Asymptotically Optimal Multistage Hypothesis Tests

Thesis by

Jay L. Bartroff

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2004
(Defended 18 May 2004)

© 2004
Jay L. Bartroff
All Rights Reserved

Acknowledgements

I would like to express my deepest gratitude to my advisor, Prof. Gary Lorden, for the wealth of opportunities he has given me – the chance to write this thesis under his supervision being just one of many. His enthusiastic support, guidance, patience, and weekly pep talks made this work possible.

I would like to thank the other members of my thesis committee – Prof. Emmanuel Candes, Prof. Robert Sherman, and Prof. David Wales – for sharing their time and knowledge with me. I would also like to acknowledge the other professors I've been fortunate enough to learn from while at Caltech, including Tom Wolff, Nikolai Makarov, David Assaf, Dirk Hundertmark, Bill Bing, and Clint Dodd.

I've had a blast at Caltech, thanks largely to the friends I've had here. These include Kimball Martin, Carlos Salazar-Lazaro, Don Chang, Jen Johnson, Gene Short, Lou Madsen, Gary Leskowitz, Rob Peters, Harvey Newmark, Hannes Helgason, Tom Lo, Cheryl Van Buskirk, Nadeem Moghul, Dave & Tia Kennedy, and Dionna Roper.

Most of all I would like to thank God and my family. My parents, Jack & Barbara, and my sister, Jeana, have been a tireless source of love and support; I dedicate this thesis to them.

Abstract

This thesis investigates variable stage size multistage hypothesis testing in three different contexts, each building on the previous.

We first consider the problem of sampling a random process in stages until it crosses a predetermined boundary at the end of a stage – first for Brownian motion and later for a sum of i.i.d. random variables. A multistage sampling procedure is derived and its properties are shown to be not only sufficient but also necessary for asymptotic optimality as the distance to the boundary goes to infinity.

Next we consider multistage testing of two simple hypotheses about the unknown parameter of an exponential family. Tests are derived, based on optimal multistage sampling procedures, and are shown to be asymptotically optimal.

Finally we consider multistage testing of two separated composite hypotheses about the unknown parameter of an exponential family. Tests are derived, based on optimal multistage tests of simple hypotheses, and are shown to be asymptotically optimal. Numerical simulations show marked improvement over group sequential sampling in both the simple and composite hypotheses contexts.

Contents

Chapter 1

Introduction

1.1 Background

Sequential hypothesis testing has been a source of interesting problems since its inception in the late 1940's. Some highlights are Wald's [32] seminal book, Chernoff's [3] development of asymptotic considerations, Schwarz's [28] theory of asymptotic shape of Bayes stopping regions for exponential families, Kiefer & Sacks' [12] extension of Chernoff's and Schwarz's work to general distributions and hypotheses, and Lorden's [20, 23] use of one-sided SPRT's that are o(cost per observation)-Bayes. The majority of the sequential literature involves tests that take data in a "one at a time" fashion, and their optimality properties are proven under the assumption that sampling costs are proportional to average sample size. But in practice it is often much more costly to carry out n single experiments than one experiment of size n. Hence a criticism of sequential testing – and perhaps a barrier to more practical applications of it – is that, in real-world situations, it is often more natural to take data in groups or *stages*.

An early example of a such a *multistage* procedure is Stein's two-stage extension of the Student's t-test [31], whose power is independent of the variance, estimated in the initial stage. This idea of using an initial stage to estimate the true state of nature and hence fix a design parameter of the procedure that follows has been used in two-stage procedures of Wald [33], Sobel [1], Hall [13] and others (see, e.g., [15]).

Schmitz [27] and Morgan & Cressie [7, 24] have proved general existence results for a large class of multistage problems. In particular, the theorems of Schmitz show that optimal multistage sampling strategies share the fundamental "renewal-type" property of optimal stopping strategies [5]: at each stage an optimal procedure behaves as if it were starting from scratch, but with the problem's parameters appropriately updated by the data already obtained. Such general results do not, however, tell us anything specific about the optimal tests and certainly not how to apply them without resorting to backward induction-type computer algorithms or artificial truncations.

The most general investigation of variable stage size multistage hypothesis testing is by Lorden [22]. Modelled after the sequential likelihood ratio tests of Schwarz [28], Lorden's tests essentially "do what the best fully-sequential test would do" in as few stages as possible. Lorden showed for simple hypotheses and separated composite hypotheses about the parameter of an exponential family that, except in a degenerate case, three stages are necessary and sufficient to achieve a sample size that is asymptotically the same as the best fully-sequential tests.

Pocock [25], DeMets [8, 9] and others have considered multistage testing explicitly for applications to medical clinical trials. These studies are more concerned with practical issues that arise in multistage medical trials than with mathematical optimality however. The methods proposed are largely ad hoc and incorporate severe restrictions, like an ad hoc number of stages and a fixed stage size. Moreover, these authors propose no alternative to the constant stage size, or *group sequential*, paradigm currently used in clinical trials.

1.2 Summary

In broad terms, this thesis investigates the structure of efficient multistage hypothesis tests in a general setting that allows variable stage size. Specifically, we consider three different but closely related problems, for which we now give a brief motivation.

A common theme in sequential hypothesis testing is that testing composite hypotheses can often be reduced to testing simple hypotheses. For example, Kiefer & Sacks [12], Lorden [20, 23], Schwarz [28], and Weiss [35] have all used this technique to reduce asymptotic optimality considerations for testing composite hypotheses to a "simple vs. simple" hypothesis test once a substantial number of observations have been taken - namely, a test of the estimated true state of nature versus the estimated true state restricted to the opposing hypothesis. Moreover, testing simple hypotheses can typically be reduced to a boundary crossing problem. For example, in testing simple hypotheses Lorden [19, 20] showed that minimizing a linear combination of sampling and error costs can be achieved asymptotically by performing a "one-sided" test, minimizing sampling costs under one hypothesis and error costs under the other, which is in turn equivalent to sampling until the likelihood ratio crosses a fixed boundary. These examples seem to suggest the following informal hierarchy:

Testing composite hypotheses

reduces to

Testing simple hypotheses

reduces to

Boundary crossing problem

The three multistage problems considered in this thesis are precisely the three levels of this hierarchy, studied in the reverse order. In Chapter 2 we consider the problem of sampling in stages a random process with known drift - first Brownian motion and later a sum of i.i.d. random variables - until it crosses a predetermined boundary, $a > 0$, at the end of a stage. The optimal, or *Bayes*, procedure is defined to be that which minimizes the *risk*, defined as a linear combination of the expected total sample size and expected number of stages. Since no closed-form Bayes solution exists, we study the problem as $a \to \infty$. We derive a family of sampling procedures around the principle of comparing the expected overshoot over the boundary, a, to the ratio, h, of the cost per stage to the cost per observation. In striking contrast with group sequential sampling, the stage sizes of these procedures decrease roughly as a sequence of successive square roots, with probability approaching one. The average number of stages used by these tests turns out to be determined by the asymptotic relationship of h to the *critical functions*,

$$h_m(a) = a^{(1/2)^m} (\log a)^{1/2-(1/2)^m},$$

which also play a key role in characterizing the number of stages, m, required by an optimal procedure. We prove not only that these sampling procedures minimize the risk to first order as $a \to \infty$, but also that their global properties are necessary for any efficient procedure. We prove these claims first for Brownian motion and then extend them to sums of i.i.d. random variables from a large class of distributions that allows large deviation and Central Limit Theorem-type approximations.

In Chapter 3 we use the optimal multistage sampling procedures of Chapter 2 to derive efficient multistage tests of simple hypotheses about the unkown parameter of an exponential family of densities. First we consider *one decision tests* of simple hypotheses, i.e., tests that aim to stop sampling and reject the alternative hypothesis as soon as possible if the null hypothesis is true, but want to continue sampling without ever stopping if the alternative hypothesis true. We define the *risk* in this case as a linear combination of the sampling cost under the null hypothesis and the probability of ever stopping under the alternative hypothesis. We show that one decision tests that are essentially the optimal multistage sampling procedures of Chapter 2 minimize this risk to second order as the costs per observation and per stage approach zero. Using combinations of these one decision tests we derive (ordinary) two decision tests of simple hypotheses and show that they asymptotically minimize the integrated risk to second order. A small-sample procedure based on these tests is proposed, and its improvement over group sequential sampling is illustrated by a numerical simulation of testing

$$\mu = -1/4 \quad \text{vs.} \quad \mu = 1/4,$$

where μ is the mean of i.i.d. normally distributed random variables with variance one.

In Chapter 4 we extend to a continuous parameter setting the ideas developed in

Chapter 3 and, using the optimal simple hypothesis tests as a guide, we design tests of composite hypotheses of the form

$$H_0 : \underline{\theta} \leq \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta_0 < \theta_1 \leq \theta \leq \overline{\theta}$$

about the parameter θ of an exponential family of densities. For a loss function w, vanishing on (θ_0, θ_1) and positive and bounded on $[\underline{\theta}, \theta_0] \cup [\theta_1, \overline{\theta}]$, and a prior Lebesgue density λ_0, continuous, positive, and bounded on $[\underline{\theta}, \overline{\theta}]$, we show that our tests minimize

$$\int_{\underline{\theta}}^{\overline{\theta}} [E_\theta(c \cdot N + d \cdot M) + w(\theta)P_\theta(\text{error})]\lambda_0(\theta)d\theta$$

to second order as the costs per observation and per stage, c and d, approach zero. Here N and M are the total number of observations and stages, respectively. Whereas the simple hypotheses problem of Chapter 3 naturally reduces to the boundary crossing problem of Chapter 2, unfortunately this composite hypotheses problem is not sufficiently well-approximated by the simple hypotheses problem to clarify considerations of second order optimality until "right before the final stage." Hence, proving that our test behaves optimally in the time leading up to the final stage requires quite intricate and technical arguments. These arguments make much use of Laplace-type expansions of the stopping risk originated by Schwarz [29] and strengthened by Lorden [23], as well as generalizations of the tools developed in Chapter 2 for proving stage-wise bounds on the random process as it is being sampled by our procedure. A small-sample procedure is also proposed, which performs significantly better than group sequential sampling in a numerical simulation of the problem of testing

$$-1 \leq \mu \leq -1/4 \quad \text{vs.} \quad 1/4 \leq \mu \leq 1,$$

where μ is the mean of i.i.d. normally distributed random variables with variance one.

1.3 Preliminaries

In this section we briefly introduce sequential hypothesis testing and give some preliminaries to the main results. For a more general introduction, see Chernoff [4], Govindarajulu [11], and Siegmund [30].

Let $X_1, X_2, \ldots$ be i.i.d. random variables with density function f. Suppose it is desired to test the hypotheses

$$H_0 : f = f_0 \quad \text{vs.} \quad H_1 : f = f_1 \tag{1.1}$$

for given densities f_0, f_1. Classical tests of these hypotheses would choose a sample size before the data are taken, then somehow decide between the hypotheses based on the observed data. It is possible to reach a decision earlier without sacrificing accuracy, however, if the data are observed sequentially and the total sample size, N, is a function of the data as they are observed and is therefore a random variable. Such random variables are called stopping times:

Definition 1.1. A random variable N taking values in $\{0, 1, 2, \ldots, \infty\}$ is a *stopping time* with respect to the sequence $X_1, X_2, \ldots$ if for every $n \geq 1$, the event $\{N = n\}$ depends only on $X_1, \ldots, X_n$ and the event $\{N = 0\}$ does not depend on the X_i.

Tests of hypotheses such as (1.1) whose sample size is determined by a stopping time N are called *sequential tests.* Note that $N \equiv k$ is allowed - i.e., fixed sample size tests satisfy this definition. An example is the Sequential Probability Ratio Test (SPRT), developed by Wald [32] during World War II. Letting

$$l_n = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)},$$

the SPRT is defined by choosing constants $0 < A < B < \infty$ and sampling until

$$A < l_n < B$$

is violated. Specifically, the SPRT will stop sampling at time

$$N = \inf\{n \geq 1 : l_n \notin (A, B)\}$$

and

$$\text{reject } H_0 \text{ if } \quad l_N \geq B$$
$$\text{reject } H_1 \text{ if } \quad l_N \leq A.$$

The values A, B determine the relevant error probabilities, $P_0(\text{reject } H_0)$ and $P_1(\text{reject } H_1)$.

Wald and Wolfowitz [34] showed that the SPRT is the best possible test of the hypotheses (1.1) in the following strong sense.

Theorem 1.2 (Wald and Wolfowitz). *Among all tests of the hypotheses (1.1) for which*

$$P_0(\textit{reject } H_0) \leq \alpha \quad \textit{and} \quad P_1(\textit{reject } H_1) \leq \beta$$

and

$$E_0 N < \infty \quad \textit{and} \quad E_1 N < \infty, \tag{1.2}$$

the SPRT with error probabilities α, β minimizes both $E_0 N$ and $E_1 N$ simultaneously.

Remark. Lorden [21] showed that the assumption (1.2) is superfluous.

Wald [32] developed the following fundamental tools to compute the operating characteristics of the SPRT.

Theorem 1.3 (Wald's Equation). *Let $X_1, X_2, \ldots$ be i.i.d. with mean $\mu = EX_1$. For any stopping time N with $EN < \infty$,*

$$E\left(\sum_{i=1}^{N} X_i\right) = \mu EN.$$

Theorem 1.4 (Wald's Likelihood Identity). *Let $X_1, X_2, \ldots$ be i.i.d. with density f, g under P_f, P_g, respectively, and let*

$$l_n = \prod_{i=1}^{n} \frac{f(X_i)}{g(X_i)},$$

the likelihood ratio. For an arbitrary event A (measurable with respect to the σ-algebra generated by N),

$$P_f(A \cap \{N < \infty\}) = E_g(l_N; A \cap \{N < \infty\}).$$

Results analogous to Theorems 1.3 and 1.4 hold for Brownian motion; see, e.g., [30].

Chapter 2

Optimal Multistage Sampling

Many problems in theoretical and applied statistics involve observing a random process until it crosses a predetermined boundary. We consider a version of this classical problem in which a random process, first Brownian motion and later a sum of i.i.d. random variables, is sampled in stages until it exceeds a boundary $a > 0$ at the end of a stage. As an example consider periodic monitoring of a pollutant in a water supply. There is a critical level for the pollutant above which some action must be taken but below which one will only decide when to test again, basing that decision on the current level.

If one incurs a fixed cost for each unit sampled and an additional fixed cost for each stage, then a natural measure of the performance of a multistage sampling procedure is the sum of these costs upon first crossing the boundary. In this chapter we describe a family of sampling procedures and show they are first-order optimal as $a \to \infty$.

Many aspects of the boundary-crossing or "first-exit" problem are well-studied. The powerful methods of renewal theory address successive "exits" and the time between such events (see [10], pages 358-388). Lorden [18] obtained sharp, uniform bounds for the excess over the boundary of random walks. Siegmund [30] discusses further applications in sequential analysis.

Schmitz [27] and Morgan & Cressie [7, 24] have proven general existence results for a large class of multistage sampling problems. In particular, the theorems of Schmitz show that a Bayes sampling strategy does exist for the problem considered here and that the optimum has the "renewal-type" property that at each stage it behaves as if it were starting from scratch, given the data so far. But these authors do not propose specific procedures, and though there is an extensive literature dealing with fully-sequential (one-at-a-time) sampling, there have been few investigations of the performance of procedures that vary the sample size from stage to stage.

The families of procedures, δ_m and $\hat{\delta}_m$, constructed below are shown to be first-order asymptotically optimal in Theorems 2.8 and 2.15. They have variable stage sizes which decrease roughly as a sequence of successive square roots, while the average number of stages required is determined by the ratio of the cost per stage divided to the cost per unit time in relation to a family of critical functions, h_m, defined below. These critical functions define "critical bands" - i.e., regions of the first quadrant which are closely related to how close any efficient procedure can be to the boundary after each stage of sampling; Lemmas 2.7 and 2.14 give precise lower bounds on this distance. Theorems 2.9 and 2.16 then provide converse statements to the optimality of $\delta_m, \hat{\delta}_m$, showing that any competing procedure must use at least as many stages, and the sooner it deviates from the "schedules" of Lemmas 2.7 and 2.14, the worse its performance.

2.1 Procedures for Brownian Motion

Let $X(t)$ be Brownian motion with known drift $\mu > 0$ and variance one per unit time. Define a *multistage sampling rule* T to be a sequence of nonnegative random variables $(T_1, T_2, \ldots)$ such that, for $k \geq 1$

$$T_{k+1} \cdot 1\{T_1 + \cdots + T_k \leq t\} \in \mathcal{E}_t \quad \text{for all } t \geq 0, \tag{2.1}$$

where $\mathcal{E}_t$ is the class of all random variables determined by $\{X(s) : s \leq t\}$. The interpretation of (2.1) is that by the time $T^k \equiv T_1 + \cdots + T_k$, the end of the first k stages, an observer who knows the values $\{X(s) : s \leq T^k\}$ also knows the value of T_{k+1}, the size of the $(k+1)$st stage. By a convenient abuse of notation, we will also let T denote the total sampling time, T^M, where $M = \inf\{m \geq 1 : X(T^m) \geq a\}$, the total number of stages required to cross the boundary a. We will then describe a *multistage sampling procedure* by the pair $\delta = (T, M)$. When there is no confusion as to which sampling procedure is being used, the shorthand $X_k = X(T^k)$, $X_0 = 0$ will be employed. We will also write $T(a), M(a)$ when we wish to emphasize the initial distance to the boundary, a.

Let $c, d > 0$ denote the cost per unit time and cost per stage, respectively, and consider the problem of finding the multistage sampling procedure (T, M) that minimizes

$$c \cdot ET + d \cdot EM.$$

Dividing through by c, this is seen to be equivalent to minimizing

$$ET + h \cdot EM, \tag{2.2}$$

where $h = d/c$. By Wald's equation,

$$ET = EX(T)/\mu = a/\mu + E(X(T) - a)/\mu \geq a/\mu, \tag{2.3}$$

so the procedure that minimizes

$$E(T - a/\mu) + h \cdot EM \tag{2.4}$$

also minimizes (2.2), and using (2.4) instead of (2.2) will also lead to a more refined asymptotic theory.

To describe a procedure that asymptotically minimizes (2.4) to first-order, it suffices to consider sequences $\{(a, h)\}$ such that $a \to \infty$. We are interested in problems where optimal procedures use a bounded number of stages and it turns out that this requires

$$h > a^{\varepsilon}$$

for some $\varepsilon > 0$. It will turn out that good procedures use m stages (almost always) if, as $a \to \infty$,

$$a^{(1/2)^m}(\log a)^{1/2-(1/2)^m} \ll h \ll a^{(1/2)^{m-1}}(\log a)^{1/2-(1/2)^{m-1}}, \tag{2.5}$$

where "$\ll$" means asymptotically of smaller order. We therefore define the *critical functions*

$$h_m(a) = a^{(1/2)^m}(\log a)^{1/2-(1/2)^m}$$

for $m = 1, 2, \ldots$ and $a \geq 1$, with $h_0(a) \equiv a$. An essentially complete description of how to achieve asymptotic optimality is thus given by showing how to proceed in two cases. The case defined by (2.5) is called $\{(a,h)\}$ being in the *mth critical band.* The other case is

$$h \sim Qh_m(a)$$

for some $Q \in (0,\infty)$, which we refer to as $\{(a,h)\}$ being on the *boundary between critical bands m and $m+1$.*

It will prove convenient in the sequel to treat h as a function of a. To translate the above formulation into these terms, let $\mathcal{B}_m^o$ be the class of positive functions $h(a)$ such that $\{(a,h(a))\}$ is in the mth critical band (for every sequence of a's approaching ∞) and let $\mathcal{B}_m^+$ be the class of positive functions $h(a)$ such that $\{(a,h(a))\}$ is on the boundary of critical bands m and $m+1$ (for every sequence of a's approaching ∞). That is,

$$\begin{aligned}
\mathcal{B}_m^o &\equiv \{h : (0,\infty) \to (0,\infty) | \quad h_m \ll h \ll h_{m-1}\}, \\
\mathcal{B}_m^+ &\equiv \{h : (0,\infty) \to (0,\infty) | \quad h \sim Qh_m, \text{ some } Q \in (0,\infty)\},
\end{aligned}$$

and let $\mathcal{B}_m = \mathcal{B}_m^o \cup \mathcal{B}_m^+$. Our notation reflects that, as $a \to \infty$, the average number of stages of an efficient procedure approaches

$$\begin{aligned}
m &\quad \text{if } h \in \mathcal{B}_m^o \\
m + \eta &\quad \text{if } h \in \mathcal{B}_m^+,
\end{aligned}$$

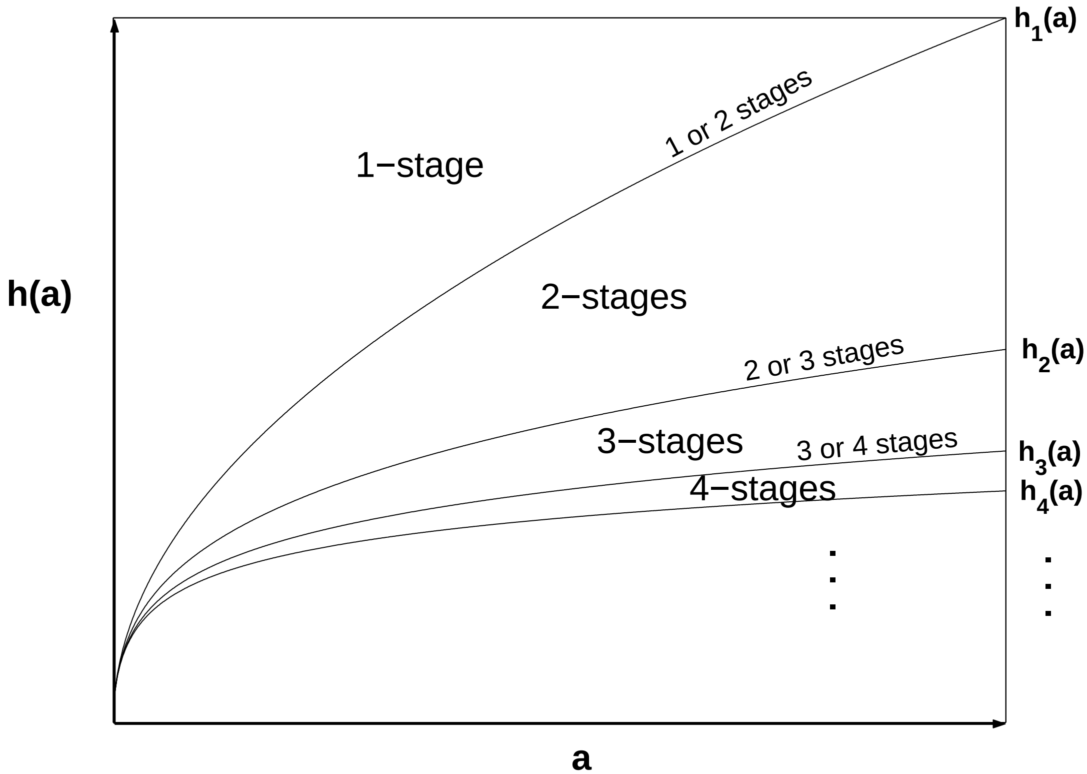

Figure 1.

where $\eta \in (0,1)$ is a function of $\lim_{a\to\infty} h(a)/h_m(a)$; see Figure 1. Finally, we define the *risk* of a procedure $\delta = (T, M)$ to be

$$R(\delta) = E(T - a/\mu) + h(a)EM \tag{2.6}$$

for a given $h(a) \in \mathcal{B}_m$, some $m \geq 1$. By (2.3), the definition of risk (2.6) is equivalent to the expectation of a linear combination of the so-called "overshoot," $X(T) - a$, and the number of stages used. Define the Bayes procedure $\delta^* = (T^*, M^*)$ to be one that achieves $R^* = \inf_\delta R(\delta)$. Dependence on a will usually be suppressed to simplify notation.

A convenient way of parametrizing stage sizes is by the probability of stopping at the end of a stage. Thus, for $a > 0$, $p \in (0,1)$, and z_p the upper p-quantile of the standard normal distribution, let $t(p,a)$ be the unique solution of

$$\frac{a-\mu t(p,a)}{\sqrt{t(p,a)}} = z_p. \tag{2.7}$$

The probability of being across a boundary a units away at the end of a stage of size $t(p,a)$ is p; in this sense we will refer to the *stopping probability* of a stage. A simple computation gives

$$t(p,a) = a/\mu - \frac{z_p\sqrt{4a\mu + z_p^2} - z_p^2}{2\mu^2}.$$

Letting Φ and ϕ denote the standard normal distribution function and density, define

$$\Delta(z) \equiv \int_z^\infty \Phi(-x)dx = \phi(z) - \Phi(-z)z.$$

The function Δ will appear often in calculations of the expected overshoot or undershoot of a random process. For example,

$$\begin{aligned}
E[X(t(p,a)) - a; X(t(p,a)) \geq a] &= \int_a^\infty P(X(t(p,a)) > x)dx \quad \text{(integration by parts)} \\
&= \int_{z_p}^\infty \Phi(-z)\sqrt{t(p,a)}dz \quad \text{(change of variables)} \\
&= \Delta(z_p)\sqrt{t(p,a)} \\
&\sim \Delta(z_p)\sqrt{a/\mu}
\end{aligned}$$

as $a \to \infty$, provided $z_p = o(\sqrt{a})$; we will use relations like these below without further comment.

2.1.1 Geometric Sampling

If (T, M) is the procedure that samples with stopping probability $p \in (0,1)$, constant across the stages, then $T_k = t(p, a - X_{k-1}) \cdot 1\{X_{k-1} < a\}$ and M is a geometric random variable with mean $1/p$. We will thus refer to (T, M) as *geometric sampling with probability* p. Although p is constant across the stages, we do allow p to vary with a, the initial distance to the boundary.

Not only is geometric sampling an interesting random process in its own right, but it has also been conjectured that optimal multistage procedures share its stationarity property. While Theorem 2.9 will show this is not true, geometric sampling will prove a useful tool in designing the final stages of our optimal procedures in the next section.

Lemmas 2.1 and 2.2 establish some fundamental upper bounds on the behavior of geometric sampling.

Lemma 2.1. *Let* $p \in (0,1)$, $q = 1-p$, *and*

$$g(a) \equiv \frac{\Delta(z_q)}{qz_p} \cdot (a - \mu t(p,a)) = \frac{\Delta(z_q)}{2\mu q}(\sqrt{4a\mu + z_p^2} - z_p). \tag{2.8}$$

If (T, M) *is geometric sampling with probability* p, *then*

$$ET - a/\mu \leq \begin{cases} \frac{q\Delta(z_p)}{\mu\Delta(z_q)} \cdot g(a) + \mu^{-1}\sum_{k\geq 2} g^{(k)}(a)q^k, & \text{if } p \leq 1/2 \\ \frac{q\Delta(z_p)}{\mu\Delta(z_q)} \sum_{k\geq 1} g^{(k)}(a)q^{k-1}, & \text{if } p \geq 1/2, \end{cases} \tag{2.9}$$

where $g^{(k)}$ *denotes the* k*th iterate of* g.

Proof. First we will prove

$$E(a - X_k | M > k) \leq g^{(k)}(a) \quad \text{for all } k \geq 0. \tag{2.10}$$

The $k = 0$ case is trivial and we have

$$\begin{aligned} E(a - X_{k+1}|M > k+1, X_k) &= E[(a - X_k) - (X_{k+1} - X_k)|M > k+1, X_k] \\ &= \Delta(z_q)\sqrt{t(p, a - X_k)}/q \\ &= \Delta(z_q)\frac{(a - X_k) - \mu t(p, a - X_k)}{qz_p} \\ &= g(a - X_k). \end{aligned}$$

g is increasing and concave, so by Jensen's inequality and the induction hypothesis

$$\begin{aligned} E(a - X_{k+1}|M > k+1) &= E(g(a - X_k)|M > k+1) \\ &\leq g(E(a - X_k|M > k+1)) \\ &= g(E(a - X_k|M > k)) \qquad (2.11) \\ &\leq g(g^{(k)}(a)) = g^{(k+1)}(a), \end{aligned}$$

proving (2.10). In (2.11) we use that $E(a - X_k|M > k+1) = E(a - X_k|M > k)$; this is true since the value of X_k and the number of additional stages required to cross the boundary are independent, as long as $X_k < a$.

We now prove (2.9). Let $p \leq 1/2$. $E(T_1|M \geq 1) = t(p, a)$ and for $k \geq 2$,

$$\begin{aligned} E(T_k|M \geq k) &= E(t(p, a - X_{k-1})|M > k-1) \\ &\leq \mu^{-1} E(a - X_{k-1}|M > k-1) \quad \text{(by virtue of } p \leq 1/2) \\ &\leq \mu^{-1} g^{(k-1)}(a) \end{aligned}$$

by (2.10). Using these two relations

$$E(T|M = m) = \sum_{k=1}^{m} E(T_k|M = m) = \sum_{k=1}^{m} E(T_k|M \geq k) \leq t(p, a) + \mu^{-1} \sum_{k=2}^{m} g^{(k-1)}(a),$$

since $E(T_k|M = m) = E(T_k|M \geq k)$ for any $m \geq k$ as discussed above. Thus

$$\begin{aligned} ET &= E(E(T|M)) \\ &\leq t(p, a) + \mu^{-1} \sum_{m \geq 2} q^{m-1} p \sum_{k=2}^{m} g^{(k-1)}(a) \\ &= t(p, a) + \mu^{-1} \sum_{k \geq 1} g^{(k)}(a) q^k \quad \text{(by reversing order of summation)} \\ &= a/\mu + \frac{q\Delta(z_p)}{\mu\Delta(z_q)} \cdot g(a) + \mu^{-1} \sum_{k \geq 2} g^{(k)}(a) q^k, \end{aligned}$$

using the relation between g and $t(p, \cdot)$ in (2.8).

