Nearly Optimal Sequential Tests for Exponential Families I

Gary Lorden

Abstract

Abstract The problem of testing separated hypothesis $\theta \leq \theta_0$ and $\theta \geq \theta_1$ about the parameter θ of a one-dimensional exponential family is considered. It is shown that the generalized sequential likelihood ratio tests (GSLRT's) can be chosen so as to attain the Bayes risk to within $o(c)$ as c, the cost per observation, goes to zero. It is assumed that the prior distribution of θ has a continuous positive (Lebesgue) density on an interval $[a, b]$ in the interior of the natural parameter space. The Type I and Type II error probabilities of the GSLRT's are evaluated asymptotically as a consequence of a general theorem on asymptotic boundary-crossing probabilities. Sufficient conditions for families of tests to be $o(c)$-Bayes are given and examples of such procedures base on stopping risks or mixtures of likelihood ratios are given.

0 Introduction and Summary

Consider the problem of testing two separated hypotheses $a \leq \theta \leq \theta_0$ and $b \geq \theta \geq \theta_1$ sequentially based on independent and identically distributed random variables $X_1, X_2, X_3, \ldots$ whose distribution belongs to an exponential family of densities

$$f_\theta(x) = \exp(\theta x - \psi(\theta)),$$

with respect to a non-degenerate σ-finite measure, ν. Assume that a and b are interior points of the interval of θ's for which $\exp(\theta x)$ is ν-integrable, the so-called natural parameter space. Then ψ is infinitely differentiable on $[a, b]$, $\psi'(\theta) = E_\theta X_1$, and $\psi''(\theta) = \mathrm{Var}_\theta X_1$, so that ψ is strictly convex.

Letting $f_{\theta n} = f_\theta(X_1) \cdots f_\theta(X_n)$ ($f_{\theta 0} \equiv 1$), it is easy to evaluate under θ,

$$I(\theta, \theta') = E_\theta \log\left(\frac{f_{\theta 1}}{f_{\theta' 1}}\right) = (\theta - \theta')\psi'(\theta) - (\psi(\theta) - \psi(\theta'))$$

Define $\widehat{\theta}_n \in [a, b]$ as the value of θ maximizing the likelihood, $f_{\theta n}$, on $[a, b]$. A family of generalized sequential likelihood ratio tests (GSLRT's), one for each $\gamma > 0$, is then defined by the following rule:

$$\begin{aligned} &\text{stop and reject } \theta \leq \theta_0 \text{ if } f_{\theta_0 n} \leq \gamma h_0(\widehat{\theta}_n) f_{\widehat{\theta}_n n} \text{ and } \widehat{\theta}_n \geq \theta_2, \\ &\text{stop and reject } \theta \geq \theta_1 \text{ if } f_{\theta_1 n} \leq \gamma h_1(\widehat{\theta}_n) f_{\widehat{\theta}_n n} \text{ and } \widehat{\theta}_n \leq \theta_2, \end{aligned} \tag{1}$$

where θ_2 satisfies $I(\theta_2, \theta_0) = I(\theta_2, \theta_1)$ and h_0, h_1 are positive and continuous on $[\theta_2, b]$, $[a, \theta_2]$, respectively. Either hypothesis may be rejected if both sets of conditions are met. Theorem 2 of Section 3 establishes that as $\gamma \to 0$, the probability when θ_0 is true that the GSLRT (1) rejects $\theta \le \theta_0$ is asymptotic (assuming ν is non-lattice) to

$$\gamma\sqrt{\log \gamma^{-1}} \int_{\theta_2}^{b} \frac{L(\theta,\theta_0)}{I(\theta,\theta_0)} h_0(\theta) \left(\frac{\psi''(\theta)}{2\pi I(\theta,\theta_0)}\right)^{1/2} d\theta,$$

where $L(\theta, \theta_0)$ is defined in (2) Section 1. (The factor L/I is the necessary correction for excess over the boundary, as discussed in Sections 1 and 3, and may be roughly approximated or dispensed with in applications, as usual.) By the obvious monotonicity of the OC curve, the above integral is in fact the asymptotic Type I error, and a similar integral expresses the Type II error.

By a trite extension of Theorem 1 of Lorden (1972), accommodating non-constant h_0, h_1, the family of GSLRT's in (1) is seen to have an appealing optimality property as $\gamma \to 0$. Letting $n(\theta, \gamma)$ denote the infimum of the expected sample size at θ over all tests whose Type I and Type II error probabilities are less than or equal to those of the GSLRT (whose sample size is denoted by $\widehat{N}(\gamma)$),

$$E_\theta \widehat{N}(\gamma) \le n(\theta,\gamma)\left(1 + O\left(\frac{\log\log\gamma^{-1}}{\log\gamma^{-1}}\right)\right)$$

simultaneously (and uniformly) for all θ in $[a, b]$.

Besides the error probability approximations, the main purpose of the present paper is to demonstrate the near-optimality of GSLRT's in terms of more refined asymptotics in a Bayes context like that of Schwarz's (1962) study of asymptotic shapes. Assume a cost per observation $0 < c < 1$, and

(A1) A bounded non-negative loss function, $W(\cdot)$, vanishing on (θ_0, θ_1), continuous from the left (right) at θ_0 (resp. θ_1), and bounded away from zero on $[a, \theta_0]$ and $[\theta_1, b]$,

and

(A2) A prior density, $\lambda_0(\cdot)$ (with respect to Lebesgue measure), continuous and positive on $[a, b]$.

(Loss functions vanishing at θ_0 or θ_1 can be handled, too, as discussed in the remark following the proof of Theorem 1.)

By Theorem 1, any family of GSLRT's defined by (1) with $\gamma = o(\sqrt{\log c^{-1}})$ attains the Bayes integrated risk to within $O(c)$ as $c \to 0$ for <u>every</u> loss function W satisfying (A1) and every prior density continuous and positive on $[a_1, b_1] \subset [a, b]$ satisfying $(a_1, b_1) \supset [\theta_0, \theta_1]$. Since the Bayes risk is well-known to be of order $c\log c^{-1}$, this implies an asymptotic efficiency of $1 - O(1/\log c^{-1})$ in terms of integrated risk.

Theorem 1 also shows that for fixed W and λ_0 one can improve these asymptotic Bayes results from $O(\cdot)$ to $o(\cdot)$ by choosing $\gamma = c\sqrt{\log c^{-1}}$ and making the right choice of h_0 and h_1,

$$h_i(\theta) = \frac{\sqrt{2\pi}\lambda_0(\theta)|\psi'(\theta) - \psi'(\theta_i)|}{\lambda_0(\theta_i)W(\theta_i)L(\theta,\theta_i)(I(\theta,\theta_i)\psi''(\theta))^{1/2}}, \quad i = 0, 1.$$

Geometrical information along the lines of Schwarz (1962) is given in Corollary 3 of Section 2, which describes asymptotically in the plane of (n, S_n) $(S_n = X_1 + \ldots + X_n)$ the continuation region $\mathcal{B}^*(c)$, of a Bayes solution with respect to a prior density λ_0

satisfying (A2). In the sector where $\psi'(a) < S_n/n < \psi'(b)$, corresponding to $a < \widehat{\theta} < b$, the region $\mathcal{B}^*(c)$ is bounded within two regions whose boundaries differ by $O(1)$ – and this is improved to $o(1)$ uniformly (i.e. pointwise convergence) in sectors of the form $\psi'(\theta_2)+\varepsilon < S_n/n < \psi'(b)-\varepsilon$. More complicated $O(1)$ approximations to $\mathcal{B}^*(c)$ outside the sector where $a < \widehat{\theta} < b$ are readily obtainable from the results of Sections 1 and 2, but are of limited interest for the asymptotic theory because the prior probability of exiting from $\mathcal{B}^*(c)$ outside the sector goes to zero with c, as shown in the proof of Theorem 1.

Corollary 2 of Section 2 gives sets of sufficient conditions for families of procedures to be $o(c)$-Bayes under (A1) and (A2). Examples of families satisfying these conditions are given, including procedures which stop when the stopping risk for a certain prior (not λ_0) is less than a constant times c and procedures which stop when an appropriately chosen mixture of likelihood ratios is more than a constant over $c\log c^{-1}$. It is interesting that these approaches to defining tests, which are seemingly quite different from the GSLRT approach, yield tests which are essentially equivalent asymptotically: their stopping boundaries, as well as those of the Bayes procedure and the GSLRT, converge to the same limit within sectors excluding $\widehat{\theta} = a$, θ_2, and b and their Type I and Type II error probabilities have the same asymptotic expressions.

Later papers in this series will extend the theory to k-hypothesis tests and open-ended tests, study by computer the accuracy of error probability approximations and the efficiency of tests in the case of small samples, and develop another approach to $o(c)$-Bayes tests that works for a large class of priors supported on the entire natural parameter space. The present investigation uses many of the mathematical ideas that have been developed in the asymptotic theory of sequential analysis pioneered by Chernoff (1959), Schwarz (1962), and Kiefer and Sacks (1963). Schwarz studied the shape and first-order rate of growth of the Bayes continuation regions for a more general class of priors than that in (A2), finding them to be identical with those of simple computable procedures involving maximum likelihood ($h_i(\theta) \equiv 1$ and $\gamma = c$ in (1)). It remained for Wong (1968) to prove that these attractive procedures actually attain the Bayes risk asymptotically. Fushimi (1965, 1967) obtained second-order approximations to the Bayes regions in special cases and Schwarz (1969) generalized his results by obtaining to within $O(1)$ the asymptotic expansion of the stopping risk along any ray from the origin in the (n, S_n) plane. (Expansions like Schwarz's are given in the next section and are accurate to $o(1)$ uniformly over certain sectors of the (n, S_n) plane, as required for the proof of Theorem 1.) Both Schwarz's and Fushimi's results did not establish that any improvement in the integrated risk of Schwarz's procedures would result from using the second-order correction and, indeed, left undetermined the right choice of coefficient for the second-order term. The latter uncertainty was due to the lack of any improvement in Schwarz's approximation of $\mathcal{B}^*(c)$ within constant-stopping-risk boundaries of order c and $c\log c^{-1}$, whose second-order approximations differ by $\log\log c^{-1}$. Lorden (1967), using the methods of Kiefer and Sacks (1963), improved this constant-stopping-risk approximation in a general setting, obtaining an upper bound M^*c. But the assumptions of that paper require separation of the indifference region from the hypotheses, e.g. vanishing prior density near the ends of (θ_0, θ_1). Lemma 3 of Section 3 remedies this situation by establishing that (A1) and (A2) are, indeed, sufficient for the existence of an M^* such that a Bayes procedure continues sampling whenever the stopping risk exceeds M^*c. The first result showing that Schwarz-type procedures actually attain second-order efficiency,

$$1 - O\left(\frac{\log\log c^{-1}}{\log c^{-1}}\right),$$

was obtained by Gloria Zerdy in her thesis (1975) for the case of the normal mean with prior density supported by the real line.

Although restricted by the assumed separation of the indifference interval from the hypotheses, Lorden's (1967) results showed that the Bayes risk could be attained to within $O(c)$ by stopping when the stopping risk is less than a constant times c, a procedure proposed by Kiefer and Sacks. The practical difficulty in using such a procedure is, of course, the need to compute the stopping risk at each stage. This kind of computational difficulty can be overcome in many cases by using instead mixtures of likelihood ratios, a method suggested by Wald and extended into the elegant theory of open-ended tests developed by Robbins (1970) and Robbins and Siegmund (1970). The present work makes considerable use of the methods and results of the recent refinements of that theory (Pollak and Siegmund, 1975) and (Lai and Siegmund, 1977). Also of importance is the "stopping problem" approach used by Bickel and Yahav (1967, 1968, 1969) in their general theory of asymptotically pointwise optimal (A.P.O.) sequential tests and estimators.

The present paper is an extension to a continuous parameter setting of the ideas developed in Lorden (1977), where simple $o(c)$-Bayes procedures were proposed for general problems involving only finitely many parameter values. Like that paper, the present considerations hinge on showing that the problem of when to stop is sufficiently well approximated asymptotically by reducing considerations to one-sided sequential probability ratio tests (SPRT's) of the maximum likelihood θ against the most likely θ in a hypothesis which is close to rejection.

1 Preliminaries and asymptotics

Probability and expectation under θ will be denoted by P_θ and E_θ, respectively. The symbols P_{λ_0} and E_{λ_0} will be similarly used for the λ_0-mixture of θ-probabilities and expectations. Let $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$ denote a sequence of sigma-fields with respect to which all stopping rules and terminal decision rules are assumed defined. $\mathcal{F}_0$ is trivial and for $n \geq 1$, $X_1, \ldots, X_n$ are $\mathcal{F}_n$-measurable.

It will be convenient to represent the testing problem geometrically as follows. Consider the half-plane

$$\{(t,s)| \quad 1 \leq t < \infty, \quad -\infty < s < \infty\}$$

and for a given prior density, λ_0, associate each point (t,s) in the half-plane with a posterior density

$$\lambda_{s,t}(y) = \frac{\lambda_0(y)\exp(ys - t\psi(y))}{\int_a^b \lambda_0(\theta)\exp(\theta s - t\psi(\theta))d\theta}$$

having the same support as λ_0. Thus, if the first n observations have sum S_n, then the point (n, S_n) in the (t,s) plane will be associated with λ_{n,S_n}, which is indeed the usual posterior density. The notation λ_n will be used as a shorthand for λ_{n,S_n}.

It will also prove useful to define a "continuous" analog of $f_{\theta n} = \exp(\theta S_n - n\psi(\theta))$, namely

$$f_{\theta t} = \exp(\theta s - t\psi(\theta)).$$

The definition of the maximum likelihood estimator is also extended to the entire (t, s) half-plane by setting

$$\widehat{\theta}(s,t) = \begin{cases} b & \text{if} \quad s > t\psi'(b) \\ a & \text{if} \quad s < t\psi'(a) \\ \text{the unique solution of } s = t\psi'(\theta), \text{ otherwise.} \end{cases}$$

Assume for convenience that ψ and ψ' vanish at θ_2 and that $\theta_2 = 0$, i.e.

$$\theta_0 < 0 < \theta_1, \quad \psi(0) = \psi'(0) = 0, \quad \text{and} \quad \psi(\theta_0) = \psi(\theta_1).$$

This standardization essentially involves subtracting $E_{\theta_2}X$ from the X's and θ_2 from the θ's and has the convenient feature that $\text{sgn}(s) = \text{sgn}(\widehat{\theta}) = \text{sgn}(f_{\theta_1 t}/f_{\theta_0 t} - 1)$. The maximum likelihood estimator based on S_n at time n is $\widehat{\theta}_n = \widehat{\theta}(n, S_n)$. Routine computation shows that if $\psi'(a) \leq s/t \leq \psi'(b)$ then

$$\frac{f_{\widehat{\theta}t}}{f_{\theta t}} = \exp(tI(\widehat{\theta}, \theta)) \quad \text{for all} \quad \theta,$$

where $\widehat{\theta} = \widehat{\theta}(s,t)$.

Define the posterior risk of rejecting $\theta \leq \theta_0$,

$$Y_0(s,t) = \frac{\int_a^{\theta_0} \exp(ys - t\psi(y))\lambda_0(y)W(y)dy}{\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy},$$

and the posterior risk of rejecting $\theta \geq \theta_0$,

$$Y_1(s,t) = \frac{\int_{\theta_1}^{b} \exp(ys - t\psi(y))\lambda_0(y)W(y)dy}{\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy}.$$

Then the <u>stopping risk</u> at (s, t) is

$$r(\lambda_{s,t}) = \min(Y_0(s,t), Y_1(s,t)).$$

The integrated risk of a test procedure δ having stopping time N with respect to the prior λ_0 and cost per observation $0 < c < 1$ is

$$r_c(\lambda_0, \delta) = \int_a^b \left[cE_\theta N + W(\theta)P_\theta(\delta \text{ makes wrong decision})\right] \lambda_0(\theta)d\theta.$$

All the procedures considered in this paper may be regarded as defining continuation regions in the (t, s) half-plane, with their complements partitioned into sets where one or the other terminal decision is chosen. For all such procedures, the posterior risk $r(\lambda_{s,t}, \delta)$ is well-defined in the entire (t, s) half-plane as the $\lambda_{s,t}$-integrated risk of the test that begins at the point (t, s) and proceeds as indicated by the δ continuation region and terminal decision regions.

