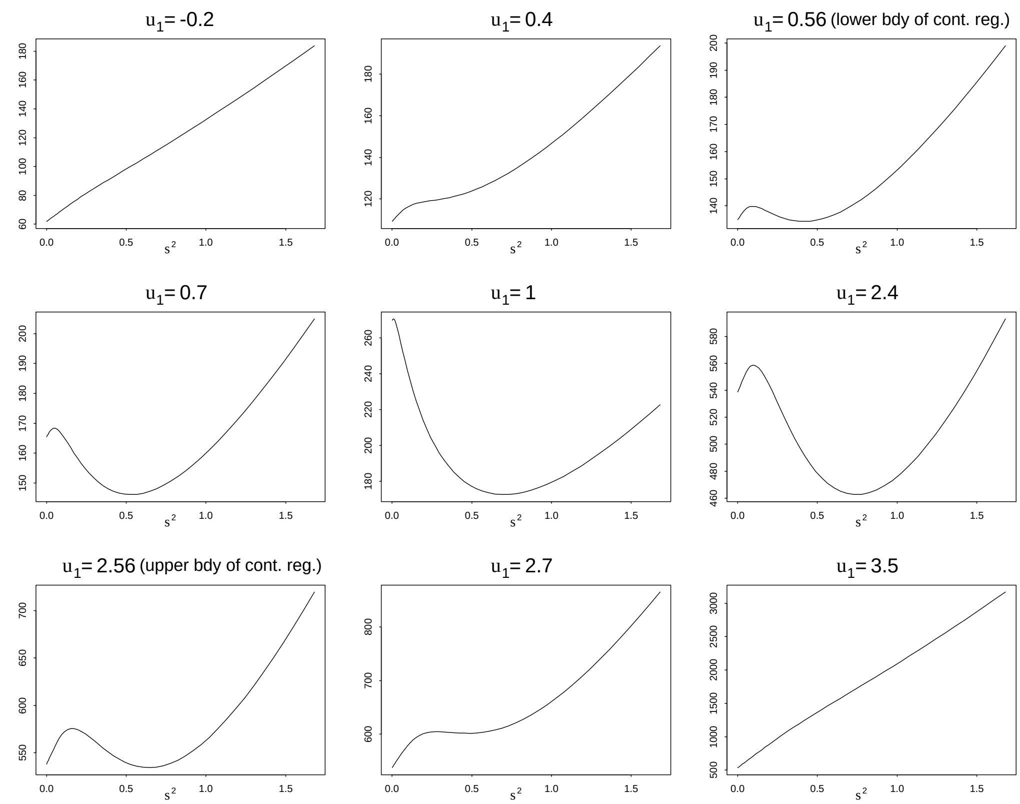

Figure 3: Graphs of $r_2(\tau_1, u_1, \cdot)$ for the optimal procedure corresponding to Table 1, for various u_1. The optimal procedure has first look-time $\tau_1 = 0.42$. Time units have been scaled so that $\tau_{\text{fix}} = 1.0$. The Lagrange multipliers are $\lambda_0 = 388,\ \lambda_1 = 210$ corresponding to $\alpha = 0.025,\ \beta = 0.1$. For each u_1, the global minimum of $r_2(\tau_1, u_1, \cdot)$ is greater than zero if and only if u_1 is in the continuation region, i.e., $0.56 < u_1 < 2.56$.

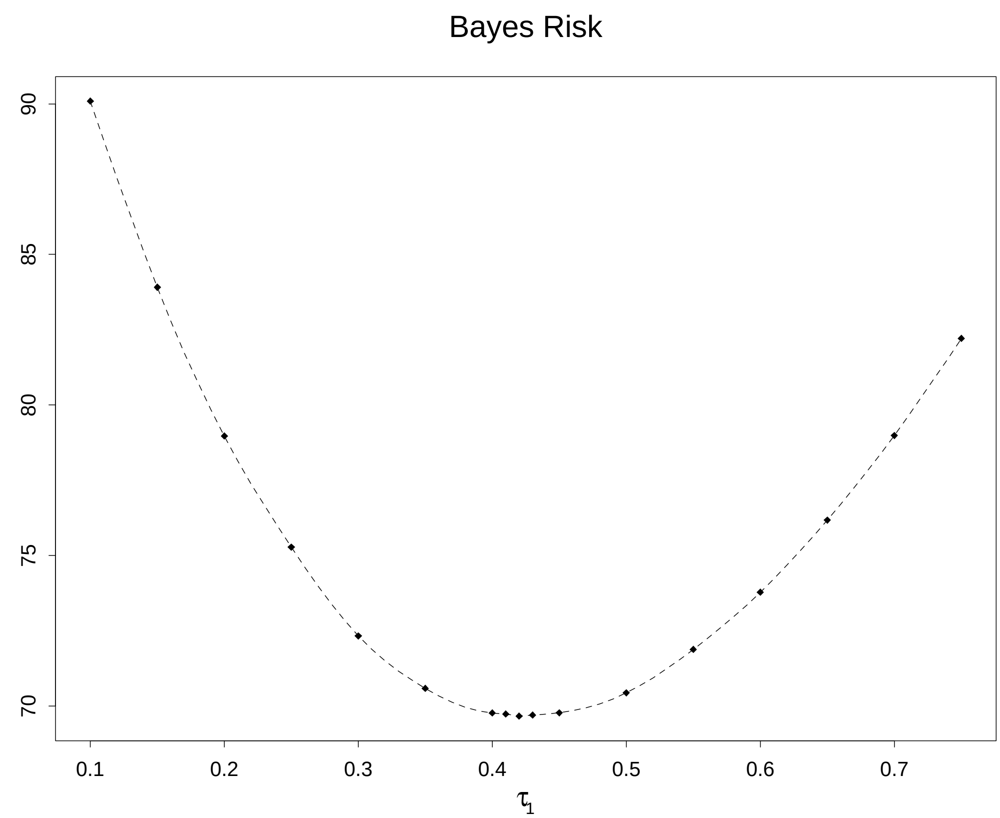

Figure 4: Graph of the Bayes risk, corresponding to Table 1, of $r(\delta_{\tau_1,\alpha_0,\beta_0}, L)$ for various values of the first look-time τ_1. Time units have been scaled so that $\tau_{\text{fix}} = 1.0$. The dashed line is a smooth interpolating spline.