Now let $p \geq 1/2$. Then $z_p \leq 0$ and consequently $t(p, \cdot)$ is concave, so using Jensen's inequality and (2.10),

$$\begin{aligned} E(T_k|M \geq k) &= E[t(p, a - X_{k-1})|M > k-1] \\ &\leq t(p, E[a - X_{k-1}|M > k-1]) \leq t(p, g^{(k-1)}(a)) \end{aligned}$$

and, as computed above,

$$ET = E(E(T|M)) \leq \sum_{m \geq 1} q^{m-1} p \sum_{k=1}^{m} t(p, g^{(k-1)}(a)) = a/\mu + \frac{q\Delta(z_p)}{\mu\Delta(z_q)} \sum_{k \geq 1} g^{(k)}(a) q^{k-1},$$

again using (2.8) for the final step. □

Lemma 2.2. *Let (T, M) be geometric sampling with probability $1/2 \leq p(a) \to 1$ and let Y be an arbitrary random variable. There is a $K < \infty$ such that*

$$E(X_{M(Y)} - Y|Y > 0) \leq K|z_{p(a)}| \cdot (\sqrt{E(Y|Y > 0)} \vee |z_{p(a)}|).$$

Remark. The lemma will frequently be used in the following form: If Y and $p(a) \to 1$ are such that $|z_{p(a)}| = o\left(\sqrt{E(Y|Y > 0)}\right)$, then

$$E(X_{M(Y)} - Y; Y > 0) = O\left(|z_{p(a)}|\sqrt{E(Y|Y > 0)}\right).$$

Proof. Let $y > 0$ and g be as in Lemma 2.1 with $p = p(a)$ and $q = 1 - p$. A simple computation shows that g has a unique positive fixed point $y^* = \Delta(z_q)\phi(z_p)/(\mu q^2)$

such that $g(y) \le (y \vee y^*)$. Then

$$\begin{aligned} E(X_{M(y)} - y) &= \mu(ET(y) - y/\mu) \quad \text{(Wald's equation)} \\ &\le \frac{q\Delta(z_p)}{\Delta(z_q)} \sum_{k=1}^{\infty} g^{(k)}(y) q^{k-1} \quad \text{(by Lemma 2.1)} \qquad (2.12) \\ &\le \frac{q\Delta(z_p)}{\Delta(z_q)} \cdot \begin{cases} \sum_{k=1}^{\infty} g(y) q^{k-1} = g(y)/p \le 2g(y) & \text{for } y > y^*, \\ \sum_{k=1}^{\infty} y^* q^{k-1} = y^*/p \le 2y^* & \text{for } y \le y^*, \end{cases} \\ &\qquad \text{(since } g(y) \le (y \vee y^*)) \\ &\le 2 \cdot \frac{q\Delta(z_p)}{\Delta(z_q)} \cdot (g(y) \vee y^*). \qquad (2.13) \end{aligned}$$

Now

$$\sqrt{4y\mu + z_p^2} - z_p \le 2\sqrt{y\mu} + 2|z_p| \le 4(\sqrt{y\mu} \vee |z_p|) \le K_1(\sqrt{y} \vee |z_p|)$$

with $K_1 = 4(\sqrt{\mu} \vee 1)$. Also, $\Delta(z_p) \sim |z_p|$ as $p \to 1$, so

$$\frac{q\Delta(z_p)}{\Delta(z_q)} \cdot g(y) = \frac{\Delta(z_p)}{2\mu}(\sqrt{4y\mu + z_p^2} - z_p) \le K_2|z_p|(\sqrt{y} \vee |z_p|),$$

with $K_2 = 3K_1/(4\mu) < \infty$, say. Also,

$$\begin{aligned} \frac{q\Delta(z_p)}{\Delta(z_q)} \cdot y^* &= \frac{q\Delta(z_p)}{\Delta(z_q)} \cdot \frac{\Delta(z_q)\phi(z_p)}{\mu q^2} = \frac{\Delta(z_p)\phi(z_p)}{\mu q} \\ &\sim \frac{|z_p|\phi(z_p)}{q\mu} \\ &\sim \frac{z_p^2}{\mu} \quad \text{(since } \phi(z_p) \sim q|z_p| \text{ as } p \to 1) \\ &\le K_3|z_p|(\sqrt{y} \vee |z_p|), \end{aligned}$$

with $K_3 = \mu^{-1}$. Plugging these estimates into (2.13), we have

$$E(X_{M(y)} - y) \le K|z_p|(\sqrt{y} \vee |z_p|)$$

for all $y > 0$ with $K = 3 \max K_i$, say, and thus

$$\begin{aligned} E(X_{M(Y)} - Y|Y > 0) &\leq K|z_p|[E(\sqrt{Y}|Y > 0) \vee |z_p|] \\ &\leq K|z_p|(\sqrt{E(Y|Y > 0)} \vee |z_p|), \end{aligned}$$

where this last step uses Jensen's inequality since the square root is concave. □

2.1.2 The Procedures δ_m and $\hat{\delta}_m$

In this section we define two families of procedures, δ_m and $\hat{\delta}_m$, and prove some properties which will later be used to prove them first-order optimal under different assumptions about h - namely, δ_m is optimal when $h \in \mathcal{B}_m^o$ and $\hat{\delta}_m$ is optimal when $h \in \mathcal{B}_m^+$.

Given any positive function h, define $\delta_1(h)$ to be geometric sampling with probability $p^{(1)}(a)$, where $p^{(1)} : (0, \infty) \to (0, 1)$ is any function satisfying $0 < \varepsilon \leq p^{(1)}(a) \to 1$ as $a \to \infty$ in such a way that $z_{p^{(1)}(a)} = o(h(a)/\sqrt{a})$. (The choice of $p^{(1)}(a)$ will not be reflected in the notation). For $m = 1, 2, \ldots$, define $\delta_{m+1}(h)$ to have first stage stopping probability $\Phi(-\sqrt{\log(a/h^2(a) + 1)})$, followed (if necessary) by $\delta_m(h \circ f^{-1})$, where $f(x) \equiv (2/\sqrt{\mu})\sqrt{x \log(x + 1)}$.

Given a constant $p \in (0, 1)$, define $\hat{\delta}_1(p)$ to have first stage stopping probability p, followed (if necessary) by geometric sampling with probability $\hat{p}^{(1)}(a)$, where $\hat{p}^{(1)} : (0, \infty) \to (0, 1)$ is any function satisfying $0 < \varepsilon \leq \hat{p}^{(1)}(a) \to 1$ as $a \to \infty$ in such a way that $z_{\hat{p}^{(1)}(a)} = o(a^{1/4})$. (Again, the choice of $\hat{p}^{(1)}(a)$ will be suppressed in notation). Define $\hat{\delta}_{m+1}(p)$ to have first stage stopping probability $\Phi(-\sqrt{(1 - 2^{-m}) \log a})$, followed (if necessary) by $\hat{\delta}_m(p)$. Note that the value of the constant p is "passed through" for $m > 1$ in the sense that the mth stage of $\hat{\delta}_m(p)$ begins $\hat{\delta}_1(p)$, unless of course the boundary is crossed during the first $m - 1$ stages.

The next two propositions establish the operating characteristics of δ_m and $\hat{\delta}_m$.

Proposition 2.3. *Let m be a positive integer and $h \in \mathcal{B}_m^o$. If $(T^{(m)}, M^{(m)}) = \delta_m(h)$,* *then, as $a \to \infty$,*

$$\begin{aligned} E(T^{(m)} - a/\mu) &= o(h(a)), && (2.14) \\ EM^{(m)} &\to m. && (2.15) \end{aligned}$$

Remark. The restriction $h = o(h_0)$ in the $m = 1$ case is a device that simplifies the proof but is unnecessary in the sense that if $h_0(a) = a = O(h(a))$, then, for suitably chosen $\tilde{h}(a) = o(a)$, the proposition ensures $EM^{(1)} \to 1$ and $E(T^{(1)} - a/\mu) = o(\tilde{h}(a)) = o(h(a))$ for $(T^{(1)}, M^{(1)}) = \delta_1(\tilde{h})$.

Proof. We prove a slightly stronger statement by induction on m. In addition to (2.14) and (2.15), we show that if $0 < b < \infty$, then

$$\begin{aligned} \sup_{a \le b} E(T^{(m)} - a/\mu) &< \infty, && (2.16) \\ \sup_{a \le b} EM^{(m)} &< \infty. && (2.17) \end{aligned}$$

Also, without loss of generality we assume h is non-decreasing. Otherwise, we could replace $h(a)$ by $\underline{h}(a) \equiv \inf_{x \ge a} h(x)$ in what follows, since $\underline{h}$ is non-decreasing and bounded above by h.

The procedure $\delta_1(h)$ is geometric sampling with probability $p^{(1)}(a) \to 1$. If $h \in \mathcal{B}_1^o$, then $h(a) = o(a)$ whence $z_{p^{(1)}} = o(h(a)/\sqrt{a}) = o(\sqrt{a})$, and so Lemma 2.2 with Wald's equation show that

$$E(T^{(1)} - a/\mu) = E(X_{M^{(1)}} - a)/\mu = O(|z_{p^{(1)}}|\sqrt{a}) = o((h(a)/\sqrt{a}) \cdot \sqrt{a}) = o(h(a)),$$

as well as that (2.16) holds for $m = 1$. The relation $EM^{(1)} = 1/p^{(1)}(a)$ implies (2.15) and (2.17) for $m = 1$.

Now assume $h \in \mathcal{B}_{m+1}^o$ and let $(T^{(m+1)}, M^{(m+1)}) = \delta_{m+1}(h)$. Let $z_1 = \sqrt{\log(a/h^2(a) + 1)}$ and $p_1 = \Phi(-z_1)$. Obviously $\lim_{a\to\infty} h(f^{-1}(a))/h_m(a) = \lim_{a\to\infty} h(a)/h_m(f(a))$ and, using the definitions of h_m and f,

$$\begin{aligned} h_m(f(a)) &= O((a\log a)^{(1/2)^{m+1}}(\log(a\log a))^{1/2-(1/2)^m}) \\ &= O(a^{(1/2)^{m+1}}(\log a)^{1/2-(1/2)^{m+1}}) = O(h_{m+1}(a)) = o(h(a)). \quad (2.18) \end{aligned}$$

Thus $h_m(a) = o(h(f^{-1}(a)))$ and a similar argument gives $h(f^{-1}(a)) = o(h_{m-1}(a))$, so that $h \circ f^{-1} \in \mathcal{B}^o_m$. Then, by the induction hypothesis, (2.14)-(2.17) hold with $(T^{(m)}, M^{(m)}) = \delta_m(h \circ f^{-1})$. Now

$$EM^{(m+1)}(a) = 1 + E(M^{(m)}(a - X_1); X_1 < a)$$

and so (2.17) holds for $m + 1$ since it holds for m. Further, using the induction hypothesis and letting $C = (2\sqrt{\mu})^{-1}$, $Y = a - X_1$,

$$\begin{aligned} EM^{(m+1)}(a) &= 1 + E(M^{(m)}(Y); Y > Cz_1\sqrt{a}) + E(M^{(m)}(Y); 0 < Y \le Cz_1\sqrt{a}) \\ &= 1 + m(1 + o(1))P(Y > Cz_1\sqrt{a}) + O(1)P(0 < Y \le Cz_1\sqrt{a}) \\ &= (m + 1) + o(1), \end{aligned}$$

since

$$\begin{aligned} &P(0 < Y \le Cz_1\sqrt{a}) \le P(Y \le Cz_1\sqrt{a}) \\ &\le\ 1 - \Phi\left(\frac{a - \mu t(p_1, a) - Cz_1\sqrt{a}}{\sqrt{t(p_1, a)}}\right) \\ &=\ 1 - \Phi(z_1 - Cz_1\sqrt{\mu}(1 + o(1))) \quad \text{(by (2.7) and since } \sqrt{t(p_1, a)} \sim \sqrt{a/\mu}) \\ &\le\ 1 - \Phi(z_1/4) \to 0. \quad (2.19) \end{aligned}$$

Next we estimate $E(T^{(m+1)} - a/\mu)$. Let $C' = 2/\sqrt{\mu}$. Using Wald's equation and the definition of δ_{m+1},

$$\begin{aligned}
&\mu E(T^{(m+1)} - a/\mu) = E(X_{M^{(m+1)}} - a) \\
&= E(X_{M^{(m+1)}} - a; M^{(m+1)} = 1) + E(X_{M^{(m)}(Y)} - Y; Y > C' z_1 \sqrt{a}) \\
&\quad + E(X_{M^{(m)}(Y)} - Y; 0 < Y \leq C' z_1 \sqrt{a}) \\
&\equiv A_1(a) + A_2(a) + A_3(a),
\end{aligned}$$

and to show that (2.14) and (2.16) hold for $m+1$ it suffices to show the A_i satisfy the same bounds for $i = 1, 2, 3$. We have

$$\begin{aligned}
A_1(a) &= E(X_1 - a; X_1 \geq a) = \Delta(z_1)\sqrt{t(p_1, a)} \\
&\sim (\phi(z_1)/z_1^2)\sqrt{a/\mu} \quad \text{(since } \Delta(z) \sim \phi(z)/z^2 \text{ as } z \to \infty) \\
&= O(h(a)/z_1^2) = o(h(a)).
\end{aligned}$$

Also, the existence of the first moment of X_1 implies $A_1(a)$ is bounded for bounded values of a.

Let $\varphi(y) = E(X_{M^{(m)}(y)} - y)$ for $y > 0$. By the induction hypothesis

$$\varphi(y) = o(h(f^{-1}(y))) \quad \text{as } y \to \infty, \tag{2.20}$$

$$\sup_{y \leq y_0} \varphi(y) < \infty \quad \text{for all } y_0 < \infty. \tag{2.21}$$

By a routine computation, $E(Y; Y > C' z_1 \sqrt{a}) = O(\phi(z_1)\sqrt{a}) = O(h(a))$. Let $K < \infty$ be such that $E(Y; Y > C' z_1 \sqrt{a}) \leq K h(a)$ for large a. Let $\varepsilon > 0$. Using (2.20), we have

$$\varphi(y) = o(h(f^{-1}(y))) = o(h_{m-1}(y)) = o(y), \tag{2.22}$$

since the $m = 1$ case is the largest, asymptotically. Thus assume a is large enough so that $\varphi(y) \le (\varepsilon/K)y$ when $y > C' z_1\sqrt{a}$. Then

$$\begin{aligned} A_2(a) &= E(\varphi(Y); Y > C'z_1\sqrt{a}) \le (\varepsilon/K)E(Y; Y > C'z_1\sqrt{a}) \\ &\le (\varepsilon/K)Kh(a) = \varepsilon h(a), \end{aligned}$$

showing $A_2(a) = o(h(a))$. Also, (2.21) and (2.22) imply that there are constants C_1, a_1 such that

$$A_2(a) \le C_1 + E(Y; Y > a_1)$$

for all a, and the latter is finite for bounded values of a by the same argument used on $A_1(a)$.

The condition (2.21) implies $A_3(a) = E(\varphi(Y); 0 < Y \le C'z_1\sqrt{a})$ is bounded for bounded values of a, so to show $A_3(a) = o(h(a))$ it suffices to show

$$\tilde{A}_3(a) \equiv E(\varphi(Y); a_0 < Y < C'z_1\sqrt{a}) = o(h(a)),$$

for any constant a_0. Let $\varepsilon > 0$ and choose a_0 such that

$$\varphi(y) \le \varepsilon h(f^{-1}(y)) \;\text{ for } y > a_0, \tag{2.23}$$

by virtue of (2.20). Now h and f^{-1} are both non-decreasing, so $h \circ f^{-1}$ is non-decreasing also, and since $C'z_1\sqrt{a} \le f(a)$ we have

$$\tilde{A}_3(a) \le \varepsilon h(f^{-1}(C'z_1\sqrt{a}))P(a_0 < Y \le C'z_1\sqrt{a}) \le \varepsilon h(f^{-1}(f(a))) = \varepsilon h(a),$$

showing $\tilde{A}_3(a) = o(h(a))$. □

Before proving bounds on the operating characteristics of $\hat{\delta}_m$ in Proposition 2.5, we introduce the following positive constants and prove a property of them in Lemma 2.4. For $m \ge 1$ define

$$\kappa_m = \kappa_m(\mu) = \mu^{-2+(1/2)^m} \prod_{i=1}^{m-1} [(1/2)^{m-1-i} - (1/2)^{m-1}]^{(1/2)^{i+1}}. \tag{2.24}$$

Lemma 2.4. *For $m \geq 1$, as $a \to \infty$*

$$\kappa_m h_m(\sqrt{(1-2^{-m})/\mu \cdot a \log a}) \sim \kappa_{m+1} h_{m+1}(a).$$

Proof.

$$\begin{aligned}
\log(\kappa_{m+1}/\kappa_m) &= -(1/2)^{m+1}\log\mu + \sum_{i=1}^{m}(1/2)^{i+1}\log[(1/2)^{m-i} - (1/2)^m] \\
&\quad - \sum_{i=1}^{m-1}(1/2)^{i+1}\log[(1/2)^{m-1-i} - (1/2)^{m-1}] \\
&= -(1/2)^{m+1}\log\mu + (1/2)^{m+1}\log[1-2^{-m}] - \sum_{i=1}^{m-1}(1/2)^{i+1}\log 2 \\
&= (1/2)^{m+1}\log(1-2^{-m})/\mu + (1/2 - (1/2)^m)\log(1/2).
\end{aligned}$$

On the other hand, letting $a' = \sqrt{(1-2^{-m})/\mu \cdot a \log a}$,

$$\begin{aligned}
\log(h_m(a')/h_{m+1}(a)) &= (1/2)^m \log a' + (1/2 - (1/2)^m)\log\log a' \\
&\quad -(1/2)^{m+1}\log a - (1/2 - (1/2)^{m+1})\log\log a \\
&= (1/2)^{m+1}[\log(1-(1/2)^m)/\mu + \log a + \log\log a] \\
&\quad +(1/2 - (1/2)^m)[\log(1/2) + \log\log a + o(1)] \\
&\quad -(1/2)^{m+1}\log a - (1/2 - (1/2)^{m+1})\log\log a \\
&= (1/2)^{m+1}\log(1-2^{-m})/\mu + (1/2-(1/2)^m)\log(1/2) + o(1) \\
&= \log(\kappa_{m+1}/\kappa_m) + o(1)
\end{aligned}$$

so that $h_m(a')/h_{m+1}(a) \to \kappa_{m+1}/\kappa_m$. □

We will use the notation $f \lesssim g$ for $f \leq (1+o(1)) \cdot g$.

Proposition 2.5. *Let $m \geq 1$ and $p \in (0,1)$ a constant. If $(T^{(m)}, M^{(m)}) = \hat{\delta}_m(p)$,* *then, as $a \to \infty$,*

$$E(T^{(m)} - a/\mu) \lesssim \Delta(z_p)\kappa_m h_m(a), \tag{2.25}$$

$$EM^{(m)} \to m + 1 - p. \tag{2.26}$$

Proof. As in the proof of the previous proposition, we prove a slightly stronger claim by induction. In addition to (2.25) and (2.26), we show that if $0 < b < \infty$, then

$$\sup_{a \le b} E(T^{(m)} - a/\mu) < \infty \tag{2.27}$$

$$\sup_{a \le b} EM^{(m)} < \infty. \tag{2.28}$$

Let $(T^{(1)}, M^{(1)}) = \hat{\delta}_1(p)$. By Wald's equation we have

$$\mu E(T^{(1)} - a/\mu) = E(X_1 - a; M^{(1)} = 1) + E(X_{M^{(1)}} - a; M^{(1)} > 1) \tag{2.29}$$

$$= \Delta(z_p)\sqrt{a/\mu}(1 + o(1)) + E(X_{M^{(1)}} - a; M^{(1)} > 1). \tag{2.30}$$

Letting (T', M') be the geometric sampling with probability $\hat{p}^{(1)}(a)$ that follows the first stage of $\hat{\delta}_1(p)$, Lemma 2.2 implies that

$$E(X_{M^{(1)}} - a; M^{(1)} > 1) = E(X_{M'(a-X_1)} - (a - X_1); X_1 < a)$$

$$\le K|z_{\hat{p}^{(1)}}|\sqrt{E(a - X_1 | X_1 < a)} \quad (K < \infty) \tag{2.31}$$

$$= O(|z_{\hat{p}^{(1)}}|a^{1/4}) = o(\sqrt{a}). \tag{2.32}$$

Substituting (2.32) into (2.30) gives

$$E(T^{(1)} - a/\mu) = \Delta(z_p)\mu^{-3/2}\sqrt{a} + o(\sqrt{a}) = \Delta(z_p)\kappa_1 h_1(a) + o(h_1(a)),$$

while (2.29) and (2.31) show that $E(T^{(1)} - a/\mu)$ is bounded for bounded values of a. The relation $EM^{(1)} = 1 + (1-p)/p_2 \to 2 - p$ establishes (2.26) and (2.28) for $m = 1$.

Fix $m \ge 1$ and let $(T^{(m+1)}, M^{(m+1)}) = \hat{\delta}_{m+1}(p)$. Also let $z_1 = \sqrt{(1 - 2^{-m}) \log a}$,

$p_1 = \Phi(-z_1)$, and suppose $\varepsilon > 0$. We have

$$\mu E(T^{(m+1)} - a/\mu) = E(X_1 - a; M^{(m+1)} = 1) + E(X_{M^{(m+1)}} - a; M^{(m+1)} > 1) \quad (2.33)$$

and

$$\begin{aligned} E(X_1 - a; M^{(m+1)} = 1) &= \Delta(z_p)\sqrt{t(p,a)} \sim O(\sqrt{a}\phi(z_1)/z_1^2) \quad (2.34) \\ &= O(a^{(1/2)^{m+1}}/z_1^2) = o(h_{m+1}(a)), \quad (2.35) \end{aligned}$$

by substituting the value of z_1. Thus we can assume a is large enough so that

$$E(X_1 - a; M^{(m+1)} = 1) \leq \varepsilon\Delta(z_p)\kappa_{m+1}\mu h_{m+1}(a). \quad (2.36)$$

For $y > 0$ define $\varphi(y) = E(X_{M^{(m)}(y)} - y)$. By the induction hypothesis and Wald's equation there are constants C_1, y_1 such that

$$\varphi(y) \leq \begin{cases} C_1, & \text{if } 0 < y \leq y_1 \\ \Delta(z_p)\kappa_m\mu h_m(y)(1+\varepsilon), & \text{if } y_1 < y. \end{cases}$$

Then, letting $Y = a - X_1$,

$$\begin{aligned} &E(X_{M^{(m+1)}} - a; M^{(m+1)} > 1) = E(\varphi(Y); Y > 0) \\ &\quad \leq C_1 P(Y \leq y_1) + \Delta(z_p)\kappa_m\mu(1+\varepsilon)E(h_m(Y); Y > y_1). \quad (2.37) \end{aligned}$$

Note that h_m is concave and satisfies $h_m(a + o(a)) \sim h_m(a)$ as $a \to \infty$. Routine computations give

$$P(Y > y_1) \to 1 \text{ and } E(Y; Y > y_1) \sim z_1\sqrt{a/\mu} \quad (2.38)$$

as $a \to \infty$, so, by Jensen's inequality,

$$\begin{aligned} \kappa_m E(h_m(Y); Y > y_1) &\le \kappa_m P(Y > y_1) h_m(E(Y|Y > y_1)) \\ &\sim \kappa_m h_m(z_1 \sqrt{a/\mu}) \sim \kappa_{m+1} h_{m+1}(a), \end{aligned} \tag{2.39}$$

this last by Lemma 2.4. Thus assume a is large enough so that

$$\begin{aligned} \kappa_m E(h_m(Y); Y > y_1) &\le (1+\varepsilon)\kappa_{m+1} h_{m+1}(a), \\ C_1 P(Y \le y_1) &\le \varepsilon \Delta(z_p) \kappa_{m+1} \mu h_{m+1}(a). \end{aligned}$$

Plugging these estimates into (2.37) and combining with (2.36) gives

$$\begin{aligned} E(T^{(m+1)} - a/\mu) &\le [\varepsilon + \varepsilon + (1+\varepsilon)^2] \Delta(z_p) \kappa_{m+1} h_{m+1}(a) \\ &\le (1+5\varepsilon) \Delta(z_p) \kappa_{m+1} h_{m+1}(a). \end{aligned}$$

Since ε was arbitrary, this shows that (2.25) holds for $m+1$. For bounded intervals of a, the equality in (2.34) shows $E(X_{M^{(m+1)}} - a; M^{(m+1)} = 1)$ is bounded while (2.37) and (2.39) show $E(X_{M^{(m+1)}} - a; M^{(m+1)} > 1)$ is also bounded, and hence $E(T^{(m+1)} - a/\mu)$ is bounded.

Let $\psi(y) = EM^{(m)}(y)$ for $y > 0$ and let $\varepsilon > 0$. By the induction hypothesis there are positive constants C_2, y_2 such that

$$\begin{aligned} \psi(y) \le C_2 \quad &\text{if} \quad 0 < y \le y_2, \\ |\psi(y) - (m+1-p)| \le \varepsilon/3 \quad &\text{if} \quad y_2 < y. \end{aligned}$$

As with the first part of (2.38), $P(Y \le y_2) \to 0$. So assume a is large enough so that

$$P(Y \le y_2) \le (\varepsilon/3) \min\{(m+1-p)^{-1}, C_2^{-1}\}.$$

Then

$$
\begin{aligned}
&|EM^{(m+1)} - (m+2-p)| \\
&\quad = |1 + E(\psi(Y); 0 < Y \le y_2) + E(\psi(Y); Y > y_2) - (m+2-p)| \\
&\quad \le E(|\psi(Y) - (m+1-p)|; Y > y_2) + C_2 P(Y \le y_2) + (m+1-p)P(Y \le y_2) \\
&\quad \le \varepsilon/3 + \varepsilon/3 + \varepsilon/3. \qquad (2.40)
\end{aligned}
$$

This shows that $EM^{(m+1)} \to m+2-p$ and (2.40), with the induction hypothesis, shows that $EM^{(m+1)}$ is bounded for bounded values of a. □

2.1.3 Optimality of δ_m and $\hat{\delta}_m$

For $a \ge y^2 > 0$ define $F_y(a) = \sqrt{a \log(a/y^2)}$. If h is a positive function, then for a such that $h^2(a) \le a$ define

$$F^{(k)}_{h(a)}(a) = F^{(k)}_y(a)|_{y=h(a)}.$$

Note that $h(\cdot)$ is not iterated, e.g. $F^{(2)}_{h(a)}(a) = F_{h(a)}(F_{h(a)}(a)) \ne F_{h(F_{h(a)}(a))}(F_{h(a)}(a))$.

The next lemma shows that, when $h \in \mathcal{B}_m$, square roots of the iterates $F^{(k-1)}_{h(a)}(a)$ are roughly constant multiples of the critical functions h_k. The constants themselves are given by the solutions of the following recurrence relation. For $1 \le k \le m$ define C^m_k to be the unique solution of

$$C^m_{k+1} = \sqrt{C^m_k} \cdot [(1/2)^{k-1} - (1/2)^{m-1}]^{1/4}; \quad C^m_1 = 1. \qquad (2.41)$$

After taking logarithms, solving (2.41) amounts to solving a difference equation. This computation gives

$$C^m_k = \prod_{i=1}^{k-1} \left[(1/2)^{k-1-i} - (1/2)^{m-1}\right]^{(1/2)^{i+1}}, \qquad (2.42)$$

where it is understood that an empty product equals 1. Note also that

$$\kappa_m = (1/\mu)^{2-(1/2)^m} C_m^m. \tag{2.43}$$

Lemma 2.6. *If $h \in \mathcal{B}_m^+$, then*

$$\sqrt{F_{h(a)}^{(k-1)}(a)} \sim C_k^m h_k(a) \quad \text{as } a \to \infty, \text{ for } 1 \le k \le m. \tag{2.44}$$

If $h \in \mathcal{B}_m^o$, then

$$C_k^{m-1} \lesssim \frac{\sqrt{F_{h(a)}^{(k-1)}(a)}}{h_k(a)} \lesssim C_k^m \quad \text{as } a \to \infty, \text{ for } 1 \le k < m. \tag{2.45}$$

Proof. Let F^k denote $F_{h(a)}^{(k)}(a)$. First we prove (2.44) by induction on k. For $k=1$,

$$\sqrt{F^0} = \sqrt{a} = 1 \cdot \sqrt{a} = C_1^m \cdot h_1(a).$$

Now assume $2 \le k+1 \le m$, $\sqrt{F^{k-1}} \sim C_k^m h_k(a)$, and let $Q = \lim h/h_m \in (0,\infty)$. Observe that

$$\begin{aligned} \log\left(\frac{F^{k-1}}{h(a)^2}\right) &\sim \log\left(\frac{(C_k^m h_k(a))^2}{(Qh_m(a))^2}\right) \\ &\sim \log\left(\frac{h_k(a)^2}{h_m(a)^2}\right) \\ &\sim \log\left(\frac{a^{(1/2)^{k-1}}(\log a)^{1-(1/2)^{k-1}}}{a^{(1/2)^{m-1}}(\log a)^{1-(1/2)^{m-1}}}\right) \\ &\sim [(1/2)^{k-1} - (1/2)^{m-1}]\log a, \end{aligned} \tag{2.46}$$

so

$$\begin{aligned}
\sqrt{F^k} &= \left\{F^{k-1}\log(F^{k-1}/h(a)^2)\right\}^{1/4} \\
&\sim \left\{(C_k^m h_k(a))^2[(1/2)^{k-1}-(1/2)^{m-1}]\log a\right\}^{1/4} \\
&= \sqrt{C_k^m}\cdot a^{(1/2)^{k+1}}(\log a)^{1/4-(1/2)^{k+1}}[(1/2)^{k-1}-(1/2)^{m-1}]^{1/4}(\log a)^{1/4} \\
&= \sqrt{C_k^m}\cdot[(1/2)^{k-1}-(1/2)^{m-1}]^{1/4}h_{k+1}(a) \\
&= C_{k+1}^m h_{k+1}(a), \qquad (2.47)
\end{aligned}$$

by (2.41).

Next we prove (2.45) by induction on k. The $k=1$ case is again easy since

$$C_1^{m-1} = \frac{\sqrt{F^0}}{h_1} = C_1^m = 1$$

for any $m \geq 2$. Now assume $2 \leq k+1 < m$ and that (2.45) holds for k. Then, since $h_m \ll h \ll h_{m-1}$,

$$\log\left(\frac{F^{k-1}}{h(a)^2}\right) \lesssim \log\left(\frac{(C_k^m h_k(a))^2}{h_m(a)^2}\right) \sim [(1/2)^{k-1}-(1/2)^{m-1}]\log a,$$

by the same argument leading to (2.46). Then, by repeating the argument leading to (2.47) with $\lesssim$ in place of $\sim$,

$$\sqrt{F^k} \lesssim \sqrt{C_k^m}\cdot[(1/2)^{k-1}-(1/2)^{m-1}]h_{k+1}(a) = C_{k+1}^m h_{k+1}(a),$$

by (2.41). The other bound is similar:

$$\log\left(\frac{F^{k-1}}{h(a)^2}\right) \gtrsim \log\left(\frac{(C_k^{m-1} h_k(a))^2}{h_{m-1}(a)^2}\right) \sim [(1/2)^{k-1}-(1/2)^{m-2}]\log a,$$

and so

$$\sqrt{F^k} \gtrsim \sqrt{C_k^{m-1}}\cdot[(1/2)^{k-1}-(1/2)^{m-2}]h_{k+1}(a) = C_{k+1}^{m-1} h_{k+1}(a),$$

by replacing m by $m-1$ in (2.47) and (2.41). $\square$

The next lemma establishes a lower bound on how close any efficient procedure can be to the boundary after each of the first $m-1$ stages when $h \in \mathcal{B}_m$.

Lemma 2.7. *Assume that $h \in \mathcal{B}_m$. If $\delta = (T, M)$ is any procedure such that $R(\delta) = O(h(a))$, then*

$$\frac{a - X_k}{(1/\mu)^{1-2^{-k}} F_{h(a)}^{(k)}(a)} \geq 1 \quad \textit{in probability as } a \to \infty \tag{2.48}$$

for $k = 0, 1, \ldots, m-1$.