The numbers $L(\theta, \theta')$, which play a significant role in what follows, are defined as in Lorden (1977): for θ, θ' belonging to the natural parameter space,

$$L(\theta, \theta') = \exp\left(-\sum_{n=1}^{\infty} n^{-1}\left[P_\theta(f_{\theta n} \leq f_{\theta' n}) + P_{\theta'}(f_{\theta' n} < f_{\theta n})\right]\right), \tag{2}$$

a formula which derives from the work of Spitzer (1956).

Note that, since $P_\theta(f_{\theta n} = f_{\theta' n}) = P_{\theta'}(f_{\theta' n} = f_{\theta n})$, evidently $L(\theta, \theta') = L(\theta', \theta)$. Furthermore, as shown in Lorden (1977),

$$0 < L(\theta, \theta') \leq \min(1, I(\theta, \theta')) \quad \text{if} \quad \theta \neq \theta'$$

and stopping at the smallest $n \geq 0$ (or ∞ if there is no n) such that

$$v f_{\theta' n} L(\theta, \theta') \leq u f_{\theta n}$$

minimizes

$$u E_\theta N + v P_{\theta'}(N < \infty) = E_\theta \left(uN + v \frac{f_{\theta' N}}{f_{\theta N}} \right)$$

among all extended stopping times, for arbitrary $u, v > 0$, a fact which is needed for Lemma 6. Further properties of the L-numbers needed in the present context are the following:

(i) $L(\theta, \theta') \downarrow$ in θ' for $\theta' \leq \theta$ and $\nearrow$ in θ' for $\theta' \geq \theta$.

(ii) $L(\theta, \theta')$ is continuous in each variable.

To verify these properties, let $L_n(\theta, \theta')$ denote the nth term of the series used to define $L(\theta, \theta')$ and write (using $1\{\}$ for indicator function)

$$L_n(\theta, \theta') = E_\theta \left[1 \left\{ \frac{f_{\theta' n}}{f_{\theta n}} \geq 1 \right\} + \frac{f_{\theta' n}}{f_{\theta n}} 1 \left\{ \frac{f_{\theta' n}}{f_{\theta n}} < 1 \right\} \right] = E_\theta \exp \left(-\log^+ \frac{f_{\theta n}}{f_{\theta' n}} \right).$$

To verify (i) for $\theta' \leq \theta$, note that this last is increasing in θ'. (For $\theta'' \in (\theta', \theta)$, $\log f_{\theta n} > \log f_{\theta'' n}$ implies $\log f_{\theta'' n} > \log f_{\theta' n}$ by the concavity of the log-likelihood function.) Property (i) for $\theta' \geq \theta$ is verified similarly. Also, recalling that $L(\theta, \theta') = L(\theta', \theta)$, this last and the bounded convergence theorem show that property (ii) holds with $L(\theta, \theta')$ replaced by $L_n(\theta, \theta')$. It follows that for $\theta \neq \theta'$,

$$\lim_{\varepsilon \downarrow 0} \left(\sum_{n=1}^{\infty} n^{-1} L_n(\theta, \theta' - \varepsilon) - \sum_{n=1}^{\infty} n^{-1} L_n(\theta, \theta' + \varepsilon) \right) = 0 \tag{3}$$

because the tails of these series are seen to be uniformly small for ε in a neighborhood of 0 as a consequence of the monotonicity of the $L_n(\theta, \cdot)$'s. Relation (3) combined with the monotonicity property, (i), establishes that, for every θ, $L(\theta, \theta')$ is continuous at $\theta' \neq \theta$, and continuity at $\theta' = \theta$ follows from the fact that $\lim_{\theta' \to \theta} \sup L(\theta, \theta') \leq \lim_{\theta' \to \theta} \sup I(\theta, \theta') = 0$.

The L-numbers play a different role in asymptotic evaluations of error probabilities. The fact that the ratios $L(\theta, \theta')/I(\theta, \theta')$ and $L(\theta', \theta)/I(\theta', \theta)$ are the right correction factors to Wald's error probability bounds for SPRT's was established in Siegmund (1975) using different notation based on the work of Spitzer (1956) (see also the discussion in Lorden (1977). A generalization to mixture stopping rules was given in Lai and Siegmund (1977) and the latter result is the basis for the asymptotic evaluation of error probabilities in Section 3.

The next lemma provides the tools for analyzing the stopping risk asymptotically.

Lemma 1. *Assume λ_0 satisfies (A2). As $\zeta + t \to \infty$,*

$$\int_a^\theta \exp((y - \theta)\zeta - tI(\theta, y))\lambda_0(y) dy \sim \lambda_0(\theta) \int_a^\theta \exp((y - \theta)\zeta - \frac{1}{2} t\psi''(\theta)(y - \theta)^2) dy \tag{4}$$

uniformly for $a < \theta \le b$, $\zeta \ge \zeta_0$ *(arbitrary), and* $t \ge 1$. *Under the additional restriction* $\theta \ge a + \varepsilon > a$

$$\int_a^\theta \exp((y-\theta)\zeta - tI(\theta,y))\lambda_0(y)dy \sim \lambda_0(\theta)\int_a^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy \tag{5}$$

$$= \lambda_0(\theta)\int_{-\infty}^0 \exp(x\zeta - \frac{1}{2}t\psi''(\theta)x^2)dx$$

uniformly.

Proof: Choose $M > 1$ large enough so that ψ'', $1/\psi''$, ψ''', and λ_0 are bounded above by M on $[a,b]$. Then, letting $u = \zeta + t$,

$$\left| I(\theta,y) - \frac{1}{2}\psi''(\theta)(y-\theta)^2 \right| \le M|y-\theta|^3 \le Mu^{-6/5} \tag{6}$$

provided that θ, $y \in [a,b]$ and $|y-\theta| \le u^{-2/5}$. Furthermore, if $y \le \theta - u^{-2/5}$ and $u^{2/5} > 4M^2$, then using (6)

$$I(\theta,y) \ge I(\theta,\theta-u^{-2/5}) \ge \frac{1}{2}\psi''(\theta)u^{-4/5} - Mu^{-6/5} \ge \frac{u^{-4/5}}{4M}. \tag{7}$$

In the left-hand side of (4) write

$$\int_a^\theta = \int_a^{\max(a,\theta-u^{-2/5})} + \int_{\max(a,\theta-u^{-2/5})}^\theta .$$

Assuming from now on that $u^{2/5} > 4M^2$, (7) implies

$$0 \le \int_a^{\max(a,\theta-u^{-2/5})} \le M(b-a)\exp\left(-\left[u^{-2/5}(\zeta+\zeta_0^-) + \frac{u^{-4/5}t}{4M} - (b-a)\zeta_0^-\right]\right)$$

$$\le M(b-a)\exp\left(-\frac{u^{1/5}}{4M} + (b-a)\zeta_0^-\right). \tag{8}$$

Using (6)

$$\int_{\max(a,\theta-u^{-2/5})}^\theta = \xi \int_{\max(a,\theta-u^{-2/5})}^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy,$$

where

$$e^{-Mu^{-1/5}} \min_{[\theta-u^{-2/5},\theta]} \lambda_0(y) \le \xi \le e^{-Mu^{-1/5}} \max_{[\theta-u^{-2/5},\theta]} \lambda_0(y).$$

Since λ_0 is uniformly continuous and positive on $[a,b]$, $\xi \sim \lambda_0(\theta)$ uniformly in θ, ζ, and t as $u \to \infty$ and therefore

$$\int_{\max(a,\theta-u^{-2/5})}^\theta \sim \lambda_0(\theta)\int_{\max(a,\theta-u^{-2/5})}^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy \tag{9}$$

$$\sim \lambda_0(\theta)\int_a^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy,$$

the last relation being verified as follows. In the non-trivial case where $a < \theta - u^{-2/5}$,

$$\int_a^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy \geq \int_{\theta - u^{-2/5}}^\theta \exp(Mu(y-\theta))dy = \frac{1-e^{-Mu^{3/5}}}{Mu}, \tag{10}$$

whereas the integral from a to $\theta - u^{-2/5}$ can be estimated as in (8) and seen to be uniformly of smaller order. Similarly, note from (8) - (10) that the contribution to the left-hand side of (4) from $[a, \max(a, \theta - u^{-2/5})]$ is negligible, so that (9) implies (4).

To prove (5), choose M to satisfy the additional requirement $M > \varepsilon/2$, write $\int_a^\theta \exp((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2)dy = \int_{-\infty}^\theta - \int_{-\infty}^a$, and note that since $\theta - a \geq \varepsilon$,

$$\int_{-\infty}^a \leq \int_{-\infty}^a \exp\left((y-\theta)\frac{\varepsilon u}{2M}\right)dy \leq \frac{2M}{\varepsilon u}\exp\left(-\frac{\varepsilon^2 u}{2M}\right),$$

whereas $\int_{-\infty}^\theta$ is by (7) at least $1/2Mu$ for large u.

Remark. By a similar argument it can be shown that as $t - \zeta \to \infty$

$$\int_\theta^b \exp\left((y-\theta)\zeta - tI(\theta,y)\right)\lambda_0(y)dy \sim \lambda_0(\theta)\int_\theta^b \exp\left((y-\theta)\zeta - \frac{1}{2}t\psi''(\theta)(y-\theta)^2\right)dy \tag{11}$$

uniformly for $a \leq \theta < b$, $\zeta \leq \zeta_1$ (arbitrary), and $t \geq 1$, and that under the additional restriction $\theta \leq b - \varepsilon < b$,

$$\int_\theta^b \exp\left((y-\theta)\zeta - tI(\theta,y)\right)\lambda_0(y)dy \sim \lambda_0(\theta)\int_\theta^b \exp(x\zeta - \frac{1}{2}t\psi''(\theta)x^2)dx. \tag{12}$$

By applying Lemma 1 and the analogous relations (11) and (12), many asymptotic relations can be derived which are useful in the sequel. For instance, using the identity

$$(y-\theta)s - t(\psi(y)-\psi(\theta)) = (y-\theta)(s - t\psi'(\theta)) - tI(\theta,y)) \tag{13}$$

and setting $\zeta = s - t\psi(\theta)$ leads to

$$\begin{aligned}\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy &= \exp(\theta s - t\psi(\theta))\int_a^b \exp((y-\theta)\zeta - tl(\theta,y))\lambda_0(y)dy \\ &= \exp(\theta s - t\psi(\theta))\left(\int_a^\theta + \int_\theta^b\right) \\ &\sim \lambda_0(\theta)\exp(\theta s - t\psi(\theta))\int_{-\infty}^\infty \exp(x\zeta - \frac{1}{2}t\psi''(\theta)x^2dx \\ &= \left(\frac{2\pi}{t\psi''(\theta)}\right)^{1/2}\lambda_0(\theta)exp(\theta s - t\psi(\theta))\end{aligned} \tag{14}$$

as $t \to \infty$, uniformly for s, t, θ satisfying

$$a + \varepsilon \leq \theta \leq b - \varepsilon \quad \text{and} \quad \zeta_0 \leq s - t\psi'(\theta) \leq \zeta_1.$$

In particular, if $\psi'(a) + \varepsilon \leq s/t \leq \psi'(b) - \varepsilon$, then setting $\theta = \widehat{\theta}$, so that $s = t\psi'(\theta)$, leads to the relation

$$\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy \sim \left(\frac{2\pi}{t\psi''(\widehat{\theta})}\right)^{1/2}\lambda_0(\widehat{\theta})\exp(\widehat{\theta}s - t\psi(\widehat{\theta})) \tag{15}$$

as $t \to \infty$.

To evaluate asymptotically the integral in (15) for all $s/t \geq b\psi'(a)+\varepsilon$ (and similarly for $s/t \leq b\psi'(b)-\varepsilon$), write it as the sum of integrals over $[a, \widehat{\theta}]$ and $[\widehat{\theta}, b]$, use the identity

$$(y-\widehat{\theta})s - t(\psi(y)-\psi(\widehat{\theta})) = (y-\widehat{\theta})(s-t\psi'(b))^+ - tl(\widehat{\theta}, y)$$

and apply (5) and, if $\widehat{\theta} < b$, (11), to obtain

$$\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy \sim$$
$$\lambda_0(\widehat{\theta})\exp(\widehat{\theta}s - t\psi(\widehat{\theta}))\int_\infty^b \exp((y-\widehat{\theta})(s-t\psi'(b))^+ - \frac{1}{2}t\psi''(\widehat{\theta})(y-\widehat{\theta})^2)dy.$$

Upon completing the square, the integral can be expressed using Mills' Ratio $\rho^*(x) = \Phi(-x)/\phi(x)$. Using this result for $\widehat{\theta} \geq (a+b)/2$ and the obvious analog for $\widehat{\theta} \leq (a+b)/2$ yields the conclusion that as $|s|+t \to \infty$, uniformly for all s and $t > 1$,

$$\int_a^b \exp(ys - t\psi(y))\lambda_0(y) \sim \frac{\lambda_0(\widehat{\theta})\exp(s\widehat{\theta} - t\psi(\widehat{\theta}))}{\sqrt{t\psi''(\widehat{\theta})}} \quad \times$$
$$\begin{cases} \rho^*\left(\frac{|s-t\psi'(a)|}{\sqrt{t\psi''(a)}}\right), & \text{if} \quad s \leq t\psi'(a) \\ \sqrt{2\pi}\Phi\left(\sqrt{t\psi''(\widehat{\theta})}\min(b-\widehat{\theta}, \widehat{\theta}-a)\right), & \text{if} \quad t\psi'(a) \leq s \leq t\psi'(b) \\ \rho^*\left(\frac{|s-t\psi'(b)|}{\sqrt{t\psi''(b)}}\right), & \text{if} \quad s \geq t\psi'(b). \end{cases} \quad (16)$$

When the full strength of (16) is not needed in the sequel, it is convenient to use relations like

$$\log \int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy = \log\left[\frac{\sqrt{2\pi}\lambda_0(\widehat{\theta})\exp(\widehat{\theta}s - t\psi(\widehat{\theta}))}{\sqrt{2\pi}(s-t\psi'(b))^+ + \sqrt{t\psi''(\widehat{\theta})}}\right] + O(1), \quad (17)$$

which holds uniformly for $s/t \geq \psi'(a)+\varepsilon$ as $\max(s,t) \to \infty$ and with $o(1)$ uniformly for $\psi'(a)+\varepsilon \leq s/t \leq \psi'(b)-\varepsilon$ by (15). Relation (17) is easily derived from (16) using the facts that the value of Φ in (16) is between 1/2 and 1 and that $(1+\sqrt{2\pi}x)\rho^*(x)$ has positive upper and lower bounds on $[0,\infty]$ (being positive and continuous with positive limits at 0 and ∞).

Using relation (5) with λ replaced by λW, which like λ is continuous from the left and positive at θ_0 (the only value of θ needed),

$$\int_a^{\theta_0} \exp(ys - t\psi(y))\lambda_0(y)W(y)dy$$
$$\sim \lambda_0(\theta_0)W(\theta_0)\exp(\theta_0 s - t\psi(\theta_0))\rho\left(\frac{s-t\psi'(\theta_0)}{\sqrt{t\psi''(\theta_0)}}\right)(t\psi''(\theta_0))^{-1/2}$$
$$\sim \frac{\lambda_0(\theta_0)W(\theta_0)\exp(\theta_0 s - t\psi(\theta_0))}{s-t\psi'(\theta_0)} \quad (18)$$

as $(s-t\psi'(\theta_0))/\sqrt{t\psi''(\theta_0)} \to \infty$; hence, as $\max(s,t) \to \infty$ for $s/t \geq \psi'(\theta_0)+\varepsilon$.