Proof. Let $G_k(a) = (1/\mu)^{1-2^{-k}} F_{h(a)}^{(k)}(a)$. Given $\varepsilon > 0$, let

$$V_k = \{a - X_k \geq (1-\varepsilon) G_k(a)\}.$$

The $k = 0$ case is trivial since (2.48) is equivalent to $a \geq a$. Fix $1 \leq k < m$ and assume that $P(V_{k-1}) \to 1$. Let

$$\zeta_k = \frac{a - X_{k-1} - \mu T_k}{\sqrt{T_k}}.$$

Note that

$$\begin{aligned} h(a)^2 &= o(h_{m-1}(a)^2) \quad (h \in \mathcal{B}_m) \\ &= o(F_{h(a)}^{(m-2)}(a)) \quad \text{(by Lemma 2.6)} \\ &= o(G_{m-2}(a)) \\ &= o(G_{k-1}(a)) \end{aligned}$$

since $k - 1 \leq m - 2$. Thus $G_{k-1}/h^2 \to \infty$ and so does $\log(G_{k-1}/h^2)$. With this, we claim

$$P(\zeta_k \geq \sqrt{\log(G_{k-1}(a)/h^2(a))} - 1 | V_{k-1}) \to 1. \tag{2.49}$$

Let $\zeta(a) = \sqrt{\log(G_{k-1}(a)/h^2(a))} - 1$ and $U = \{\zeta_k < \zeta(a)\}$. If (2.49) were to fail there would be a constant $\eta > 0$ and a sequence of a's approaching ∞ on which $P(U|V_{k-1}) > \eta$. Then

$$\begin{aligned}\mu R(\delta) &\geq \mu E(T - a/\mu) = E(X_M - a) \geq E[(X_M - a)1\{M = k\}; U \cap V_{k-1}] \\ &= E[\Delta(\zeta_k)\sqrt{t(\Phi(-\zeta_k), a - X_{k-1})}; U \cap V_{k-1}]. \qquad (2.50)\end{aligned}$$

The function inside the expectation in (2.50) is decreasing in both ζ_k and X_{k-1}, hence

$$\mu R(\delta) \geq \Delta(\zeta(a))\sqrt{t(\Phi(-\zeta(a)), (1-\varepsilon)G_{k-1}(a))} \cdot P(U \cap V_{k-1}). \qquad (2.51)$$

By assumption, $P(U|V_{k-1}) \geq \eta$ and $P(V_{k-1}) \to 1$, so

$$P(U \cap V_{k-1}) \geq \eta/2, \qquad (2.52)$$

say, for large enough a. Also

$$\begin{aligned}\Delta(\zeta(a))\sqrt{t(\Phi(-\zeta(a)), (1-\varepsilon)G_{k-1}(a))} &\sim \frac{\phi(\zeta(a))}{\zeta^2(a)}\sqrt{(1-\varepsilon)G_{k-1}(a)/\mu} \\ &\geq \varepsilon' h(a)\frac{\exp(\sqrt{\log(G_{k-1}(a)/h^2(a))} - 1/2)}{(\sqrt{\log(G_{k-1}(a)/h^2(a))} - 1)^2} \quad (\varepsilon' > 0) \\ &= h(a)/o(1). \qquad (2.53)\end{aligned}$$

Plugging (2.52) and (2.53) into (2.51) gives $h(a) = o(R(\delta))$, which contradicts our assumption that $R(\delta) = O(h(a))$. Hence, (2.49) must hold. Then

$$\begin{aligned}P(V_k|U' \cap V_{k-1}) &= P(a - X_k \geq (1-\varepsilon)G_k(a)|U' \cap V_{k-1}) \\ &= P\left(\frac{(X_k - X_{k-1}) - \mu T_k}{\sqrt{T_k}} \leq \frac{a - X_{k-1} - (1-\varepsilon)G_k(a) - \mu T_k}{\sqrt{T_k}}\,\middle|\, U' \cap V_{k-1}\right). \qquad (2.54)\end{aligned}$$

On V_{k-1},

$$\begin{aligned}\frac{a - X_{k-1} - (1-\varepsilon)G_k(a) - \mu T_k}{\sqrt{T_k}} &= \zeta_k - \frac{2\mu(1-\varepsilon)G_k(a)}{\sqrt{4\mu(a - X_{k-1}) + \zeta_k^2} - \zeta_k} \\ &\geq \zeta_k - \frac{2\mu(1-\varepsilon)G_k(a)}{\sqrt{4\mu(1-\varepsilon)G_{k-1}(a) + \zeta_k^2} - \zeta_k},\end{aligned}$$

which is increasing in ζ_k. Hence, on U',

$$\begin{aligned}&\zeta_k - \frac{2\mu(1-\varepsilon)G_k(a)}{\sqrt{4\mu(1-\varepsilon)G_{k-1}(a) + \zeta_k^2} - \zeta_k} \geq \zeta(a) - \frac{2\mu(1-\varepsilon)G_k(a)}{\sqrt{4\mu(1-\varepsilon)G_{k-1}(a) + \zeta^2(a)} - \zeta(a)} \\ &= \zeta(a) - \frac{2\mu(1-\varepsilon)G_k(a)}{\sqrt{4\mu(1-\varepsilon)G_{k-1}(a)}}(1 + o(1)) \\ &= \zeta(a) - \sqrt{(1-\varepsilon)\log(F_{h(a)}^{(k-1)}(a)/h^2(a))}(1 + o(1)) \\ &\sim (1 - \sqrt{1-\varepsilon})\sqrt{\log(G_{k-1}(a)/h^2(a))} \equiv \gamma(a) \to \infty.\end{aligned}$$

Substituting this back into (2.54) gives

$$P(V_k | U' \cap V_{k-1}) \geq 1 - [\gamma(a)/2]^{-2} \to 1$$

by Chebyshev's inequality. Thus $P(V_k) \geq P(V_k | U' \cap V_{k-1}) P(U' \cap V_{k-1}) \to 1$ since $P(U' \cap V_{k-1}) \to 1$ by the induction hypothesis and (2.49), finishing the induction and proving the lemma. □

Next we prove the optimality of δ_m and $\hat{\delta}_m$.

Theorem 2.8. *If $h \in \mathcal{B}_m^o$, then*

$$R(\delta_m(h)) \sim m h(a) \sim R^*. \tag{2.55}$$

If $h \in \mathcal{B}_m^+$, then

$$R(\hat{\delta}_m(p^*)) \sim \left[m + 1 - p^* + \frac{\Delta(z_{p^*})\kappa_m}{Q} \right] h(a) \sim R^*, \tag{2.56}$$

where $Q = \lim_{a\to\infty} h(a)/h_m(a) \in (0,\infty)$ *and* p^* *is the unique solution of the equation*

$$\frac{p^*}{\phi(z_{p^*})} = \frac{Q}{\kappa_m}.$$

Proof. Assume that $h \in \mathcal{B}_m^o$. Proposition 2.3 implies $R(\delta_m(h)) \sim mh(a)$. By the Bayes property, $R^* \le R(\delta_m(h)) = O(h)$ and so Lemma 2.7 applies to δ^*. Then, letting X_k^* denote the δ^*-sampled process,

$$\begin{aligned} R^* &\ge h(a)EM^* \\ &\ge h(a)mP(M^* \ge m) \\ &= h(a)mP(a - X_{m-1}^* > 0) \\ &\ge h(a)mP\left(a - X_{m-1}^* \ge (1/2)(1/\mu)^{1-2^{-(m-1)}} F_{h(a)}^{(m-1)}(a)\right) \\ &\sim mh(a) \quad \text{(by Lemma 2.7)} \\ &\sim R(\delta_m(h)) \ge R^*, \end{aligned}$$

proving (2.55).

If $h \in \mathcal{B}_m^+$ with $h(a)/h_m(a) \to Q \in (0,\infty)$, then Proposition 2.5 shows that

$$\begin{aligned} R(\hat{\delta}_m(p^*)) &\lesssim \Delta(z_{p^*})\kappa_m h_m(a) + (m+1-p^*)h(a) \\ &\sim [\Delta(z_{p^*})\kappa_m/Q + m + 1 - p^*]h(a). \end{aligned}$$

Again $R^* \le R(\hat{\delta}_m(p^*)) = O(h(a))$, so Lemma 2.7 applies and we have

$$P(a - X_{m-1}^* \ge (1-\varepsilon)(1/\mu)^{1-2^{-(m-1)}} F_{h(a)}^{(m-1)}(a)) \to 1$$

for any $\varepsilon > 0$. Fix such an ε. Let $(T^{*(m)}, M^{*(m)})$ denote the continuation of δ^* after the $(m-1)$st stage, i.e.,

$$\begin{aligned} M^{*(m)} &= M^* - (1\{M^* \ge 1\} + \cdots + 1\{M^* \ge m-1\}), \\ T^{*(m)} &= T^* - (T_1^* + \cdots + T_{m-1}^*). \end{aligned}$$

For $y > 0$ define

$$\varphi(y) = E[\mu^{-1}(X_{M^{*(m)}} - y) + h(a)M^{*(m)} | a - X^*_{m-1} = y].$$

We will show below that $\varphi(y)$ is non-decreasing in y. Let

$$\gamma(a) = (1-\varepsilon)(1/\mu)^{1-2^{-(m-1)}} F^{(m-1)}_{h(a)}(a).$$

We now compute a lower bound for $\varphi(\gamma(a))$. Letting

$$p = P(M^{*(m)} = 1 | a - X^*_{m-1} = \gamma(a)),$$

$$\begin{aligned}
\mu^{-1}&E(X_{M^{*(m)}} - (a - X^*_{m-1}) | a - X^*_{m-1} = \gamma(a)) \\
&\geq \mu^{-1}E[(X_{M^{*(m)}} - (a - X^*_{m-1}))1\{M^{*(m)} = 1\} | a - X^*_{m-1} = \gamma(a)] \\
&= \mu^{-1}\Delta(z_p)\sqrt{t(p, \gamma(a))} \\
&\sim \mu^{-1}\Delta(z_p)\sqrt{\gamma(a)/\mu} \\
&= \mu^{-1}\Delta(z_p)\sqrt{(1-\varepsilon)(1/\mu)^{2-2^{-m+1}} F^{(m-1)}_{h(a)}(a)} \\
&\sim \Delta(z_p)\sqrt{1-\varepsilon} \cdot (1/\mu)^{2-2^{-m}} C^m_m h_m(a) \quad \text{(by Lemma 2.6)} \\
&\sim \frac{\Delta(z_p)\kappa_m\sqrt{(1-\varepsilon)}}{Q} h(a),
\end{aligned} \tag{2.57}$$

this last by (2.43) and $h \sim Qh_m$. Also, $E(M^{*(m)} | a - X^*_{m-1} = \gamma(a)) \geq 2 - p$, and combining this with (2.57) gives

$$\varphi(\gamma(a)) \geq \left[\frac{\Delta(z_p)\kappa_m\sqrt{(1-\varepsilon)}}{Q} + (2-p)\right] h(a)(1 + o(1)).$$

Letting $Y = a - X^*_{m-1}$ and $V = \{Y \geq \gamma(a)\}$, we have

$$\begin{aligned} R^* &= E[\mu^{-1}(X_{M^*} - a) + h(a)M^*] \\ &\geq E[\mu^{-1}(X_{M^{*(m)}} - Y) + h(a)(m - 1 + M^{*(m)}); V] \\ &= E(\varphi(Y); V) + (m-1)h(a)P(V) \\ &\geq \varphi(\gamma(a))P(V) + (m-1)h(a)P(V) \quad (\varphi \text{ non-decreasing}) \\ &\gtrsim \left[\frac{\Delta(z_p)\kappa_m\sqrt{(1-\varepsilon)}}{Q} + (m+1-p)\right] h(a). \end{aligned} \tag{2.58}$$

Using calculus, it can be shown that the expression in brackets in (2.58) achieves its unique minimum when $p = p^*(\varepsilon)$, the unique solution of

$$\frac{p^*(\varepsilon)}{\phi(z_{p^*(\varepsilon)})} = \frac{Q}{\kappa_m\sqrt{1-\varepsilon}}.$$

Thus,

$$R^* \geq \left[\frac{\Delta(z_{p^*(\varepsilon)})\kappa_m\sqrt{(1-\varepsilon)}}{Q} + (m+1-p^*(\varepsilon))\right] h(a)(1+o(1)).$$

This holds for all $\varepsilon > 0$, so by a standard asymptotic technique (e.g., [6], p. 188), there is a sequence $\varepsilon_a \to 0$ for which it holds. Moreover, $p^*(\varepsilon_a) \to p^*(0) = p^*$, which proves (2.56).

Finally, we show that $\varphi(\cdot)$ is non-decreasing. Fix $a > 0$ and let $0 < y \leq y'$. Let $(T'^{(m)}, M'^{(m)})$ denote the continuation of δ^* after the $(m-1)$st stage that uses the same stopping probability at each stage as $(T^{*(m)}, M^{*(m)})$ when starting from $a - X^*_{m-1} = y'$. Then

$$E(M'^{(m)}|a - X^*_{m-1} = y) = E(M^{*(m)}|a - X^*_{m-1} = y') \tag{2.59}$$

and, letting

$$p_1 = P(M^{*(m)} = 1|a - X^*_{m-1} = y') = P(M'^{(m)} = 1|a - X^*_{m-1} = y),$$

$$\begin{aligned} E[(X_{M'(m)} - y)1\{M'^{(m)} = 1\}|a - X^*_{m-1} = y] &= \Delta(z_{p_1})\sqrt{t(p_1, y)} \\ &\leq \Delta(z_{p_1})\sqrt{t(p_1, y')} \quad (y \leq y') \\ &= E[(X_{M^*(m)} - y')1\{M^{*(m)} = 1\}|a - X^*_{m-1} = y']. \end{aligned}$$

Similar arguments inductionly give

$$E[(X_{M'(m)} - y)1\{M'^{(m)} > 1\}|a - X^*_{m-1} = y] \leq E[(X_{M^*(m)} - y')1\{M^{*(m)} > 1\}|a - X^*_{m-1} = y'],$$

and these last two bounds show

$$E(X_{M'(m)} - y|a - X^*_{m-1} = y) \leq E(X_{M^*(m)} - y'|a - X^*_{m-1} = y'). \tag{2.60}$$

Then

$$\begin{aligned} \varphi(y) &\leq E[\mu^{-1}(X_{M'(m)} - y) + h(a)M'^{(m)}|a - X^*_{m-1} = y] \quad \text{(optimality of } (T^{*(m)}, M^{*(m)})) \\ &\leq E[\mu^{-1}(X_{M^*(m)} - y') + h(a)M^{*(m)}|a - X^*_{m-1} = y'] \quad \text{(by (2.59) and (2.60))} \\ &= \varphi(y'), \end{aligned}$$

finishing the proof. □

The final theorem of this section is a converse to Theorem 2.8, showing that good procedures must behave like $\delta_m, \hat{\delta}_m$ in not only the sense that m stages are necessary when $h \in \mathcal{B}_m$, but also that the sooner a procedure deviates from the "schedule" of Lemma 2.7, the worse its performance.

Theorem 2.9. *Assume that $h \in \mathcal{B}_m$ and let*

$$\delta_m = \begin{cases} \delta_m(h), & \text{if } h \in \mathcal{B}^o_m \\ \hat{\delta}_m(p^*), & \text{if } h \in \mathcal{B}^+_m. \end{cases}$$

If $\delta = (T, M)$ is a procedure such that there is a sequence $a_i \to \infty$ with

$$P(a_i - X_k \leq (1-\varepsilon)(1/\mu)^{1-2^{-k}} F^{(k)}_{h(a_i)}(a_i)) \quad \textit{bounded below 1} \tag{2.61}$$

for some $1 \le k < m$ *and* $\varepsilon > 0$*, then there is* $C > 0$ *such that*

$$R(\delta) - R(\delta_m) \ge C \cdot h_{k^*}(a_i) \to \infty, \tag{2.62}$$

where k^* *is the smallest* k *for which (2.61) holds. In particular, (2.62) holds if* $P(M \ge m) \not\to 1$.

Proof. Let $V_k = \{a - X_k \le (1-\varepsilon)(1/\mu)^{1-2^{-k}} F^{(k)}_{h(a)}(a)\}$. By arguments of the type used in the proof of Lemma 2.7, there is an $\eta > 0$ such that

$$\begin{aligned} R(\delta) &\ge \mu^{-1} E(X_M - a_i) \\ &\ge \mu^{-1} E(X_M - a_i; \{M = k^*\} \cap V_{k^*-1}) \\ &\ge \Delta(z_\eta)\sqrt{t(\eta, (1-\varepsilon)(1/\mu)^{1-2^{-k}} F^{(k^*-1)}_{h(a_i)}(a_i))} \cdot \eta \\ &\ge C\sqrt{F^{(k^*-1)}_{h(a_i)}(a_i)} \ge C' h_{k^*}(a_i), \end{aligned}$$

for appropriately chosen $C, C' > 0$, where this last inequality uses Lemma 2.6. By Theorem 2.8, $R(\delta_m) = O(h(a)) = o(h_{k^*}(a))$ since $k^* < m$, proving (2.62). If $P(V_k) \to 1$ for all $1 \le k < m$, then $P(M \ge m) \ge P(V_{m-1}) \to 1$, proving the second assertion. □

2.2 Procedures for i.i.d. Random Variables

In this section we extend the sampling procedures and techniques of the first half of this chapter to procedures for discrete, i.i.d. data. Specifically, let $X_1, X_2, \ldots$ be i.i.d. from a distribution whose characteristic function is analytic in some neighborhood of the origin. For example, the one-parameter exponential family considered in Chapters 3 and 4 satisfies this requirement. Assume the common mean μ is positive and, since the problem is not changed by multiplying the X_i and the boundary $a > 0$ by a positive constant, we assume without loss of generality that $\mathrm{Var} X_i = 1$.

Define a *multistage stopping rule* N to be a sequence of non-negative integer valued random variables $(N_1, N_2, \ldots)$ such that

$$N_{k+1} \cdot 1\{N_1 + \cdots + N_k = n\} \in \mathcal{E}_n \quad \text{for all } n \geq 1, \tag{2.63}$$

where $\mathcal{E}_n$ is the class of all random variables determined by $X_1, \ldots, X_n$. By analogy with the continuous case in Section 2.1, the interpretation of the measurability requirement (2.63) is that by the time $N^k \equiv N_1 + \cdots + N_k$, the end of the first k stages, an observer who knows the values $X_1, \ldots, X_{N^k}$ also knows N_{k+1}, the size of the $(k+1)$st stage. We will also let N denote the total sample size, N^M, where $M = \inf\{m \geq 1 : X_1 + \cdots + X_{N^m} \geq a\}$. We will denote a *(discrete) multistage sampling procedure* by the pair $\delta = (N, M)$. When there is no confusion as to which sampling procedure is being used, the simplifying notation $S_k \equiv X_1 + \cdots + X_{N^k}$, $S_0 \equiv 0$ will be employed. We will write $N(a), M(a)$ when we wish to emphasize the initial distance to the boundary, a. Given a positive function h, we again define the *risk* of a procedure $\delta = (N, M)$ to be

$$R(\delta) = E(N - a/\mu) + h(a)EM$$

and the Bayes procedure $\delta^* = (N^*, M^*)$ to be that which achieves $R^* \equiv \inf_\delta R(\delta)$. (We shall continue to suppress the dependence on a in notation.) We define the problem analogously as for Brownian motion: to sample $X_1, X_2, \ldots$ in stages until $S_k \geq a$, with the aim of minimizing the risk.

The procedures of the previous section were designed around the principle of comparing expected overshoot over the boundary, often in the large deviation range, with the ratio of cost per stage to cost per unit sample. To use these ideas on discrete data, we need a way of estimating the expected overshoot of a sum of random variables. Let $\Sigma_n = X_1 + \cdots + X_n$ and $\{a_n\}$ an arbitrary sequence. If the X_i are i.i.d. $N(\mu, 1)$ then it is a simple computation to show

$$E(\Sigma_n - a_n; \Sigma_n > a_n) = \sqrt{n} \cdot \Delta\left(\frac{a_n - n\mu}{\sqrt{n}}\right),$$

where

$$\Delta(z) \equiv \int_z^\infty \Phi(-x)dx = \phi(z) - \Phi(-z)z.$$

Since the distribution of $(\Sigma_n - n\mu)/\sqrt{n}$ approaches the standard normal distribution as n gets large even if the X_i are not normals, then one might conjecture that

$$E(\Sigma_n - a_n; \Sigma_n > a_n) \sim \sqrt{n} \cdot \Delta\left(\frac{a_n - n\mu}{\sqrt{n}}\right) \quad \text{as } n \to \infty \tag{2.64}$$

as long as the boundary a_n is not too far in the tail of the sum's distribution. The next lemma gives general conditions under which this is true.

Lemma 2.10. *If a_n is such that*

$$\lim_{n\to\infty} \frac{a_n - n\mu}{\sqrt{n}} \in (-\infty, \infty) \quad \textit{or} \quad n^{1/6} \gg \frac{a_n - n\mu}{\sqrt{n}} \to \infty$$

as $n \to \infty$, then (2.64) holds.

The idea of the proof is to approximate the distribution of Σ_n by the normal distribution in the large deviations range of the tail and use a cruder bound, based on Schwarz' inequality, for the remaining tail.

Proof. Let $T_n = (\Sigma_n - n\mu)/\sqrt{n}$ and $b_n = (a_n - n\mu)/\sqrt{n}$. Then

$$E(\Sigma_n - a_n; \Sigma_n > a_n) = \sqrt{n}E(T_n - b_n; T_n > b_n) = \sqrt{n}\int_{b_n}^\infty P(T_n > x)dx,$$

using the familiar "integration by parts" formula

$$E(Y; Y > y) = yP(Y > y) + \int_y^\infty P(Y > x)dx, \tag{2.65}$$

which holds whenever EY exists. Hence to show that (2.64) holds it suffices to show

$$\int_{b_n}^\infty P(T_n > x)dx \sim \Delta(b_n).$$

First assume $b_n \to \infty$ such that $b_n = o(n^{1/6})$. Choose $c_n \to \infty$ such that $b_n + \varepsilon \leq$

$c_n = o(n^{1/6})$, some $\varepsilon > 0$. Observe that

$$\begin{aligned}
\frac{\phi(c_n)}{\Delta(b_n)} &\sim b_n^2 \frac{\phi(c_n)}{\phi(b_n)} \quad (\text{since } \Delta(x) \sim \phi(x)/x^2 \text{ as } x \to \infty) \\
&= b_n^2 \exp[-c_n^2/2 + b_n^2/2] \\
&= b_n^2 \exp[-(1/2)(c_n - b_n)(c_n + b_n)] \\
&\leq b_n^2 \exp[-(\varepsilon/2)(c_n + b_n)] \to 0. \qquad (2.66)
\end{aligned}$$

Write

$$\int_{b_n}^{\infty} P(T_n > x)dx = \int_{b_n}^{c_n} P(T_n > x)dx + \int_{c_n}^{\infty} P(T_n > x)dx.$$

By Theorem XVI.7.1 of [10], $P(T_n > x) \sim \Phi(-x)$ for large x satisfying $x = o(n^{1/6})$. Thus

$$\int_{b_n}^{c_n} P(T_n > x)dx \sim \int_{b_n}^{c_n} \Phi(-x)dx = \Delta(b_n) - \Delta(c_n) \sim \Delta(b_n), \qquad (2.67)$$

since $\Delta(c_n) \leq \phi(c_n) = o(\Delta(b_n))$ by (2.66). For the other term,

$$\int_{c_n}^{\infty} P(T_n > x)dx = E(T_n; T_n > c_n) - c_n P(T_n > c_n) \qquad (2.68)$$

by (2.65) and, using Mills' ratio and (2.66),

$$c_n P(T_n > c_n) \sim c_n \Phi(-c_n) \sim \phi(c_n) = o(\Delta(b_n)).$$

The other piece is

$$\begin{aligned}
E(T_n; T_n > c_n) &= E(T_n 1\{T_n > c_n\}) \\
&\leq \sqrt{ET_n^2 \cdot E1\{T_n > c_n\}^2} \quad (\text{Schwarz' inequality}) \\
&= \sqrt{1 \cdot P(T_n > c_n)} \qquad (2.69) \\
&\sim \sqrt{\Phi(-c_n)} \\
&\sim \sqrt{\phi(c_n)/c_n} \\
&= o(\Delta(b_n)),
\end{aligned}$$

by an argument like that leading to (2.66), replacing $c_n^2/2$ by $c_n^2/4$. These last two estimates give $\int_{c_n}^{\infty} P(T_n > x)dx = o(\Delta(b_n))$ and combining this with (2.67) gives $\int_{b_n}^{\infty} P(T_n > x)dx \sim \Delta(b_n)$, finishing the proof of this case.

Now assume $b_n \to b \in (-\infty, \infty)$. Suppose $\varepsilon > 0$; we will show

$$\left| \int_{b_n}^{\infty} P(T_n > x)dx - \Delta(b_n) \right| \leq \varepsilon$$

for large n. Since $\Delta(b')$, $b'P(T_n > b')$, and $\sqrt{P(T_n > b')}$ all approach 0 as $b' \to \infty$, we can choose $b' > b$ such that all these are less than $\varepsilon/4$ when n is at least some arbitrary, fixed n_o. First write

$$\int_{b_n}^{\infty} P(T_n > x)dx = \int_{b_n}^{b'} P(T_n > x)dx + \int_{b'}^{\infty} P(T_n > x)dx.$$

Using the Berry-Esseen Theorem,

$$\begin{aligned} \int_{b_n}^{b'} P(T_n > x)dx &= [1 + O(1/\sqrt{n})] \int_{b_n}^{b'} \Phi(-x)dx \\ &= [1 + O(1/\sqrt{n})][\Delta(b_n) - \Delta(b')] \end{aligned}$$

and so

$$\begin{aligned} \left| \int_{b_n}^{b'} P(T_n > x)dx - \Delta(b_n) \right| &\leq \Delta(b') + O(1/\sqrt{n})[\Delta(b_n) - \Delta(b')] \\ &\leq \varepsilon/4 + O(1/\sqrt{n}) \cdot O(1) \\ &\leq \varepsilon/4 + \varepsilon/4 = \varepsilon/2 \end{aligned} \tag{2.70}$$

for sufficiently large n. Then

$$\begin{aligned} &\left| \int_{b_n}^{\infty} P(T_n > x)dx - \Delta(b_n) \right| \leq \left| \int_{b_n}^{b'} P(T_n > x)dx - \Delta(b_n) \right| + \left| \int_{b'}^{\infty} P(T_n > x)dx \right| \\ &\leq \varepsilon/2 + E(T_n; T_n > b') + b'P(T_n > b') \quad \text{(by (2.70) and (2.68))} \\ &\leq \varepsilon/2 + \sqrt{P(T_n > b')} + \varepsilon/4 \quad \text{(by (2.69))} \\ &\leq \varepsilon/2 + \varepsilon/4 + \varepsilon/4 = \varepsilon, \end{aligned}$$

finishing the proof. □

Recall our definition

$$t(p, a) = a/\mu - \frac{z_p\sqrt{4a\mu + z_p^2} - z_p^2}{2\mu^2}$$

as the unique solution of

$$\frac{a - \mu t(p, a)}{\sqrt{t(p, a)}} = z_p,$$

so that the probability of Brownian motion being across a boundary a units away at the end of a stage of size $t(p, a)$ is p. Recall also that geometric sampling is defined as sampling so that this stopping probability is constant across the stages. We now extend this definition to discrete, possibly non-Gaussian data as follows. Define *(discrete) geometric sampling with probability* p to be the procedure (N, M) such that

$$\begin{aligned} N_k &\equiv \lceil t(p, a - S_{k-1}) \rceil 1\{S_{k-1} < a\}, \quad k \geq 1 \\ M &\equiv \inf\{m \geq 1 : S_m \geq a\}. \end{aligned}$$

Note that when the X_i are not Gaussian, we do not know a priori that the *true* stopping probability is close to p, nor that M behaves like a geometric random variable in any sense. However, we will see that both of these are true below by the Central Limit Theorem and large deviations theory. Our next lemma establishes upper bounds on discrete geometric sampling when the stopping probability approaches 1 as $a \to \infty$.

Lemma 2.11. *Let (N, M) be discrete geometric sampling with probability $p(a)$. There is a constant $p_o \in (1/2, 1)$ such that if $p(a) \geq p_o$ and $p(a) \to 1$ as $a \to \infty$, then*

$$EN - a/\mu \lesssim |z_{p(a)}|\sqrt{a}/\mu^{3/2} \tag{2.71}$$

$$EM \to 1 \tag{2.72}$$

and

$$\sup_{a\le b} EN - a/\mu \;<\; \infty \tag{2.73}$$
$$\sup_{a\le b} EM \;<\; \infty \tag{2.74}$$

for any $b<\infty$.

Proof. First we prove the statements regarding EM. Let $p>1/2$, $x>0$, $\Sigma_n = X_1+\cdots+X_n$, and $n(p,x)=\lceil t(p,x)\rceil$. Write

$$\begin{aligned}\left|P(\Sigma_{n(p,x)}<x)-(1-p)\right| \le & \left|P(\Sigma_{n(p,x)}<x)-\Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right)\right| \\ & + \left|\Phi(z_p)-\Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right)\right|.\end{aligned} \tag{2.75}$$

By the Berry-Esseen Theorem there is a constant C_1 such that

$$\left|P(\Sigma_{n(p,x)}<x)-\Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right)\right| \le \frac{C_1}{\sqrt{n(p,x)}}. \tag{2.76}$$

Since $n(p,x)\ge t(p,x)$ we have

$$\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}} \le \frac{x-\mu t(p,x)}{\sqrt{t(p,x)}} = z_p.$$

Then, using the inequality

$$\Phi(x)-\Phi(y) \le \phi(x)(x-y) \quad \text{for } y\le x\le 0,$$

$$\begin{aligned}
\left|\Phi(z_p) - \Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right)\right| &= \Phi(z_p) - \Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right) \\
&\le \phi(z_p)\left[z_p - \frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right] \\
&\le \phi(z_p)\left[z_p - \frac{x-\mu(t(p,x)+1)}{\sqrt{t(p,x)+1}}\right] \quad (\text{since } n(p,x) \le t(p,x)+1) \\
&= \phi(z_p)\left[z_p - \frac{x-\mu t(p,x)}{\sqrt{t(p,x)}} \cdot \frac{\sqrt{t(p,x)}}{\sqrt{t(p,x)+1}} + \frac{\mu}{\sqrt{t(p,x)+1}}\right] \\
&= \phi(z_p)\left[z_p\left(1 - \frac{\sqrt{t(p,x)}}{\sqrt{t(p,x)+1}}\right) + \frac{\mu}{\sqrt{t(p,x)+1}}\right]. \qquad (2.77)
\end{aligned}$$

Since $t(p,x) \to \infty$ as $p \to 1$ and

$$1 - \frac{\sqrt{t(p,x)}}{\sqrt{t(p,x)+1}} \le \frac{1}{2t(p,x)} = o(1/\sqrt{t(p,x)}),$$

from (2.77) we get that there is $C_2 < \infty$, $p_o \in (1/2, 1)$ such that

$$\left|\Phi(z_p) - \Phi\left(\frac{x-\mu n(p,x)}{\sqrt{n(p,x)}}\right)\right| \le C_2 \frac{\phi(z_p)}{\sqrt{n(p,x)}} \qquad (2.78)$$

for $p \ge p_o$. Combining (2.76) and (2.78) into (2.75), we have

$$P(\Sigma_{n(p,x)} < x) \le 1 - p + (C_1 + C_2)\frac{\phi(z_p)}{\sqrt{n(p,x)}} \le 1 - p + C_3 \frac{\phi(z_p)}{|z_p|},$$

some $C_3 < \infty$, since $\sqrt{n(p,x)} \ge \sqrt{t(p,x)} \ge |z_p|/(2\mu)$. Then

$$\begin{aligned}
P(M > k+1 | M > k) &= P(\Sigma_{n(p,a-S_k)} < a - S_k | a - S_k > 0) \\
&\le 1 - p + C_3 \frac{\phi(z_p)}{|z_p|}. \qquad (2.79)
\end{aligned}$$

Plugging $p = p(a)$ into this and assuming a is large enough so that

$$1 - p(a) + C_3 \frac{\phi(z_{p(a)})}{|z_{p(a)}|} \le 1/2, \qquad (2.80)$$

we have $P(M > k+1|M > k) \leq 1/2$ for all $k \geq 1$, and hence

$$\begin{aligned} P(M > k) &= P(M > k|M > k-1) \cdots P(M > 2|M > 1)P(M > 1) \\ &\leq (1/2)^{k-1} P(M > 1). \end{aligned}$$

This gives

$$EM = 1 + P(M > 1) + P(M > 2) + \cdots \leq 1 + 2P(M > 1) \to 1 \tag{2.81}$$

as $a \to \infty$ since, by (2.79),

$$P(M > 1) \leq 1 - p(a) + C_3 \frac{\phi(z_{p(a)})}{|z_{p(a)}|} \to 0,$$

proving (2.72). If we choose p_o large enough so that (2.80) holds for all $a > 0$, then (2.81) shows that $EM \leq 3$ for all $a > 0$, proving (2.74).

Next we estimate EN. Let $p = p(a)$. We have $N_1 = n(p, a) \leq t(p, a) + 1$. For N_2, consider $E(a - S_1|S_1 < a)$. Let $S_1' = \sum_{i=1}^{N_1}(2\mu - X_i)$ and $a' = 2N_1\mu - a$. Note that $2\mu - X_1, 2\mu - X_2, \ldots$ are i.i.d. with mean μ and variance 1, and

$$\zeta_a \equiv \frac{a' - \mu N_1}{\sqrt{N_1}} = \frac{\mu N_1 - a}{\sqrt{N_1}} = |z_p| + o(1) = o(a^{1/6}).$$

Hence Lemma 2.10 applies and

$$E(a - S_1|S_1 < a) = E(S_1' - a'|S_1' > a') \sim \sqrt{N_1}\Delta(\zeta_a)/\Phi(-\zeta_a),$$

using $P(S_1' > a') \sim \Phi(-\zeta_a)$ by large deviations. Since $\zeta_a \sim |z_p|$ and $\sqrt{N_1} \sim \sqrt{a/\mu}$, this shows

$$E(a - S_1|S_1 < a) \sim \sqrt{a/\mu}\Delta(z_{1-p})/(1 - p).$$

Now let $k \geq 2$. Since $t(p, \cdot)$ is increasing and concave,

$$\begin{aligned} E(N_2; M = k) &= E(N_2|M = k)P(M = k) \\ &\leq E[t(p, a - S_1) + 1|M = k]P(M = k) \\ &\leq [t(p, E[a - S_1|M = k]) + 1]P(M = k) \\ &= [t(p, E[a - S_1|S_1 < a]) + 1]P(M = k) \qquad (2.82) \\ &= [t(p, \sqrt{a/\mu} \cdot \Delta(z_{1-p})(1 - p)^{-1}(1 + o(1))) + 1]P(M = k). \qquad (2.83) \end{aligned}$$

In (2.82) we use

$$E[a - S_1|M = k] = E[a - S_1|M > 1] = E[a - S_1|S_1 < a];$$

this is true since the number of additional stages required to cross the boundary is independent of S_1, provided $S_1 < a$.

To estimate N_i for $i > 2$ we will bound $E[N_{i+1} - N_i]$. Let $2 < i < k$.

$$\begin{aligned} E[N_{i+1} - N_i; M = k] &\leq E[t(p, a - S_i) + 1 - t(p, a - Si - 1); M = k] \\ &= E[t(p, a - S_{i-1} - (S_i - S_{i-1})) - t(p, a - S_{i-1}); M = k] + P(M = k). (2.84) \end{aligned}$$

Since

$$0 < \frac{d}{dx} t(p, x) \leq 2/\mu$$

for all $x > 0$ when $p \geq 1/2$,

$$t(p, x - y) - t(p, x) \leq 2y^-/\mu$$

and thus (2.84) becomes

$$E[N_{i+1} - N_i; M = k] \leq (2/\mu)E[(S_i - S_{i-1})^-; M = k] + P(M = k). \qquad (2.85)$$

Recall that $\Sigma_n = X_1 + \cdots + X_n$ and let $\varphi(n) = E(-\Sigma_n; -\Sigma_n > 0)$. For large n

$$\begin{aligned}\varphi(n) &= E(\Sigma_n - 0; -\Sigma_n > 0) \\ &\leq E[-\Sigma_n - (-n\mu + n^{9/14}); -\Sigma_n > (-n\mu + n^{9/14})] \quad (\text{since } -n\mu + n^{9/14} < 0) \\ &\sim \sqrt{n} \cdot \Delta(n^{1/7}) \sim \sqrt{n} \cdot \frac{\phi(n^{1/7})}{n^{2/7}} \lesssim e^{-n^{1/7}},\end{aligned}$$

this last line using Lemma 2.10 since

$$\frac{-n\mu + n^{9/14} - n \cdot E(-X_1)}{\sqrt{n}} = n^{1/7} = o(n^{1/6}).$$

Thus

$$\varphi(n) \leq (3/2)e^{-n^{1/7}}, \tag{2.86}$$

say, for large n. Now

$$\begin{aligned}E[(S_i - S_{i-1})^-; M = k] &= E[(S_i - S_{i-1})^-; M > i]P(M = k|M > i) \\ &= E[\varphi(n(p, a - S_{i-1})); S_{i-1} < a]P(M = k|M > i) \qquad (2.87)\end{aligned}$$

using the same conditioning argument as above. Also, for all $x > 0$

$$n(p, x) \geq \frac{|z_p|}{\mu} \geq \frac{|z_{p_o}|}{\mu} \equiv \underline{n}.$$

Combining (2.85)-(2.87),

$$\begin{aligned}E(N_{i+1} - N_i; M = k) &\leq (2/\mu) \cdot (3/2)e^{-\underline{n}^{1/7}} P(S_{i-1} < a)P(M = k|M > i) + P(M = k) \\ &= (3/\mu)e^{-\underline{n}^{1/7}} P(M \geq i)P(M = k|M > i) + P(M = k). \qquad (2.88)\end{aligned}$$

By (2.79) there is a constant C such that the probability of crossing the boundary at each stage is at least $p - C\phi(\mu\underline{n})/\underline{n}$. Assuming p_o (and hence $\underline{n}$) are large enough so that

$$(3/\mu)e^{-\underline{n}^{1/7}} \leq 1/2 \leq p_o - C\phi(\mu\underline{n})/\underline{n},$$

by (2.88) we have

$$E(N_{i+1} - N_i; M = k) \le (1/2) \cdot (1/2)^{i-1} \cdot (1/2)^{k-i-1} + (1/2)^{k-1} = (1/2)^{k-2}$$

and thus

$$E(N_i; M = k) \le E(N_2; M = k) + (i-2)(1/2)^{k-2}$$

for $2 \le i \le k$. Combining this with (2.83) we have

$$\begin{aligned}
EN &= N_1 + \sum_{k\ge 2}\sum_{i=2}^{k} E(N_i; M = k) \\
&\le N_1 + \sum_{k\ge 2}[(k-1)E(N_2; M = k) + (1/2)^{k-1}(k-2)(k-1)] \\
&\le t(p,a) + 1 + [t(p, \sqrt{a/\mu} \cdot \Delta(z_{1-p})(1-p)^{-1}(1+o(1))) + 1]\sum_{k\ge 2}(k-1)P(M = k) \\
&\quad + \sum_{k\ge 2}(1/2)^{k-1}(k-2)(k-1) \\
&= t(p,a) + 1 + [t(p, \sqrt{a/\mu} \cdot \Delta(z_{1-p})(1-p)^{-1}(1+o(1))) + 1](EM-1) + 2 \\
&\le t(p,a) + t(p, \sqrt{a/\mu} \cdot \Delta(z_{1-p})(1-p)^{-1}(1+o(1)))(EM-1) + 5, \qquad (2.89)
\end{aligned}$$

using $EM - 1 \le 2$. By (2.79),

$$\begin{aligned}
EM - 1 &= P(M > 1) + P(M > 2) + \cdots \\
&\le P(M > 1)[1 + C_3\phi(z_p)/|z_p| + (C_3\phi(z_p)/|z_p|)^2 + \cdots] \\
&= P(M > 1)[1 - C_3\phi(z_p)/|z_p|]^{-1}.
\end{aligned}$$

We know $P(M > 1) \sim 1 - p$ by large deviations, and $[1 - C_3\phi(z_p)/|z_p|]^{-1} \to 1$, so $EM - 1 \sim 1 - p$. Then

$$\begin{aligned}
t(p, \sqrt{a/\mu} \cdot \Delta(z_{1-p})(1-p)^{-1}(1+o(1)))(EM-1) &\sim \sqrt{a}/\mu^{3/2} \cdot \Delta(z_{1-p})(1-p)^{-1} \cdot (1-p) \\
&= \sqrt{a}/\mu^{3/2} \cdot \Delta(z_{1-p}) = o(\sqrt{a}).
\end{aligned}$$

Plugging this and the estimate

$$t(p, a) = a/\mu + |z_p|\sqrt{a}/\mu^{3/2} + o(|z_p|\sqrt{a})$$

into (2.89) we get that for large a

$$EN \leq a/\mu + |z_p|\sqrt{a}/\mu^{3/2} + o(|z_p|\sqrt{a}),$$

which is (2.71).

For small a, $|z_p|$ is bounded so $t(p, a)$ is as well. N_1 is thus bounded, whence $E(a - S_1|S_1 < a)$ is bounded and thus so is $t(p, E(a - s_a|S_1 < a))$. Then, for any $b < \infty$,

$$\sup_{a \leq b} EN \leq \sup_{a \leq b} \{t(p, a) + 1 + [t(p, E[a - S_1|S_1 < a]) + 1](EM - 1)\} < \infty,$$

which is (2.73), completing the proof. □

2.2.1 The Discrete Procedures δ_m and $\hat{\delta}_m$

In this section we describe two families of sampling procedures, δ_m and $\hat{\delta}_m$, and establish their operating characteristics. In the next section we will see that these properties are enough to make them first-order optimal, δ_m when $h \in \mathcal{B}_m^o$ and $\hat{\delta}_m$ when $h \in \mathcal{B}_m^+$. These procedures are defined analogously to those for Brownian motion in Section 2.1.2, with minor modifications to account for discrete data. The proofs of their operating characteristics are similar to those in Section 2.1.2, but significant additional Central Limit Theorem-type arguments are required.

Let $C = 2/\sqrt{\mu}$ and $f(a) = C\sqrt{a\log(a+1)}$. Given a positive function h, define $\delta_1(h)$ to be geometric sampling with probability $p_o \leq p_1^{(1)}(a) \to 1$ such that

$$\left|z_{p_1^{(1)}(a)}\right| = o[(h(a)/\sqrt{a}) \wedge a^{1/6}],$$

where p_o is that given by Lemma 2.11. Define $\delta_{m+1}(h) = (N^{(m+1)}, M^{(m+1)})$ to have first stage $N_1^{(m+1)} = \lceil t(p_1^{(m+1)}, a)\rceil$, where

$$z_{p_1^{(m+1)}} = C\sqrt{\log(a/h(a)^2+1)},$$

followed on $\{S_1 < a\}$ by $(N^{(m)}(a-S_1), M^{(m)}(a-S_1))$, where $(N^{(m)}, M^{(m)}) = \delta_m(h \circ f^{-1})$.

For $p \in (0,1)$, define $\hat{\delta}_1(p) = (N^{(1)}, M^{(1)})$ to have first stage $N_1^{(1)} = \lceil t(p,a)\rceil$, followed (on $\{S_1 < a\}$) by geometric sampling with probability $\hat{p}_2^{(1)}(a-S_1)$, where

$$z_{\hat{p}_2^{(1)}(y)} = (-\sqrt{\log(y+1)} \wedge z_{p_o})$$

and p_o is that given by Lemma 2.11. Define $\hat{\delta}_{m+1}(p) = (N^{(m+1)}, M^{(m+1)})$ to have first stage $N_1^{(m+1)} = \lceil t(\hat{p}_1^{(m+1)}(a), a)\rceil$, where

$$z_{\hat{p}_1^{(m+1)}(a)} = \sqrt{(1-2^{-m})\log(a+1)},$$

followed (if necessary) by $(N^{(m)}(a-S_1), M^{(m)}(a-S_1))$, where $(N^{(m)}, M^{(m)}) = \hat{\delta}_m(p)$.

Proposition 2.12. *If $h \in \mathcal{B}_m^o$ then $(N^{(m)}, M^{(m)}) = \delta_m(h)$ satisfies*

$$EN^{(m)} - a/\mu = o(h(a)) \tag{2.90}$$
$$EM^{(m)} \to m \tag{2.91}$$

as $a \to \infty$.

Proof. We prove a slightly stronger statement by induction on m: in addition to (2.90) and (2.91) we show that if $b < \infty$, then

$$\sup_{a \le b} EN^{(m)} - a/\mu < \infty \tag{2.92}$$
$$\sup_{a \le b} EM^{(m)} < \infty. \tag{2.93}$$

For $m = 1$, $(N^{(1)}, M^{(1)})$ is discrete geometric sampling, and Lemma 2.11 shows that (2.91) and the boundedness properties of $N^{(1)}, M^{(1)}$ hold, and as well as

$$EN^{(1)} - a/\mu = O(|z_p|\sqrt{a}),$$

where $|z_p| = o(h(a)/\sqrt{a} \wedge a^{1/6})$. By this restriction on z_p,

$$EN^{(1)} - a/\mu \leq o(h(a)/\sqrt{a} \cdot \sqrt{a}) = o(h(a)),$$

so (2.45) holds as well, completing the $m = 1$ case.

Now assume $h \in \mathcal{B}^o_{m+1}$. Let $C = 2/\sqrt{\mu}$ and define $f(a) = C\sqrt{a \log(a+1)}$, whose inverse is well-defined since f is increasing. It was shown in the proof of Proposition 2.3 (see (2.18)) that

$$h \circ f \in \mathcal{B}^o_m. \tag{2.94}$$

Now

$$EM^{(m+1)}(a) = 1 + E[M^{(m)}(a - S_1); S_1 < a]$$

so $EM^{(m+1)}(a)$ is bounded for small a since $EM^{(m)}(a)$ is by the induction hypothesis. Further, letting $z_1 = C\sqrt{\log(a/h(a)^2 + 1)}$, $C' = (2\sqrt{\mu})^{-1}$, and $Z = (S_1 - \mu N_1)/\sqrt{N_1}$, observe that

$$\begin{aligned}
P(0 < a - S_1 \leq C' z_1\sqrt{a}) &\leq P(S_1 > a - C' z_1 \sqrt{a}) \\
&= P\left(Z > \frac{a - \mu N_1}{\sqrt{N_1}} - \frac{C' z_1 \sqrt{a}}{\sqrt{N_1}}\right) \\
&= P\left(Z > z_1 + o(1) - \frac{C' z_1 \sqrt{a}}{\sqrt{N_1}}\right) \\
&= P(Z > z_1[1 - C'\sqrt{\mu}](1 + o(1))) \quad (\text{since } \sqrt{N_1} \sim \sqrt{a/\mu}) \\
&= P(Z > (z_1/2)(1 + o(1))) \\
&\leq (z_1/4)^{-2} \to 0
\end{aligned} \tag{2.95}$$

by Chebyshev's inequality. Then

$$\begin{aligned} EM^{(m+1)}(a) &= 1 + E[M^{(m)}(a - S_1); a - S_1 > C' z_1 \sqrt{a}] \\ &\quad + E[M^{(m)}(a - S_1); 0 < a - S_1 \leq C' z_1 \sqrt{a}] \\ &= 1 + m(1 + o(1))P(a - S_1 > C' z_1 \sqrt{a}) + E[M^{(m)}(a - S_1); 0 < a - S_1 \leq C' z_1 \sqrt{a}], \end{aligned}$$

and so

$$\left| EM^{(m+1)}(a) - (m+1) \right| \leq m \cdot o(1) + O(1) \cdot P(0 < a - S_1 \leq C' z_1 \sqrt{a}) = o(1)$$

as $a \to \infty$. We've shown that (2.91) and (2.93) hold for $m + 1$.

Next we handle $EN^{(m+1)}$. Using Wald's equation

$$\begin{aligned} \mu E(N^{(m+1)} - a/\mu) &= E(S_{M^{(m+1)}} - a) \\ &= E(S_{M^{(m+1)}} - a; M^{(m+1)} = 1) + E(S_{M^{(m+1)}} - a; M^{(m+1)} > 1) \\ &= E(S_{M^{(m+1)}} - a; M^{(m+1)} = 1) + E(S_{M^{(m+1)}} - a; a - S_1 > C z_1 \sqrt{a}) \\ &\quad + E(S_{M^{(m+1)}} - a; 0 < a - S_1 \leq C z_1 \sqrt{a}) \\ &\equiv A_1 + A_2 + A_3. \end{aligned}$$

To show that (2.90) and (2.92) hold for $m + 1$ it suffices to show the A_i satisfy the same bounds. Note that $A_1 = E(S_1 - a; S_1 \geq a)$ and

$$\frac{a - \mu N_1^{(m+1)}}{\sqrt{N_1^{(m+1)}}} \sim z_1 = o(a^{1/6})$$

so by Lemma 2.10,

$$A_1 \sim \sqrt{N_1^{(m+1)}} \Delta(z_1) \sim \sqrt{N_1^{(m+1)}} \frac{\phi(z_1)}{z_1^2} = O(\sqrt{a}) \cdot O(h(a)/\sqrt{a})/z_1^2 = o(h(a)). \quad (2.96)$$

For small values of a, $N_1^{(m+1)}(a)$ is bounded, hence $A_1 < \infty$ by the existence of the first moment of X_1.

Let $\varphi(y) = E(S_{M^{(m)}(y)} - y)$ for $y > 0$. By the induction hypothesis and (2.94) we know

$$\varphi(y) = o(h(f^{-1}(y))) \tag{2.97}$$

and that $\varphi(y)$ is bounded for bounded values of y. Note that

$$h(f^{-1}(y)) = o(h_{m-1}(y)) = o(y) \tag{2.98}$$

since the $m = 1$ gives the largest asymptotically. Let $\varepsilon > 0$ and $Y = a - S_1$. (2.97) and (2.98) imply that for large a,

$$\begin{aligned} A_2 &= E[\varphi(Y); Y > Cz_1\sqrt{a}] \\ &\leq \varepsilon E[Y; Y > Cz_1\sqrt{a}] \\ &= \varepsilon(E[Y - Cz_1\sqrt{a}; Y > Cz_1\sqrt{a}] + Cz_1\sqrt{a}P(Y > Cz_1\sqrt{a})). \end{aligned} \tag{2.99}$$

Let $S_1' = \sum_{i=1}^{N_1}(2\mu - X_i)$ and $a' = Cz_1\sqrt{a} - a + 2\mu N_1^{(m+1)}$ so that (2.99) becomes

$$A_2 \leq \varepsilon(E[S_1' - a'; S_1' > a'] + Cz_1\sqrt{a}P(S_1' > a').$$

Note that $2\mu - X_1, 2\mu - X_2, \ldots$ are i.i.d. with mean μ, variance 1, and that

$$\frac{a' - \mu N_1^{(m+1)}}{\sqrt{N_1^{(m+1)}}} = \frac{Cz_1\sqrt{a}}{\sqrt{N_1^{(m+1)}}} - \frac{a - \mu N_1^{(m+1)}}{\sqrt{N_1^{(m+1)}}} \sim 2z_1 - z_1 = z_1 = o(a^{1/6}), \tag{2.100}$$

so by Lemma 2.10,

$$E(S_1' - a'; S_1' > a') \sim \sqrt{N_1^{(m+1)}}\Delta(z_1) \sim \sqrt{N_1}\frac{\phi(z_1)}{z_1^2} = o(h(a)) \tag{2.101}$$

by (2.96). By large deviations ([10], Theorem XVI.7.1) and (2.100) $P(S_1' > a') \sim$

$\Phi(-z_1)$ so

$$\begin{aligned} Cz_1\sqrt{a}P(S_1' > a') \sim Cz_1\sqrt{a}\cdot\Phi(-z_1) &\sim Cz_1\sqrt{a}\cdot\frac{\phi(z_1)}{z_1} \\ &= C\sqrt{a}\cdot O(h(a)/\sqrt{a}) = O(h(a)), \end{aligned}$$

hence there is $C'' < \infty$ such that

$$Cz_1\sqrt{a}P(S_1' > a') \le C''h(a) \tag{2.102}$$

for large a. Plugging (2.101) and (2.102) into (2.99) we have $A_2 \le o(h(a) + \varepsilon C''h(a)$. Since ε was arbitrary and independent of C'', this shows $A_2 = o(h(a))$. For small values of a, (2.97) and (2.98) imply that there are constants C_1, a_1 such that

$$A_2 \le C_1 + E(Y; Y > a_1)$$

and the latter is finite by the same argument used on A_1, showing that A_2 is bounded for small values of a.

A_3 is bounded for small values of a by virtue of (2.97). To show $A_3 = o(h(a))$ it thus suffices to show

$$\tilde{A}_3 \equiv E(\varphi(Y); a_o < Y < Cz_1\sqrt{a}) = o(h(a))$$

for any constant a_o. Let a_o be such that

$$\varphi(y) \le h(f^{-1}(y)) \quad \text{for } y > a_o. \tag{2.103}$$

Then

$$\tilde{A}_3 = E(\varphi(Y); a_o < Y \le C'z_1\sqrt{a}) + E(\varphi(Y); C'z_1\sqrt{a} < Y \le Cz_1\sqrt{a}) \tag{2.104}$$

and $h \circ f^{-1}$ is non-decreasing, so (2.103) implies

$$E(\varphi(Y); a_o < Y \leq C' z_1 \sqrt{a}) \leq h(f^{-1}(C' z_1 \sqrt{a})) P(0 < Y \leq C' z_1 \sqrt{a}) \leq h(a) \cdot o(1),$$

by (2.95) and since $C' z_1 \sqrt{a} < f(a)$ for large a. Given $\varepsilon > 0$, let a be large enough so that

$$\begin{aligned} E(\varphi(Y); a_o < Y \leq C' z_1 \sqrt{a}) &\leq (\varepsilon/2) h(a) \\ \varphi(y) &\leq (\varepsilon/2) h(f^{-1}(y)) \end{aligned}$$

whenever $y > C' z_1 \sqrt{a}$. Plugging these into (2.104) gives

$$\tilde{A}_3 \leq (\varepsilon/2) h(a) + (\varepsilon/2) h(f^{-1}(C z_1 \sqrt{a})) \leq (\varepsilon/2) h(a) + (\varepsilon/2) h(a) = \varepsilon h(a)$$

which shows $\tilde{A}_3 = o(h(a))$ and hence that $A_3 = o(h(a))$, completing the induction step and the proof. □

Next we establish the operating characteristics of $\hat{\delta}_m(p)$.

Proposition 2.13. *Let $p \in (0,1)$, $m \geq 1$, and κ_m as in (2.24). Then $(N^{(m)}, M^{(m)}) = \hat{\delta}_m(p)$ satisfy*

$$\begin{aligned} EN^{(m)} - a/\mu &\lesssim \Delta(z_p) \kappa_m h_m(a) && (2.105) \\ EM^{(m)} &\to m + 1 - p && (2.106) \end{aligned}$$

as $a \to \infty$.

Proof. As in the proof of the previous proposition, we will prove a slightly stronger statement by induction on m. In addition to (2.105) and (2.106), we will show that if $b < \infty$, then

$$\begin{aligned} \sup_{a \le b} EN^{(m)} - a/\mu &< \infty \\ \sup_{a \le b} EM^{(m)} &< \infty. \end{aligned}$$

By Wald's equation,

$$\mu E(N^{(1)} - a/\mu) = E(S_{M^{(1)}} - a; M^{(1)} = 1) + E(S_{M^{(1)}} - a; M^{(1)} > 1).$$

Now

$$\frac{a - \mu N_1^{(1)}}{\sqrt{N_1^{(1)}}} \to z_p,$$

a constant, so by Lemma 2.10

$$E(S_{M^{(1)}} - a; M^{(1)} = 1) \sim \sqrt{N_1^{(1)}} \cdot \Delta(z_p) \sim \sqrt{a/\mu} \cdot \Delta(z_p), \tag{2.107}$$

since $\sqrt{N_1^{(1)}} \sim \sqrt{a/\mu}$. For $y > 0$ let $(N'(y), M'(y))$ be the discrete geometric procedure with probability $p_2(y)$ that follows the first stage when $S_1 < a$. By Lemma 2.11 we know

$$\varphi(y) \equiv EN'(y) - y/\mu = O(|z_{p_2(y)}|\sqrt{y}), \qquad \text{and} \qquad \sup_{y \le x} \varphi(y) > \infty \tag{2.108}$$

$$\psi(y) \equiv EM'(y) = 1 + o(1), \qquad \text{and} \qquad \sup_{y \le x} \psi(y) > \infty \tag{2.109}$$

for any $x < \infty$. Then, letting $Y = a - S_1$, $E(S_{M^{(1)}} - a; M^{(1)} > 1) = E(\varphi(Y); Y > 0)$. By (2.108) there are constants y_o, C_o, C_1 such that

$$\varphi(y) \le \begin{cases} C_o, & 0 < y \le y_o \\ C_1 |z_{p_2}| \sqrt{y}, & y > y_o. \end{cases}$$

We may also assume y_o is large enough so that $|z_{p_2(y)}| = \sqrt{\log y}$ for $y > y_o$. Then, using concavity of $y \mapsto \sqrt{y \log y}$ with Jensen's inequality,

$$\begin{aligned} E(\varphi(Y); Y > 0) &\leq C_o + C_1 E(\sqrt{Y \log Y}; Y > y_o) \\ &\leq C_o + C_1 \sqrt{E(Y|Y > y_o) \log E(Y|Y > y_o)}, \end{aligned} \tag{2.110}$$

and

$$E(Y|Y > y_o) = P(Y > y_o)^{-1} E(Y; Y > y_o) = O(\sqrt{a})$$

by an argument similar to the one leading to (2.107). Plugging this into (2.110) gives

$$E(S_{M^{(1)}} - a; M^{(1)} > 1) = E(\varphi(Y); Y > 0) \leq O(a^{1/4}\sqrt{\log a}) = o(\sqrt{a}),$$

and combining this with (2.110) gives

$$EN^{(1)} - a/\mu = \Delta(z_p)\mu^{-3/2}\sqrt{a} + o(\sqrt{a}) = \Delta(z_p)\kappa_1 h_1(a) + o(h_1(a)).$$

For small values of a, $N_1^{(1)}$ is bounded and so $E(S_{M^{(1)}} - a; M^{(1)} = 1)$ is bounded as well. Similarly, $E(Y|Y > y_o) = P(Y > y_o)^{-1} E(a - S_1; S_1 < a - y_o)$ is bounded and so $E(S_{M^{(1)}} - a; M^{(1)} > 1) = E(\varphi(Y); Y > 0)$ is bounded as well, by the relation (2.110).

To handle $M^{(1)}$ we write

$$EM^{(1)} = 1 + E(M^{(1)} - 1; M^{(1)} > 1) = 1 + E(\psi(Y); Y > 0).$$

Given $\varepsilon > 0$, by (2.109) there are constants C_2, y_2 such that

$$\psi(y) \leq \begin{cases} C_2, & 0 < y \leq y_2 \\ 1 + \varepsilon, & y > y_2. \end{cases}$$

Then

$$EM^{(1)} \leq 1 + C_2 P(0 < Y \leq y_2) + (1 + \varepsilon) P(Y > y_2). \tag{2.111}$$

Since

$$\frac{a - \mu N_1^{(1)} - y_2}{\sqrt{N_1^{(1)}}}, \frac{a - \mu N_1^{(1)}}{\sqrt{N_1^{(1)}}} \to z_p$$

as $a \to \infty$, $P(0 < Y \leq y_2) \to 0$ and $P(Y > y_2) \to 1 - p$ by the Central Limit Theorem. Thus assume a is large enough so that $P(0 < Y \leq y_2) \leq \varepsilon/C_2$ and $P(Y > y_2) \leq 1 - p + \varepsilon$. Then

$$EM^{(1)} \leq 1 + \varepsilon + (1 + \varepsilon)(1 - p + \varepsilon) \leq 2 - p + 4\varepsilon$$

and a similar argument shows

$$EM^{(1)} \geq 1 + (1 - p)(1 - \varepsilon) \geq 2 - p - 2\varepsilon$$

for large a. Since ε was arbitrary, this implies $EM^{(1)} \to 2 - p$. $EM^{(1)}$ is also clearly bounded for small values of a; e.g. (2.111) holds for all $a > 0$ and shows $EM^{(1)} \leq 1 + C_2 + (1 + \varepsilon)$. This completes the $m = 1$ case.

Next we consider $(N^{(m+1)}, M^{(m+1)}) = \hat{\delta}_{m+1}(p)$. By Wald's equation

$$\begin{aligned} \mu E(N^{(m+1)} - a/\mu) &= E(S_{M^{(m+1)}} - a; M^{(m+1)} = 1) + E(S_{M^{(m+1)}} - a; M^{(m+1)} > 1) \\ &= E(S_1 - a; S_1 \geq a) + E(S_{M^{(m+1)}} - a; M^{(m+1)} > 1). \end{aligned}$$

Letting $z_1 = \sqrt{(1 - 2^{-m}) \log(a + 1)}$, by definition of $N_1^{(m+1)}$,

$$\frac{a - \mu N_1^{(m+1)}}{\sqrt{N_1^{(m+1)}}} \sim z_1 = o(a^{1/6})$$

so Lemma 2.10 applies and

$$\begin{aligned} E(S_1 - a; S_1 \ge a) &\sim \sqrt{N_1^{(m+1)}} \cdot \Delta(z_1) \sim \sqrt{a/\mu} \cdot \frac{\phi(z_1)}{z_1^2} \\ &= O\left(\sqrt{a} \cdot \frac{\exp[-(1/2 - (1/2)^{m+1}) \log a]}{\log a}\right) \\ &= O\left(\frac{a^{(1/2)^{m+1}}}{\log a}\right) \\ &= o(a^{(1/2)^{m+1}}) = o(h_{m+1}(a)). \end{aligned} \tag{2.112}$$

For $y > 0$ define

$$\varphi_m(y) = E(S_{M^{(m)}(y)} - y), \quad \psi_m(y) = EM^{(m)}(y).$$

Let $\varepsilon > 0$. By the induction hypothesis and Wald's equation there are constants C_3, y_3 such that

$$\varphi_m(y) \le \begin{cases} C_3, & 0 < y \le y_3 \\ \mu\Delta(z_p)\kappa_m h_m(y)(1+\varepsilon), & y > y_3. \end{cases} \tag{2.113}$$

Thus

$$\begin{aligned} &E(S_{M^{(m+1)}} - a; M^{(m+1)} > 1) = E(\varphi_m(Y); Y > 0) \\ &\le \; C_3 P(0 < Y \le y_3) + \mu\Delta(z_p)\kappa_m(1+\varepsilon)E(h_m(Y); Y > y_3). \end{aligned} \tag{2.114}$$

Since $h_m(\cdot)$ is concave, we apply Jensen's inequality to get

$$E(h_m(Y); Y > y_3) \le P(Y > y_3) h_m(E[Y; Y > y_3] P(Y > y_3)^{-1}) \tag{2.115}$$

and claim $E[Y;Y>y_3] \sim z_1\sqrt{a/\mu}$ as $a \to \infty$. This is true since

$$\begin{aligned} E[Y;Y>y_3] &= E[a-S_1;a-S_1>y_3] \\ &= E[a-S_1]-E[a-S_1;a-S_1\le y_3] \\ &= a-\mu N_1^{(m+1)}+E[S_1-(a-y_3);S_1\ge a-y_3]-y_3P(a-S_1\le y_3) \end{aligned}$$

and

$$\frac{a-y_3-\mu N_1^{(m+1)}}{\sqrt{N_1^{(m+1)}}} \sim z_1 = o(a^{1/6})$$

so Lemma 2.10 applies and

$$E[S_1-(a-y_3);S_1\ge a-y_3] \sim \sqrt{N_1^{(m+1)}}\cdot\Delta(z_1) \sim \sqrt{a/\mu}\cdot o(1) = o(\sqrt{a}).$$

Also, $a-\mu N_1^{(m+1)} \sim z_1\sqrt{N_1^{(m+1)}} \sim z_1\sqrt{a/\mu}$, so

$$E[Y;Y>y_3] = z_1\sqrt{a/\mu}(1+o(1))+o(\sqrt{a})+O(1) \sim z_1\sqrt{a/\mu}$$

as claimed. Note also that $P(Y>y_3)\to 1$ by the Central Limit Theorem, whence we may assume that a is large enough so that, by (2.115),

$$E(h_m(Y);Y>y_3) \le (1+\varepsilon)h_m(z_1\sqrt{a/\mu}).$$

Assuming that a is large enough so that also

$$h_{m+1}(a) \ge \frac{C_3}{\mu\varepsilon\Delta(z_p)\kappa_{m+1}} \ge \frac{C_3}{\mu\varepsilon\Delta(z_p)\kappa_{m+1}}\cdot P(0<Y\le y_3),$$

we have, by (2.114),

$$\begin{aligned} &E(S_{M^{(m+1)}}-a;M^{(m+1)}>1) \le \mu\varepsilon\Delta(z_p)\kappa_{m+1}h_{m+1}(a)+\mu\Delta(z_p)(1+\varepsilon)^2\kappa_m h_m(z_1\sqrt{a/\mu}) \\ &\le \mu\varepsilon\Delta(z_p)\kappa_{m+1}h_{m+1}(a)+\mu\Delta(z_p)(1+\varepsilon)^3\kappa_{m+1}h_{m+1}(a) \quad \text{(by Lemma 2.4)} \\ &\le (1+8\varepsilon)\mu\Delta(z_p)\kappa_{m+1}h_{m+1}(a), \end{aligned}$$

and hence

$$EN^{(m+1)} - a/\mu = \mu^{-1}E(S_{M^{(m+1)}} - a) \leq (1 + 8\varepsilon + o(1))\Delta(z_p)\kappa_{m+1}h_{m+1}(a)$$

by (2.112). Since ε was arbitrary, this shows $EN^{(m+1)} - a/\mu \lesssim \Delta(z_p)\kappa_{m+1}h_{m+1}(a)$, as claimed.