Since

$$\psi'(\theta_0) < \frac{\psi(b) - \psi(\theta_0)}{b - \theta_0} < \psi'(b)$$

by the convexity of ψ, it is routinely verified that if $s/t > \psi'(b)$ then $\widehat{\theta} = b$ and, keeping this in mind, it is easy to see that

$$\frac{1}{b - \theta_0} \leq \frac{s - t\psi'(\theta_0)}{(\widehat{\theta} - \theta_0)s - t(\psi(\widehat{\theta}) - \psi(\theta_0))} \leq \frac{\psi'(\widehat{\theta}) - \psi'(\theta_0)}{I(\widehat{\theta}, \theta_0)},$$

the latter inequality because the indicated ratio is decreasing in s. For $\psi'(\theta_0) + \varepsilon \leq s/t \leq \psi'(b)$ equality holds in the right-hand inequality. Using these facts it follows from (18) that

$$\begin{aligned}\log \int_a^{\theta_0} &\exp(ys - t\psi(y))\lambda_0(y)W(y)dy \\ &= \log\left(\frac{\lambda_0(\theta)W(\theta_0)I(\widehat{\theta}, \theta_0)\exp(\theta_0 s - t\psi(\theta_0))}{(\psi'(\widehat{\theta}) - \psi'(\theta_0))((\widehat{\theta} - \theta_0)s - t(\psi(\widehat{\theta}) - \psi(\theta_0)))}\right) + O(1) \quad (19)\end{aligned}$$

uniformly for $s/t \geq \psi'(\theta_0) + \varepsilon$ as $\max(s, t) \to \infty$ and with $o(1)$ uniformly for $\psi'(\theta_0) + \varepsilon \leq s/t \leq \psi'(b)$.

Define

$$\ell_0(s, t) = \exp\left[(\theta_0 - \widehat{\theta})s - t(\psi(\theta_0) - \psi(\widehat{\theta}))\right],$$

the likelihood ratio for rejecting $\theta \leq \theta_0$, where $\widehat{\theta} = \widehat{\theta}(s, t)$. Combining (17) and (18),

$$\begin{aligned}\log Y_0(s,t)^{-1} &= \log \ell_0(s,t)^{-1} - \log \frac{\sqrt{2\pi}(s - t\psi'(b))^+ + \sqrt{t\psi''(\widehat{\theta})}}{(s - t\psi'(b))^+ + t(\psi'(\widehat{\theta}) - \psi'(\theta_0))} + O(1) \quad (20) \\ &\geq \log \ell_0(s,t)^{-1} + O(1)\end{aligned}$$

uniformly for $s/t \geq \psi'(\theta_0) + \varepsilon$ as $\max(s, t) \to \infty$, and inspection of (20) shows that

$$\begin{aligned}\log Y_0(s,t)^{-1} &\leq \log \ell_0(s,t)^{-1} + \frac{1}{2}\log t + O(1) \quad (21) \\ &\leq \log \ell_0(s,t)^{-1} + \frac{1}{2}\log\log \ell_0(s,t)^{-1} + O(1)\end{aligned}$$

since $t \leq I(\widehat{\theta}, \theta_0)^{-1} \log \ell_0(s, t)^{-1}$ and $I(\widehat{\theta}, \theta_0)$ has a positive lower bound for $s/t \geq \psi'(\theta_0) + \varepsilon$.

As a consequence of (20) and (21),

$$\log Y_0(s,t)^{-1} \sim \log \ell_0(s,t)^{-1} \quad (22)$$

uniformly for $s/t \geq \psi'(\theta_0) + \varepsilon$ as $\max(s, t) \to \infty$. Note that since $\max(s, t) \geq$ (const.) $\log \ell_0(s, t)^{-1}$, (22) holds as $\log \ell_0(s, t)^{-1} \to \infty$ and also as $Y_0(s, t) \to 0$.

Also, if $s \geq 0$ then

$$\ell_0(s, t) \leq \exp(\theta_0 s - t\psi(\theta_0)) \leq \exp(-t\psi(\theta_0)),$$

since $\theta s - t\psi(\theta)$ vanishes at 0 and is therefore non-negative at $\widehat{\theta}$, its maximum. By (20), then,

$$\log Y_0(s,t)^{-1} \geq t\psi(\theta_0) + O(1) \quad \text{as } t \to \infty \text{ for } s \geq 0. \quad (23)$$

Now, the denominator inside the brackets in (17) can be written in the form

$$\sqrt{t}\left(\sqrt{\psi'(\widehat{\theta})}+\sqrt{2\pi t}(s/t-\psi'(b))^{+}\right)$$

and the factor $\sqrt{t}$ can be substituted for, using the relation

$$I(\widehat{\theta},\theta_0)t=\left(1+\frac{b-\theta_0}{I(b,\theta_0)}\left(\frac{s}{t}-\psi'(b)\right)^{+}\right)^{-1}\log \ell_0(s,t)^{-1},$$

so that (17) and (18) yield

$$\log Y_0(s,t)^{-1}=\log \ell_0(s,t)^{-1}+\frac{1}{2}\log\log \ell_0(s,t)^{-1}-\log G_0((s-t\psi'(b))^{+},t) \\ -\log H_0(\widehat{\theta})+O(1), \quad (24)$$

where

$$G_0((s-t\psi'(b))^{+},t)=\frac{1+\sqrt{2\pi}\left(\frac{s-t\psi'(b)}{\sqrt{t\psi''(b)}}\right)^{+}}{\left(1+\frac{b-\theta_0}{I(b,\theta_0)}(\frac{s}{t}-\psi'(b))^{+}\right)^{1/2}}$$

and

$$H_0(\widehat{\theta})=\frac{(I(\widehat{\theta},\theta_0)\psi''(\widehat{\theta}))^{1/2}\lambda_0(\theta_0)W(\theta_0)}{\sqrt{2\pi}\lambda_0(\widehat{\theta})(\psi'(\widehat{\theta})-\psi'(\theta_0))},$$

and this relation holds uniformly for $s/t\ge\psi'(\theta_0)+\varepsilon$ as $\max(s,t)\to\infty$ and with $o(1)$ uniformly for $\psi'(\theta_0)+\varepsilon\le s/t\le\psi'(b)-\varepsilon$. Furthermore, by (22)

$$\log Y_0(s,t)^{-1}-\frac{1}{2}\log\log Y_0(s,t)^{-1}=\log \ell_0(s,t)^{-1}-\log G_0((s-t\psi'(b))^{+},t) \\ -\log H_0(\widehat{\theta})+O(1) \quad (25)$$

and a $o(1)$ version holds, with the same stipulations as in (24).

Asymptotic evaluation of the reciprocal mixed likelihood ratio,

$$\ell_0(s,t)=\frac{\exp(\theta_0 s-t\psi(\theta_0)}{\int_a^b\exp(ys-t\psi(y))\lambda_0(y)dy},$$

follows easily from the fact that

$$\frac{Y_0(s,t)}{\ell_0(s,t)}=\int_a^{\theta_0}\exp((y-\theta_0)s-t(\psi(y)-\psi(\theta_0)))\lambda_0(y)W(y)dy,$$

whence by (19)

$$\log \ell_0(s,t)^{-1}=\log Y_0(s,t)^{-1}-\log\log \ell_0(s,t)^{-1}+\log\frac{\lambda_0(\theta_0)W(\theta_0)I(\widehat{\theta},\theta_0)}{\psi'(\widehat{\theta})-\psi'(\theta_0)}+O(1),$$

and using (22)

$$\log \ell_0(s,t)^{-1}=\log Y_0(s,t)^{-1}-\log\log Y_0(s,t)^{-1}+\log\frac{\lambda_0(\theta_0)W(\theta_0)I(\widehat{\theta},\theta_0)}{\psi'(\widehat{\theta})-\psi'(\theta_0)}+O(1) \quad (26)$$

uniformly for $s/t\ge\psi'(\theta_0)$ as $\max(s,t)\to\infty$ and with $o(1)$ uniformly for $\psi'(\theta_0)+\varepsilon\le s/t\le\psi'(b)$ (and also as $Y_0(s,t)\to 0$ or $\ell_0(s,t)\to 0$).

Let $\lambda_{s,t}([\theta,\theta'])$ denote the $\lambda_{s,t}$-probability of the interval $[\theta,\theta']$. The next result provides an estimate which is useful in Section 2.

Lemma 2. *Under (A1) and (A2), for arbitrary* $s_1 \le 0$ *and* $\theta^* \in (\theta_0, 0)$ *there exist* $A > 0$ *and* $0 < \rho < 1$ *such that*

$$\lambda_{s,t}([a, \theta^*]) \le A(Y_0(s,t))^\rho$$

for $t \ge 1$, $s \ge s_1$.

Proof: Choose $\theta^* < v < w < 0$. Using the convexity of ψ it is easy to see that for all $s \ge 0$, $t \ge 1$, letting $\rho = (\psi(\theta^*) - \psi(v))/(\psi(a) - \psi(v))$ (and noting $0 < \rho < 1$)

$$(v - \theta^*)s - t(\psi(v) - \psi(\theta^*)) \ge \rho\{(v - a_1)s - t(\psi(v) - \psi(a_1))\}, \tag{27}$$

where a_1 is as defined in the paragraph following (A2). Also, for $s_1 \le s < 0$, $t \ge 1$, the right-hand side of (27) can exceed the left-hand side by at most $[\rho(v-a_1)-(v-\theta^*)]s_1 = B > 0$, and combining the two cases leads to

$$\frac{f_{vt}}{f_{\theta^* t}} \exp(B) \ge \left(\frac{f_{vt}}{f_{a_1 t}}\right)^\rho,$$

or, equivalently,

$$f_{\theta^* t} \le \exp(B) f_{a_1 t}^\rho f_{vt}^{1-\rho} \tag{28}$$

for $s \ge s_1$, $t \ge 1$.

Consider first the region where either $s \ge 0$, or else $t \ge s_1/\psi'(w)$ and $s \ge s_1$. For t, s in this region, the likelihood function $f_{\theta t} = \exp(\theta s - t\psi(\theta))$ is increasing for $\theta \le w$ because $\widehat{\theta} \ge w$. Hence,

$$\int_{\theta_0}^{\theta^*} f_{\theta t}\lambda_0(\theta)d\theta \le f_{\theta^* t}, \quad \int_{a_1}^{\theta_0} f_{\theta t}\lambda_0(\theta)W(\theta)d\theta \ge f_{a_1 t}\int_{a_1}^{\theta_0} \lambda_0(\theta)W(\theta)d\theta > 0,$$
$$\text{and} \quad \int_v^w f_{\theta t}\lambda_0(\theta)d\theta \ge f_{vt}\lambda_0([v,w]) > 0,$$

which combine with (28) to yield

$$\int_{\theta_0}^{\theta^*} f_{\theta t}\lambda_0(\theta)d\theta \le A_1 \left(\int_{a_1}^{\theta_0} f_{\theta t}\lambda_0(\theta)W(\theta)d\theta\right)^\rho \left(\int_v^w f_{\theta t}\lambda_0(\theta)d\theta\right)^{1-\rho}$$

for an obvious choice of $A_1 > 0$. Increasing the domain of the last integral from $[v, w]$ to $[a_1, b_1]$,

$$\lambda_{s,t}([\theta_0, u]) \le A_1(Y_0(s,t))^\rho.$$

Clearly $\lambda_{s,t}([a_1, \theta_0]) \le (\inf_{[a_1,\theta_0]} W)^{-1} Y_0(s,t)$ and, hence, allowing for the possibility that $Y_0 > 1$,

$$\lambda_{s,t}([a_1, \theta_0]) \le \left(1 + (\inf_{[a_1,\theta_0]} W)^{-1}\right)(Y_0(s,t))^\rho$$

and addition produces the stated result.

In the region where $s_1 \le s < 0$ and $1 \le t < s_1/\psi'(w)$ it is easy to see that $Y_0(s,t)$ is bounded away from 0, so that choosing A large enough makes the inequality hold here also, proving the lemma.

2 Main results

An important property of the Bayes procedures is given by

Lemma 3. *Under (A1) and (A2) there exists an $M^* > 1$ such that if the stopping risk, $r(\lambda_{s,t})$, exceeds $M^* c$, then the risk of continuing at (t, s) is smaller than the stopping risk.*

Proof: It will first be established that if $|s|$ is sufficiently large, the sign of s determines which of $Y_0(s,t)$, $Y_1(s,t)$ attains $r(\lambda_{s,t})$. Note that

$$\frac{Y_1(s,t)}{Y_0(s,t)} = \frac{\int_{\theta_1}^{b} \exp(ys - t\psi(y))\lambda_0(y)W(y)dy}{\int_{a}^{\theta_0} \exp(ys - t\psi(y))\lambda_0(y)W(y)dy}, \tag{29}$$

which is clearly increasing in s since $\theta_0 < \theta_1$. Fix

$$s = s_2 > (\theta_1 - \theta_0)^{-1} \left(\log \frac{\lambda_0(\theta_1)W(\theta_1)|\psi'(\theta_0)|}{\lambda_0(\theta_0)W(\theta_0)\psi'(\theta_1)} \right)^{-}.$$

Recall that $\psi(\theta_0) = \psi(\theta_1)$ and note that by (18) and its analog for the numerator of (29), as $t \to \infty$,

$$\begin{aligned}\log \frac{Y_1(s_2,t)}{Y_0(s_2,t)} &= (\theta_1 - \theta_0)s_2 + \log \frac{\lambda_0(\theta_1)W(\theta_1)}{\lambda_0(\theta_0)W(\theta_0)} + \log \frac{s_2 - t\psi'(\theta_0)}{t\psi'(\theta_1) - s_2} + o(1) \\ &\geq (\theta_1 - \theta_0)s_2 + \log \frac{\lambda_0(\theta_1)W(\theta_1)|\psi'(\theta_0)|}{\lambda_0(\theta_0)W(\theta_0)\psi'(\theta_1)} + o(1),\end{aligned}$$

which by the choice of s_2 is positive for large t – say, for $t \geq t_0$. Thus, since the ratio in (29) is increasing in s, it is evidently greater than 1 for $s \geq s_2$ provided that $t \geq t_0$. For $1 \leq t < t_0$, obvious estimates show that this ratio tends to $+\infty$ uniformly as $s \to +\infty$. Hence, there is an $s_1 \geq s_2$ such that if $s \geq s_1$ then the ratio is greater than 1 for all $t \geq 1$. Combining this with a similar result for negative s, there exists an s_0 such that

$$|s| \geq s_0 \quad \Rightarrow \quad \operatorname{sgn}(Y_1(s,t) - Y_0(s,t)) = \operatorname{sgn} s. \tag{30}$$

For the remainder of the proof, assume that $r(\lambda_{s,t}) = Y_0(s,t)$, the other case being similar. Then $s \geq -s_0$ and, choosing $\theta^* \in (\theta_0, 0)$, Lemma 2 gives

$$\lambda_{s,t}([a, \theta^*]) \leq A(Y_0(s,t))^{\rho}, \tag{31}$$

where $0 < \rho < 1$.

To obtain an upper bound on the risk of continuing, consider the continuation $\delta(c)$ that stops as soon as the stopping risk is less than c and chooses a terminal decision minimizing the a posteriori risk. Since the error part of the risk of this continuation is less than c, it suffices to show that

$$c \int_a^b E_\theta N(c) \lambda_{s,t}(\theta) d\theta \leq Y_0(s,t) - c,$$

where $N(c)$ is the number of observations taken by $\delta(c)$, or, equivalently,

$$\int_a^b E_\theta N(c) \lambda_{s,t}(\theta) d\theta \leq c^{-1} Y_0(s,t) - 1, \tag{32}$$

if $c^{-1}Y_0(s,t) > M^* > 1$ (suitably chosen).