For small values of a, z_1 and hence $N_1^{(m+1)}$ are bounded and so $E(S_{M^{(m+1)}} - a; M^{(m+1)} = 1) = E(S_1 - a; S_1 \geq a)$ is bounded as well. Similarly, $E(a - S_1; S_1 < a)$ is bounded and so (2.114) and (2.115) show that $E(S_{M^{(m+1)}} - a; M^{(m+1)} > 1)$ is bounded too.

Next we consider $M^{(m+1)}$.

$$EM^{(m+1)} = 1 + E(M^{(m+1)} - 1; M^{(m+1)} > 1) = 1 + E(\psi_m(Y); Y > 0),$$

where again $Y = a - S_1$. Given $\varepsilon > 0$, by the induction hypothesis there are constants C_4, y_4 such that

$$\psi_m(y) \leq \begin{cases} C_4, & 0 < y \leq y_4 \\ (1+\varepsilon)(m+1-p), & y > y_4, \end{cases}$$

and thus

$$EM^{(m+1)} \leq 1 + C_4P(0 < Y \leq y_4) + (1+\varepsilon)(m+1-p)P(Y > y_4). \tag{2.116}$$

$P(0 < Y \leq y_4) \to 0$ as $a \to \infty$ by a now routine Central Limit Theorem argument, so assume a is large enough so that $P(0 < Y \leq y_4) \leq \varepsilon/C_4$. Then

$$EM^{(m+1)} \leq 1 + \varepsilon + (1+\varepsilon)(m+1-p) = (1+\varepsilon)(m+2-p).$$

By a similar argument,

$$EM^{(m+1)} \geq 1 + (1-\varepsilon)^2(m+1-p) \geq (1-\varepsilon)^2(m+2-p)$$

for large enough a. These two bounds show $EM^{(m+1)} \to m+2-p$ since ε was arbitrary. For small values of a, $EM^{(m+1)}$ is bounded; e.g. (2.116) holds for all $a > 0$ and shows that $EM^{(m+1)} \le 1 + C_4 + (1+\varepsilon)(m+1-p)$. This completes the $m+1$ step and hence the proof. □

2.2.2 Optimality of δ_m and $\hat{\delta}_m$

In this section we prove our main results for i.i.d. sampling procedures: that δ_m (resp. $\hat{\delta}_m$) is first-order optimal when $h \in \mathcal{B}_m^o$ (resp. $h \in \mathcal{B}_m^+$). Again, the proofs are similar in spirit to those for Brownian motion in Section 2.1.3, but additional Central Limit Theorem-type arguments are needed.

Before getting to the main results in Theorem 2.15, we provide in the next lemma a bound on how close any efficient procedure can be to the boundary after each of the first $m-1$ stages of sampling when $h \in \mathcal{B}_m$. This is the discrete analog of Lemma 2.7.

Lemma 2.14. *If $h \in \mathcal{B}_m$ and δ is a procedure such that $R(\delta) = O(h(a))$, then*

$$\frac{a - S_k}{(1/\mu)^{1-(1/2)^k} F_{h(a)}^{(k)}(a)} \ge 1 \quad \textit{in probability as } a \to \infty$$

for $0 \le k < m$.

Proof. Let F^k denote $F_{h(a)}^{(k)}(a)$ and $G_k = (1/\mu)^{1-(1/2)^k} F^k$. Choose $0 < \varepsilon < 1$ and let $V_k = \{a - S_k \ge (1-\varepsilon)G_k\}$; we will show

$$P(V_k) \to 1 \quad \text{as } a \to \infty, \text{ for } 0 \le k < m. \tag{2.117}$$

The $k = 0$ case is trivial since $V_0 = \{a \ge (1-\varepsilon)a\}$. Assume that $1 \le k < m$ and $P(V_{k-1}) \to 1$. Let

$$\zeta_k = \frac{a - S_{k-1} - \mu N_k}{\sqrt{N_k}}.$$

We claim

$$P(\zeta_k \geq \sqrt{\log(G_{k-1}/h(a)^2)} - 1|V_{k-1}) \to 1. \tag{2.118}$$

Let $U = \{\zeta_k < \sqrt{\log(G_{k-1}/h(a)^2)} - 1\}$. If (2.118) were to fail there would be an $\eta > 0$ and a sequence of a's approaching ∞ on which $P(U|V_{k-1}) \geq \eta$. Then

$$\begin{aligned} ES_M - a &\gtrsim E[(S_k - a)1\{M = k\}|U \cap V_{k-1}] \cdot \eta \\ &= E[(S_k - S_{k-1} - (a - S_{k-1}))1\{M = k\}|U \cap V_{k-1}] \cdot \eta \end{aligned}$$

and

$$\frac{a - S_{k-1} - \mu N_k}{\sqrt{N_k}} = \zeta_k = O(\sqrt{\log a}) = o(a^{1/6})$$

on U, so Lemma 2.10 applies and we have

$$\begin{aligned} ES_m - a &\gtrsim E[\sqrt{N_k}\Delta(\zeta_k)|U \cap V_{k-1}] \cdot \eta \\ &= E[(2\mu)^{-1}(\sqrt{4\mu(a - S_{k-1}) + \zeta_k^2} - \zeta_k)\Delta(\zeta_k)|U \cap V_{k-1}] \cdot \eta. \end{aligned}$$

The expression inside the expectation is a decreasing function of ζ_k, hence

$$ES_m - a \gtrsim E[(2\mu)^{-1}(\sqrt{4\mu(a - S_{k-1}) + \zeta^2} - \zeta)\Delta(\zeta)|U \cap V_{k-1}] \cdot \eta\Big|_{\zeta = \sqrt{\log(G_{k-1}/h(a)^2)} - 1}. \tag{2.119}$$

Now

$$\begin{aligned} G_{k-1} \propto F_{h(a)}^{(k-1)}(a) &\gtrsim [(C_k^m)^2 \wedge (C_k^{m-1})^2]h_k(a)^2 \quad \text{(by Lemma 2.6)} \\ &\propto h_k(a)^2 \\ &\geq h_{m-1}(a)^2 \quad (\text{since } m - 1 \geq k) \\ &\gg h(a)^2, \end{aligned}$$

since $h \in \mathcal{B}_m$, so $G_{k-1}/h(a)^2 \to \infty$ and hence so does $\sqrt{\log(G_{k-1}/h(a)^2)} - 1$. Then, using $\Delta(\zeta) \sim \phi(\zeta)/\zeta^2$ as $\zeta \to \infty$, (2.119) becomes

$$\begin{aligned} ES_m - a &\gtrsim \sqrt{\frac{(1-\varepsilon)G_{k-1}}{\mu}} \cdot \frac{\phi(\sqrt{\log(G_{k-1}/h(a)^2)} - 1)}{(\sqrt{\log(G_{k-1}/h(a)^2)} - 1)^2} \cdot \eta \\ &\geq \eta' \sqrt{G_{k-1}} \cdot \frac{(h(a)/\sqrt{G_{k-1}}) \exp[\sqrt{\log(G_{k-1}/h(a)^2)}]}{(\sqrt{\log(G_{k-1}/h(a)^2)} - 1)^2} \quad \text{(some } \eta' > 0) \\ &= h(a)/o(1), \end{aligned}$$

which would imply $h(a) = o(R(\delta))$ on this sequence, contradicting our assumption $R(\delta) = O(h(a))$. Hence, (2.118) must hold.

Now $P(V_k) \geq P(V_k|U' \cap V_{k-1})P(U' \cap V_{k-1})$ and $P(U' \cap V_{k-1}) \to 1$ by the induction hypothesis and (2.118), so to show $P(V_k) \to 1$ it suffices to show

$$P(V_k|U' \cap V_{k-1}) \to 1, \tag{2.120}$$

which we do now. We have

$$\begin{aligned} &P(V_k|U' \cap V_{k-1}) = P(a - S_k \geq (1-\varepsilon)G_k|U' \cap V_{k-1}) \\ &\quad = P\left(\frac{S_k - S_{k-1} - \mu N_k}{\sqrt{N_k}} \leq \frac{a - S_{k-1} - \mu N_k - (1-\varepsilon)G_k}{\sqrt{N_k}} \,\middle|\, U' \cap V_{k-1}\right) \end{aligned} \tag{2.121}$$

and

$$\begin{aligned} \frac{a - S_{k-1} - \mu N_k - (1-\varepsilon)G_k}{\sqrt{N_k}} &= \zeta_k - \frac{(1-\varepsilon)G_k}{\sqrt{N_k}} \\ &= \zeta_k - \frac{(1-\varepsilon)G_k}{(2\mu)^{-1}(\sqrt{4\mu(a - S_{k-1}) + \zeta_k^2} - \zeta_k)} \\ &\geq \zeta_k - \frac{(1-\varepsilon)G_k}{(2\mu)^{-1}(\sqrt{4\mu(1-\varepsilon)G_{k-1} + \zeta_k^2} - \zeta_k)} \end{aligned}$$

on V_{k-1}. This last in an increasing function of ζ_k for large enough a, so on U',

$$\frac{a - S_{k-1} - \mu N_k - (1-\varepsilon)G_k}{\sqrt{N_k}} \geq \sqrt{\log(G_{k-1}/h(a)^2)} - 1 - \frac{(1-\varepsilon)G_k}{(1-\varepsilon)^{1/4}\sqrt{(1-\varepsilon)G_{k-1}/\mu}}. \tag{2.122}$$

Now

$$\begin{aligned}\frac{G_k}{\sqrt{G_{k-1}/\mu}} &= \frac{(1/\mu)^{1-2^{-k}} F_{h(a)}(F^{k-1})}{\sqrt{(1/\mu)^{1-2^{-(k-1)}} F^{k-1}/\mu}} \quad \text{(by definition of } G_k\text{)} \\ &= \frac{\sqrt{F^{k-1}\log(F^{k-1}/h(a)^2)}}{\sqrt{F^{k-1}}} \\ &= \sqrt{\log(F^{k-1}/h(a)^2)} = \sqrt{\log(G_{k-1}/h(a)^2)} + o(1).\end{aligned}$$

Plugging this back into (2.122) gives

$$\begin{aligned}\frac{a - S_{k-1} - \mu N_k - (1-\varepsilon)G_k}{\sqrt{N_k}} &\geq [1-(1-\varepsilon)^{1/4}]\sqrt{\log(G_{k-1}/h(a)^2)} - 3/2 \\ &\equiv \gamma(a) \to \infty\end{aligned}$$

as $a \to \infty$, so (2.121) becomes

$$\begin{aligned}P(V_k|U' \cap V_{k-1}) &\geq P\left(\frac{S_k - S_{k-1} - \mu N_k}{\sqrt{N_k}} \leq \gamma(a) \middle| U' \cap V_{k-1}\right) \\ &\geq 1 - O(\gamma(a)^{-2}) \to 1,\end{aligned}$$

by Chebyshev's inequality. This proves (2.120) and finishes the proof of the lemma.

□

Next we prove the optimality of δ_m, $\hat{\delta}_m$.

Theorem 2.15. *If $h \in \mathcal{B}_m^o$, then*

$$R(\delta_m(h)) \sim m \cdot h(a) \sim R^*. \tag{2.123}$$

If $h \in \mathcal{B}_m^+$, then

$$R(\hat{\delta}_m(p^*)) \sim \left[m + 1 - p^* + \frac{\Delta(z_{p^*})\kappa_m}{Q}\right] h(a) \sim R^*, \tag{2.124}$$

where $Q \equiv \lim_{a\to\infty} h(a)/h_m(a) \in (0,\infty)$ *and* p^* *is the unique solution of the equation*

$$\frac{p^*}{\phi(z_{p^*})} = \frac{Q}{\kappa_m}.$$

Proof. First assume that $h \in \mathcal{B}_m^o$. Proposition 2.12 implies $R(\delta_m(h)) \sim mh(a) = O(h(a))$ and so, by the Bayes property, $R^* \le R(\delta_m(h)) = O(h(a))$ as well. Hence Lemma 2.14 applies to $\delta^* = (N^*, M^*)$ and, letting S_k^* denote the δ^*-sampled process,

$$\begin{aligned} R^* &\ge h(a)EM^* \\ &\ge h(a)mP(M^* \ge m) \\ &\ge h(a)mP\left(a - S_{m-1}^* \ge (1/2)\cdot(1/\mu)^{1-(1/2)^{m-1}} F_{h(a)}^{(m-1)}(a)\right) \\ &= h(a)m(1+o(1)) \quad \text{(by Lemma 2.14)} \\ &= R(\delta_m(h)) \ge R^*, \end{aligned}$$

which gives (2.123).

To handle the boundary case we must work a bit harder. Assume that $h/h_m \to Q \in (0,\infty)$. Let $0 < \varepsilon < 1$ and

$$V = \left\{a - S_{m-1}^* \ge (1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}} F_{h(a)}^{(m-1)}(a)\right\}.$$

By Proposition 2.13,

$$\begin{aligned} R(\hat{\delta}_m(p^*)) &\lesssim \Delta(z_{p^*})\kappa_m h_m(a) + (m+1-p^*)h(a) \\ &\sim \left[\frac{\Delta(z_{p^*})\kappa_m}{Q} + m + 1 - p^*\right] h(a). \end{aligned} \tag{2.125}$$

In particular, $R(\hat{\delta}_m(p^*)) = O(h(a))$, so $R^* \le R(\hat{\delta}_m(p^*)) = O(h(a))$ and hence $P(V) \to 1$ by Lemma 2.14. Let $g(a)$ be an arbitrary nonnegative function of a and define

$$\begin{aligned} U(g(a)) &= \left\{a - S_{m-1}^* = (1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}} F_{h(a)}^{(m-1)}(a) + g(a)\right\} \\ \rho(g(a)) &= E(N^* - a/\mu | U(g(a))) + h(a)E(M^* | U(g(a))), \end{aligned}$$

where it is understood that by conditioning on $U(g(a))$ we mean the optimal continuation from $(1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}}F_{h(a)}^{(m-1)}(a)+g(a)$ with the appropriately adjusted parameters. Then

$$R^* \geq [E(N^*-a/\mu|V)+h(a)E(M^*|V)]P(V) \gtrsim \inf_g \rho(g(a)), \tag{2.126}$$

where the infimum is taken over all nonnegative functions g. Let $n(g(a))$ denote the value of N_m^* on $U(g(a))$; note that we may assume that this is not randomized by virtue of the stationarity property of the Bayes procedure. Let

$$z(g(a)) = \frac{(1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}}F_{h(a)}^{(m-1)}(a)+g(a)-\mu n(g(a))}{\sqrt{n(g(a))}}. \tag{2.127}$$

We now show that we only need to consider $g(a)$ for which $z(g(a))$ is bounded in the infimum in (2.126). That is,

$$\inf_g \rho(g(a)) = \inf_{g\in\mathcal{C}} \rho(g(a)), \tag{2.128}$$

where $\mathcal{C} \equiv \{g : z(g(a)) = O(1)\}$. If $g \notin \mathcal{C}$, then $\limsup_{a\to\infty} z(g(a)) = \infty$ so there is a sequence of a's approaching ∞ on which

$$\begin{aligned} P(M^*=m|U(g(a))) &= P(S_m^* \geq a|U(g(a))) \\ &= P\left(\frac{S_m^*-S_{m-1}^*-\mu N_m^*}{\sqrt{N_m^*}} \geq \frac{a-S_{m-1}^*-\mu N_m^*}{\sqrt{N_m^*}}\middle| U(g(a))\right) \\ &= P\left(\frac{S_m^*-S_{m-1}^*-\mu N_m^*}{\sqrt{N_m^*}} \geq z(g(a))\middle| U(g(a))\right) \\ &\leq z(g(a))^{-2} \to 0, \end{aligned} \tag{2.129}$$

using Chebyshev's inequality. Thus

$$\rho(g(a)) \geq h(a)E(M^*|U(g(a))) \gtrsim h(a)(m+1). \tag{2.130}$$

Some calculus shows that the function

$$p \mapsto \frac{\Delta(z_p)\kappa_m}{Q} + m + 1 - p$$

achieves its unique minimum at $p = p^*$, the unique solution of

$$\frac{p^*}{\phi(z_{p^*})} = \frac{Q}{\kappa_m}.$$

Hence

$$m + 1 = \lim_{p \to 0} \left[\frac{\Delta(z_p)\kappa_m}{Q} + m + 1 - p \right] \geq \eta + \left[\frac{\Delta(z_{p^*})\kappa_m}{Q} + m + 1 - p^* \right], \qquad (2.131)$$

some $\eta > 0$, giving

$$\begin{aligned} \rho(g(a)) - R(\hat{\delta}_m(p^*)) &\gtrsim h(a)(m+1) - h(a) \left[\frac{\Delta(z_{p^*})\kappa_m}{Q} + m + 1 - p^* \right] \\ &\qquad \text{(by (2.130) and (2.125))} \\ &\geq \eta h(a) \to \infty \end{aligned}$$

and hence $\rho(g(a)) > R(\hat{\delta}_m(p^*)) \geq R^*$. Thus, by replacing g on any such subsequence by a function for which $z(g(a))$ is bounded, we construct a function in $\mathcal{C}$ that dominates g, and whence (2.128) holds.

Now let $g \in \mathcal{C}$. Since $z(g(a))$ is bounded, by Lemma 2.10 and Wald's equation we have

$$
\begin{aligned}
E(N^* - a/\mu | U(g(a))) &= \mu^{-1} E(S_{M^*} - a | U(g(a))) \\
&\geq \mu^{-1} E[(S_{M^*} - a) 1\{M^* = m\} | U(g(a))] \\
&\sim \mu^{-1} \sqrt{n(g(a))} \cdot \Delta(z(g(a))) \\
&\sim \mu^{-1} \sqrt{\frac{(1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}} F_{h(a)}^{(m-1)}(a) + g(a)}{\mu}} \cdot \Delta(z(g(a))) \quad \text{(by (2.127))} \\
&\geq \mu^{-1} \sqrt{\frac{(1-\varepsilon)(1/\mu)^{1-(1/2)^{m-1}} F_{h(a)}^{(m-1)}(a)}{\mu}} \cdot \Delta(z(g(a))) \\
&\sim \sqrt{1-\varepsilon} \cdot (1/\mu)^{2-(1/2)^m} C_m^m h_m(a) \cdot \Delta(z(g(a))) \quad \text{(by Lemma 2.6)} \\
&\sim \sqrt{1-\varepsilon} \cdot \kappa_m h(a) Q^{-1} \cdot \Delta(z(g(a))), \qquad (2.132)
\end{aligned}
$$

this last using $\kappa_m = (1/\mu)^{2-(1/2)^m} C_m^m$. Let $p(g(a)) = \Phi(-z(g(a)))$. By the relation (2.129),

$$P(M^* = m | U(g(a))) \sim \Phi(-z(g(a))) = p(g(a))$$

follows from the Central Limit Theorem, since we know $n(g(a)) \to \infty$ by the relation (2.127). This implies

$$E[M^* | U(g(a))] \gtrsim m + 1 - p(g(a))$$

and combining this with (2.132) gives

$$
\begin{aligned}
R^* &\gtrsim \inf_{g \in \mathcal{C}} \rho(g(a)) \\
&\gtrsim \inf_{g \in \mathcal{C}} \left[\frac{\kappa_m \Delta(z(g(a))) \sqrt{1-\varepsilon}}{Q} + m + 1 - p(g(a)) \right] h(a) \\
&\geq \inf_{g \in \mathcal{C}} \left[\frac{\kappa_m \Delta(z(g(a)))}{Q} + m + 1 - p(g(a)) \right] h(a) \sqrt{1-\varepsilon} \\
&= \inf_{p \in (0,1)} \left[\frac{\kappa_m \Delta(z_p)}{Q} + m + 1 - p \right] h(a) \sqrt{1-\varepsilon} \\
&= \left[\frac{\kappa_m \Delta(z_{p^*})}{Q} + m + 1 - p^* \right] h(a) \sqrt{1-\varepsilon}.
\end{aligned}
$$

This argument holds for all $\varepsilon > 0$, so by a now routine asymptotic argument, there is a sequence $\varepsilon_a \to 0$ for which it holds. Then

$$\begin{aligned} R^* &\gtrsim \left[\frac{\kappa_m \Delta(z_{p^*})}{Q} + m + 1 - p^*\right] h(a)\sqrt{1-\varepsilon_a} \\ &\sim \left[\frac{\kappa_m \Delta(z_{p^*})}{Q} + m + 1 - p^*\right] h(a) \\ &\gtrsim R(\hat{\hat{\delta}}_m(p^*)) \quad \text{(by (2.125))} \\ &\geq R^*, \end{aligned}$$

which proves (2.124) and completes the proof. □

Our final theorem of this chapter is a type of converse to Theorem 2.15, showing that the properties of $\delta_m, \hat{\hat{\delta}}_m$ established in Propositions 2.12, 2.13, and Lemma 2.14 are not only sufficient but necessary. Moreover, Theorem 2.16 gives a precise lower bound on the risk inefficiency of any procedure that deviates from the "schedule" of Lemma 2.14. This is the discrete analog of Theorem 2.9.

Theorem 2.16. *Assume that $h \in \mathcal{B}_m$ and let*

$$\delta_m = \begin{cases} \delta_m(h), & \text{if } h \in \mathcal{B}_m^o \\ \hat{\hat{\delta}}_m(p^*), & \text{if } h \in \mathcal{B}_m^+. \end{cases}$$

If $\delta = (N, M)$ is a procedure such that there is a sequence $a_i \to \infty$ with

$$P(a_i - S_k \leq (1-\varepsilon)(1/\mu)^{1-2^{-k}} F_{h(a_i)}^{(k)}(a_i)) \quad \text{bounded below 1} \tag{2.133}$$

for some $1 \leq k < m$ and $\varepsilon > 0$, then there is $C > 0$ such that

$$R(\delta) - R(\delta_m) \geq C \cdot h_{k^*}(a_i) \to \infty, \tag{2.134}$$

where k^* is the smallest k for which (2.133) holds. In particular, (2.134) holds if $P(M \geq m) \not\to 1$.

Proof. Let $V_k = \{a - S_k \leq (1-\varepsilon)(1/\mu)^{1-2^{-k}} F_{h(a)}^{(k)}(a)\}$. By repeating the argument leading to (2.119) there is an $\eta > 0$ such that

$$\begin{aligned} R(\delta) &\geq \mu^{-1} E(S_M - a_i) \\ &\geq \mu^{-1} E(S_M - a_i; \{M = k^*\} \cap V_{k^*-1}) \\ &\geq \eta \cdot \sqrt{F_{h(a_i)}^{(k^*-1)}(a_i)} \cdot \Delta(z_\eta) \\ &\geq C h_{k^*}(a_i), \end{aligned}$$

some $C > 0$, since $\eta\Delta(z_\eta) > 0$ and $h_{k^*} = O(\sqrt{F_{h(a_i)}^{(k^*-1)}})$ by Lemma 2.6. By Theorem 2.15,

$$R(\delta_m) = O(h(a)) = o(h_{k^*}(a))$$

since $k^* < m$, proving (2.134). Since

$$P(V_k) \to 1 \quad \text{for all } 1 \leq k < m \Rightarrow P(M \geq m) \to 1,$$

if $P(M \geq m) \not\to 1$ then there is some $k^* < m$ for which $P(V_{k^*}) \not\to 1$ and hence (2.134) holds, proving the second assertion. □

Chapter 3

Multistage Tests of Simple Hypotheses

In this chapter we use the multistage sampling procedures of Chapter 2 to design efficient multistage tests of simple hypotheses in two different settings. In Section 3.1 we consider tests that have just one terminal decision and are designed to have a large sample size under the alternative hypothesis. In Section 3.2 we use these so-called one decision tests to design efficient two decision tests concerning members of a one-dimensional exponential family. In both settings the resulting procedures share the global properties of the multistage sampling procedures discussed in Chapter 2. The stage sizes decrease roughly as a sequence of successive square roots, while the average number of stages required is determined by the asymptotics of the ratio of the cost per stage to cost per observation, involving the critical functions h_m.

Let $X_1, X_2, \ldots$ be i.i.d. with a density belonging to an exponential family

$$f(x|\theta) = \exp(\theta x - \psi(\theta)) \tag{3.1}$$

with respect to some non-degenerate σ-finite measure. Let f_0 and f_1 be two distinct members of this family whose corresponding parameter values, θ_0 and θ_1, lie in the interior of the natural parameter space. Then ψ is infinitely differentiable at θ_0, θ_1, $\psi'(\theta_i) = E_i X_1$, and $\psi''(\theta_i) = \text{Var}_i X_1$ for $i = 0, 1$, where E_i, Var_i denote expectation, variance under f_i. Let

$$l_n = \prod_{i=1}^{n} \frac{f_0(X_i)}{f_1(X_i)},$$

the likelihood ratio, and let

$$I_i = E_i \log\left(\frac{f_i(X_1)}{f_{1-i}(X_1)}\right), \quad i = 0, 1,$$

the Kullback-Leibler information numbers.

3.1 One Decision Tests

Consider the problem of deciding between f_0 and f_1 by sampling the X_i in stages. Suppose also that if f_0 is the true density, sampling costs are being incurred and so we want to stop sampling as soon as possible and reject the hypothesis $f = f_1$. On the other hand, if f_1 is the true density sampling costs nothing and our preferred action is to observe $X_1, X_2, \ldots$ ad infinitum. As an example, suppose a new drug is being marketed under the hypothesis that its side effects are insignificant. Physicians prescribing the drug record and report on the side effects and if they appear unacceptably high ($f = f_0$), this must be announced and the drug withdrawn from use. But as long as the hypothesis of insignificant side effects ($f = f_1$) remains tenable, no action is required.

Specifically, define a *one decision test of* f_0 *vs.* f_1 to be a pair (N, M) such that $N = (N_1, N_2, \ldots)$ is a sequence of nonnegative integer-valued random variables satisfying the measurability requirement (2.63), which essentially requires that the size of the $(k+1)$st stage, N_{k+1}, is determined by the data obtained in the first k stages. $N^k \equiv N_1 + \cdots + N_k$ should be interpreted as the sample size through the kth stage and $M \equiv \inf\{m \geq 1 : N_m = 0\}$, the number of stages. By a convenient abuse of notation, we also let N denote N^M, the total sample size. If one pays costs per observation and per stage under f_0, plus a cost for terminating sampling under f_1, then a natural measure of the performance of a one decision test of f_0 vs. f_1 is the expected sum of these costs. Hence we define the *risk* of a one decision test of f_0 vs. f_1 to be

$$R(N, M) = cE_0N + dE_0M + P_1(N < \infty), \tag{3.2}$$

where $c, d > 0$ and P_i is probability under f_i. Let (N^*, M^*) denote the Bayes test, that which achieves risk $R^* \equiv \inf_{(N,M)} R(N, M)$. Note that a "one decision test of f_0 vs. f_1" may only reject $f = f_1$.

In this section we derive a family of one decision tests and show they minimize the risk to second-order as $c, d \to 0$. As one may expect from (3.2), the notion of "efficiency" depends heavily on the rates at which c and d approach 0. To simplify our bookkeeping, we assume that d is the independent variable and that $c = c(d)$, though this choice is arbitrary. Recall that the critical functions were defined as

$$h_m(a) = a^{(1/2)^m}(\log a)^{1/2-(1/2)^m} \text{ for } m \geq 1, \quad h_0(a) = a,$$

and we say the sequence $\{(a, h)\}$ is

$$\begin{array}{ll} \text{in the } m\text{th critical band} & \text{if } h_m(a) \ll h \ll h_{m-1}(a) \\ \text{on the boundary between critical bands } m,\, m+1 & \text{if } \lim h/h_m(a) \in (0, \infty). \end{array}$$

It will turn out that efficient tests will use m stages (almost always) if

$$h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1})$$

as $d \to 0$. Proceeding by analogy with Chapter 2, we thus give an essentially complete description of the problem while assuming $\{(\log d^{-1}, d/c)\}$ is either in the mth critical band or on the boundary between critical bands m and $m+1$ (for every sequence of d's approaching zero), for some $m \geq 1$. Thus we define

$$\begin{aligned} \mathcal{B}^o_m(d) &= \{c : (0,1) \to (0,1) | \quad h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1})\}, \\ \mathcal{B}^+_m(d) &= \left\{ c : (0,1) \to (0,1) \middle| \quad \frac{d/c}{h_m(\log d^{-1})} \to Q, \text{ some } Q \in (0, \infty) \right\}, \end{aligned}$$

and we assume in our main results that

$$c \in \mathcal{B}_m(d) \equiv \mathcal{B}^o_m(d) \cup \mathcal{B}^+_m(d) \tag{3.3}$$

for some $m \geq 1$. Note that $c \in \mathcal{B}_m(d)$ implies $h_m(\log d^{-1}) = O(d/c)$, hence a consequence of this assumption is that $d/c \to \infty$ as $d \to 0$, which we shall assume throughout this chapter. Indeed, if $d/c \leq B < \infty$, then it is not hard to show that a fully-sequential test minimizes the risk (3.2) to second-order. Since our main interest here is variable stage size multistage procedures, we can be sure the assumption (3.3) does not exclude any interesting cases.

Since the "decision" aspect of a one decision test is trivial, any multistage sampling procedure can be used as a one decision test. In particular, we will be interested in using multistage sampling procedures to sample the log-likelihood process

$$\log(f_0(X_1)/f_1(X_1)),\ \log(f_0(X_2)/f_1(X_2)),\ \ldots$$

until $\sum \log(f_0(X_i)/f_1(X_i))$ exceeds a predetermined boundary. The only slight technicality to overcome is that multistage sampling procedures were defined in Chapter 2 with respect to random processes with unit variance. To remedy this, we simply transform the log-likelihood process to have variance one under E_0: let

$$C = (|\theta_0 - \theta_1|\sqrt{\psi''(\theta_0)})^{-1} > 0$$

and

$$Y_i = C \log(f_0(X_i)/f_1(X_i)) = \frac{(\theta_0 - \theta_1)X_i - \psi(\theta_0) + \psi(\theta_1)}{|\theta_0 - \theta_1|\sqrt{\psi''(\theta_0)}}, \tag{3.4}$$

so that

$$E_0 Y_i = C I_0 \quad \text{and} \quad \mathrm{Var}_0 Y_i = \frac{(\theta_0 - \theta_1)^2 \mathrm{Var}_0 X_i}{|\theta_0 - \theta_1|^2 \psi''(\theta_0)} = 1.$$

Whenever we use a multistage sampling procedure as a one decision test below, we will always mean with respect to $Y_1, Y_2, \ldots$.

The following lemma shows that the Bayes one decision test is essentially a one-sided likelihood ratio test, stopping only if the likelihood ratio exceeds a boundary determined by the parameter values.

Lemma 3.1. *There exists* $a^* = \log d^{-1} + o(1)$ *such that*

$$\log l_{N^*} \geq a^*. \tag{3.5}$$

Proof. By Wald's likelihood ratio identity we can write

$$\begin{aligned} R^* &= \inf_{(N,M)} \{cE_0 N + dE_0 M + P_1(N < \infty)\} \\ &= \inf_{(N,M)} E_0[cN + dM + l_N^{-1} 1\{N < \infty\}]. \end{aligned}$$

Suppose that the Bayes procedure has observed $X_1, \ldots, X_n$ in m stages. By the Bayes property we know that (N^*, M^*) will stop at this point only if the stopping risk is no greater than the continuation risk, i.e., only if

$$\begin{aligned} cn + dm + l_n^{-1} &\leq cn + dm + \inf_{(N,M):N\geq 1} E_0[cN + dM + l_n^{-1} l_N^{-1} 1\{N < \infty\}] \\ \Leftrightarrow \quad 1 &\leq \inf_{(N,M):N\geq 1} E_0[l_n(cN + dM) + l_N^{-1} 1\{N < \infty\}], \end{aligned} \tag{3.6}$$

where it is understood that such infimums are taken over all continuations and the expectation is conditional on $X_1, \ldots, X_n$. For $t > 0$ define

$$\rho(t) = \inf_{(N,M):N\geq 1} E_0[t(cN + dM) + l_N^{-1} 1\{N < \infty\}],$$

so that (3.6) implies

$$\rho(l_{N^*}) \geq 1. \tag{3.7}$$

Note that $\rho(t)$ (as a function of t) is the infimum of a set of lines, each of slope at least $c + d$, by virtue of the restriction of the infimum to the class of all (N, M) such that N (and hence M) are at least one. Thus $\rho(t)$ is continuous, strictly increasing, and satisfies $\rho(t) \geq t(c+d)$, so that

$$\rho(t) \geq 1 \quad \text{when} \quad t \geq (c+d)^{-1}. \tag{3.8}$$

If (N', M') is the procedure that samples with constant stage size one (i.e., fully-sequential sampling) and an appropriately chosen boundary, then it is well-known (see, e.g., [20]) that

$$1 > P_1(N' < \infty) = E_0[l_{N'}^{-1} 1\{N' < \infty\}] \quad \text{and} \quad E_0 N' = E_0 M' < \infty$$

and hence

$$\rho(t) \leq t(c+d)E_0 N' + P_1(N' < \infty) < 1$$

for sufficiently small t. Since $\rho(\cdot)$ is continuous and increasing, this last and (3.8) imply that there is a unique number, call it e^{a^*}, such that $\rho(e^{a^*}) = 1$. Then

$$\begin{aligned} \log l_{N^*} &= \log \rho^{-1}(\rho(l_{N^*})) \\ &\geq \log \rho^{-1}(1) \quad \text{(by (3.7))} \\ &= \log e^{a^*} \quad (\text{since } \rho(e^{a^*}) = 1) \\ &= a^*, \end{aligned}$$

establishing (3.5).

To show that $a^* = \log d^{-1} + o(1)$, let Y_i be as in (3.4) and $(N, M) = \delta_1(h)$, the multistage sampling procedure described in Section 2.2 with $h(a) \equiv a^{3/2}$ and boundary $a \equiv C \log(d/c)$. Since $\sqrt{a} \ll h(a) \ll a$, by Proposition 2.12

$$E_0 N - a(CI_0)^{-1} = o(h(a)) \quad \text{and} \quad E_0 M = 1 + o(1).$$

Observe that

$$l_N^{-1} = \exp[-C^{-1}(Y_1 + \cdots + Y_N)] \leq \exp[-C^{-1}a] = c/d$$

so that

$$\begin{aligned}
\rho(t) &\leq E_0[t(cN + dM) + l_N^{-1} 1\{N < \infty\}] \\
&\leq tc[E_0(N - a(CI_0)^{-1}) + a(CI_0)^{-1} + (d/c)E_0 M] + E_0 l_N^{-1} \\
&\leq tc[o(h(a)) + a(CI_0)^{-1} + (d/c)(1 + o(1))] + c/d \\
&= tc[o(d/c) + d/c(1 + o(1))] + c/d \\
&= td(1 + o(1)) + c/d.
\end{aligned}$$

This implies

$$\rho(t) \leq 1 \text{ when } t \leq d^{-1}(1 + o(1)), \tag{3.9}$$

and so

$$\begin{aligned}
a^* &= \log e^{a^*} = \log \rho^{-1}(1) \quad (\text{since } \rho(e^{a^*}) = 1) \\
&\geq \log \rho^{-1}(\rho(d^{-1}(1 + o(1)))) \quad (\text{by (3.9)}) \\
&= \log(d^{-1}(1 + o(1))) = \log d^{-1} + o(1).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
a^* &= \log \rho^{-1}(1) \leq \log \rho^{-1}(\rho([c + d]^{-1})) \quad (\text{by (3.8)}) \\
&= \log(c + d)^{-1} = \log d^{-1} + o(1)
\end{aligned}$$

since $d/c \to \infty$, establishing $a^* = \log d^{-1} + o(d)$. □

Before proving our main result of this section in Theorem 3.2, we consolidate our notation a bit. The following function provides the coefficient of the second-order term in the Bayes risk for both the $c \in \mathcal{B}_m^o(d)$ and $c \in \mathcal{B}_m^+(d)$ cases. For $m = 1, 2, \ldots$ and $Q, \mu > 0$ define

$$u_m(Q, \mu) = m + 1 - p^* + \frac{\Delta(z_{p^*})\kappa_m(\mu)}{Q},$$

where $p^* = p^*(m, Q, \mu)$ is the unique solution of

$$\frac{p^*}{\phi(z_{p^*})} = \frac{Q}{\kappa_m(\mu)}. \tag{3.10}$$

Now fix $\mu > 0$. Note that $p^* \to 1$ as $Q \to \infty$, so $\Delta(z_{p^*}) \sim |z_{p^*}|$ as $Q \to \infty$. Also,

$$Q = \frac{p^*}{\phi(z_{p^*})}\kappa_m(\mu) \sim \frac{\kappa_m(\mu)}{\phi(z_{p^*})}$$

as $Q \to \infty$, so

$$\frac{\Delta(z_{p^*})\kappa_m(\mu)}{Q} \sim |z_{p^*}|\phi(z_{p^*}) \to 0$$

and hence

$$\lim_{Q\to\infty} u_m(Q, \mu) = m.$$

Thus we can extend our definition of u_m to all $Q \in (0, \infty]$ by setting

$$u_m(\infty, \mu) \equiv \lim_{Q\to\infty} u_m(Q, \mu) = m.$$

Theorem 3.2 shows that the asymptotically optimal multistage sampling procedures derived in Chapter 2 are second-order optimal as one decision tests. Said another way, Lemma 3.1 tells us that efficient one decision tests are essentially likelihood ratio tests and the part of the risk (3.2) due to error is of smaller order than the sampling costs, which we already know our multistage sampling procedures minimize.

Theorem 3.2. *Assume $c \in \mathcal{B}_m(d)$ and let*

$$Q = \lim_{d\to 0} \frac{d/c}{h_m(C \log d^{-1})} \in (0, \infty]$$

and $p^* = p^*(m, Q, CI_0)$ as in (3.10). Let $\delta_m, \hat{\delta}_m$ be the multistage sampling procedures defined in Section 2.2.1 and

$$(N, M) = \begin{cases} \delta_m(d/c), & \text{if } c \in \mathcal{B}_m^o(d) \\ \hat{\delta}_m(p^*), & \text{if } c \in \mathcal{B}_m^+(d) \end{cases}$$

applied to $Y_1, Y_2, \ldots$ *with boundary* $a = C \log d^{-1}$. *Then*

$$R(N,M) = cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, CI_0) + o(d) \tag{3.11}$$

$$R^* = cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, CI_0) + o(d) \tag{3.12}$$

as $d \to 0$.

Proof. Since $R^* \le R(N,M)$, it suffices to prove (3.11) with "$\le$" and (3.12) with "$\ge$." Assume first that $c \in \mathcal{B}_m^o(d)$, i.e.,

$$h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1}). \tag{3.13}$$

Note that in our notation, $Q = \infty$ and hence $u_m(Q, CI_0) = m$. Let $a^* = \log d^{-1} + o(1)$ be that given by Lemma 3.1. Then

$$h_k(Ca^*) \sim C^{(1/2)^k} h_k(a^*) \propto h_k(\log d^{-1} + o(1)) \sim h_k(\log d^{-1})$$

since $(d/dx)h_k(x)$ is bounded for large x, thus

$$h_m(Ca^*) \ll d/c \ll h_{m-1}(Ca^*) \tag{3.14}$$

by (3.13). By Lemma 3.1 we know that (N^*, M^*) stops iff $l_{N^*} \ge e^{a^*}$, so by comparing (N^*, M^*) with the Bayes multistage sampling procedure with boundary Ca^* in the $\mathcal{B}_m^o$ case (because of (3.14)) of Theorem 2.15,

$$\begin{aligned} R^* &= cE_0N^* + dE_0M^* + P_1(N^* < \infty) \\ &\ge c[E_0(N^* - a^*/I_0) + (d/c)E_0M^*] + ca^*/I_0 \\ &\ge c[m(d/c) + o(d/c)] + cI_0^{-1}(\log d^{-1} + o(1)) \quad \text{(by Theorem 2.15)} \\ &= cI_0^{-1} \log d^{-1} + d \cdot m + o(d) \\ &= cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, CI_0) + o(d). \end{aligned} \tag{3.15}$$

By (3.13) we can also apply the $\mathcal{B}_m^o$ case of Theorem 2.15 to (N, M) to get

$$E_0(N - I_0^{-1}\log d^{-1}) + (d/c)E_0 M \le m(d/c) + o(d/c). \tag{3.16}$$

Then

$$\begin{aligned} R(N,M) &= cE_0N + dE_0M + P_1(N<\infty) \\ &= c[E_0(N - I_0^{-1}\log d^{-1}) + (d/c)E_0M] + cI_0^{-1}\log d^{-1} + P_1(N<\infty) \\ &\le c[m(d/c) + o(d/c)] + cI_0^{-1}\log d^{-1} + P_1(N<\infty) \quad \text{(by (3.16))} \\ &= cI_0^{-1}\log d^{-1} + d\cdot m + o(d) + P_1(N<\infty) \\ &= cI_0^{-1}\log d^{-1} + d\cdot u_m(Q, CI_0) + o(d) + P_1(N<\infty), \end{aligned} \tag{3.17}$$

so to show (3.11) holds it suffices to show $P_1(N < \infty) = o(d)$. Now the right hand side of (3.16) is obviously $O(d/c)$, so we can apply Lemma 2.14 to $S_k \equiv Y_1 + \cdots + Y_{N^k}$ (with $C \log d^{-1}$ in place of a and CI_0 in place of μ) to get

$$P_0(C\log d^{-1} - S_{m-1} \ge (1/2)(CI_0)^{-1+(1/2)^{m-1}} F_{d/c}^{(m-1)}(C\log d^{-1})) \to 1$$

as $d \to 0$. Let U be the above event and note that on U,

$$\begin{aligned} C\log d^{-1} - S_{m-1} &\ge (1/2)(CI)^{-1+(1/2)^{m-1}} F_{d/c}^{(m-1)}(C\log d^{-1}) \\ &\ge (1/2)^2(CI)^{-1+(1/2)^{m-1}}[C_m^m h_m(C\log d^{-1})]^2 \quad \text{(by Lemma 2.6)} \\ &\ge \eta h_m(C\log d^{-1})^2, \end{aligned}$$

$\eta > 0$. On U, the mth stage of (N, M) begins geometric sampling with probability of crossing the boundary approaching one (under P_0). Then, letting

$$\rho_m = \frac{S_m - S_{m-1} - CI_0N_m}{\sqrt{N_m}},$$

$$P_0(S_m \geq C \log d^{-1} + \sqrt{h_m(C \log d^{-1})}|U)$$
$$= P_0\left(\rho_m \geq \frac{C \log d^{-1} - S_{m-1} - CI_0N_m}{\sqrt{N_m}} + \sqrt{\frac{h_m(C\log d^{-1})}{N_m}}\middle| U\right) \to 1$$

if

$$h_m(C \log d^{-1}) \ll N_m \tag{3.18}$$

on U, since

$$\frac{C \log d^{-1} - S_{m-1} - CI_0N_m}{\sqrt{N_m}} \to -\infty$$

by definition of (N, M). But (3.18) holds since

$$N_m \geq \frac{C \log d^{-1} - S_{m-1}}{CI_0} \geq \frac{\eta h_m(C \log d^{-1})^2}{CI_0} \gg h_m(C \log d^{-1}) \tag{3.19}$$

on U. Thus let

$$V = U \cap \left\{S_m \geq C \log d^{-1} + \sqrt{h_m(C \log d^{-1})}\right\}$$

so that

$$P_0(V) = P_0(S_m \geq C \log d^{-1} + \sqrt{h_m(C \log d^{-1})}|U) \cdot P_0(U) \to 1 \cdot 1.$$

Note that $l_n = \exp(C^{-1} \sum_1^n Y_i)$, so that by Wald's likelihood identity and letting V' denote the compliment of V,

$$\begin{aligned}
P_1(N < \infty) &= E_0(l_N^{-1}; N < \infty) \leq E_0 l_N^{-1} \\
&= E_0[\exp(-C^{-1}S_m); V] + E_0[\exp(-C^{-1}S_M); V'] \\
&\leq \exp(-\log d^{-1} - C^{-1}\sqrt{h_m(C \log d^{-1})}) + E_0[\exp(-\log d^{-1}); V'] \\
&\qquad \text{(by definition of } V \text{ and since } S_M \geq C \log d^{-1}) \\
&= d \cdot \exp(-C^{-1}\sqrt{h_m(C \log d^{-1})}) + d \cdot P_0(V') \\
&= d \cdot o(1) + d \cdot o(1) = o(d),
\end{aligned}$$

proving (3.11) in the $c \in \mathcal{B}_m^o(d)$ case.

Now assume $c \in \mathcal{B}_m^+(d)$. By the same arguments leading to (3.15) and (3.17) but using the boundary cases of the appropriate results,

$$\begin{aligned} R^* &\geq cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, CI_0) + o(d) \\ &\geq R(N, M) - P_1(N < \infty), \end{aligned}$$

so it again suffices to show $P_1(N < \infty) = o(d)$. Let U be as above and

$$\begin{aligned} W_1 &= \left\{ S_m \geq C \log d^{-1} + \sqrt{h_m(C \log d^{-1})} \right\}, \\ W_2 &= \left\{ S_m \leq C \log d^{-1} - \sqrt{h_m(C \log d^{-1})} \right\}, \\ W_3 &= \left\{ S_{m+1} \geq C \log d^{-1} + (h_m(C \log d^{-1}))^{1/5} \right\}, \text{ and} \\ W &= (U \cap W_1) \sqcup (U \cap W_2 \cap W_3). \end{aligned}$$

We will show $P_0(W) \to 1$ as $d \to 0$, which will allow us to say that the log-likelihood ratio is far enough beyond the boundary at the end of the mth stage (on W_1) or at the end of the $(m+1)$st stage (on W_3) that $P_1(N < \infty) = o(d)$.

$$\begin{aligned} P_0(U \cap W_1) &= P_0(W_1 | U) P_0(U) \sim P_0(W_1 | U) \\ &= P_0 \left(\rho_m \geq \frac{C \log d^{-1} - S_{m-1} - CI_0 N_m}{\sqrt{N_m}} + \sqrt{\frac{h_m(C \log d^{-1})}{N_m}} \middle| U \right) \end{aligned}$$

and

$$\frac{C \log d^{-1} - S_{m-1} - CI_0 N_m}{\sqrt{N_m}} \to z_{p^*}$$

by definition of (N, M). Then

$$P_0(U \cap W_1) \to p^* \qquad (3.20)$$

by the Central Limit Theorem if $\sqrt{h_m(C \log d^{-1})} \ll \sqrt{N_m}$ on U, which holds by

(3.19). To handle the other piece, first write

$$P_0(U \cap W_2 \cap W_3) = P_0(U)P_0(W_2|U)P_0(W_3|U \cap W_2) \sim P_0(W_2|U)P_0(W_3|U \cap W_2).$$

We have $P(W_2|U) \to 1 - p^*$ by an argument similar to the one showing (3.20). Also

$$\begin{aligned} P_0(W_3|U \cap W_2) &= P_0\left(\rho_{m+1} \geq \frac{C \log d^{-1} - S_m - CI_0N_{m+1}}{\sqrt{N_{m+1}}} + \frac{(h_m(C\log d^{-1}))^{1/5}}{\sqrt{N_{m+1}}} \middle| U \cap W_2\right) \\ &\to 1 \end{aligned}$$

since

$$\sqrt{N_{m+1}} \geq \sqrt{\frac{C \log d^{-1} - S_m}{CI_0}} \geq \frac{(h_m(C\log d^{-1}))^{1/4}}{\sqrt{CI_0}} \gg (h_m(C\log d^{-1}))^{1/5}$$

and

$$\frac{C \log d^{-1} - S_m - CI_0N_{m+1}}{\sqrt{N_{m+1}}} \to -\infty$$

on $U \cap W_2$ since the $(m+1)$st stage of (N, M) begins geometric sampling with probability of crossing the boundary approaching one. Combining these estimates we have $P_0(U \cap W_2 \cap W_3) \to 1 - p^*$ and combining this with (3.20) shows

$$P_0(W) = P_0(U \cap W_1) + P_0(U \cap W_2 \cap W_3) \to p^* + 1 - p^* = 1.$$

With this in hand, and noting that, on W,

$$S_M - C\log d^{-1} \geq \sqrt{h_m(C\log d^{-1})} \wedge (h_m(C\log d^{-1}))^{1/5} = (h_m(C\log d^{-1}))^{1/5},$$

$$\begin{aligned}
P_1(N<\infty) &= E_0(l_N^{-1}; N<\infty) \le E_0 l_N^{-1} \\
&= E_0[\exp(-C^{-1}S_M); W] + E_0[\exp(-C^{-1}S_M); W'] \\
&\le \exp(-\log d^{-1} - C^{-1}(h_m(C\log d^{-1}))^{1/5}) + E_0[\exp(-\log d^{-1}); W'] \\
&= d\cdot\exp(-C^{-1}(h_m(C\log d^{-1}))^{1/5}) + d\cdot P_0(W') \\
&= d\cdot o(1) + d\cdot o(1) = o(d),
\end{aligned}$$

finishing the boundary case and the proof. □

3.2 Tests of Two Simple Hypotheses

In this section we use the optimal one decision tests from the previous section to derive optimal multistage tests of two simple hypotheses. Again assume f_0, f_1 are two distinct densities from the exponential family (3.1). Consider the problem of deciding between f_0 and f_1 by sampling $X_1, X_2, \ldots$ in stages while incurring a cost per observation, a cost per stage, and a penalty for making the wrong decision. More specifically, define a *test* of the hypotheses

$$H_0 : f_0 \quad \text{vs.} \quad H_1 : f_1$$

to be a triple (N, M, D), where $N = (N_1, N_2, \ldots)$ is a sequence of nonnegative integer-valued random variables satisfying the measurability requirement (2.63), $M \equiv \inf\{m \ge 1 : N_m = 0\}$, and D takes values in $\{0, 1\}$. N_k should be interpreted as the size of the kth stage, $N^k \equiv N_1 + \cdots + N_k$ the sample size through the kth stage, M the number of stages, and D the "decision," i.e., the choice of i such that $H_i : f_i$ is deemed correct. By a convenient abuse of notation, we let N also denote N^M, the total sample size.

Define the *integrated risk* of a test $\delta = (N, M, D)$ with respect to prior π and loss parameters w_i to be

$$r(\delta) = \sum_{i=0}^{1} \pi_i [cE_i N + dE_i M + w_i P_i(D = 1 - i)],$$

where $c, d > 0$. To avoid trivialities we assume $\pi_i, w_i > 0$. Let $\delta^* = (N^*, M^*, D^*)$ denote the Bayes test, that which achieves integrated risk $r^* \equiv \inf_\delta r(\delta)$.

We describe a family of tests and show they minimize the integrated risk to second-order as $d \to 0$. We continue to assume that $c \in \mathcal{B}_m(d)$, some $m \geq 1$. Extending the notation of the previous section, for $i = 0, 1$ define

$$C_i = (|\theta_0 - \theta_1| \sqrt{\psi''(\theta_i)})^{-1} > 0$$

and

$$Y_j^{(i)} = C_i \log(f_i(X_j)/f_{1-i}(X_j)) \quad \text{for } j = 1, 2, \ldots$$

so that

$$E_i Y_j^{(i)} = C_i I_i \quad \text{and} \quad \mathrm{Var}_i Y_j^{(i)} = 1.$$

Whenever we speak of a one decision test of f_i vs. f_{1-i} (i.e., a test which chooses f_i as the correct density) below, we will always mean the one defined with respect to $Y_1^{(i)}, Y_2^{(i)}, \ldots$.

Our first lemma gives us a lower bound on the integrated risk of δ^* by comparing it to the best one decision tests.

Lemma 3.3. *If $c \in \mathcal{B}_m(d)$, then*

$$cE_0 N^* + dE_0 M^* + P_1(D^* = 0) \geq cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, C_0 I_0) - o(d) \tag{3.21}$$

as $d \to 0$, where

$$Q \equiv \lim_{d \to 0} \frac{d/c}{h_m(C_0 \log d^{-1})} \in (0, \infty].$$

Remark. The lemma actually holds for any test (N, M, D) such that $l_N \leq K_1 d$ on $\{D = 1\}$ for some constant K_1, since this is the only property of the Bayes test used in the proof, though we will not need this full strength in what follows. The lemma also holds of course with the indices $0, 1$ reversed.

Proof. The idea of the proof is to compare the left hand side of (3.21) with the Bayes risk of Theorem 3.2 by extending δ^* to a one decision test of f_0 vs. f_1 on the event $\{D^* = 1\}$. Let

$$N = M = \inf\{n \geq 1 : l_n \geq d^{-2}\},$$

i.e., fully-sequential sampling with boundary d^{-2} for the likelihood ratio. Then define

$$\begin{aligned} N' &= N^* + N \cdot 1\{D^* = 1\} \\ M' &= M^* + M \cdot 1\{D^* = 1\}. \end{aligned}$$

(N', M') coincides with δ^* on $\{D^* = 0\}$, but continues with the one decision procedure (N, M) on $\{D^* = 1\}$, and is hence a one decision procedure itself. Since

$$\{N' < \infty\} = \{D^* = 0\} \sqcup \{D^* = 1, N < \infty\},$$

we have

$$\begin{aligned} &cE_0 N^* + dE_0 M^* + P_1(D^* = 0) \\ &\quad = c[E_0 N' - E_0(N; D^* = 1)] + d[E_0 M' - E_0(M; D^* = 1)] \\ &\qquad + P_1(N' < \infty) - P_1(D^* = 1, N < \infty) \\ &\quad = [cE_0 N' + dE_0 M' + P_1(N' < \infty)] \\ &\qquad - [cE_0(N; D^* = 1) + dE_0(M; D^* = 1) + P_1(D^* = 1, N < \infty)] \\ &\quad \equiv R_1 - R_2. \end{aligned}$$

By Theorem 3.2,

$$\begin{aligned} R_1 &= cE_0 N' + dE_0 M' + P_1(N' < \infty) \\ &\geq cI^{-1} \log d^{-1} + d \cdot u_m(Q, C_0 I_0) - o(d), \end{aligned}$$

so to show that (3.21) holds it suffices to show $R_2 = o(d)$.

We can write

$$R_2 \le [cE_0(N|D^* = 1) + dE_0(M|D^* = 1)]P_0(D^* = 1) + P_1(N < \infty|D^* = 1). \quad (3.22)$$

By Theorem 1 of [18],

$$\begin{aligned} E_0(N|D^* = 1) = E_0(M|D^* = 1) &\le I_0^{-1} \log d^{-2} + I_0^{-2} E_0 \left[\left(\log \frac{f_0(X_1)}{f_1(X_1)} \right)^+ \right]^2 \\ &= O(\log d^{-1}) + O(1) = O(\log d^{-1}). \end{aligned}$$

By Lemma 3.4, which follows, there exists $K_1 < \infty$ such that $l_{N^*} \le K_1 d$ on $\{D^* = 1\}$. Using this and Wald's likelihood identity,

$$\begin{aligned} P_0(D^* = 1) &= E_1(l_{N^*}; D^* = 1, N^* < \infty) \\ &\le E_1(K_1 d; D^* = 1, N^* < \infty) \le K_1 d = O(d). \end{aligned}$$

Combining these two estimates gives

$$\begin{aligned} [cE_0(N|D^* = 1) + dE_0(M|D^* = 1)]P_0(D^* = 1) &= [c \cdot O(\log d^{-1}) + d \cdot O(\log d^{-1})]O(d) \\ &= O(d^2 \log d^{-1}). \qquad (3.23) \end{aligned}$$

Now, by definition of (N, M),

$$P_1(N < \infty|D^* = 1) = E_0(l_N^{-1} 1\{N < \infty\}|N > 0) \le E_0(d^2 1\{N < \infty\}|N > 0) \le d^2.$$

Plugging this and (3.23) into (3.22),

$$\begin{aligned} R_2 &\le O(d^2 \log d^{-1}) + d^2 \\ &= O(d^2 \log d^{-1}) \\ &= d \cdot O(d \log d^{-1}) = d \cdot o(1) = o(d), \end{aligned}$$

finishing the proof. □

The next lemma shows, by considering stopping risk concerns, that the Bayes test is roughly a likelihood ratio test.

Lemma 3.4. *There is a constant* $K_1 > 0$ *such that*

$$\begin{aligned} l_{N^*} &\leq K_1 d \quad \text{on} \quad \{D^* = 1\}, \\ l_{N^*} &\geq (K_1 d)^{-1} \quad \text{on} \quad \{D^* = 0\}. \end{aligned} \tag{3.24}$$

Conversely, there is a constant $K_2 > 0$ *such that* δ^* *stops after the* k*th stage of sampling and*

$$\begin{aligned} \text{rejects } H_0 \quad \text{if} \quad l_{N^{*k}} &\leq K_2 d, \\ \text{rejects } H_1 \quad \text{if} \quad l_{N^{*k}} &\geq (K_2 d)^{-1}. \end{aligned} \tag{3.25}$$

Proof. For $i = 0, 1$ and $k \geq 1$ let

$$r_{ik} = \frac{w_i \pi_i f_i(X_1, \ldots, X_{N^{*k}})}{\sum_{j=0}^{1} \pi_j f_j(X_1, \ldots, X_{N^{*k}})},$$

the posterior risk of rejecting H_i after the kth stage. Note that we can write these in terms of likelihood ratios:

$$r_{0k} = \frac{w_0 \pi_0 l_{N^{*k}}}{\pi_0 l_{N^{*k}} + \pi_1}, \quad r_{1k} = \frac{w_1 \pi_1}{\pi_0 l_{N^{*k}} + \pi_1}. \tag{3.26}$$

Also, let $r_k = r_{0k} \wedge r_{1k}$, the stopping risk after the kth stage.

The Bayes procedure stops sampling if the stopping risk is less than all possible continuation risks. One possible continuation is fully-sequential sampling. By Lemma 2 of [17] there is a constant $K^* < \infty$ such that a Bayes procedure can only stop when the continuation risk of fully-sequential sampling is less than K^* times the cost per observation - $c + d$ in this case. Thus, when δ^* stops,

$$r_{M^*} \leq K^*(c + d) \leq 2K^* d$$

meaning $r_{0M^*} \le 2K^*d$ or $r_{1M^*} \le 2K^*d$. If $r_{0M^*} \le 2K^*d$, then by the first relation in (3.26) and some simple algebra

$$l_{N^*} \le \frac{\pi_1 \cdot 2K^*d}{\pi_0(w_0 - 2K^*d)} \le \frac{4\pi_1 K^*}{\pi_0 w_0} d \tag{3.27}$$

for small enough d. Clearly $r_{0M^*} < r_{1M^*}$ in this case so we can be sure $D^* = 1$. Otherwise, $r_{1M^*} \le 2K^*d$ so that, similarly,

$$l_{N^*} \ge \frac{\pi_1(w_1 - 2K^*d)}{\pi_0 \cdot 2K^*d} \ge \frac{\pi_1 w_1}{4\pi_0 K^*} d^{-1}$$

for small enough d and $D^* = 0$. We see from this last and (3.27) that (3.24) holds with

$$K_1 = \frac{4\pi_1 K^*}{\pi_0 w_0} \vee \frac{4\pi_0 K^*}{\pi_1 w_1}.$$

Since each additional stage of sampling costs at least $c + d > d$, δ^* will stop after the kth stage of sampling if $r_k \le d$. If

$$l_{N^*k} \le \frac{\pi_1}{\pi_0 w_0} d, \tag{3.28}$$

then (3.26) and some algebra show

$$d \ge \frac{w_0 \pi_0 l_{N^*k}}{\pi_0 l_{N^*k} + \pi_1} = r_{0k}$$

and hence δ^* will stop. Also clearly $r_{0M^*} < r_{1M^*}$ so we can be sure δ^* rejects H_0. Similarly, if

$$l_{N^*k} \ge \frac{\pi_1 w_1}{\pi_0} d^{-1} \tag{3.29}$$

then

$$d \ge \frac{w_1 \pi_1}{\pi_0 l_{N^*k} + \pi_1} = r_{1k},$$

so δ^* will stop and reject H_1. Thus, we see from (3.28) and (3.29) that (3.25) holds with

$$K_2 = \frac{\pi_1}{\pi_0 w_0} \wedge \frac{\pi_0}{\pi_1 w_1}.$$

□

Next, we define a test δ and prove its optimality. For this, we consider separately two cases of the relationship between f_0 and f_1 in the exponential family (3.1). The first case, considered in Section 3.2.1, is when $I_0 = I_1$ and $\text{Var}_0 X_i = \text{Var}_1 X_i$. This is a symmetric case in the sense that the two corresponding one decision tests dictate the same initial stage size, and hence they can be applied simultaneously. This case is of interest because it contains, most notably, the Normal mean problem, i.e.,

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu = \mu_1,$$

where μ is the mean of Normal random variables with known variance, and the symmetric Bernoulli case,

$$H_0 : p = 1/2 - \beta \quad \text{vs.} \quad H_1 : p = 1/2 + \beta,$$

where p is the probability of success in a Bernoulli trial. If $I_0 \neq I_1$, the nature of the Bayes test is fundamentally different. In this case, considered in Section 3.2.2, the two initial stages given by the one decision tests are of different order of magnitude, and hence cannot be applied simultaneously. This gives rise to a necessary "exploratory" first stage. The remaining case, where $I_0 = I_1$ and $\text{Var}_0 X_i \neq \text{Var}_1 X_i$ is at present unsolved, but the popular examples contained in the former and the generality of the latter make our analysis sufficient for most practical purposes.

3.2.1 Case I: $I_0 = I_1$ and $\text{Var}_0 X_i = \text{Var}_1 X_i$

Assume $c \in \mathcal{B}_m(d)$. Let $(N^{(0)}, M^{(0)})$ be the one decision test of f_0 vs. f_1 described in Theorem 3.2 and let $(N^{(1)}, M^{(1)})$ be the corresponding one decision test of f_1 vs. f_0. Under the assumptions $I_0 = I_1$ and $\text{Var}_0 X_i = \text{Var}_1 X_i$, the two procedures $(N^{(0)}, M^{(0)})$ and $(N^{(1)}, M^{(1)})$ dictate the same first stage size. Define the first stage of $\delta = (N, M, D)$ to be this common first stage size,

$$N_1 \equiv N_1^{(0)} = N_1^{(1)}.$$

If $l_{N_1} \geq 1$, continue with $(N^{(0)}, M^{(0)})$, stopping the first time $l_{N^k} \geq d^{-1}$ to reject H_1, as dictated by $(N^{(0)}, M^{(0)})$, or $l_{N^k} \leq d$ to reject H_0. Otherwise, $l_{N_1} < 1$ so continue with $(N^{(1)}, M^{(1)})$ similarly.