Since λ_0 is continuous and positive at θ_0, it is at least ε on $[\theta_0, \theta_0 + 2\varepsilon]$ for some $0 < \varepsilon < -\theta_0/2$. Now, if $s \geq 0$, then the likelihood function $f_{\theta t}$ is increasing for $\theta \leq 0$ and thus

$$Y_0(s,t) \leq \frac{(\sup \lambda_0 W) f_{\theta_0 t}}{\lambda_0([\theta_0+\varepsilon, \theta_0+2\varepsilon]) f_{\theta_0+\varepsilon,t}} \leq \frac{(\sup \lambda_0 W) f_{\theta_0 t}}{\varepsilon^2 f_{\theta_0+\varepsilon,t}} \leq \frac{(\sup \lambda_0 W)}{\varepsilon^2} \exp(-tI(\theta_0+\varepsilon, \theta_0)).$$

Using this and a similar estimate for $Y_1(s,t)$ in case $s \leq 0$, it is seen that the stopping risk goes to zero exponentially in t as $t \to \infty$, uniformly in s. Therefore, there is a $B > 0$ such that

$$N(c) \leq B(1 + \log c^{-1}), \tag{33}$$

whence by (31)

$$\int_a^{\theta^*} E_\theta N(c) \lambda_{s,t}(\theta) d\theta \leq AB(Y_0(s,t))^\rho (1 + \log c^{-1}) \leq AB\rho^{-1}(c^{-1}Y_0(s,t))^\rho. \tag{34}$$

To estimate $E_\theta N(c)$ for $\theta^* \leq \theta \leq b$, define the stopping times

$$N(\theta, \gamma) = \inf\{n \mid (\theta - \theta_0)S_n - n(\psi(\theta) - \psi(\theta_0)) \geq \gamma\}$$

for $\gamma \geq 0$ and $\theta \in (\theta^*, b]$. By Wald's equation and the upper hound on excess over the boundary in Lorden (1970)

$$E_\theta N(\theta, \gamma) \leq \frac{\gamma}{I(\theta, \theta_0)} + 1 + \frac{(\theta - \theta_0)^2 \psi''(\theta)}{(I(\theta, \theta_0))^2} \leq B^*(1 + \gamma) \tag{35}$$

for $\theta^* \leq \theta \leq b$.

To verify that for all $\theta \in (\theta^*, b]$

$$N(c) \leq N(\theta, \gamma(\theta)) \quad \text{where} \quad \gamma(\theta) = \frac{\theta - \theta_0}{\theta^* - \theta_0} \log \frac{c^{-1}Y_0(s,t)}{\lambda_{s,t}([\theta^*, \theta])}, \tag{36}$$

it suffices to show that at time $n = N(\theta, \gamma(\theta))$

$$\int_a^{\theta_0} \exp\{(y - \theta_0)S_n - n(\psi(y) - \psi(\theta_0))\} \lambda_{s,t}(y) W(y) dy \leq c \int_a^b \exp[(y - \theta_0)S_n - n(\psi(y) - \psi(\theta_0))] \lambda_{s,t}(y) dy. \tag{37}$$

For $\theta^* \leq y \leq \theta$ the bracketed quantity on the right equals at time $n = N(\theta, \gamma(\theta))$

$$(y - \theta_0)\left[S_n - n\frac{\psi(y) - \psi(\theta_0)}{y - \theta_0}\right] \geq (y - \theta_0)\left[S_n - n\frac{\psi(\theta) - \psi(\theta_0)}{\theta - \theta_0}\right] \geq (y - \theta_0)\frac{\gamma(\theta)}{\theta - \theta_0} \geq \frac{\theta^* - \theta_0}{\theta - \theta_0}\gamma(\theta) = \log \frac{c^{-1}Y_0(s,t)}{\gamma_{s,t}([\theta^*, \theta])},$$

by the convexity of ψ and the definition of $N(\theta, \gamma(\theta))$. Hence, the right- hand side of (37) is at least $Y_0(s,t)$. The left-hand side of (37) is indeed smaller because at time

$n = N(\theta, \gamma(\theta))$ the concave function $yS_n - n\psi(y)$ is larger at $y = \theta$ than at $y = \theta_0$ and, hence, is even smaller for $y < \theta_0$, so that the quantity in braces is negative.

From (35) and (36) it follows that for $\theta \in (\theta^*, b]$

$$E_\theta N(c) \leq B^* \left(1 + \frac{b - \theta_0}{\theta^* - \theta_0} \log \frac{c^{-1} Y_0(s,t)}{\lambda_{s,t}([\theta^*, \theta])}\right),$$

whence

$$\int_{\theta^*}^{b} E_\theta N(c) \lambda_{s,t}(\theta) d\theta \leq B^* \left(1 + \frac{b - \theta_0}{\theta^* - \theta_0} \log c^{-1} Y_0(s,t)\right) + B^* \int_{\theta^*}^{b} \left(\log \lambda_{s,t}([\theta^*, \theta])^{-1}\right) \lambda_{s,t}(\theta) d\theta.$$

This last integral by the change of variable $u = \lambda_{s,t}([\theta^*, \theta])$ is seen to equal the integral of $\log u^{-1} du$ from 0 to $\lambda_{s,t}([\theta^*, b])$, which is less than 1. Combining these estimates with (34) yields

$$\int_{a}^{b} E_\theta N(c) \lambda_{s,t}(\theta) d\theta \leq AB\rho^{-1}(c^{-1} Y_0(s,t))^\rho + B^{**}(2 + \log(c^{-1} Y_0(s,t))), \tag{38}$$

where B^{**} is B times $(b - \theta_0)/(\theta^* - \theta_0)$. To prove (32), note that the right-hand side of (38) is of smaller order than $c^{-1} Y_0(s,t)$ as the latter tends to infinity. Hence, there exists an $M^* \geq 1$ such that for $c^{-1} Y_0(s,t) > M^*$ (38) implies (32) and the lemma is proved.

Lemma 4. *For every n, as $t \to \infty$*

$$\log Y_0(s + S_n, t + n)^{-1} = \log Y_0(s,t)^{-1} + (\widehat{\theta} - \theta_0) S_n - n(\psi(\widehat{\theta}) - \psi(\theta_0)) + o(1) \tag{39}$$

uniformly for $|S_n| \leq A < \infty$ and $\psi'(\theta_0) + \varepsilon \leq s/t \leq \psi'(b) - \varepsilon$, where $\widehat{\theta} = \widehat{\theta}(s,t)$.

Proof: The conditions on S_n and s/t imply that for $t \geq 2\varepsilon^{-1} A + n$

$$\psi'(\theta_0) + \frac{\varepsilon}{2} \leq \frac{s + S_n}{t + n} \leq \psi'(b) = \frac{\varepsilon}{2}.$$

Thus, (18) applies to show that the logarithm of the ratio of the numerator of $Y_0(s + Sn, t + n)$ to that of $Y_0(s,t)$ equals

$$\theta_0 S_n - n\psi(\theta_0) - \log\left(1 + \frac{S_n - n\psi'(\theta_0)}{s - t\psi'(\theta_0)}\right) + o(1) = \theta_0 S_n - n\psi(\theta_0) + o(1)$$

uniformly as $t \to \infty$. Also, (14) applies to the denominators of the risks since $s = t\psi'(\widehat{\theta})$ and $-A - n\psi'(b) \leq s + S_n - (t+n)\psi'(\widehat{\theta}) \leq A - n\psi'(\theta_0)$. Therefore, the logarithm of the ratio of the denominator of $Y_0(s + Sn, t + n)$ to that of $Y_0(s + t)$ evidently equals $\widehat{\theta} S_n - n\psi(\widehat{\theta}) + o(1)$ uniformly as $t \to \infty$ and the lemma follows.

For every $n = 1, 2, \ldots$,

$$\rho_n(\theta_1, \theta_2) = E_0 |f_{\theta_1 n} - f_{\theta_2 n}| \tag{40}$$

defines a metric on the natural parameter space which is continuous in each variable. Clearly

$$\rho_n(\theta_1, \theta_2) = E_{\theta_1} \left| \frac{f_{\theta_2 n}}{f_{\theta_1 n}} - 1 \right| = E_{\theta_2} \left| \frac{f_{\theta_1 n}}{f_{\theta_2 n}} - 1 \right| \tag{41}$$

and if A is $\mathcal{F}_n$-measurable and N is a stopping time $\leq n$ on A, then $N1\{A\}$ is $\mathcal{F}_n$-measurable and

$$|E_{\theta_1} N1\{A\} - E_{\theta_2} N1\{A\}| = |E_0 N1\{A\} f_{\theta_1 n} - E_0 N1\{A\} f_{\theta_2 n}| \\ \leq |E_0 N1\{A\}(f_{\theta_1 n} - f_{\theta_2 n})| \leq n\rho_n(\theta_1, \theta_2). \quad (42)$$

Define the family of tests $\{\delta(c)\}$ by the rules

(i) stop and reject $\theta \leq \theta_0$ if $s \geq 0$ and

$$L(\widehat{\theta}, \theta_0) H_0(\widehat{\theta}) G_0((s - t\psi'(b))^+, t)\ell_0(s,t) \leq c\sqrt{\log c^{-1}},$$

(ii) stop and reject $\theta \geq \theta_1$ if $s < 0$ and

$$L(\widehat{\theta}, \theta_1) H_1(\widehat{\theta}) G_1((s - t\psi'(a))^-, t)\ell_1(s,t) \leq c\sqrt{\log c^{-1}},$$

where ℓ_0, G_0, and H_0 are as in (24) and ℓ_1, G_1, and H_1 are defined similarly. As a consequence of (25) in the $O(1)$ version and its analog for $Y_1(s,t)$ there exists an $M^{**} > 0$ such that $\delta(c)$ stops whenever the stopping risk is $\leq M^{**}c$. Using the same relations, it is clear that there is an $\widetilde{M} > 0$ such that $\delta(c)$ continues if the stopping risk exceeds $\widetilde{M}c$, and furthermore the posterior risk of $\delta(c)$'s terminal decision is at most $\widetilde{M}c$. (In the sequel, it is assumed that M^*, satisfying Lemma 3, is chosen to be at least $\max(\widetilde{M}, L_0(0,\theta_0)^{-1}, L_0(0,\theta_1)^{-1})$, whence the above stopping-risk property of $\delta(c)$ as well as the similar property of the Bayes procedures can be stated in terms of the same M^*.)

Also define the family of GSLRT's $\{\widehat{\delta}(c)\}$ by the modification of (i) and (ii) obtained by omitting G_0 and G_1.

The main lemma needed for the proof of Theorem 1 is

Lemma 5. *Fix* $0 < b_1 < b_2 < b$. *Let*

$$\mathcal{B}_1(c) = \{(t,s) \mid b_1 \leq \widehat{\theta}(s,t) \leq b_2 \quad \textit{and} \quad Y_0(s,t) \leq M^*c\}.$$

Assume (A1) and (A2) hold and let $\delta^*(c)$ *denote a Bayes solution. As* $c \to 0$,

$$r_c(\lambda_{s,t}, \delta^*(c)) = cR(\widehat{\theta}, c^{-1}r(\lambda_{s,t})) + o(c),$$

and

$$r_c(\lambda_{s,t}, \delta(c)) = cR(\widehat{\theta}, c^{-1}r(\lambda_{s,t})) + o(c),$$

uniformly for $(t,s) \in \mathcal{B}_1(c)$, *where*

$$R(\theta, v) = inf_{N \geq 0} E_\theta \left(N + v\frac{f_{\theta_0 N}}{f_{\theta N}} \right).$$

Proof: Consider a prior $\lambda = \lambda_{s,t} \in \mathcal{B}_1(c)$ and write

$$r_c(\lambda, \delta(c)) = \int_a^b E_\theta \left[cN(c) + \widetilde{r}(\lambda_{N(c)}) \right] \lambda(\theta) d\theta, \quad (43)$$

where $\widetilde{r}$ denotes the (not necessarily optimal) posterior risk of the decision made by δ. Consider $0 < \varepsilon < \min(b_1, b - b_2)$ to be chosen later and note that since the log-likelihood function $\theta s - t\psi(\theta)$ associated with $\lambda = \lambda_{s,t}$ is concave with maximum at

$\widehat{\theta} = \widehat{\theta}(s,t)$

$$\lambda([a,\widehat{\theta}-\varepsilon]) = \frac{\int_a^{\widehat{\theta}-\varepsilon} \exp(ys - t\psi(y))\lambda_0(y)dy}{\int_a^b \exp(ys - t\psi(y))\lambda_0(y)dy} \le \frac{f_{\widehat{\theta}-\varepsilon,t}}{(1/2)\varepsilon\underline{\lambda_0} f_{\widehat{\theta}-\varepsilon/2,t}} \le 2(\varepsilon\underline{\lambda_0})^{-1}\left(\frac{f_{\widehat{\theta}-\varepsilon,t}}{f_{\widehat{\theta}t}}\right)^{1/2}$$
$$= 2(\varepsilon\underline{\lambda_0})^{-1}e^{-(1/2)tI(\widehat{\theta},\widehat{\theta}-\varepsilon)} \le 2(\varepsilon\underline{\lambda_0})^{-1}e^{-tm_1\varepsilon^2}$$

where $\underline{\lambda_0} = \min \lambda_0$ and $m_1 = (1/2)\min\psi''$ on $[0,b_2]$. Using (21) and the fact that $(t,s) \in \mathcal{B}_1(c)$,

$$\log Y_0(s,t)^{-1} \le \frac{3}{2}\log \ell_0(s,t)^{-1} + O(1) \le \frac{3}{2}tI(b_2,\theta_0) + O(1) \tag{44}$$

uniformly in s as $t \to \infty$, which combines with the estimate for $\lambda([a,\widehat{\theta}-\varepsilon])$ and a similar estimate for $\lambda([\widehat{\theta}+\varepsilon,b])$ to yield the existence of positive B and ρ such that

$$\lambda((\widehat{\theta}-\varepsilon,\widehat{\theta}+\varepsilon)) \ge 1 - B\varepsilon^{-1}c^{\rho\varepsilon^2} \tag{45}$$

for all $(t,s) \in \mathcal{B}(c)$. (Note that, as $c \to \infty$, $(t,s) \in \mathcal{B}(c)$ implies $Y_0(s,t) \to 0$, so that $\max(s,t) \to \infty$ and, hence, $t \to \infty$.) Since $\delta(c)$ stops by the time the stopping risk is less than $M^{**}c$, the bound in (33) applies, and using also the fact that $\delta(c)$ chooses a decision having risk at most M^*c, the stopping time $N(c)$ of $\delta(c)$ satisfies

$$E_\theta[cN(c) + \widetilde{r}(\lambda_{N(c)})] \le B'c(1+\log c^{-1}), \tag{46}$$

whence by (43) and (45)

$$\left| r_c(\lambda,\delta(c)) - \int_{\widehat{\theta}-\varepsilon}^{\widehat{\theta}+\varepsilon} E_\theta[cN(c) + \widetilde{r}(\lambda_{N(c)})]\lambda(\theta)d\theta \right|$$
$$\le BB'\varepsilon^{-1}c^{1+\rho\varepsilon^2}(1+\log c^{-1}) = o(c) \tag{47}$$

for all $\lambda = \lambda_{s,t}$ such that $(t,s) \in \mathcal{B}_1(c)$, upon choosing $\varepsilon = \varepsilon(c) = (\log c^{-1})^{-1/3}$, which goes to zero with c.