Theorem 3.5. *If $I_0 = I_1$, $Var_0 X_i = Var_1 X_i$, and $c \in \mathcal{B}_m(d)$, then*

$$\begin{aligned} r(\delta) &= cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, C_0 I_0) + o(d) \\ r^* &= cI_0^{-1} \log d^{-1} + d \cdot u_m(Q, C_0 I_0) + o(d) \end{aligned} \tag{3.30}$$

as $d \to 0$, where

$$Q \equiv \lim_{d \to 0} \frac{d/c}{h_m(C_0 \log d^{-1})} \in (0, \infty].$$

Proof. Let $I = I_0 = I_1$ and note that the assumption of equal variances implies $C_0 = C_1$, so let C denote this common value. Since $r^* \leq r(\delta)$, it suffices to establish (3.30) with "$\leq$" and (3.30) with "$\geq$," which we do first. We have

$$\begin{aligned} r^* &= \sum_{i=0}^{1} \pi_i [cE_i N^* + dE_i M^* + w_i P_i(D^* = 1 - i)] \\ &= \sum_{i=0}^{1} [\pi_i c E_i N^* + \pi_i d E_i M^* + \pi_{1-i} w_{1-i} P_{1-i}(D^* = i)] \\ &= \sum_{i=0}^{1} \pi_{1-i} w_{1-i} [c_i E_i N^* + d_i E_i M^* + P_{1-i}(D^* = i)], \end{aligned} \tag{3.31}$$

where

$$c_i = \frac{\pi_i}{\pi_{1-i} w_{1-i}} c, \quad d_i = \frac{\pi_i}{\pi_{1-i} w_{1-i}} d.$$

Note that $d_i / c_i = d/c$ and

$$h_m(\log d_i^{-1}) = h_m(\log d^{-1} + O(1)) \sim h_m(\log d^{-1}).$$

Thus, if

$$h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1})$$

then

$$h_m(\log d_i^{-1}) \sim h_m(\log d^{-1}) \ll d_i/c_i \ll h_{m-1}(\log d_i^{-1}) \sim h_{m-1}(\log d^{-1}),$$

while if

$$\lim_{d \to 0} \frac{d/c}{h_m(\log d^{-1})} \in (0, \infty),$$

then

$$\lim_{d_i \to 0} \frac{d_i/c_i}{h_m(\log d_i^{-1})} = \lim_{d \to 0} \frac{d/c}{h_m(\log d^{-1})} \in (0, \infty).$$

This shows that $c_i \in \mathcal{B}_m(d_i)$. Moreover,

$$\lim_{d_i \to 0} \frac{d_i/c_i}{h_m(C \log d_i^{-1})} = \lim_{d \to 0} \frac{d/c}{h_m(C \log d^{-1})} = Q \in (0, \infty],$$

so by Lemma 3.3

$$\begin{aligned} c_i E_i N^* + d_i E_i M^* + P_{1-i}(D^* = i) &\geq c_i I^{-1} \log d_i^{-1} + d_i \cdot u_m(Q, CI) + o(d_i) \\ &= c_i I^{-1} \log d^{-1} + d_i \cdot u_m(Q, CI) + o(d). \end{aligned}$$

Plugging this into (3.31),

$$\begin{aligned} r^* &\geq \sum_{i=0}^{1} \pi_{1-i} w_{1-i} [c_i I^{-1} \log d^{-1} + d_i \cdot u_m(Q, CI) + o(d)] \\ &= \sum_{i=0}^{1} \pi_i [c I^{-1} \log d^{-1} + d \cdot u_m(Q, CI) + o(d)] \\ &= c I^{-1} \log d^{-1} + d \cdot u_m(Q, CI) + o(d), \end{aligned}$$

since $\pi_0 + \pi_1 = 1$.

Next we handle (3.30). Let $(N, M, D) = \delta$ and for an arbitrary event A let

$$r(\delta; A) = \sum_{i=0}^{1} \pi_i [cE_i(N; A) + dE_i(M; A) + w_i P_i(D = 1 - i, A)]$$

and obviously $r(\delta; A) + r(\delta; A') = r(\delta)$. Let $l^k = l_{N^k}$, the likelihood ratio after the kth stage. Let

$$\begin{aligned} A_0 &= \left\{ \left| \log l^1 - IN_1 \right| \leq C^{-1} \sqrt{N_1} \log N_1 \right\} \\ A_1 &= \left\{ \left| \log l^1 + IN_1 \right| \leq C^{-1} \sqrt{N_1} \log N_1 \right\}. \end{aligned}$$

Let $(N^{(0)}, M^{(0)})$ be the one decision test of f_0 vs. f_1 in the definition of δ. The following six bounds are proved in Lemma 3.6, which follows this proof:

$$\begin{aligned} cE_0(N; A_0) &\leq cE_0 N^{(0)} + o(d) && (3.32) \\ dE_0(M; A_0) &\leq dE_0 M^{(0)} + o(d) && (3.33) \\ P_0(D = 1, A_0) &= o(d) && (3.34) \\ cE_1(N; A_0) &= o(d) && (3.35) \\ dE_1(M; A_0) &= o(d) && (3.36) \\ P_1(D = 0, A_0) &\leq P_1(N^{(0)} < \infty) + o(d). && (3.37) \end{aligned}$$

Using these bounds

$$\begin{aligned} r(\delta; A_0) &= \sum_{i=0}^{1} \pi_i [cE_i(N; A) + dE_i(M; A) + w_i P_i(D = 1 - i, A)] \\ &\leq \pi_0 [cE_0 N^{(0)} + dE_0 M^{(0)} + o(d)] + \pi_1 [w_1 P_1(N^{(0)} < \infty) + o(d)] \\ &= \pi_1 w_1 [c_0 E_0 N^{(0)} + d_0 E_0 M^{(0)} + P_1(N^{(0)} < \infty)] + o(d) \\ &\leq \pi_1 w_1 [c_0 I^{-1} \log d_0^{-1} + d_0 \cdot u_m(Q, CI) + o(d_0)] + o(d) \quad \text{(by Theorem 3.2)} \\ &= \pi_0 [cI^{-1} \log d^{-1} + d \cdot u_m(Q, CI)] + o(d) && (3.38) \end{aligned}$$

and the same argument with the indices reversed gives

$$r(\delta; A_1) \leq \pi_1[cI^{-1}\log d^{-1} + d \cdot u_m(Q, CI)] + o(d). \tag{3.39}$$

Now we consider $r(\delta; A_0' \cap A_1')$. Let $A = A_0' \cap A_1'$. The bounds

$$\begin{aligned} cE_0(N; A) &= o(d) && (3.40)\\ dE_0(M; A) &= o(d) && (3.41)\\ P_0(D = 1, A) &= o(d), && (3.42) \end{aligned}$$

are also proved in the next lemma, along with their equivalents with indices reversed, and thus $r(\delta; A) = o(d)$. Combining this with (3.38) and (3.39) gives

$$\begin{aligned} r(\delta) &= r(\delta; A_0) + r(\delta; A_1) + r(\delta; A) \\ &\leq \sum_{i=0}^{1} \pi_i[cI^{-1}\log d^{-1} + d \cdot u_m(Q, CI)] + o(d) \\ &= cI^{-1}\log d^{-1} + d \cdot u_m(Q, CI) + o(d), \end{aligned}$$

finishing the proof. □

Lemma 3.6. *Under the assumptions of Theorem 3.5, the bounds (3.32)-(3.37) and (3.40)-(3.42) hold.*

Proof. Let $B = \{\log l^k > -\log d^{-1} \text{ for all } k = 1, \ldots, M\}$ and note that δ and $(N^{(0)}, M^{(0)})$ coincide on $A_0 \cap B$ since $\log l^1 \geq IN_1 - C\sqrt{N_1}\log N_1 > 0$ for small d on A_0 and $\log l^k$ never crosses the lower boundary $-\log d^{-1}$ on B. Recall the definition

$$t_\mu(p, a) = a/\mu - \frac{z_p\sqrt{4a\mu + z_p^2} - z_p^2}{2\mu^2}$$

and that the stages of our multistage sampling procedures, and hence the one decision tests and δ, are defined in terms of $t_\mu(p,a)$. First we prove the crude bound

$$E_i(N|U) = O(\log d^{-1}) \quad \text{for any } U \text{ such that } E_i(M|U) = O(1), \tag{3.43}$$

$i = 0, 1$. In the $c \in \mathcal{B}_m^o(d)$ [resp. $c \in \mathcal{B}_m^+(d)$] case, the mth [resp. $(m+1)$st] stage of δ begins geometric sampling, in which the size of each stage is bounded by

$$\begin{aligned} \lceil t_{CI}(p, (C\log d^{-1} - \sum Y_i) \vee (\sum Y_i + C\log d^{-1}))\rceil &\leq \lceil t_{CI}(p, 2C\log d^{-1})\rceil \\ &= \frac{2C\log d^{-1}}{CI} + o(\log d^{-1}) \\ &= O(\log d^{-1}), \end{aligned}$$

where $p \to 1$ but slowly enough so that $|z_p| = O(\log\log d^{-1})$. Similarly, $\lceil t_{CI}(p, 2C\log d^{-1})\rceil$ also bounds the first m stages, but where p goes to zero for the first $m-1$ stages and approaches a limit in $(0,1)$ for the mth stage of the boundary case. In either case, p is bounded below 1. Hence, these initial stages are $O(\log d^{-1})$ as well, since $t_{CI}(p,a)$ is nondecreasing in p. Thus, the size of each stage of δ is uniformly $O(\log d^{-1})$ and therefore

$$E_i(N|U) \leq O(\log d^{-1})E_i(M|U) = O(\log d^{-1}),$$

proving (3.43).

Clearly $E_0(M|A_0 \cap B') = O(1)$, so using this crude bound and Wald's likelihood identity,

$$P_0(A_0 \cap B') \leq P_0(B') = E_1(l^M; B') \leq E_1(d; B') \leq d$$

and $E_0(N; A_0 \cap B) \leq E_0 N^{(0)}$ since δ and $(N^{(0)}, M^{(0)})$ coincide on $A_0 \cap B$, so that

$$\begin{aligned} cE_0(N; A_0) &= cE_0(N; A_0 \cap B) + cE_0(N; A_0 \cap B') \\ &\leq cE_0 N^{(0)} + c \cdot O(d\log d^{-1}) \\ &= cE_0 N^{(0)} + o(c) = cE_0 N^{(0)} + o(d), \end{aligned}$$

which proves (3.32). Similarly, $E_0(M; A_0 \cap B) \leq E_0 M^{(0)}$ and $E_0(M|A_0 \cap B') = O(1)$, so that

$$\begin{aligned} dE_0(M; A_0) &\leq dE_0(M; A_0 \cap B) + dE_0(M|A_0 \cap B')P_0(A_0 \cap B') \\ &\leq dE_0 M^{(0)} + d \cdot O(1) \cdot d = dE_0 M^{(0)} + o(d), \end{aligned}$$

proving (3.33). Letting $\gamma(d) = IN_1 - C^{-1}\sqrt{N_1}\log N_1$,

$$\begin{aligned} P_0(D = 1, A_0) &\leq P_0(D = 1|A_0) = P_0(l^M \leq -\log d^{-1}|\log l^1 \geq \gamma(d)) \\ &\leq \exp[-(\log d^{-1} + \gamma(d))] \\ &= de^{-\gamma(d)} = o(d), \end{aligned}$$

proving (3.34). Since $\gamma \sim IN_1 \sim \log d^{-1}$ we have

$$\begin{aligned} P_1(A_0) &= E_0(l_1^{-1}; \log l_1 \geq \gamma(d)) \leq E_0(e^{-\gamma(d)}; \log l_1 \geq \gamma(d)) \\ &\leq e^{-\gamma(d)} \leq \exp[-(1/2)\log d^{-1}] = \sqrt{d}. \end{aligned}$$

Also $E_1(N|A_0) = O(\log d^{-1})$ by (3.43) so

$$cE_1(N; A_0) = cE_1(N|A_0)P_1(A_0) \leq c\sqrt{d} \cdot O(\log d^{-1}) = c \cdot o(1) = o(d),$$

proving (3.35). $E_1(M|A_0) = O(1)$ and clearly $P_1(A_0) \to 0$, so

$$dE_1(M; A_0) = dE_1(M|A_0)P_1(A_0) = d \cdot O(1) \cdot o(1) = o(d),$$

proving (3.36). Since δ and $(N^{(0)}, M^{(0)})$ coincide on $A_0 \cap B$,

$$P_1(D = 0, A_0 \cap B) = P_1(N^{(0)} < \infty, A_0 \cap B) \leq P_1(N^{(0)} < \infty).$$

Also

$$P_1(D = 0, A_0 \cap B') = E_0[(l^M)^{-1}; D = 0, A_0 \cap B'] \leq E_0[d; D = 0, A_0 \cap B'] \leq dP_0(B') = o(d)$$

since clearly $P_0(B') \to 0$. Combining these two gives

$$P_1(D=0; A_0) = P_1(D=0; A_0 \cap B) + P_1(D=0; A_0 \cap B') \le P_1(N^{(0)} < \infty) + o(d),$$

proving (3.37).

Now

$$P_0(A) \le P_0(A_0') = P_0(\log l^1 < \gamma(d)) = P_0\left(\frac{-\log l^1 + IN_1}{C^{-1}\sqrt{N_1}} > \frac{IN_1 - \gamma(d)}{C^{-1}\sqrt{N_1}}\right)$$

and

$$\frac{IN_1 - \gamma(d)}{C^{-1}\sqrt{N_1}} = \log N_1 = o(N_1^{1/6}),$$

so by large deviations and Mills' ratio

$$P_0(A) \le \Phi(-\log N_1)(1 + o(1)) \sim \frac{\phi(\log N_1)}{\log N_1} = O\left(\frac{\exp[-(1/2)(\log\log d^{-1})^2]}{\log\log d^{-1}}\right),$$

since $IN_1 \sim \log d^{-1}$ implies $\log N_1 = \log\log d^{-1} + O(1)$. Thus

$$\begin{aligned} cE_0(N; A) &= cE_0(N|A)P_0(A) \\ &\le c \cdot O(\log d^{-1}) \cdot O\left(\frac{\exp[-(1/2)(\log\log d^{-1})^2]}{\log\log d^{-1}}\right) \\ &= c \cdot o(1) = o(d), \end{aligned}$$

proving (3.40). It's not hard to see that $E_0(M|A) = O(1)$, so

$$E_0(M; A) = d \cdot O(P_0(A)) = d \cdot o(1) = o(d),$$

which is (3.41). Finally, since $l^M \le d$ on $\{D = 1\}$,

$$P_0(D = 1, A) = E_1(l^M; D = 1, A) \le dP_1(A) = o(d),$$

proving (3.42). □

3.2.2 Case II: $I_0 \neq I_1$

Let $I_0 < I_1$. Define $\delta = (N, M, D)$ for this case as follows. Let $(N^{(1)}, M^{(1)})$ be the one decision test of f_1 vs. f_0 described in Theorem 3.2. Let

$$Q_0 = \lim_{d \to 0} \frac{d/c}{h_m\left(C_0\left(1 - I_0/I_1\right)\log d^{-1}\right)} \in (0, \infty] \tag{3.44}$$

and let $(\dot{N}^{(0)}, \dot{M}^{(0)})$ be the one decision test of f_0 vs. f_1 described in Theorem 3.2, but with parameters

$$\frac{\pi_0 l_{N_1}}{\pi_1 w_1} \cdot c, \quad \frac{\pi_0 l_{N_1}}{\pi_1 w_1} \cdot d, \quad p^*(m, Q_0, C_0 I_0)$$

in place of c, d, p^*. Define the first stage of δ to be the first stage of $(N^{(1)}, M^{(1)})$, i.e., $N_1 \equiv N_1^{(0)}$. If $l_{N_1} < 1$, continue with $(N^{(1)}, M^{(1)})$, stopping the first time $l_{N^k} \leq d$ to reject H_0 (as dictated by $(N^{(1)}, M^{(1)})$) or $l_{N^k} \geq d^{-1}$ to reject H_1. Otherwise, $l_{N_1} \geq 1$ so begin $(\dot{N}^{(0)}, \dot{M}^{(0)})$, stopping the first time $l_{N^k} \geq d^{-1}$ to reject H_1 (as dictated by $(\dot{N}^{(0)}, \dot{M}^{(0)})$) or $l_{N^k} \leq d$ to reject H_0.

Theorem 3.7. *If $I_0 < I_1$ and $c \in \mathcal{B}_m(d)$, then*

$$\begin{aligned} r(\delta) &= \pi_0[cI_0^{-1}\log d^{-1} + d(1 + u_m(Q_0, C_0 I_0))] \\ &\quad + \pi_1[cI_1^{-1}\log d^{-1} + d \cdot u_m(Q_1, C_1 I_1)] + o(d) \end{aligned} \tag{3.45}$$

$$\begin{aligned} r^* &= \pi_0[cI_0^{-1}\log d^{-1} + d(1 + u_m(Q_0, C_0 I_0))] \\ &\quad + \pi_1[cI_1^{-1}\log d^{-1} + d \cdot u_m(Q_1, C_1 I_1)] + o(d) \end{aligned} \tag{3.46}$$

as $d \to 0$, where Q_0 is as in (3.44) and

$$Q_1 \equiv \lim_{d \to 0} \frac{d/c}{h_m(C_1 \log d^{-1})} \in (0, \infty].$$

In particular, $r(\delta) \leq r^* + o(d)$.

Proof. Let $l^k = l_{N^k}$,

$$T = \{t > 0 : |\log t - I_0 N_1| \le C_0^{-1}\sqrt{N_1}\log N_1\},$$

and $A_0 = \{l^1 \in T\}$. Let $\dot{\delta}_0 = (\dot{N}^{(0)}, \dot{M}^{(0)}, \dot{D}^{(0)})$ denote the continuation of δ after its first stage when $l^1 \ge 1$, and let $(\ddot{N}^{(0)}, \ddot{M}^{(0)})$ denote the one decision test of f_0 vs. f_1 that coincides with $\dot{\delta}_0$ except that $\dot{\delta}_0$ may stop before $(\ddot{N}^{(0)}, \ddot{M}^{(0)})$ and reject H_0 when the likelihood ratio crosses the lower boundary. We will write $(\dot{N}^{(0)}(l^1), \dot{M}^{(0)}(l^1), \dot{D}^{(0)}(l^1))$ and $(\ddot{N}^{(0)}(l^1), \ddot{M}^{(0)}(l^1))$ when we wish to emphasize the dependence on the value of l^1.

Using the bounds (3.34)-(3.36),

$$\begin{aligned} r(\delta; A_0) &= \pi_0 c E_0(N; A_0) + \pi_0 d E(M; A_0) + \pi_1 w_1 P_1(D = 0, A_0) + o(d) \\ &= E_0[\pi_0 c N + \pi_0 d M + \pi_1 w_1 (l^M)^{-1} \cdot 1\{D = 0\}; A_0] + o(d) \\ &= E_0[\pi_0 c \dot{N}^{(0)}(l^1) + \pi_0 d \dot{M}^{(0)}(l^1) + \pi_1 w_1 (l^1)^{-1} (l^{\dot{M}^{(0)}(l^1)})^{-1} \cdot 1\{\dot{D}^{(0)} = 0\}; A_0] \\ &\quad + \pi_0 c N_1 + \pi_0 d + o(d) \\ &= E_0[\varphi(l^1); l^1 \in T] + \pi_0 c N_1 + \pi_0 d + o(d), \end{aligned} \tag{3.47}$$

where

$$\varphi(t) \equiv \pi_0[c E_0 \dot{N}^{(0)}(t) + d E_0 \dot{M}^{(0)}(t)] + \pi_1 w_1 t^{-1} P_1(\dot{D}^{(0)}(t) = 0). \tag{3.48}$$

For $t \in T$,

$$\dot{N}^{(0)}(t) \le \ddot{N}^{(0)}(t) \quad \text{and} \quad \dot{M}^{(0)}(t) \le \ddot{M}^{(0)}(t) \tag{3.49}$$

since $\dot{\delta}_0$ coincides with $(\ddot{N}^{(0)}, \ddot{M}^{(0)})$ except that $\dot{\delta}_0$ may stop early by crossing the lower boundary. Also,

$$\{\dot{D}^{(0)}(t) = 0\} \subseteq \{\ddot{N}^{(0)}(t) < \infty\} \quad \text{for } t \in T \tag{3.50}$$

since the lower boundary cannot be crossed on $\{\dot{D}^{(0)}(t) = 0\}$, hence the two procedures coincide exactly. Thus, for $t \in T$,

$$\begin{aligned}
\varphi(t) &\leq \pi_0 c E_0 \ddot{N}^{(0)}(t) + \pi_0 d E_0 \ddot{M}^{(0)}(t) + \pi_1 w_1 t^{-1} P_1(\ddot{N}^{(0)}(t) < \infty) \quad \text{(by (3.49) and (3.50))} \\
&= \pi_1 w_1 t^{-1}[c' E_0 \ddot{N}^{(0)}(t) + d' E_0 \ddot{M}^{(0)}(t) + P_1(\ddot{N}^{(0)}(t) < \infty)], \qquad (3.51)
\end{aligned}$$

where

$$c' \equiv \frac{\pi_0 t}{\pi_1 w_1} c, \quad d' \equiv \frac{\pi_0 t}{\pi_1 w_1} d.$$

We now show that $c' \in \mathcal{B}_m(d')$ uniformly for $t \in T$. Note that $d'/c' = d/c$ and, for $t \in T$,

$$\begin{aligned}
\log(d')^{-1} &= \log d^{-1} - \log t + O(1) \\
&= \log d^{-1} - I_0 N_1 + o(N_1) \quad (\text{since } \log t \sim I_0 N_1 \text{ on } T) \\
&= \log d^{-1} - (I_0/I_1)\log d^{-1} + o(\log d^{-1}) \quad (\text{since } N_1 \sim I_1^{-1} \log d^{-1}) \\
&\sim (1 - I_0/I_1)\log d^{-1},
\end{aligned}$$

and this holds uniformly on T. Thus

$$h_m(\log(d')^{-1}) \sim h_m((1 - I_0/I_1)\log d^{-1}) \sim (1 - I_0/I_1)^{(1/2)^m} h_m(\log d^{-1}),$$

so if

$$h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1}),$$

then

$$h_m(\log(d')^{-1}) \ll d'/c' \ll h_{m-1}(\log(d')^{-1}),$$

and if

$$\lim_{d \to 0} \frac{d/c}{h_m(\log d^{-1})} \in (0, \infty),$$

then

$$\lim_{d'\to 0} \frac{d'/c'}{h_m(\log(d')^{-1})} = \lim_{d\to 0} \frac{d/c}{h_m((1-I_0/I_1)\log d^{-1})}$$
$$= (1-I_0/I_1)^{-(1/2)^m} \cdot \lim_{d\to 0} \frac{d/c}{h_m(\log d^{-1})} \in (0,\infty).$$

This shows that $c' \in \mathcal{B}_m(d')$ and, moreover,

$$\lim_{d\to 0} \frac{d'/c'}{h_m(C_0 \log(d')^{-1})} = \lim_{d\to 0} \frac{d/c}{h_m(C_0(1-I_0/I_1)\log d^{-1})} = Q_0 \in (0,\infty],$$

so by Theorem 3.2,

$$c'E_0\ddot{N}^{(0)}(t)+d'E_0\ddot{M}^{(0)}(t)+P_1(\ddot{N}^{(0)}(t)<\infty) \le c'I_0^{-1}\log(d')^{-1}+d'u_m(Q_0,C_0I_0)+o(d').$$

Plugging this into (3.51),

$$\begin{aligned} \varphi(t) &\le \pi_1 w_1 t^{-1}[c'I_0^{-1}\log(d')^{-1} + d'u_m(Q_0,C_0I_0) + o(d')] \\ &= \pi_1 w_1 t^{-1}[c'I_0^{-1}(\log d^{-1} - \log t + O(1)) + d'u_m(Q_0,C_0I_0) + o(d')] \\ &= \pi_0[cI_0^{-1}\log d^{-1} + d\cdot u_m(Q_0,C_0I_0)] - \pi_0 c I_0^{-1}\log t + o(d) \end{aligned}$$

uniformly on T, and plugging this into (3.47),

$$r(\delta;A_0) \le \pi_0[cI_0^{-1}\log d^{-1}+d(1+u_m(Q_0,C_0I_0))]+\pi_0 cI_0^{-1}[I_0N_1-E(\log l^1;l^1\in T)]+o(d). \tag{3.52}$$

Since $E_0 \log l^1 = I_0N_1$ and $P_0(A_0) \to 1$ quickly, one may suspect that

$$E_0(\log l^1;A_0) = I_0N_1 + o(1). \tag{3.53}$$

Assuming this holds, (3.52) becomes

$$r(\delta;A_0) \le \pi_0[cI_0^{-1}\log d^{-1} + d(1+u_m(Q_0,C_0I_0))] + o(d). \tag{3.54}$$

To see why (3.53) is true, first use Wald's equation and write

$$E_0(\log l^1; A_0) = I_0 N_1 - E_0(\log l^1; \log l^1 > I_0 N_1 + C_0^{-1}\sqrt{N_1}\log N_1) \\ - E_0(\log l^1; \log l^1 < I_0 N_1 - C_0^{-1}\sqrt{N_1}\log N_1); \tag{3.55}$$

we will show that these last two terms are $o(1)$. Letting $\Sigma_n = Y_1^{(0)} + \cdots + Y_n^{(0)}$ and $\gamma = C_0 I_0 N_1 + \sqrt{N_1}\log N_1$,

$$\begin{aligned} E_0(\log l^1; \log l^1 > I_0 N_1 + C_0^{-1}\sqrt{N_1}\log N_1) &= C_0^{-1} E_0[\Sigma_{N_1} - \gamma; \Sigma_{N_1} > \gamma] + \gamma \cdot P_0(\Sigma_{N_1} > \gamma) \\ &= O\left(\sqrt{N_1}\frac{\phi(\log N_1)}{(\log N_1)^2}\right) + O(N_1)\cdot O\left(\Phi(-\log N_1)\right) = o(1) \end{aligned}$$

by Lemma 2.10 and a routine large deviations argument. The other term in (3.55) is handled similarly, establishing (3.53).

Letting $A_1 = \{|\log(1/l^1) - I_1 N_1| \le C_1^{-1}\sqrt{N_1}\log N_1\}$ and repeating arguments from the proof of of Theorem 3.5 gives

$$\begin{aligned} r(\delta; A_1) &\le \pi_1[cI_1^{-1}\log d^{-1} + d\cdot u_m(Q_1, C_1 I_1)] + o(d) \quad \text{and} \\ r(\delta; A_0' \cap A_1') &= o(d). \end{aligned}$$

Combining these with (3.54) gives (3.45).

Next we show (3.46) with "$\ge$." Let $l^{*k} = l_{N^{*k}}$, $T^* = \{t > 0 : |\log t - I_0 N_1^*| \le C_0^{-1}\sqrt{N_1^*}\log N_1^*\}$, $A_0^* = \{l^{*1} \in T^*\}$, and

$$r_i^* = \pi_i(cE_i N^* + dE_i M^*) + \pi_{1-i} w_{1-i} P_{1-i}(D^* = i), \quad i = 0, 1.$$

Since δ^* follows its first stage with the optimal continuation, denoted by $(\dot{N}^*, \dot{M}^*, \dot{D}^*)$,

$$\begin{aligned} r_0^* &= E_0[E_0[\pi_0(cN^* + dM^*) + \pi_1 w_1 (l^{*M^*})^{-1} 1\{D^* = 0\} | l^{*1}]] \\ &= E_0[\pi_0(cE_0\dot{N}^*(l^{*1}) + dE_0\dot{M}^*(l^{*1})) + \pi_1 w_1 (l^{*1})^{-1} P_1(\dot{D}^*(l^{*1}) = 0)] \\ &\quad + \pi_0(cN_1^* + d) \end{aligned} \tag{3.56}$$

where we again write $(\dot{N}^*(l^{*1}), \dot{M}^*(l^{*1}), \dot{D}^*(l^{*1}))$ to reflect the dependence on the value of l^{*1}. Define

$$\begin{aligned} \varphi^*(t) &\equiv \pi_0[cE_0\dot{N}^*(t) + dE_0\dot{M}^*(t)] + \pi_1 w_1 t^{-1} P_1(\dot{D}^*(t) = 0) \\ &= \pi_1 w_1 t^{-1}[c' E_0 \dot{N}^*(t) + d' E_0 \dot{M}^*(t) + P_1(\dot{D}^*(t) = 0)]. \end{aligned}$$

It will be shown below that $N_1^* \sim I_1^{-1} \log d^{-1}$. Assuming this, the same arguments that showed $c' \in \mathcal{B}_m(d')$ when $t \in T$ (but with N_1^* in place of N_1) hold here for $t \in T^*$, and also

$$\lim_{d \to 0} \frac{d'/c'}{h_m(C_0 \log(d')^{-1})} = Q_0 \in (0, \infty].$$

Then by Lemma 3.3, for $t \in T^*$,

$$\begin{aligned} \varphi^*(t) &\geq \pi_1 w_1 t^{-1}[c' I_0^{-1} \log(d')^{-1} + d' u_m(Q_0, C_0 I_0) + o(d')] \\ &= \pi_0[cI_0^{-1} \log d^{-1} + d \cdot u_m(Q_0, C_0 I_0)] - \pi_0 c I_0^{-1} \log t + o(d) \end{aligned} \tag{3.57}$$

and this holds uniformly on T^*. Plugging this back into (3.56),

$$\begin{aligned} r_0^* &= E_0 \varphi^*(l^{*1}) + \pi_0(cN_1^* + d) \\ &\geq E_0[\varphi^*(l^{*1}); A_0^*] + \pi_0(cN_1^* + d) \quad (\text{since } \varphi^* \geq 0) \\ &\geq \pi_0[cI_0^{-1} \log d^{-1} + d \cdot u_m(Q_0, C_0 I_0] P_0(A_0^*) - \pi_0 c I_0^{-1} E_0[\log l^{*1}; A_0^*] \\ &\quad + \pi_0(cN_1^* + d) + o(d) \end{aligned} \tag{3.58}$$

by (3.57). The same argument that leads to (3.53) shows that $E_0[\log l^{*1}; A_0^*] = I_0 N_1^* + o(1)$ and a routine large deviations argument shows $1 - P_0(A_0^*) = O(\Phi(-\log N_1^*))$. Plugging these two estimates into (3.58) gives

$$r_0^* \geq \pi_0[cI_0^{-1} \log d^{-1} + d(1 + u_m(Q_0, C_0 I_0))] + o(d).$$

A straightforward application of Lemma 3.3 gives

$$r_1^* \geq \pi_1[cI_1^{-1}\log d^{-1} + d \cdot u_m(Q_1, C_1 I_1)] + o(d)$$

and adding these last two gives (3.46).

All that remains is to verify that $N_1^* \sim I_1^{-1}\log d^{-1}$. Suppose that

$$\underline{L} \equiv \liminf_{d\to 0} \frac{N_1^*}{\log d^{-1}} < I_1^{-1}. \tag{3.59}$$

Then there is a sequence of d's approaching 0 on which the lim inf is achieved, and by repeating the above arguments on this sequence

$$r_0^* \geq \pi_0[cI_0^{-1}\log d^{-1} + d(1 + u_m(Q_0', C_0 I_0))] + o(d), \tag{3.60}$$

where

$$\begin{aligned} Q_0' &\equiv \lim_{d\to 0} \frac{d/c}{h_m(C_0(1 - I_0\underline{L})\log d^{-1})} \\ &= \lim_{d\to 0} \frac{d/c}{\left(\frac{1-I_0\underline{L}}{1-I_0/I_1}\right)^{(1/2)^m} h_m(C_0(1 - I_0/I_1)\log d^{-1})} \\ &= Q_0 \cdot \left(\frac{1 - I_0/I_1}{1 - I_0\underline{L}}\right)^{(1/2)^m} \in (0, \infty]. \end{aligned}$$

Note further that $Q_0' < Q_0$ by this last. By reversing indices and repeating this argument, conditioning on $\{|\log(1/l^{*1}) - I_1 N_1^*| \leq C_1^{-1}\sqrt{N_1^*}\log N_1^*\}$ instead of A_0^*, we obtain

$$\begin{aligned} r_1^* &\geq \pi_1[cI_1\log d^{-1} + d(1 + u_m(Q_1', C_1 I_1))] + o(d) \\ &\geq \pi_1[cI_1\log d^{-1} + d(m+1)] + o(d)) \end{aligned} \tag{3.61}$$

since $u_m \geq m$, where

$$Q_1' \equiv \lim_{d\to 0} \frac{d/c}{h_m(C_1(1 - I_1\underline{L})\log d^{-1})} \in (0, \infty].$$

Then, using (3.45), (3.60), and (3.61), we would have

$$\begin{aligned} r^* - r(\delta) &= r_0^* + r_1^* - r(\delta) \\ &\geq d\left\{\pi_0[u_m(Q_0', C_0 I_0) - u_m(Q_0, C_0 I_0)] + \pi_1[m+1-u_m(Q_1, C_1 I_1)]\right\} - o(d). \end{aligned}$$

Now since $u_m(\cdot, C_0 I_0)$ is decreasing and $Q_0' < Q_0$,

$$u_m(Q_0', C_0 I_0) - u_m(Q_0, C_0 I_0) > 0.$$

Also $m + 1 - u_m(Q_1, C_1 I_1) > 0$ since $u_m < m + 1$. Hence there exists $\varepsilon > 0$ such that $r^* - r(\delta) \geq \varepsilon d - o(d) > 0$ for sufficiently small d. This obviously contradicts $r^* \leq r(\delta)$, so (3.59) cannot hold.