By the strict convexity of ψ

$$\psi'(\theta) - \frac{\psi(\theta)-\theta_0}{\theta-\theta_0}$$

is positive for $0 \le \theta \le b$ and, being continuous, has a positive minimum, say 2η. Since ψ' is uniformly continuous on $[\theta_0, b]$, there is a positive $\varepsilon_1 < \min(b_1, b-b_2)$ such that

$$\psi'(\theta) - \psi'(\theta - 2\varepsilon_1) < \eta \quad \text{for} \quad 0 \le \theta \le b, \tag{48}$$

whence

$$\psi'(\theta-2\varepsilon_1) - \frac{\psi(\theta)-\theta_0}{\theta-\theta_0} > \eta > 0 \quad \text{for} \quad 0 \le \theta \le b.$$

Using this relation and the monotonicity of the difference quotient yields for $\gamma > 0$ and $b_1 \le \widehat{\theta} \le b_2$

$$S_n \ge \frac{\gamma}{-\theta_0} + n(\psi'(\widehat{\theta}-\varepsilon_1) - \eta)$$
$$\Rightarrow S_n \ge \frac{\gamma}{y-\theta_0} + n\frac{\psi(y)-\psi(\theta_0)}{y-\theta_0} \quad \text{for} \quad 0 \le y \le \widehat{\theta}+\varepsilon_1. \tag{49}$$

Choosing $0 < h < -a, \eta/(\max_{[a,b]}\psi'')$, by (49) and Chebyshev's inequality, for $\widehat{\theta}-\varepsilon_1 \leq \theta \leq \widehat{\theta}+\varepsilon_1$

$$
\begin{aligned}
&P_\theta\left(S_n < \frac{\gamma}{y-\theta_0} + n\frac{\psi(y)-\psi(\theta_0)}{y-\theta_0} \quad \text{for some} \quad 0 \leq y \leq \widehat{\theta}+\varepsilon_1\right) \\
&\leq P_\theta\left(\exp(-h\{S_n + \frac{\gamma}{\theta_0} - n(\psi'(\widehat{\theta}-\varepsilon_1)-\eta)\}) \geq 1\right) \\
&\leq \exp\left(-\frac{h\gamma}{\theta_0} + n[\psi(\theta-h)-\psi(\theta)+h(\psi'(\theta-\varepsilon_1)-\eta)]\right) \\
&\leq \exp\left(-\frac{h\gamma}{\theta_0} + n\left[\frac{1}{2}h^2\max_{[a,b]}\psi'' - h\eta\right]\right) = Be^{-An} \quad (A, B > 0).
\end{aligned} \tag{50}
$$

Define the stopping time

$$
\overline{N} = \inf\left\{n \mid n \geq 1, \quad S_n \geq \frac{\log(2M^*/M^{**})}{y-\theta_0} + n\frac{\psi(y)-\psi(\theta_0)}{y-\theta_0} \quad \text{for} \quad 0 \leq y \leq \widehat{\theta}+\varepsilon_1\right\}.
$$

Since the log-likelihood function is concave, the maximum likelihood over $[a,\theta_0]$ at time $\overline{N}$ is at θ_0 and is a factor of at least $2M^*/M^{**}$ smaller than the minimum likelihood on $[0,\widehat{\theta}+\varepsilon_1]$. It follows easily that the risk of rejecting $\theta \leq \theta_0$ at time $\overline{N}$ is $\leq M^{**}Y_0(s,t)/2M^*\lambda_{s,t}([0,\widehat{\theta}+\varepsilon_1])$, which is less than M^*c by (45) for all $(t,s) \in \mathcal{B}_1(c)$ provided that $\log c^{-1} \geq (\rho\varepsilon_1^2)^{-1}\log(2B\varepsilon_1^{-1})$. Assume from now on that this last holds and that $\varepsilon = \varepsilon(c) \leq \varepsilon_1$. Then

$$
N(c) \leq \overline{N} \quad \text{for} \quad (t,s) \in \mathcal{B}_1(c). \tag{51}
$$

By the definition of $\overline{N}$ and (50), there exist $A, B > 0$ such that for $b_1 \leq \widehat{\theta} \leq b_2$

$$
P_\theta(\overline{N} > n) \leq Be^{-An} \quad \text{for} \quad \widehat{\theta}-\varepsilon \leq \theta \leq \widehat{\theta}+\varepsilon. \tag{52}
$$

For $m = 1, 2, \ldots$ define the events

$$
V_m = \left\{\overline{N} \leq m, \quad \max_{k \leq \overline{N}} |S_k| \leq m^2\right\}
$$

and the stopping times

$$
N_m = \min(m+1, \inf\{n \mid |S_n| > m^2\}).
$$

Note that

$$
V_m \subset \{\overline{N} < N_m\}. \tag{53}
$$

Also note that for $0 \leq \theta \leq b_2+\varepsilon_1$ and $0 < h < b-b_2-\varepsilon_1$

$$
\begin{aligned}
P_\theta\left(\max_{k \leq m} S_k > m^2\right) &\leq \sum_1^m P_\theta(S_k > m^2) = \sum_1^m P_\theta(\exp(h(S_k - m^2)) \geq 1) \\
&\leq \sum_1^m E_\theta \exp(h(S_k - m^2)) = \sum_1^m \exp(k\psi(\theta+h) - hm^2) \leq m\exp(m\psi(b) - hm^2),
\end{aligned}
$$

which goes to zero exponentially as $m \to \infty$. By this result and a similar estimate for $\max_{k \leq m}(-S_k)$,

$$
P_\theta\left(\max_{k \leq m} |S_k| > m^2\right) \leq B_1\exp(-A_1 m) \quad (A_1, B_1 > 0) \tag{54}
$$

for $0 \le \theta \le b_2 + \varepsilon_1$.

Using (51) and the fact that $\widetilde{r}(\lambda_{N(c)}) \le M^* c$, (47) yields for $\varepsilon = \varepsilon(c)$

$$\left| r_c(\lambda, \delta(c)) - \int_{\widehat{\theta}-\varepsilon}^{\widehat{\theta}+\varepsilon} E_\theta[(cN(c) + \widetilde{r}(\lambda_{N(c)}))1\{V_m\}]\lambda(\theta)d\theta \right| \\ \le c \sup_{\widehat{\theta}-\varepsilon \le \theta \le \widehat{\theta}+\varepsilon} E_\theta[(\overline{N} + M^*)1\{V_m'\}] + o(c).$$

Using the definitions of $\overline{N}$ and V_m,

$$E_\theta(\overline{N} + M^*)1\{V_m'\} \le E_\theta(\overline{N} + M^*)1\{\overline{N} > m\} + (m + M^*)P_\theta\left(\max_{k \le m} |S_k| > m^2\right).$$

As $m \to \infty$, this last goes to 0 uniformly for $\theta \in [\widehat{\theta} - \varepsilon, \widehat{\theta} + \varepsilon]$ and $b_1 \le \widehat{\theta} \le b_2$ by virtue of (52) and (54). Hence, combining this with the preceding estimate for $r_c(\lambda, \delta(c))$,

$$\left| r_c(\lambda, \delta(c)) - \int_{\widehat{\theta}-\varepsilon}^{\widehat{\theta}+\varepsilon} E_\theta[(cN(c) + \widetilde{r}(\lambda_{N(c)}))1\{V_m\}]\lambda(\theta)d\theta \right| \le a_m c + o(c) \qquad (55)$$

uniformly for $(t, s) \in \mathcal{B}_1(c)$, where $\varepsilon = \varepsilon(c) = (\log c^{-1})^{-1/3}$ and $\{a_m\}$ is a sequence decreasing to zero. (Note that the $o(c)$ term depends on m.)

To see that the expectation in (55) is essentially constant in θ, note that $\widetilde{r}(\lambda_{N(c)}) \le M^* c$ and, on V_m, $N(c) \le \overline{N} \le m$ so that (42) applies and, hence, for $\theta \in [\widehat{\theta} - \varepsilon, \widehat{\theta} + \varepsilon]$

$$\left| E_\theta(cN(c) + \widetilde{r}(\lambda_{N(c)}))1\{V_m\} - E_{\widehat{\theta}}(cN(c) + \widetilde{r}(\lambda_{N(c)}))1\{V_m\} \right| \\ \le cm\rho_m(\theta, \widehat{\theta}) + E_0|\widetilde{r}(\lambda_{N(c)})1\{V_m\}(f_{\theta m} - f_{\widehat{\theta} m})| \le c(m + M^*)\rho_m(\theta, \widehat{\theta}). \qquad (56)$$

If $\theta \in [0, b]$ and $\widehat{\theta} \in [b_1, b_2]$, then

$$\lim_{\varepsilon \downarrow 0} \sup_{|\theta - \widehat{\theta}| < \varepsilon} \rho_m(\theta, \widehat{\theta}) = 0. \qquad (57)$$

To verify (57), note that for $\varepsilon \le \min(-a, b - b_2)$

$$\rho_m(\theta, \widehat{\theta}) = E_{\widehat{\theta}} \left| \frac{f_{\theta m}}{f_{\widehat{\theta} m}} - 1 \right| \le E_{\widehat{\theta}}(e^{m\psi'(b)\varepsilon + \varepsilon|S_m|} - 1) < \infty,$$

since $|\theta - \widehat{\theta}| < \varepsilon$ implies that for $\theta, \widehat{\theta} \in [0, b]$

$$|(\theta - \widehat{\theta})S_m - m(\psi(\theta) - \psi(\widehat{\theta}))| \le \varepsilon|S_m| + m\psi'(b)\varepsilon$$

and for all $\widehat{\theta} \in [b_1, b_2]$ the moment generating function of S_m is finite on $[a, b - b_2]$. It follows by considering the cases $S_m \ge 0$ and $S_m < 0$ in which $\exp(m\psi'(b)\varepsilon + \varepsilon|S_m|)$ is stochastically largest when $\widehat{\theta} = b_2$ and $\widehat{\theta} = b_1$, respectively, that

$$\rho_m(\theta, \widehat{\theta}) \le E_{b_1}(e^{m\psi'(b)\varepsilon + \varepsilon|S_m|} - 1) + E_{b_2}(e^{m\psi'(b)\varepsilon + \varepsilon|S_m|} - 1)$$

which tends to zero with ε by monotone convergence, proving (57). Using (57) with $\varepsilon = \varepsilon(c) \to 0$, (56) and (55) yield

$$\left| r_c(\lambda, \delta(c)) - E_{\widehat{\theta}}[(cN(c) + \widetilde{r}(\lambda_{N(c)}))1\{V_m\} \cdot \lambda([\widehat{\theta} - \varepsilon, \widehat{\theta} + \varepsilon]) \right| \le a_m c + o(c),$$

whence

$$\left|r_c(\lambda,\delta(c)) - E_{\widehat{\theta}}[(cN(c)+\widetilde{r}(\lambda_{N(c)}))1\{V_m\}]\right| \le a_m c + o(c) \tag{58}$$

uniformly for $(t,s)\in\mathcal{B}_1(c)$, the last by (45) and (46).

For $(t,s)\in\mathcal{B}_1(c)$, $s\ge t\psi'(b_1)>0$, so that as $c\to 0$ (and, hence, $t\to\infty$), for $n=1,\dots,m$, $s+S_n\ge t\psi'(b_1)-m^2\ge s_0$ on $\{N_m>n\}$ and therefore, by (30) and the definition of $\delta(c)$, $\delta(c)$ attains the stopping risk on $\{N_m>N(c)\}$ by rejecting $\theta\le\theta_0$. Applying Lemma 4 then on $\{N_m>n\}$

$$\left|r(\lambda_n)-r(\lambda)\frac{f_{\theta_0 n}}{f_{\widehat{\theta}n}}\right|\le r(\lambda)\cdot o(1)=o(c),$$

whence, recalling that $N_m\le m+1$,

$$\left|\widetilde{r}(\lambda_{N(c)})-r(\lambda)\frac{f_{\theta_0 N(c)}}{f_{\widehat{\theta}N(c)}}\right| = o(c) \quad \text{on} \quad \{N_m>N(c)\} \tag{59}$$

uniformly for $\lambda=\lambda_{s,t}$, $(t,s)\in\mathcal{B}_1(c)$. By (58), (59), and the fact that

$$\left|r_c(\lambda,\delta(c))-E_{\widehat{\theta}}\left[\left(cN(c)+r(\lambda)\cdot\frac{f_{\theta_0 N(c)}}{f_{\widehat{\theta}N(c)}}\right)1\{V_m\}\right]\right|\le a_m c+o(c) \tag{60}$$

uniformly for $(t,s)\in\mathcal{B}_1(c)$.

Note that the only properties of $\delta(c)$ used to establish (60) were the bounds $M^{**}c$ and M^*c on its stopping regions and the posterior risk of its terminal decision, and the optimal choice of terminal decision when $s>s_0$. Hence, the same argument applies to $\delta^*(c)$ and yields

$$\left|r_c(\lambda,\delta^*(c))-E_{\widehat{\theta}}\left[\left(cN^*(c)+r(\lambda)\cdot\frac{f_{\theta_0 N^*(c)}}{f_{\widehat{\theta}N^*(c)}}\right)1\{V_m\}\right]\right|\le a_m c+o(c). \tag{61}$$

Define the modified stopping time

$$N^{**}(c)=\begin{cases}N^*(c) & \text{if } N^*(c)<N_m\\ \widetilde{N}(c) & \text{if } N^*(c)\ge N_m,\end{cases}$$

where $\widetilde{N}(c)$ is the smallest $n\ge 0$ (or ∞ if there is no n) such that

$$r(\lambda)\frac{f_{\theta_0 n}}{f_{\widehat{\theta}n}}\le\frac{c}{L(\widehat{\theta},\theta_0)}.$$

Note that by (53) and the fact that $N^*(c)\le\overline{N}$

$$N^{**}(c)=N^*(c) \quad \text{on} \quad V_m.$$

Therefore, using (61)

$$\left|r_c(\lambda,\delta^*(c))-E_{\widehat{\theta}}\left[cN^{**}(c)+r(\lambda)\frac{f_{\theta_0 N^{**}(c)}}{f_{\widehat{\theta}N^{**}(c)}}\right]\right|\le a_m c+o(c) \tag{62}$$

once it is shown that

$$E_{\widehat{\theta}}\left[\left(cN^{**}(c)+r(\lambda)\frac{f_{\theta_0 N^{**}(c)}}{f_{\widehat{\theta}N^{**}(c)}}\right)1\{V_m\}\right]=o(c). \tag{63}$$

Now, since $N^{**}(c) \le \overline{N}$ is clear using the definition of $\widetilde{N}(c)$, the left-hand side of (63) is at most

$$cE_{\widehat{\theta}}\overline{N}1\{V_m'\} + \frac{c}{L(b_1,\theta_0)}P_{\widehat{\theta}}(V_m') + M^* c P_{\widehat{\theta}}(V_m') + o(c),$$

the second term for the case where $N^*(c) \ge N_m$ and the third term plus $o(c)$ for the case where $N^*(c) < N_m$ in which, by an argument like the one for (59),

$$r(\lambda)\frac{f_{\theta_0 N^*(c)}}{f_{\widehat{\theta} N^*(c)}} = r(\lambda_{N^*(c)}) + o(c) \le M^* c + o(c).$$

By arguing as in the derivation of (55), the estimate for the left-hand side of (63) is shown to be $o(c)$, so that (63) holds and, hence, (62). By definition of $R(\theta, v)$, therefore,

$$r_c(\lambda, \delta * (c)) \ge cR(\widehat{\theta}, c^{-1}r(\lambda)) - a_m c + o(c). \tag{64}$$

To handle $\delta(c)$, consider a similar modification $N'(c)$ of $N(c)$, using $\widetilde{N}(c)$ if $N(c) \ge N_m$. Then similar estimates show that (60) implies the analog of (62),

$$\left| r_c(\lambda, \delta(c)) - E_{\widehat{\theta}}\left[cN'(c) + r(\lambda)\frac{f_{\theta_0 N'(c)}}{f_{\theta N'(c)}}\right]\right| \le a_m c + o(c). \tag{65}$$

Routine calculation shows that

$$\left|\frac{S_n + s}{t+n} - \frac{s}{t}\right| \le \frac{m^2 + m\psi'(b)}{t} \quad \text{on} \quad \{N_m > n\} \tag{66}$$

for $(t, s) \in \mathcal{B}_1(c)$. Now, (44) shows that $(t, s) \in \mathcal{B}_1(c)$ implies $t \to \infty$ uniformly as $c \to 0$. Hence, for sufficiently small c, (66) implies that $(S_n + s)/(t+n) \in (\psi'(b_1)/2, \psi'(b_2)/2 + \psi'(b)/2)$ and, applying the $o(1)$ version of (24), with $\widetilde{\theta}$ denoting $\widetilde{\theta}(s + S_n, t+n)$,

$$\begin{aligned}\log Y_0(s + S_n, t+n)^{-1} = \log \ell_0(s + S_n, t+n)^{-1} + \frac{1}{2}\log\log \ell_0(s + S_n, t+n)^{-1} \\ - \log H_0(\widetilde{\theta}) + o(1) \quad \text{on} \quad \{N_m > n\},\end{aligned}$$

uniformly for $(t, s) \in \mathcal{B}_1(c)$. The definition of $N(c)$ now shows that on $\{n \le N(c) < N_m\}$, $N(c)$ stops at n if and only if

$$\log Y_0(s + S_n, t+n)^{-1} \ge \log c^{-1} + \log L(\widetilde{\theta}, \theta_0) + o(1)$$

or, equivalently, using Lemma 4, the uniform continuity of $\log L(\cdot, \theta)$ on $[0, b]$, and (66), if and only if

$$\log r(\lambda)^{-1} + \log\frac{f_{\widehat{\theta}n}}{f_{\theta_0 n}} \ge \log c^{-1} + \log L(\widehat{\theta}, \theta_0) + o(1), \tag{67}$$

where the o(1) term is uniform for $(t, s) \in \mathcal{B}_1(c)$. Using (67) in the case where $N(c) < N_m$ and the definition of $\widetilde{N}(c)$ in case $N(c) \ge N_m$, it follows that there exists $\rho(c) \downarrow 1$ as $c \to 0$ such that

$$r(\lambda)\frac{f_{\theta_0 N'(c)}}{f_{\widehat{\theta} N'(c)}} \le \frac{c\rho(c)}{L(\widehat{\theta}, \theta_0)} \tag{68}$$

and

$$r(\lambda)\frac{f_{\theta_0 n}}{f_{\widehat{\theta}n}} > \frac{c}{\rho(c)L(\widehat{\theta}, \theta_0)} \quad \text{on} \quad \{N'(c) > n\} \tag{69}$$

for all $\lambda = \lambda_{s,t}$, $(t,s) \in \mathcal{B}_1(c)$.

It follows from (68) and (69) using Lemma 6 below that

$$E_{\widehat{\theta}}\left[N'(c) + c^{-1}r(\lambda)\frac{f_{\theta_0 N'(c)}}{f_{\widehat{\theta}N'(c)}}\right] \leq R(\widehat{\theta}, c^{-1}r(\lambda)) + \frac{2(\rho(c)-1)}{L(\widehat{\theta},\theta_0)}$$
$$= R(\widehat{\theta}, c^{-1}r(\lambda)) + o(1)$$

uniformly since $L(\widehat{\theta},\theta_0) \geq L(b_1,\theta_0)$. Hence, by (65)

$$r(\lambda,\delta(c)) \leq R(\widehat{\theta}, c^{-1}r(\lambda)) + a_m c + o(c)$$

which combines with (64) and the fact that $r(\lambda,\delta^*(c)) \leq r(\lambda,\delta(c))$ to yield

$$|r(\lambda,\delta(c)) - r(\lambda,\delta^*(c))| \leq 2a_m c + o(c) = c(2a_m + g(c,m)),$$

where $g(c,m)$ goes to zero with c for fixed m. The bound on the difference of risks, divided by c, has limit supremum at most $2a_m$ as $c \to 0$ and the lemma follows upon letting $m \to \infty$.

The next result is an extension of Lemma 1 of Lorden (1977).

Lemma 6. *Suppose that the stopping time N satisfies for some $v > 0$ and $\rho \geq 1$*

(i) $vL(\theta,\theta_0)f_{\theta_0 N} \leq \rho f_{\theta N}$ *and*

(ii) $\rho v L(\theta,\theta_0) f_{\theta_0 n} > f_{\theta n}$ *on* $\{N > n\}$, $n = 0, 1, \ldots$.

Then

$$E_\theta\left(N + v\frac{f_{\theta_0 N}}{f_{\theta N}}\right) \leq R(\theta, v) + \frac{2(\rho-1)}{L(\theta,\theta_0)}.$$

Proof: By Lemma 1 of Lorden (1977), for $w > 0$ $R(\theta,w)$ is attained by

$$T(w) = \inf\{n \mid n \geq 0, \quad wL(\theta,\theta_0)f_{\theta_0 n} \leq f_{\theta n}\}.$$

By (ii), $N \leq T(\rho v)$ and, conditioning on $\mathcal{F}_N$,

$$E_\theta\left(T(\rho v) + v\frac{f_{\theta_0 T(\rho v)}}{f_{\theta T(\rho v)}}\right) \geq E_\theta\left(N + v\frac{f_{\theta_0 N}}{f_{\theta N}}\right)$$
$$- E_\theta E_\theta\left[v\frac{f_{\theta_0 N}}{f_{\theta N}} - R\left(\theta, v\frac{f_{\theta_0 N}}{f_{\theta N}}\right)\middle| \mathcal{F}_N\right]. \quad (70)$$

Now, if $vL(\theta,\theta_0)f_{\theta_0 N} > f_{\theta N}$, then

$$v\frac{f_{\theta_0 N}}{f_{\theta N}} - R\left(\theta, v\frac{f_{\theta_0 N}}{f_{\theta N}}\right) \leq \frac{\rho}{L(\theta,\theta_0)} - R(\theta, L^{-1}(\theta,\theta_0)) = \frac{\rho-1}{L(\theta,\theta_0)}$$

by (i) and the obvious monotonicity of $R(\theta,\cdot)$. (Note that $R(\theta,L^{-1}) \geq R(\theta,w) = w$ for $w < L^{-1}$ since $T(w) \equiv 0$, whence $R(\theta,L^{-1}) \geq L^{-1}$ and, the reverse being obvious, $R(\theta,L^{-1}) = L^{-1}$.) In case $vL(\theta,\theta_0)f_{\theta_0 N} \leq f_{\theta N}$, $R(\theta, vf_{\theta_0 N}/f_{\theta N})$ is attained by stopping, hence equals $vf_{\theta_0 N}/f_{\theta N}$. Putting the cases together and using (70),

$$E_\theta\left(T(\rho v) + v\frac{f_{\theta_0 T(\rho v)}}{f_{\theta T(\rho v)}}\right) \geq E_\theta\left(N + v\frac{f_{\theta_0 N}}{f_{\theta N}}\right) - \frac{\rho-1}{L(\theta,\theta_0)}.$$

Since $T(\rho v)$ attains $R(\theta, \rho v)$,

$$E_\theta\left(T(\rho v)+\rho v\frac{f_{\theta_0 T(\rho v)}}{f_{\theta T(\rho v)}}\right)\le E_\theta\left(T(v)+\rho v\frac{f_{\theta_0 T(v)}}{f_{\theta T(v)}}\right)$$
$$\le E_\theta\left(T(v)+v\frac{f_{\theta_0 T(v)}}{f_{\theta T(v)}}\right)+(\rho-1)vE_\theta\left(\frac{f_{\theta_0 T(v)}}{f_{\theta T(v)}}\right)\le R(\theta,v)+\frac{\rho-1}{L(\theta,\theta_0)}$$

using the definition of $T(v)$, and this together with the preceding relation proves the lemma.

Theorem 1. *Under assumptions (A1) and (A2),*

$$r_c(\lambda_0,\widehat{\delta}(c))=r_c(\lambda_0,\delta^*(c))+o(c)\quad \textit{as}\quad c\to 0,$$

where $\delta^*(c)$ is a Bayes solution with respect to λ_0 and c. Furthermore, if $\{\underline{\delta}(c)\}$ is a family of GSLRT's defined by (1) with $\gamma = c\sqrt{\log c^{-1}}$, then

$$r_c(\lambda_0,\underline{\delta}(c))=r_c(\lambda_0,\delta^*(c))+O(c)\quad \textit{as}\quad c\to 0$$

for every W satisfying (A1) and every λ_0 positive and continuous on $[a_1,b_1]\subset[a,b]$ satisfying $(a_1,b_1)\supset[\theta_0,\theta_1]$.

Proof: Let $\tau(c)=\min(N(c),N^*(c))$. Conditioning on $\mathcal{F}_{\tau(c)}$,

$$\begin{aligned}r_c(\lambda_0,\delta(c))-r_c(\lambda_0,\delta^*(c))&=E_{\lambda_0}\left(r_c(\lambda_{\tau(c)},\delta(c))-r_c(\lambda_{\tau(c)},\delta^*(c))\right)\\&\le P_{\lambda_0}(\lambda_{\tau(c)}\in\widetilde{\mathcal{B}}(c))E_{\lambda_0}\left[r_c(\lambda_{\tau(c)},\delta(c))-r_c(\lambda_{\tau(c)},\delta^*(c))\big|\,\lambda_{\tau(c)}\in\widetilde{\mathcal{B}}(c)\right]\\&\quad+P_{\lambda_0}(\lambda_{\tau(c)}\notin\widetilde{\mathcal{B}}(c))E_{\lambda_0}\left[r_c(\lambda_{\tau(c)},\delta(c))-r_c(\lambda_{\tau(c)},\delta^*(c))\big|\,\lambda_{\tau(c)}\notin\widetilde{\mathcal{B}}(c)\right],\end{aligned}\tag{71}$$

where $\widetilde{\mathcal{B}}(c)=$ the set of $\lambda_{s,t}$'s such that $\widehat{\theta}(s,t)\in(b_0,-b_1)\cup(b_1,b_2)$ and $r(\lambda_{s,t})\le M^*c$ for $[b_0,-b_1]\subset(a,0)$, $[b_1,b_2]\subset(0,b)$.

Noting that $r(\lambda_{\tau(c)})\le M^*c$, it is clear that $r_c(\lambda_{\tau(c)},\delta(c))\le 2M^*c$, since the risk of error is at most M^*c and the integrated sampling cost is at most what would be required to reduce the stopping risk by a factor of M^* – and the latter is less than M^*c by the proof of Lemma 3. Using this upper bound on the posterior risk of $\delta(c)$ at time $\tau(c)$ and applying Lemma 5 and its negative θ analog to deal with the case $\lambda_{\tau(c)}\in\widetilde{\mathcal{B}}(c)$, (71) yields

$$r_c(\lambda_0,\delta(c))-r_c(\lambda_0,\delta^*(c))\le 2M^*cP_{\lambda_0}(\lambda_{\tau(c)}\notin\widetilde{\mathcal{B}}(c))+o(c),$$

so that

$$r_c(\lambda_0,\delta(c))-r_c(\lambda_0,\delta^*(c))=o(c),\tag{72}$$

once it is shown that

$$\limsup_{c\to 0}P_{\lambda_0}(\lambda_{\tau(c)}\notin\widetilde{\mathcal{B}}(c))\le 2\{1-\lambda_0((b_0,-b_1)\cup(b_1,b_2))\},\tag{73}$$

b_0, b_1, and b_2 being arbitrarily close to a, 0, and b, respectively.

To prove (73), note that by an argument similar to (50) it can be shown that for $\varepsilon>0$

$$P_\theta(|S_n-n\psi'(\theta))\ge n\varepsilon+A\quad\text{for some}\quad n\ge 1)\to 0\quad\text{as}\quad A\to\infty,$$

uniformly for $\theta \in [b_3, b_4] \subset (b_1, b_2)$. Combine with the fact that for some $\eta > 0$

$$\max(|S_{\tau(c)}|, \tau(c)) \geq \eta \log c^{-1}$$

for sufficiently small c, easily verified along the lines of (22) and the sentence following it, to conclude that

$$P_{\lambda_0}(\lambda_{\tau(c)} \notin \widetilde{\mathcal{B}}(c)) \to 0 \quad \text{as} \quad c \to 0,$$

uniformly on any subinterval $[b_3, b_4]$ of (b_1, b_2). Using this and a similar relation for negative θ, (73) follows routinely and, hence, (72) is proved. The proof of the first part of the theorem will he completed by showing that

$$r_c(\lambda_0, \widehat{\delta}(c)) \leq r_c(\lambda_0, \delta(c)) + o(c). \tag{74}$$

Comparing the definitions of $\delta(c)$ and $\widehat{\delta}(c)$, it is easy to see that G_0 and G_1 are larger than 1 when $\max(|s|, t)$ is sufficiently large. Hence, for sufficiently small c the continuation region of $\delta(c)$ contains that of $\widehat{\delta}(c)$, so that the sample sizes of the latter are less than or equal to those of $\delta(c)$. Furthermore, for $\theta \leq \theta_0$

$$P_\theta(\widehat{\delta}(c) \quad \text{rejects} \quad \theta \leq \theta_0) - P_\theta(\delta(c) \quad \text{rejects} \quad \theta \leq \theta_0) \leq P_\theta(\widehat{\theta}_{\widehat{N}(c)} = b)$$

and, using the definition of $\widehat{\delta}(c)$ and the concavity of the log-likelihood function,

$$\begin{aligned} P_\theta(\widehat{\theta}_{\widehat{N}(c)} = b) &\leq P_\theta\left(\frac{f_{\theta_0 n}}{f_{bn}} \leq \gamma \quad \text{and} \quad \widehat{\theta}_n = b \quad \text{for some} \quad n \geq 1\right) \\ &\leq P_\theta\left(\left(\frac{f_{\theta_0 n}}{f_{bn}}\right)^{\frac{b-\theta_0}{b-\theta}} \leq \gamma \quad \text{for some} \quad n \geq 1\right) \\ &\leq \gamma^{\frac{b-\theta}{b-\theta_0}} \end{aligned}$$

where $\gamma = c\sqrt{\log c^{-1}}/L(b, \theta_0)H_0(b)$. Thus,

$$\begin{aligned} &\int_a^{\theta_0} \left\{P_\theta(\widehat{\delta}(c) \quad \text{rejects} \quad \theta \leq \theta_0) - P_\theta(\delta(c) \quad \text{rejects} \quad \theta \leq \theta_0)\right\} \lambda_0(\theta) W(\theta) d\theta \\ &\qquad\qquad \leq \max_{[a,\theta_0]}(\lambda W) \cdot \frac{b-\theta_0}{\log \gamma^{-1}} \cdot \gamma \leq \text{const.}\left(\frac{c}{\sqrt{\log c^{-1}}}\right). \end{aligned}$$

Using a similar result for $\theta \geq \theta_1$ as well as the comparison of sample sizes made above, (74) is established and the first part of the theorem is proved.

To prove the $O(c)$ part of the theorem, it will suffice to show that for the Bayes decision problem on $[a_1, b_1]$ specified by W and λ_0,

$$\underline{\delta}(c) \text{ has } O(c) \text{ error risk} \tag{75}$$

and the stopping rule $\underline{N}(c)$ of $\underline{\delta}(c)$ satisfies

$$\underline{N}(c) \leq T(Qc) \quad \text{for some} \quad Q > 0 \tag{76}$$

where $T(\gamma)$ stops when the stopping risk is less than or equal to γ. To see that (75) and (76) suffice, observe that the proof of Lemma 3 requires only trite changes if the continuation considered is changed from $T(c)$ to $T(Qc)$. Hence, there is an $M^* > 1$ such that if the stopping risk equals M^*c, then it pays to continue via $T(Qc)$. Evidently the cost of reducing the stopping risk by a factor of M^*/Q is less than M^*c and, therefore,

$$E_{\lambda_{T(M^*c)}}[T(Qc) - T(M^*c)] \le M^*$$

because at time $T(M^*c)$ continuing via $T(Qc)$ takes no longer than reducing the stopping risk by a factor of M^*/Q. Taking the *a priori* expectations of both sides of the last relation yields

$$E_{\lambda_0}T(Qc) - E_{\lambda_0}T(M^*c) \le M^*,$$

whence by (76) and the fact that the Bayes stopping rule, $N^*(c)$, stops no sooner than $T(M^*c)$,

$$E_{\lambda_0}\underline{N}(c) - E_{\lambda_0}N^*(c) \le M^*.$$

This relation and (75) imply that $\underline{\delta}(c)$ is indeed $O(c)$-Bayes.

The proof will be completed by showing that the GSLRT $\underline{\delta}(c)$ satisfies the sufficient conditions (75) and (76). To prove (75), it suffices to show that

$$\delta(c) \text{ has } O(c) \text{ error risk with respect to } \underline{\lambda}, \tag{77}$$

where $\underline{\lambda}$ is the uniform density on $[a, b]$. (The loss function W defined on $[a_1, b_1]$ is understood to be extended to $[a, b]$, if necessary, so as to still satisfy (A1).) Note that (77) implies (75) because of

$$\int_{a_1}^{\theta_0} P_\theta(\underline{\delta}(c) \text{ errs})\lambda_0(\theta)W(\theta)d\theta \le (\max\lambda_0)(b-a)\int_{a_1}^{\theta_0} P_\theta(\underline{\delta}(c) \text{ errs})\underline{\lambda}(\theta)W(\theta)d\theta$$

and a similar relation for $[\theta_1, b]$. Now, the modification $\delta(c)$ of $\underline{\delta}(c)$ that uses the correction factors G_0 and G_1 when s/t is outside $[\psi'(a), \psi'(b)]$ approximates the $\underline{\lambda}$ stopping risk to within $O(1)$ and has an Mc upper bound on the $\underline{\lambda}$ posterior risk of its terminal decision, just as in the first part of the proof. Also, the same argument used for (74) shows that the $\underline{\lambda}$ error risk of $\underline{\delta}(c)$ is at most $o(c)$ larger than that of $\delta(c)$, and (77) follows, proving (75).

To prove (76), replace h_0, h_1 by a lower bound $d > 0$ in (1) and note that this change cannot decrease the sample size. This GSLRT with constant d on $[a, b]$ stops no later than one restricted to $[a_1, b_1]$ with the same constant. Finally, the latter rule is seen to stop no later than some $T(Qc)$, by comparing it with a G_0, G_1-modification, using the fact that the factors G_0 and G_1 have positive lower bounds. Thus, (76) is established and the proof of Theorem 1 is complete.

Remarks.

1. Results like those in Theorem 1 can be obtained for different W's, e.g. W vanishing at θ_0 and θ_1 with positive one-sided derivatives, $W'(\theta_0)$ and $W'(\theta_1)$. The asymptotics leading to (18) extend straightforwardly and lead in the case just mentioned to replacing $W(\theta_0)$ by $W'(\theta_0)$ and squaring the denominator in the final expression in (18). This leads to GSLRT's with $\gamma = c(\log c^{-1})^{3/2}$ (as in Fushimi's work) and $h_i(\theta)$ multiplied by $|\psi'(\widehat{\theta}) - \psi'(\theta_i)|/I(\widehat{\theta}, \theta_i)$, with $W'(\theta_i)$ replacing $W(\theta_i)$.
2. The choice of h_0, h_1 for the $o(c)$-Bayes GSLRT's is quite possibly discontinuous at $\widehat{\theta} = 0$, i.e. the upper and lower stopping boundaries meet the t-axis at different points (whose distance is constant for all c). It is easy to see that if both boundaries are extended by extending the definition of $H_0(\widehat{\theta})$ to include negative

$\widehat{\theta}$ and similarly extending $H_1(\widehat{\theta})$, then the boundaries intersect always within a strip $|s| \leq s_1$ as $c \to 0$. Using the intersecting stopping boundaries and any terminal decision rule that agrees with sgn s outside the strip (e.g. one based on the ray from the origin to the point of intersection) is still $o(c)$-Bayes because such a procedure satisfies the sufficient conditions in the remark following Corollary 2.

The next corollary is in the spirit of Schwarz's (1962) results on asymptotic shapes of Bayes continuation regions. For given λ_0, let $\mathcal{B}^*(c)$ denote the region of the (t, s) plane where the (Bayes) continuation risk is less than the stopping risk. Following Schwarz, we consider the intersection of this region with the rays $\{(t,s)|\ s = t\psi'(\widehat{\theta})\}$, $a \leq \widehat{\theta} \leq b$. Let $(\mathcal{B}^*(\widehat{\theta}, c), \psi'(\widehat{\theta})\mathcal{B}^*(\widehat{\theta}, c))$ denote a boundary point of $\mathcal{B}^*(c)$ on the $\widehat{\theta}$ ray. It is also helpful to define for arbitrary g on $[a, b]$

$$\mathcal{R}(g(\widehat{\theta})) = \{(t,s)|\ r(\lambda_{s,t}) > g(\widehat{\theta}(s,t))\}.$$

Corollary 1. *Under (A1) and (A2), as* $c \to 0$

$$\mathcal{B}^*(\widehat{\theta}, c) = A(\widehat{\theta}) + B(\widehat{\theta})\left(\log c^{-1} - \frac{1}{2}\log\log c^{-1}\right) + O(1) \tag{78}$$

uniformly for $a \leq \widehat{\theta} \leq b$, *and with* $o(1)$ *in place of* $O(1)$ *uniformly for* $\widehat{\theta} \in [a+\varepsilon, -\varepsilon] \cup [\varepsilon, b-\varepsilon]$, $\varepsilon > 0$, *where*

$$A(\widehat{\theta}) = \begin{cases} I(\widehat{\theta}, \theta_0)^{-1} \log L(\widehat{\theta}, \theta_0) H_0(\widehat{\theta}) & \textit{if } \widehat{\theta} \geq 0 \\ I(\widehat{\theta}, \theta_1)^{-1} \log L(\widehat{\theta}, \theta_1) H_1(\widehat{\theta}) & \textit{if } \widehat{\theta} < 0 \end{cases}$$

$$B(\widehat{\theta}) = \begin{cases} I(\widehat{\theta}, \theta_0)^{-1} & \textit{if } \widehat{\theta} \geq 0 \\ I(\widehat{\theta}, \theta_1)^{-1} & \textit{if } \widehat{\theta} < 0. \end{cases}$$

Remarks. $O(1)$ approximation of $\mathcal{B}^*(c)$ in the region where $s/t \notin [\psi'(a), \psi'(b)]$ is obtainable using Lemma 3 and (25).

Proof: It suffices to consider $\widehat{\theta} \geq 0$ since the other case is similar. Since

$$\log \ell_0(t\psi'(\widehat{\theta}), t)^{-1} = tI(\widehat{\theta}, \theta_0),$$

the $O(1)$ relation follows at once from (22) and the fact that $\mathcal{R}(M^*c) \subset \mathcal{B}^*(c) \subset \mathcal{R}(c)$. The $o(1)$ result requires the application of Lemma 5 and the companion fact (whose proof is essentially the same) that for $\lambda = \lambda_{s,t} \in \mathcal{B}_1(c)$ the (Bayes) continuation risk equals

$$cR_1(\widehat{\theta}, c^{-1}r(\lambda)) + o(c),$$

where R_1 is defined like R but with the restriction $N \geq 1$ in place of $N \geq 0$. Note that $v^{-1}R_1(\widehat{\theta}, v)$ is obviously strictly decreasing in v. Thus, letting $L = L(\widehat{\theta}, \theta_0)$, if $v < \rho^{-1}L^{-1}$ for some $\rho > 1$, then

$$v^{-1}R_1(\widehat{\theta}, v) \geq \rho L R_1(\widehat{\theta}, (\rho L)^{-1}) > LR_1(\widehat{\theta}, L^{-1}) + L(\rho - 1) = 1 + L(\rho - 1),$$

so that $R_1(\widehat{\theta}, v) \geq (1+\eta)v$, where $\eta > 0$ and is independent of $\widehat{\theta}$. Therefore, if $1 < c^{-1}r(\lambda) < (\rho L(\widehat{\theta}, \theta_0))^{-1}$, then

$$cR_1(\widehat{\theta}, c^{-1}r(\lambda)) - r(\lambda) \geq \eta r(\lambda) > \eta c,$$

whence the evaluation of the continuation risk above shows it to be greater than the stopping risk. Extending this, there evidently exists a $\rho(c) \downarrow 1$ as $c \downarrow 0$ such that continuation does not pay if $(t, s) \in \mathcal{B}_1(c)$ and $1 < c^{-1}r(\lambda_{s,t}) < (\rho L(\widehat{\theta}, \theta_0))^{-1}$ (and certainly not if $c^{-1}r(\lambda) \leq 1$). Using the fact that $r(\lambda) < c/L(\widehat{\theta}, \theta_0)$ implies $r(\lambda) \leq M^* c$, it follows that

$$\mathcal{B}^*(c) \subset \mathcal{R}\left(\frac{c}{\rho(c)L(\widehat{\theta}, \theta_0)}\right) \quad \text{in the sector} \quad \varepsilon \leq \widehat{\theta} \leq b - \varepsilon. \tag{79}$$

By a similar argument using the evaluation of the Bayes risk in Lemma 5, the latter is shown to be less than the stopping risk if $(t, s) \in \mathcal{B}_1(c)$ and $c^{-1}r(\lambda_{s,t}) > \rho(c)L(\widehat{\theta}, \theta_0)^{-1}$, for some $\rho(c) \downarrow 1$ as $c \downarrow 0$, which together with Lemma 3 establishes

$$\mathcal{B}^*(c) \supset \mathcal{R}\left(\frac{c\rho(c)}{L(\widehat{\theta}, \theta_0)}\right) \quad \text{in the sector} \quad \varepsilon \leq \widehat{\theta} \leq b - \varepsilon. \tag{80}$$

(The function $\rho(c)$ can obviously be chosen to satisfy both (79) and (80).) To complete the proof, it is straightforward to derive the $o(1)$ version of (78) from (79) and (80) using (25).

A variety of sequential procedures can be shown to attain the Bayes risk to within $o(c)$, like $\delta(c)$ and $\widehat{\delta}(c)$, using the following result.

Corollary 2. *Under (A1) and (A2), the following set of conditions is sufficient for a family of tests, $\{\widetilde{\delta}(c)\}$, with continuation regions, $\widetilde{\mathcal{B}}(c)$, to attain the Bayes risk to within $o(c)$ as $c \to 0$:*

(i) $\widetilde{\mathcal{B}}(c) \subset \mathcal{R}(K_1 c)$ *for some* $K_1 > 0$.

(ii) $\widetilde{\mathcal{B}}(c) \supset \mathcal{R}(K_2 c)$ *for some* $K_2 > 0$.

(iii) There is an s^* such that $\delta(c)$ chooses a terminal decision according to $sgn(S_n)$ whenever $|S_n| > s^*$ upon stopping.

(iv) Given $\varepsilon > 0$ there is a $\rho(c) \downarrow 1$ as $c \downarrow 0$ such that

$$\mathcal{R}\left(\frac{c\rho(c)}{L(\widehat{\theta})}\right) \subset \widetilde{\mathcal{B}}(c) \subset \mathcal{R}\left(\frac{c}{\rho(c)L(\widehat{\theta})}\right)$$

in the region $\{\widehat{\theta} \in (a+\varepsilon, -\varepsilon) \cup (\varepsilon, b-\varepsilon)\}$, where $L(\widehat{\theta}) = L(\widehat{\theta}, \theta_0)$ if $\widehat{\theta} \geq 0$, $= L(\widehat{\theta}, \theta_1)$ if $\widehat{\theta} < 0$.

Remark: Letting $\mathcal{D}(c)$ denote the continuation region of $\delta(c)$, it can be shown using relation (25) and its analog for negative θ that an equivalent set of conditions is obtained by using $\mathcal{D}$ instead of $\mathcal{R}$ in (i) and (ii) and changing the inclusions in (iv) to

(iv)′ $\mathcal{D}(c\rho(c)) \subset \widetilde{\mathcal{B}}(c) \subset \mathcal{D}(c\rho(c)^{-1})$.

Also, letting $\widehat{\mathcal{D}}(c)$ denote the continuation region of $\widehat{\delta}(c)$, it suffices to use in place of (ii)

(ii)′ $\widetilde{\mathcal{B}}(c) \supset \widehat{\mathcal{D}}(K_2 c)$ for some $K_2 > 0$.

Proof of Corollary 2: By a routine argument using (24) and its analog for negative θ the ratio of the posterior risks $Y_0(s, t)$ and $Y_1(s, t)$ is shown to be bounded in every strip $|s| \leq s^*$. Therefore, conditions (ii) and (iii) suffice to guarantee that the posterior risk of the terminal decision made by $\widehat{\delta}(c)$ is at most a constant times c. Examination of the argument leading to (72), which establishes the $o(c)$-optimality of $\delta(c)$, shown that properties (i), (ii), and (iv) together with the upper bound on the posterior risk of the terminal decision are sufficient.

The fact that (ii)$'$ in place of (ii) suffices is clear from the argument for (74).

It is easy to define families of $o(c)$-Bayes procedures using Corollary 2 and the asymptotic relations of Section 1, For the examples which follow, only the definition of the upper boundary is given, i.e. the condition for stopping and rejecting $\theta \leq \theta_0$ when $\widehat{\theta} \geq 0$, it being understood that a similar rule is used to reject $\theta \geq \theta_1$ when $\widehat{\theta} < 0$.

(i)

$$\frac{f_{\theta_0 n}}{f_{\widehat{\theta} n}} \sqrt{\log \frac{f_{\widehat{\theta} n}}{f_{\theta_0 n}}} \leq h_0(\widehat{\theta}) c \log c^{-1}$$

(ii)

$$Y_0(S_n, n) \leq \frac{c}{L(\widehat{\theta}, \theta_0)}$$

(iii)

$$\widetilde{Y}_0(S_n, n) \leq \frac{c}{L(\theta_1, \theta_0)} \quad \text{where}$$

$\widetilde{Y}_0$ is defined with respect to the prior

$$\widetilde{\lambda}_0(\theta) = \frac{\lambda_0(\theta)}{L(\theta) \int_a^b \frac{\lambda_0(y)}{L(y)} dy}$$

and $L(\theta) = L(\theta, \theta_0)$ if $\theta \geq 0$, $= L(\theta, \theta_1)$ if $\theta < 0$.

(iv)

$$\lambda_0(\theta_0) W(\theta_0) \leq (c \log c^{-1}) h^*(\widehat{\theta}) \int_0^b \frac{f_{\theta n}}{f_{\theta_0 n}} \lambda_0(\theta) d\theta,$$

where $h^*(\widehat{\theta}) = |\psi'(\widehat{\theta}) - \psi'(\theta_*)| / I(\widehat{\theta}, \theta_*) L(\widehat{\theta}, \theta_*)$, $\theta_* = \theta_0$ if $\widehat{\theta} \geq 0$, $= \theta_1$ if $\widehat{\theta} < 0$. (For (iv), note that (26) still holds for $s \geq 0$ if the denominator of $\ell_0(s,t)$ is redefined by integrating over $[0, b]$ rather than $[a, b]$. This is because the ratio of the integral over $[0, b]$ to that over $[a, 0]$ is increasing in s, goes to infinity with t uniformly for $s/t \geq \varepsilon > 0$ and, for $s \geq 0$ is bounded away from zero for all $t \geq 1$.)

A final example is the mixture stopping rule

(v)

$$\underline{\lambda}_0(\theta_0) W(\theta_0) \leq (c \log c^{-1}) h^*(\theta_0) \int_0^b \frac{f_{\theta n}}{f_{\theta_0 n}} \underline{\lambda}_0(\theta) d\theta,$$

where

$$\underline{\lambda}_0(\theta) = \lambda_0(\theta) h^*(\theta) \quad \text{(normalization optional).}$$

In each of the above examples, one can use intersecting (extended) boundaries as discussed in the second remark following the proof of Theorem 1. In particular the procedure of example (iii) can be modified to "stop when the λ-stopping risk is less than or equal to $c/L(\theta_0, \theta_1)$", with terminal decision based on minimizing the $\widetilde{\lambda}_0$ or λ_0 version of the posterior risk – or simply on $\text{sgn}(s)$.

3 Boundary-crossing and error probabilities

The following extension of relation (26) of Lai and Siegmund (1977) is the key result.

Theorem 2. *Suppose $g(\cdot)$ is continuous and positive on $[0, b]$ and ν is non-lattice. Then as $\xi \downarrow 0$ the following are asymptotic:*

$$P_0(\xi) = \xi \int_0^b \frac{L(\theta, \theta_0)}{I(\theta, \theta_0)} g(\theta) d\theta,$$
$$P_1(\xi) = P_{\theta_0}\left(f_{\theta_0 n} \le \xi \int_0^b f_{\theta n} g(\theta) d\theta \quad \textit{for some} \quad n \ge 1\right),$$
$$P_2(\xi) = P_{\theta_0}\left(f_{\theta_0 n} \le \xi (\log \xi^{-1})^{-1/2} g(\theta_n^+) Q(\theta_n^+) f_{\theta_n^+ n} \quad \textit{for some} \quad n \ge 1\right),$$

where $Q(\theta) = (2\pi I(\theta, \theta_0)/\psi''(\theta))^{1/2}$ and $\theta_n^+ = (\widehat{\theta}_n)^+$, the maximum likelihood estimator restricted to $[0, b]$.

Furthermore, if $g(\theta) = \lambda_0(\theta) h(\theta)$, where λ_0 and h are continuous and positive on $[0, b]$, then all of the above are asymptotic to

$$P_3(\xi) = P_{\theta_0}\left(f_{\theta_0 n} \le \xi h(\theta_n^+) \int_0^b f_{\theta n} \lambda_0(\theta) d\theta \quad \textit{for some} \quad n \ge 1\right).$$

Proof: The fact that $P_1(\xi) \sim P_0(\xi)$ is expressed (in different notation) in relation (26) of L-S (Lai and Siegmund, 1977), which obviously holds for finite measures as well as probabilities. Let P_g denote the g-mixture of P_θ's, and

$$A_n(\xi) = \left\{ f_{\theta_0 n} \le \xi \int_0^b f_{\theta n} g(\theta) d\theta \quad \text{and} \quad \theta_n^+ \notin (0, b) \right\}, \quad n = 1, 2, \ldots.$$

Then, making the usual estimate (see L-S),

$$P_{\theta_0}\left(\bigcup_{n=1}^{\infty} A_n(\xi)\right) \le \xi P_g\left(\bigcup_{n=1}^{\infty} A_n(\xi)\right) = o(\xi) \quad \text{as} \quad \xi \downarrow 0, \tag{81}$$

because $P_g(\cup A_n) \to 0$ by an argument like the one for (73). Using (81) and the fact that $P_1(\xi) \sim P_0(\xi)$ is of order ξ,

$$P_1(\xi) \sim P_{\theta_0}\left\{ f_{\theta_0 n} \le \xi \int_0^b f_{\theta n} g(\theta) d\theta \quad \text{and} \quad \theta_n^+ \in (0, b) \quad \text{for some} \quad n \ge 1 \right\}. \tag{82}$$

Consider the event in braces in (82) and note that the inequality there cannot occur for small ξ unless $\max(S_n, n)$ is large. It follows that (16) can be applied (with $a = 0$ and $\lambda_0 = g$) and, using the fact that $nI(\theta_n^+, \theta_0) = \log(f_{\theta+n^+n}/f_{\theta_0 n})$ when $0 \le \theta_n^+ \le b$, one obtains

$$P_1(\xi) \sim P_{\theta_0}\Big\{ f_{\theta_0 n} \le \xi Q(\theta_n^+) g(\theta_n^+) (\log f_{\theta+n^+n}/f_{\theta_0 n})^{-1/2} \widetilde{\Phi}(n, \theta_n^+)$$
$$\text{and} \quad \theta_n^+ \in (0, b) \quad \text{for some} \quad n \ge 1 \Big\},$$

where $\widetilde{\Phi}(m, \theta) = (\sqrt{m\psi''(\theta)} \min(b - \theta, \theta))$. Since $gQ\widetilde{\Phi}$ is bounded away from 0 and ∞, the last relation leads easily to

$$P_1(\xi) \sim P_{\theta_0}\Big(f_{\theta_0 n} \le \xi (\log \xi^{-1})^{-1/2} Q(\theta_n^+) g(\theta_n^+) \widetilde{\Phi}(n, \theta_n^+) f_{\theta_0 n}$$
$$\text{and} \quad 0 < \theta_n^+ < b \quad \text{for some} \quad n \ge 1 \Big). \tag{83}$$

Since $\widetilde{\Phi} \le 1$, (83) and $P_1(\xi) \sim P_0(\xi)$ imply

$$P_0(\xi) \lesssim P_2(\xi), \tag{84}$$

where "$\lesssim$" means that as $\xi \downarrow 0$ the lim sup of the ratio of P_0 to P_2 is at most 1.

Fix m and in the derivation of (83) replace $g(\theta)$ by $g(\theta)/\widetilde{\Phi}(m,\theta)$, denoting the new P_0 and P_1 by $\widetilde{P}_0(\xi)$ and $\widetilde{P}_1(\xi)$. Then the following analog of (83) is obtained:

$$\widetilde{P}_1(\xi) \sim P_{\theta_0}\Big(f_{\theta_0 n} \le \xi(\log \xi^{-1})^{-1/2} Q(\theta_n^+) g(\theta_n^+) \widetilde{\Phi}(m,\theta_n^+)^{-1} \widetilde{\Phi}(n,\theta_n^+) f_{\theta_n^+ n} \quad \text{and} \quad 0 < \theta_n^+ < b \quad \text{for some} \quad n \ge 1\Big).$$

It follows that

$$\widetilde{P}_1(\xi) \gtrsim P_{\theta_0}\Big(f_{\theta_0 n} \le \xi(\log \xi^{-1})^{-1/2} Q(\theta_n^+) g(\theta_n^+) f_{\theta_n^+ n} \quad \text{and} \quad 0 < \theta_n^+ < b \quad \text{for some} \quad n \ge 1\Big) \tag{85}$$

because $\widetilde{\Phi}(n,\theta)$ is increasing in n and for sufficiently small ξ the event on the right-hand side of (85) cannot occur unless $n \ge m$ (the conditions $n < m$ and $0 < \theta_n^+ < b$ obviously implying a positive lower bound on $f_{\theta_0 n}/f_{\theta_n^+ n}$).

It is clear that the right-hand side of (85) is smaller than $P_2(\xi)$ by at most the following (two-term) sum

$$\sum_{\theta=0,b} P_{\theta_0}\Big(f_{\theta_0 n} \le \xi(\log \xi^{-1})^{-1/2} Q(\theta) g(\theta) f_{\theta n} \quad \text{for some} \quad n \ge 1\Big),$$

which is clearly of order $\xi(\log \xi^{-1})^{-1/2}$ by the usual estimate. Since $P_2(\xi)$ itself is of order at least ξ by (84), evidently the right-hand side of (85) is asymptotic to $P_2(\xi)$ and, using $\widetilde{P}_1(\xi) \sim \widetilde{P}_0(\xi)$,

$$\widetilde{P}_0(\xi) \gtrsim P_2(\xi). \tag{86}$$

Comparing the definitions of $\widetilde{P}_0(\xi)$ (based on $g(\theta)/\widetilde{\Phi}(m,\theta)$) and $P_0(\xi)$ (based on $g(\theta)$), note that as $m \to \infty$, $\widetilde{\Phi}(m,\theta) \to 1$ on $(0,b)$, whence

$$\int_0^b \frac{L(\theta,\theta_0)}{I(\theta,\theta_0)} \frac{g(\theta)}{\widetilde{\Phi}(m,\theta)} d\theta \to \int_0^b \frac{L(\theta,\theta_0)}{I(\theta,\theta_0)} g(\theta) d\theta$$

by the bounded convergence theorem. Since (86) holds for every m, letting $m \to \infty$ shows that $P_0(\xi) \gtrsim P_2(\xi)$ and, the reverse having been established in (84), $P_0(\xi) \sim P_2(\xi)$.

The fact that $P_3(\xi) \sim P_1(\xi)$ if $g(\theta) = \lambda_0(\theta) h(\theta)$ is an easy consequence of the relation

$$\int_0^b \frac{f_{\theta n}}{f_{\theta_0 n}} \lambda_0(\theta) h(\theta) d(\theta) \sim h(\theta_n^+) \int_0^b \frac{f_{\theta n}}{f_{\theta_0 n}} \lambda_0(\theta) d(\theta)$$

which holds uniformly as $\max(S_n, n) \to \infty$ by (16), and also holds as either side becomes infinite (since this requires $\max(S_n, n) \to \infty$). This completes the proof of Theorem 2.

Remark: By choosing $g(\cdot)$ so that g times Q equals an arbitrary continuous positive function on $[0,b]$, one can obtain from the relation $P_2(\xi) \sim P_0(\xi)$ the asymptotic evaluation of the probability of crossing a certain class of boundaries in the (n, S_n)

plane. The result can, of course, be shifted from $[0, b]$ to an arbitrary closed subinterval of the interior of the natural parameter space. This class of boundaries has segments outside the sector where $0 < \widehat{\theta} \le b$ which are straight lines. Within the sector, variation of the choice of g produces $O(1)$ deformation of the boundaries. Note that the same class of boundaries is obtained from mixture stopping rules supported on $[0, b]$.

The asymptotic evaluation of error probabilities comes from

Corollary 3. *Under (A1) and (A2), the GSLRTs defined by* (1) *with* $\gamma = c\sqrt{\log c^{-1}}$ *satisfy*

$$P_{\theta_0}(\textit{reject } \theta \le \theta_0) \sim c \log c^{-1} \int_0^b \frac{L(\theta, \theta_0)}{I(\theta, \theta_0)} \frac{h(\theta)}{Q(\theta)} d\theta$$

as $c \to 0$ *provided that* ν *is non-lattice. A similar relation holds for* $P_{\theta_1}(\textit{error})$.

Proof: Let A and B denote the upper and lower boundaries, respectively, of the GSLRT. Using Theorem 2 and its proof along with (1),

$$\begin{aligned} P_{\theta_0}(\text{ever cross } B) &= P_{\theta_0}\left(f_{\theta_0 n} \le c\sqrt{\log c^{-1}} h(\widehat{\theta}_n) f_{\widehat{\theta}_n n} \quad \text{and} \quad \widehat{\theta}_n \ge 0 \quad \text{for some} \quad n \ge 1\right) \\ &\sim c \log c^{-1} \int_0^b \frac{L(\theta, \theta_0)}{I(\theta, \theta_0)} \frac{h(\theta)}{Q(\theta)} d\theta, \end{aligned} \tag{87}$$

and it clearly suffices to show that

$$P_{\theta_0}(\text{cross } B \text{ after crossing } A) = o(c \log c^{-1}). \tag{88}$$

Using (16) for $a = 0$ as was done in the proof of Theorem 2, it is clear that in the region of the (n, S_n) plane where $f_{\theta_0 n} \le c\sqrt{\log c^{-1}} h(\widehat{\theta}_n) f_{\widehat{\theta}_n n}$ and $\widehat{\theta}_n \ge 0$, the mixed likelihood ratio

$$\int_0^b \frac{L(\theta, \theta_0)}{I(\theta, \theta_0)} \frac{h(\theta)}{Q(\theta)} d\theta$$

is bounded below by $\eta(c \log c^{-1})$, where $\eta > 0$. Hence, by considering the time of first crossing B,

$$P_{\theta_0}(\text{cross } B \text{ after crossing } A) \le \eta^{-1} c \log c^{-1} P_{h/Q}(\text{cross } B \text{ after crossing } A), \tag{89}$$

where the probability on the right-hand side is the h/Q miture of P_θ's.

Now, for $0 < \varepsilon < b$,

$$\begin{aligned} P_{h/Q}(\text{cross } B \text{ after crossing } A) &\le \int_0^\varepsilon \frac{h}{Q} + \int_\varepsilon^b \frac{h(\theta)}{Q(\theta)} P_\theta(\text{cross } B \text{ after crossing } A) d\theta \\ &\le \int_0^\varepsilon \frac{h}{Q} + \int_\varepsilon^b \frac{h(\theta)}{Q(\theta)} P_\theta(\text{ever cross } A) d\theta \\ &\lesssim \int_0^\varepsilon \frac{h}{Q} \quad \text{as } c \to 0, \end{aligned} \tag{90}$$

since for $\theta \ge \varepsilon$ the probability of ever crossing A can be shown to go to zero with c uniformly and exponentially, as in the argument for (73). Since ε can be chosen arbitrarily small in (90), (89) and (90) imply (88), proving the corollary.

Remark: The same asymptotic formula for error probabilities applies to all the examples of $o(c)$-Bayes procedures satisfying Corollary 2 of Section 2. The derivations are obtained straightforwardly by approximating their boundaries from above and below by those of $\widehat{\delta}(c)$'s or mixture stopping rules and applying Theorem 2. The argument that the lower boundary is asymptotically negligible proceeds as in the proof for $\widehat{\delta}(c)$ given in the corollary above. An asymptotic evaluation of error probabilities was obtained by Woodroofe (1976) for Schwarz's parabolic boundaries for testing the mean drift of a Wiener process.

References

Bickel, P. J. and Yahav, J. A. (1967). Asymptotically pointwise optimal procedures in sequential analysis. In *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, volume 1, pages 401–413. University of California Press.

Bickel, P. J. and Yahav, J. A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *The Annals of Mathematical Statistics*, 39:442–456.

Bickel, P. J. and Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(4):257–276.

Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.

Fushimi, M. (1965). Properties of asymptotically optimal Bayes sequential tests. *Raag Res. Notes*, 3(90).

Fushimi, M. (1967). On the rate of convergence of asymptotically optimal Bayes tests. *Rep. Stat. Appl. Res. JUSE*, 14:1–77.

Kiefer, J. and Sacks, J. (1963). Asymptotically optimum sequential inference and design. *The Annals of Mathematical Statistics*, 34:705–750.

Lai, T. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *The Annals of Statistics*, 5:946–954.

Lorden, G. (1967). Integrated risk of asymptotically Bayes sequential tests. *The Annals of Mathematical Statistics*, 38:1399–1422.

Lorden, G. (1970). On excess over the boundary. *The Annals of Mathematical Statistics*, 41(2):520–527.

Lorden, G. (1972). Likelihood ratio tests for sequential k-decision problems. *The Annals of Mathematical Statistics*, 43:1412–1427.

Lorden, G. (1977). Nearly-optimal sequential tests for finitely many parameter values. *The Annals of Statistics*, 5:1–21.

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

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41:1397–1409.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the wiener process and sample sums. *The Annals of Mathematical Statistics*, 41:1410–1429.

Schwarz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *The Annals of Mathematical Statistics*, 33(1):224–236.

Schwarz, G. (1969). A second-order approximation to optimal sampling regions. *The Annals of Mathematical Statistics*, 40:313–315.

Siegmund, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *Journal of the Royal Statistical Society, Series B (Methodological)*, 37:394–401.

Spitzer, F. (1956). A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society*, 82:323–339.

Wong, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *The Annals of Mathematical Statistics*, 39:1244–1263.

Woodroofe, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *The Annals of Probability*, 4(1):67–80.

Zerdy, G. (1975). *Asymptotic sequential procedures for normal populations*. PhD thesis, Cornell University.