On the other hand, if

$$\eta \equiv \limsup_{d \to 0} \frac{N_1^*}{\log d^{-1}} - I_1^{-1} > 0, \tag{3.62}$$

then again on a sequence of d's approaching zero we would have

$$\begin{aligned} r^* - r(\delta) &= r_0^* + r_1^* - r(\delta) \\ &\geq r_0^* + \pi_1 c E_1 N^* - r(\delta) \\ &\geq \pi_0 c I_0^{-1} \log d^{-1} + \pi_1 c N_1^* - r(\delta) \quad \text{(by Lemma 3.3)} \\ &\geq \pi_0 c I_0^{-1} \log d^{-1} + \pi_1 c(\eta + I_1^{-1}) \log d^{-1}(1 + o(1)) - [(\pi_0/I_0 + \pi_1/I_1) c \log d^{-1} + O(d)] \\ &\qquad \text{(by (3.62) and (3.45))} \\ &= \pi_1(\eta + o(1)) \cdot c \log d^{-1} - O(d) \\ &= \pi_1(\eta + o(1)) \cdot c \log d^{-1} - o(c \log d^{-1}) > 0 \end{aligned}$$

for sufficiently small d, again a contradiction. Thus (3.62) cannot hold either, so that $N_1^* \sim I_1^{-1} \log d^{-1}$ and the proof is complete. □

3.3 A Numerical Example

The procedures δ described above in Theorems 3.2, 3.5, and 3.7 are asymptotic not only in the sense that their optimality properties are proved in the limit as $d \to 0$, but also in the sense that they are defined in terms of the rates at which $c, d \to 0$. Thus, there may be more than one small-sample procedure that are asymptotically equivalent to the above procedures and hence asymptotically optimal, among which a statistician may want to choose when designing a procedure for practical applications. In this section we describe one such small-sample procedure and give the results of a numerical experiment comparing it to group-sequential sampling.

Choose m_0^* and m_1^* to be

$$m_i^* = \inf\left\{m \geq 1 : \kappa_m(C_i I_i) h_m(C_i^{-1} \log d^{-1}) - \kappa_{m+1}(C_i I_i) h_{m+1}(C_i^{-1} \log d^{-1}) \leq d/c\right\},$$

$i = 0, 1$, and let δ be the test designed from the multistage sampling procedures $\delta_{m_i^*}(d/c)$ (the "$c \in \mathcal{B}_m^o(d)$ case" sampling procedures, as described in Section 2.63), as described in Sections 3.2.1 and 3.2.2. That is, δ has first stage the smaller of the first stages of the $\delta_{m_i^*}$, followed by the appropriate continuation, determined by whether $l^1 \geq 1$ or $l^1 < 1$.

Table 1 contains the results of a numerical experiment comparing δ with group-sequential (i.e., constant stage-size) testing of the hypotheses $\mu = .25$ vs. $\mu = -.25$, concerning the mean of normally distributed random variables with unit variance. Below $\delta_g(k)$ denotes group-sequential testing with constant stage-size k, which samples until

$$\left|\sum_j \log(f_0(X_j)/f_1(X_j))\right| \geq \log d^{-1} \tag{3.63}$$

at the end of a stage. The boundary $\log d^{-1}$ is chosen because it is the same boundary

Table 1
Numerical Results for Testing Normal Mean
$\mu = .25$ vs. $\mu = -.25$ ($d = .001$, $\pi_i = 1/2$, $w_i = 1$)

Test	EN	EM	int. risk (d)	2nd-order risk (d)
		$d/c = 1$		
δ	62.2	5.2	68.0	9.5
$\delta_g(1)$	57.5	57.5	115.0	56.5
$\delta_g(15)$	64.9	4.6	73.0	14.5
$\delta_g(30)$	76.7	2.6	80.0	21.5
		$d/c = 5$		
δ	68.3	2.9	16.7	4.2
$\delta_g(1)$	57.5	57.5	69.5	57.0
$\delta_g(22)$	72.7	3.3	18.0	5.5
$\delta_g(44)$	83.6	1.9	18.9	6.4
		$d/c = 10$		
δ	76.6	1.9	9.8	2.9
$\delta_g(1)$	57.5	57.5	63.8	57.1
$\delta_g(37)$	80.5	2.2	10.4	3.7
$\delta_g(74)$	97.6	1.3	11.2	4.5

used by δ. Indeed, recall that δ will stop sampling the first time

$$\begin{aligned} C_i \log d^{-1} &\leq \left| C_i \sum_j Y_j^{(i)} \right| = \left| C_i \sum_j \log(f_i(X_j)/f_{1-i}(X_j)) \right| \\ \Leftrightarrow \quad \log d^{-1} &\leq \left| \sum_j \log(f_0(X_j)/f_1(X_j)) \right|, \end{aligned}$$

where $i = 1\{\text{sign}(\log l^1) \leq 0\}$.

For each value of d/c, the operating characteristics of $\delta_g(k)$ are given for $k = 1$, the best possible k (determined by simulation), and two times the best possible k. Since both δ and δ_g must sample until (3.63) occurs, the cost of number of observations required for this and the first stage represents a "fixed cost" which all procedures will incur. Thus, we obtain a more accurate comparison of the efficiency due to sampling by considering the *2nd-order risk* of the procedures, defined as

$$\text{integrated risk } -(cEN^{(1)} + d),$$

where $N^{(1)}$ is the number of observations of $\delta_g(1)$.

The results show significant improvement in the integrated risk and 2nd-order risk of δ over δ_g. The size of the smallest possible 2nd-order risk is not known, so it is difficult to say how much further improvement is possible without backward induction type calculations, which remain prohibitively large in this general setting. We would expect the difference between δ and the best group sequential test to decrease for larger values of d/c, since $EM^* \to 1$ in this limit.

The procedure δ is asymptotically optimal by virtue of Theorems 3.5 and 3.7 when $c \in \mathcal{B}_m^o(d)$ since $m_i^* = m$ for sufficiently small d. This is true since

$$\begin{aligned}\kappa_m(C_i I_i) h_m(C_i \log d^{-1}) - {}&\kappa_{m+1}(C_i I_i) h_{m+1}(C_i \log d^{-1}) - d/c \\ &= (d/c) \cdot \left[O\left(\frac{h_m(\log d^{-1})}{d/c}\right) - O\left(\frac{h_{m+1}(\log d^{-1})}{d/c}\right) - 1 \right] \\ &= (d/c) \cdot [o(1) - o(1) - 1] \to -\infty,\end{aligned}$$

so

$$\kappa_m(C_i I_i) h_m(C_i \log d^{-1}) - \kappa_{m+1}(C_i I_i) h_{m+1}(C_i \log d^{-1}) \le d/c$$

and similarly

$$\kappa_k(C_i I_i) h_k(C_i \log d^{-1}) - \kappa_{k+1}(C_i I_i) h_{k+1}(C_i \log d^{-1}) > d/c$$

for all $k < m$ and for sufficiently small d. Thus m_i^* and m will coincide for sufficiently small d.

Chapter 4

Multistage Tests of Composite Hypotheses

In this chapter we extend the methods developed in Chapters 2 and 3 to the continuous setting. Consider testing the two separated composite hypotheses

$$H_0 : \underline{\theta} \le \theta \le \theta_0 \quad \text{vs.} \quad H_1 : \theta_0 < \theta_1 \le \theta \le \overline{\theta}, \tag{4.1}$$

by sampling i.i.d. random variables $X_1, X_2, \ldots$ in stages, whose distribution belongs to the exponential family of densities

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

with respect to some non-degenerate σ-finite measure. Assume that $[\underline{\theta}, \overline{\theta}]$ is contained in the interior of the natural parameter space, so that ψ is infinitely differentiable on $[\underline{\theta}, \overline{\theta}]$ and $\psi'(\theta) = E_\theta X_1$, $\psi''(\theta) = \mathrm{Var}_\theta X_1$, where $E_\theta, \mathrm{Var}_\theta$ denote expectation and variance under f_θ. We denote multistage tests of the hypotheses (4.1) by triples (N, M, D), where N is the total number of observations, M is the total number of stages, and D is the decision variable, taking values in $\{0, 1\}$. Again we assume a cost per observation c and a cost per stage d which will both approach zero at rates described below. Given a Lebesgue prior density λ_0 for the true parameter θ, positive and bounded on its support $[\underline{\theta}, \overline{\theta}]$, and a loss function $w(\theta)$ representing the penalty for a wrong decision when θ is the true value of the parameter, vanishing on

(θ_0, θ_1) and bounded away from 0 and ∞ on $[\underline{\theta}, \theta_0] \cup [\theta_1, \overline{\theta}]$, a natural measure of the performance of a procedure $\delta = (N, M, D)$ is its *integrated risk*,

$$r(\lambda_0, \delta) \equiv \int_{\underline{\theta}}^{\overline{\theta}} [cE_\theta N + dE_\theta M + w(\theta) P_\theta(\delta \text{ makes wrong decision})] \lambda_0(\theta) d\theta, \qquad (4.2)$$

where P_θ denotes probability under f_θ.

We define a family of multistage tests of the hypotheses (4.1) in Section 4.2, establish bounds on their operating characteristics, and, after a detailed analysis of the Bayes test in Section 4.3, show that they minimize the integrated risk to second-order as $c, d \to 0$. These variable stage-size procedures are similar to those considered in Chapters 2 and 3, yet the continuum of possible values of the parameter θ, which must be re-estimated at the end of each stage, makes the arguments considerably more intricate. These procedures also share a property of those of Section 3.2.2 that utilize an "exploratory" first stage – a stage whose size is a smaller order of magnitude than the first stage of any relevant simple hypothesis test. This first stage allows the the "true" parameter value to be sufficiently well estimated to design future stages.

In Section 4.4 we present the results of a numerical experiment comparing our procedure with group sequential (i.e., constant stage size) testing. The results show that these variable stage size tests significantly improve upon group sequential sampling, but also suggest that more efficient practical procedures are possible through a higher level of theoretical refinement.

As one may expect from (4.2), the nature of efficient tests depends heavily on the rates at which $c, d \to 0$. As was done is Chapter 3, we will assume that d is the independent variable and that $c = c(d)$, though this choice is arbitrary. We also continue to assume, for any sequence of d's approaching zero, that the sequence $\{(\log d^{-1}, d/c)\}$ is either in the *mth critical band*, i.e.,

$$h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1}), \qquad (4.3)$$

or on the *boundary between critical bands* m *and* $m+1$, i.e.,

$$\lim_{d\to 0} \frac{d/c}{h_m(\log d^{-1})} \in (0,\infty) \tag{4.4}$$

for some $m \geq 1$. We summarize this assumption by saying $c \in \mathcal{B}_m(d)$, where

$$\begin{aligned} \mathcal{B}_m^o(d) &\equiv \{c : (0,1) \to (0,1) \mid \quad h_m(\log d^{-1}) \ll d/c \ll h_{m-1}(\log d^{-1})\}, \\ \mathcal{B}_m^+(d) &\equiv \left\{ c : (0,1) \to (0,1) \;\middle|\; \frac{d/c}{h_m(\log d^{-1})} \to Q, \text{ some } Q \in (0,\infty) \right\}, \\ \text{and} \quad \mathcal{B}_m(d) &\equiv \mathcal{B}_m^o(d) \cup \mathcal{B}_m^+(d). \end{aligned}$$

As discussed in Sections 2.1 and 3.1, these definitions suffice to give a useful description of asymptotic optimality.

4.1 Preliminaries

Define a *test* of the hypotheses (4.1) to be a triple (N, M, D) where $N = (N_1, N_2, \ldots)$ is a sequence of stopping variables satisfying the the measurability requirement (2.63). N_k should be interpreted as the size of the kth stage and $N^k \equiv N_1 + \cdots + N_k$ the sample size through the kth stage. M is the number of stages before decision and, as a convenient abuse of notation, we also let N denote the total sample size, N^M.

Assume for convenience that

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

This standardization essentially involves subtracting $E_{\theta_2} X_1$ from the X_i and θ_2 from θ, where θ_2 is the unique solution of

$$\psi'(\theta_2) = \frac{\psi(\theta_1) - \psi(\theta_0)}{\theta_1 - \theta_0}$$

(see [2], Proposition 1.6), and it has the convenient feature that $\text{sign}(\theta) = \text{sign}(\psi'(\theta))$.

Let $S_k = X_1 + \cdots + X_k$ and, given a test (N, M, D), let $S^k = S_{N^k}$. Let $\hat{\theta}^*(n) =$

$(\psi')^{-1}(S_n/n)$, the (unrestricted) MLE of θ, and let $\hat{\theta}(n)$ denote the $[\underline{\theta}, \overline{\theta}]$-retricted MLE. We will use the shorthand $\hat{\theta}^*_k, \hat{\theta}_k$ for $\hat{\theta}^*(N^k), \hat{\theta}(N^k)$, with respect to a given test. It will prove useful to associate each point (n, S_n) with a point in the half-plane

$$\{(t,s) : 1 \le t < \infty, -\infty < s < \infty\}.$$

Thus we define the continuous analog of $\hat{\theta}$, namely

$$\hat{\theta}(s,t) \equiv \begin{cases} \overline{\theta} & \text{if } s > t\psi'(\overline{\theta}) \\ \underline{\theta} & \text{if } s < t\psi'(\underline{\theta}) \\ (\psi')^{-1}(s/t) & \text{otherwise.} \end{cases}$$

Let

$$I(\theta, \vartheta) = E_\theta \log[f_\theta(X_1)/f_\vartheta(X_1)] = (\theta - \vartheta)\psi'(\theta) - \psi(\theta) + \psi(\vartheta),$$

the Kullback-Leibler information number. Given a value θ, we will be interested in the "closest competitor" – the parameter value in the set $\{\theta_0, \theta_1\}$ minimizing $I(\theta, \cdot)$. Thus, given θ, define

$$\theta' = \begin{cases} \theta_0, & \text{if } \theta \ge 0 \\ \theta_1, & \text{if } \theta < 0. \end{cases}$$

Indeed,

$$I(\theta, \theta') = \begin{cases} \min_{\vartheta \le \theta_0} I(\theta, \vartheta), & \text{if } \theta \ge 0 \\ \min_{\vartheta \ge \theta_1} I(\theta, \vartheta), & \text{if } \theta < 0. \end{cases}$$

We will often use the convenient shorthand $I(\theta) \equiv I(\theta, \theta')$. We also define a slight extension of $I(\theta)$ that will be useful in proving convergence results near the endpoints of $[\underline{\theta}, \overline{\theta}]$, namely

$$I_\vartheta(\theta) \equiv (\theta - \theta')\psi'(\vartheta) - \psi(\theta) + \psi(\theta').$$

For functions g, continuous on $[\underline{\theta}, \overline{\theta}]$, we employ the generic notation

$$\overline{g} \equiv \max_{\theta \in [\underline{\theta}, \overline{\theta}]} g(\theta), \quad \underline{g} \equiv \min_{\theta \in [\underline{\theta}, \overline{\theta}]} g(\theta).$$

Applying this to $I(\theta)$, it is easy to see that

$$\overline{I} = I(\underline{\theta}) \vee I(\overline{\theta}), \quad \underline{I} = I(0).$$

Let

$$\ell(t, \theta) \equiv (\theta - \theta')s - t[\psi(\theta) - \psi(\theta')],$$

the continuous analog of the log-likelihood ratio of θ versus θ'. Note that dependence on s is suppressed in notation; this should not cause confusion as the value of s is often contained in the value of θ used, e.g., $\ell(t, \hat{\theta}(s,t)) = tI_{\hat{\theta}^*(s,t)}(\hat{\theta}(s,t))$. We will use the shorthand $\ell_k = \ell(N^k, \hat{\theta}_k)$, with respect to a given test.

Let

$$E_{\lambda_0}(\cdot) = \int_{\underline{\theta}}^{\overline{\theta}} E_\theta(\cdot)\lambda_0(\theta)d\theta,$$

the λ_0-mixture of θ-expectations, and $P_{\lambda_0}(\cdot) = E_{\lambda_0}1\{\cdot\}$. We associate each point (s,t) with the density

$$\lambda_{(s,t)}(\theta) \equiv \frac{\lambda_0(\theta)\exp[\theta s - t\psi(\theta)]}{\int_{\underline{\theta}}^{\overline{\theta}} \lambda_0(\vartheta)\exp[\vartheta s - t\psi(\vartheta)]d\vartheta}.$$

Note that $\lambda_{(s,t)}$ can be interpreted as a prior density, "moving forward" from (s,t), or a posterior density, since $\lambda_{(S_n,n)}$ is in fact the posterior density of θ given $X_1, \ldots, X_n$. λ_k will denote $\lambda_{(S^k,N^k)}$ with respect to a given test.

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

$$Y_0(s,t) = \frac{\int_{\underline{\theta}}^{\theta_0} w(\vartheta)\exp[\vartheta s - t\psi(\vartheta)]\lambda_0(\vartheta)d\vartheta}{\int_{\underline{\theta}}^{\overline{\theta}} \exp[\vartheta s - t\psi(\vartheta)]\lambda_0(\vartheta)d\vartheta},$$

and the posterior risk of rejecting $\theta \geq \theta_1$ by

$$Y_1(s,t) = \frac{\int_{\theta_1}^{\overline{\theta}} w(\vartheta) \exp[\vartheta s - t\psi(\vartheta)]\lambda_0(\vartheta)d\vartheta}{\int_{\underline{\theta}}^{\overline{\theta}} \exp[\vartheta s - t\psi(\vartheta)]\lambda_0(\vartheta)d\vartheta}.$$

Then the *stopping risk* at (s,t) is

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

Note that, with respect to a given test $\delta = (N, M, D)$,

$$E_{\lambda_0} r(\lambda_M) = E_{\lambda_0}[w(\theta); \delta \text{ makes wrong decision}],$$

so we may write

$$r(\lambda_0, \delta) = E_{\lambda_0}[cN + dM + r(\lambda_M)].$$

The first auxiliary lemma gives a bound on the rate of convergence of the expected inverse information number.

Lemma 4.1. *As* $n \to \infty$,

$$E_\theta I_{\hat{\theta}^*(n)}(\hat{\theta}(n))^{-1} = I(\theta)^{-1} + O(1/n) \tag{4.5}$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Remark. If $N = N(d)$ is a stopping time and $\underline{n}(d)$ a function such that $N \geq \underline{n}$ a.s. and $\underline{n}(d) \to \infty$ as $d \to 0$, then the lemma implies

$$E_\theta I_{\hat{\theta}^*(N)}(\hat{\theta}(N))^{-1} = I(\theta)^{-1} + O(1/\underline{n})$$

as $d \to 0$; the lemma will frequently be used in this form.

Proof. It suffices to prove (4.5) for all $\theta \in [\underline{\theta}, \overline{\theta}]$ since uniformity follows from continuity of $\theta \mapsto E_\theta I_{\hat{\theta}_n^*}(\hat{\theta}_n)^{-1} - I(\theta)^{-1}$ and compactness of $[\underline{\theta}, \overline{\theta}]$; see, for example,

[26], Theorem 7.25.

Let $\varphi = (\psi')^{-1}$, $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$, $\mathcal{N}$ be the natural parameter space, and $J = \psi'(\mathcal{N})$. For $x \in J$ define

$$g(x) = \begin{cases} g_1(x) \equiv I(\varphi(x))^{-1}, & x \in J_1 \equiv [\psi'(\underline{\theta}), \psi'(\overline{\theta})] \\ g_2(x) \equiv I_{\varphi(x)}(\overline{\theta})^{-1}, & x \in J_2 \equiv (\psi'(\overline{\theta}), \sup J) \\ g_3(x) \equiv I_{\varphi(x)}(\underline{\theta})^{-1}, & x \in J_3 \equiv (\inf J, \psi'(\underline{\theta})) \end{cases}$$

so that $g(\overline{X}_n) = I_{\hat{\theta}^*(n)}(\hat{\theta}(n))^{-1}$ and we can write

$$\begin{aligned} E_\theta I_{\hat{\theta}_n^*}(\hat{\theta}_n)^{-1} - I(\theta)^{-1} &= E_\theta[g(\overline{X}_n) - g(\psi'(\theta))] \\ &= \sum_{i=1}^{3} E_\theta[g_i(\overline{X}_n) - g(\psi'(\theta)); \overline{X}_n \in J_i] \equiv \sum_{i=1}^{3} A_i. \quad (4.6) \end{aligned}$$

First consider $\theta \in (\underline{\theta}, \overline{\theta})$. Since $g(\psi'(\theta)) = g_1(\psi'(\theta))$, using a Taylor series we can write

$$g_1(\overline{X}_n) - g(\psi'(\theta)) = g_1'(\psi'(\theta))(\overline{X}_n - \psi'(\theta)) + R_1(\overline{X}_n),$$

where $|R_1(\overline{X}_n)| \leq (\overline{X}_n - \psi'(\theta))^2 \overline{|g_1''|}/2$. Then

$$\begin{aligned} A_1 &= E_\theta[g_1(\overline{X}_n) - g(\psi'(\theta)); \overline{X}_n \in J_1] \\ &= E_\theta[g_1'(\psi'(\theta))(\overline{X}_n - \psi'(\theta)) + R_1(\overline{X}_n); \overline{X}_n \in J_1] \\ &= g_1'(\psi'(\theta)) E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_1] + E_\theta[R_1(\overline{X}_n); \overline{X}_n \in J_1]. \end{aligned}$$

Since $E_\theta \overline{X}_n = \psi'(\theta)$,

$$E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_1] = -E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_2 \cup J_3]$$

and

$$\begin{aligned} E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_2] &= (\psi'(\overline{\theta}) - \psi'(\theta)) P_\theta(\overline{X}_n > \psi'(\overline{\theta})) + E_\theta(\overline{X}_n - \psi'(\overline{\theta}); \overline{X}_n > \psi'(\overline{\theta})) \\ &\leq (\psi'(\overline{\theta}) - \psi'(\theta)) P_\theta(\overline{X}_n > \psi'(\overline{\theta})) + E_\theta(\overline{X}_n - a^*; \overline{X}_n > a^*), \quad (4.7) \end{aligned}$$

where $a^* \equiv \psi'(\theta) + n^{-5/14}\sqrt{\psi''(\theta)} < \psi'(\bar{\theta})$, for sufficiently large n. Using large deviations,

$$\begin{aligned}
P_\theta(\overline{X}_n > \psi'(\bar{\theta})) &= P_\theta\left(\frac{\overline{X}_n - \psi'(\theta)}{\sqrt{\psi''(\theta)/n}} > \left[\frac{\psi'(\bar{\theta}) - \psi'(\theta)}{\sqrt{\psi''(\theta)}}\right]\sqrt{n}\right) \\
&\leq P_\theta\left(\frac{\overline{X}_n - \psi'(\theta)}{\sqrt{\psi''(\theta)/n}} > n^{1/7}\right) \\
&\sim \Phi(-n^{1/7}) = o(1/n). \qquad (4.8)
\end{aligned}$$

Also, since $(na^* - n\psi'(\theta))/\sqrt{n\psi''(\theta)} = n^{1/7} = o(n^{1/6})$,

$$\begin{aligned}
E_\theta(\overline{X}_n - a^*; \overline{X}_n > a^*) &= n^{-1}E_\theta(n\overline{X}_n - na^*; n\overline{X}_n > na^*) \\
&\sim n^{-1} \cdot \frac{\phi(n^{1/7})}{n^{1/7}}\sqrt{n} = o(1/n) \qquad (4.9)
\end{aligned}$$

by Lemma 2.10. Plugging these two estimates into (4.7) gives

$$E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_2] = o(1/n)$$

and the same argument works on J_3 so we have

$$|E_\theta[\overline{X}_n - \psi'(\theta); \overline{X}_n \in J_1]| = o(1/n).$$

$$\begin{aligned}
|E_\theta[R_1(\overline{X}_n); \overline{X}_n \in J_1]| &\leq (|\overline{g_1''}|/2)E_\theta[(\overline{X}_n - \psi'(\theta))^2; \overline{X}_n \in J_1] \\
&\leq (|\overline{g_1''}|/2)\mathrm{Var}_\theta(\overline{X}_n) \\
&= (|\overline{g_1''}|/2)\psi''(\theta)/n = O(1/n), \qquad (4.10)
\end{aligned}$$

giving $|A_1| \leq O(1/n)$.

To estimate A_2 observe that, for $\overline{X}_n \in J_2$, $g_2(\overline{X}_n) \le g_2(\psi'(\overline{\theta})) = I(\overline{\theta})^{-1}$, so

$$\begin{aligned} |A_2| &= |E_\theta[g_2(\overline{X}_n) - g(\psi'(\theta)); \overline{X}_n \in J_2]| \\ &\le |I(\overline{\theta})^{-1} + I(\theta)^{-1}| P_\theta(\overline{X}_n \in J_2) \\ &= |I(\overline{\theta})^{-1} + I(\theta)^{-1}| P_\theta(\overline{X}_n > \psi'(\overline{\theta})) \\ &= o(1/n) \end{aligned} \tag{4.11}$$

by (4.8). A_3 is handled similarly and plugging into (4.6) shows that (4.5) holds for $\theta \in (\underline{\theta}, \overline{\theta})$.

Next we consider the $\theta = \overline{\theta}$ case; $\theta = \underline{\theta}$ is handled similarly. Observe that $g(\psi'(\overline{\theta})) = g_1(\psi'(\overline{\theta})) = g_2(\psi'(\overline{\theta}))$ and a simple computation verifies that $g_1'(\psi'(\overline{\theta})) = g_2'(\psi'(\overline{\theta}))$. Then, using the same expansion (4.6) and defining R_2 by analogy with R_1,

$$\begin{aligned} |A_1 + A_2| &= \left| \sum_{i=1}^{2} E_{\overline{\theta}}[g_i(\overline{X}_n) - g_i(\psi'(\theta)); \overline{X}_n \in J_i] \right| \\ &= \left| \sum_{i=1}^{2} E_{\overline{\theta}}[g_i'(\psi'(\overline{\theta}))(\overline{X}_n - \psi'(\overline{\theta})) + R_i(\overline{X}_n); \overline{X}_n \in J_i] \right| \\ &\le |g_1'(\psi'(\overline{\theta})) E_{\overline{\theta}}[\overline{X}_n - \psi'(\overline{\theta}); \overline{X}_n \in J_1 \cup J_2]| + \left| \sum_{i=1}^{2} E_{\overline{\theta}}[R_i(\overline{X}_n); \overline{X}_n \in J_i] \right| \\ &= o(1/n) + \left| \sum_{i=1}^{2} E_{\overline{\theta}}[R_i(\overline{X}_n); \overline{X}_n \in J_i] \right|, \end{aligned}$$

using the argument leading to (4.9). Repeating the argument leading to (4.10) gives

$$\left| \sum_{i=1}^{2} E_{\overline{\theta}}[R_i(\overline{X}_n); \overline{X}_n \in J_i] \right| \le O(\mathrm{Var}_{\overline{\theta}}(\overline{X}_n)) = O(1/n)$$

and hence $|A_1 + A_2| \le O(1/n)$. The same argument leading to (4.11) gives $|A_3| \le o(1/n)$ and combining this with $|A_1 + A_2| = O(1/n)$ shows that (4.5) holds at $\theta = \overline{\theta}$, as well as $\theta = \underline{\theta}$. □

The next two lemmas are Laplace-type expansions of the stopping risk due to Lorden [23].

Lemma 4.2.

$$\ell(t, \hat{\theta}(s,t)) + O(1) \leq \log Y_0(s,t)^{-1} \leq \ell(t, \hat{\theta}(s,t)) + O(\log t)$$

uniformly for $s \geq 0$ *as* $(s \vee t) \to \infty$.

Lemma 4.3. *For every* n*, as* $t \to \infty$

$$\log Y_0(s + S_n, t+n)^{-1} = \log Y_0(s,t)^{-1} + \ell(n, \hat{\theta}(s,t)) + o(1)$$

uniformly for

$$\psi'(\theta_0) + \varepsilon \leq s/t \leq \psi'(\overline{\theta}) - \varepsilon \quad \textit{and} \quad \left| \frac{s + S_n}{t+n} - \psi'(\hat{\theta}(s,t)) \right| \leq \varepsilon/2,$$

where $\varepsilon > 0$.

Remark. Lemmas 4.2 and 4.3 hold with Y_0 replaced by Y_1 and the restrictions appropriately modified for $s \leq 0$.

Define

$$a = \log d^{-1}, \quad a_k = a - \log r(\lambda_k)^{-1} \quad \text{for } k \geq 1$$

with respect to a given test. We will see below that a_k represents, after k stages of the given efficient procedure, the amount the log inverse stopping risk must further increase before stopping. The next lemma gives bounds on the difference of successive a_k for any procedure satisfying some mild bounds.

Lemma 4.4. *Let* $k \geq 1$ *and* $\delta = (N, M, D)$ *be any procedure such that there is a function* $\underline{n}(d) \to \infty$ *and a constant* $C < \infty$ *satisfying* $\underline{n}(d) \leq N_1$ *and* $N \leq a^C$ *a.s. Then, under* δ,

$$\ell(N_{k+1}, \hat{\theta}_k) - O(\log a) \leq a_k - a_{k+1} \leq \ell(N_{k+1}, \hat{\theta}_{k+1}) + O(\log a).$$

Proof. The restrictions on N allow us to write $|\log r(\lambda_i)^{-1} - \ell_i| \leq O(\log a)$ for $i = k, k+1$, by Lemma 4.2 and its analog for $Y_1(s,t)$. Using this,

$$\begin{aligned} a_{k+1} &= a - \log r(\lambda_{k+1})^{-1} \\ &\leq a - \ell_{k+1} + O(\log a) \\ &\leq a - \ell(N^{k+1}, \hat{\theta}_k) + O(\log a) \quad \text{(since } \ell_{k+1} \geq \ell(N^{k+1}, \hat{\theta}_k)) \\ &= a - \ell_k - \ell(N_{k+1}, \hat{\theta}_k) + O(\log a) \\ &\leq a_k - \ell(N_{k+1}, \hat{\theta}_k) + O(\log a), \end{aligned}$$

which gives the first inequality. On the other hand,

$$\begin{aligned} a_{k+1} &\geq a - \ell_{k+1} + O(\log a) \\ &= a - \ell(N^k, \hat{\theta}_{k+1}) - \ell(N_{k+1}, \hat{\theta}_{k+1}) + O(\log a) \\ &\geq a - \ell_k - \ell(N_{k+1}, \hat{\theta}_{k+1}) + O(\log a) \quad \text{(since } \ell_k \geq \ell(N^k, \hat{\theta}_{k+1})) \\ &\geq a_k - \ell(N_{k+1}, \hat{\theta}_{k+1}) + O(\log a), \end{aligned}$$

which gives the second inequality. □

4.2 The Tests δ_α and δ

In this section we define a test δ and prove bounds on its operating characteristics. Examining the properties of the Bayes procedure in Section 4.3 will show that δ is second-order optimal.

For $x, \sigma > 0$, let $t = t(z, x, \mu, \sigma)$ be the unique solution of

$$\frac{x - \mu t}{\sigma\sqrt{t}} = z,$$

i.e.,

$$t(z, x, \mu, \sigma) = \frac{x}{\mu} - \frac{z\sigma\sqrt{4x\mu + z^2\sigma^2} - z^2\sigma^2}{2\mu^2}$$

by a simple computation. If Z is a standard normal random variable, then

$$P(\sigma\sqrt{t}Z + \mu t \geq x) = \Phi(-z).$$

Therefore, under appropriate regularity conditions that allow Central Limit Theorem-type approximations, the probability that a random process with mean μ and variance σ^2 per unit time will be across a boundary x units away at the end of a stage of size $t(z_p, x, \mu, \sigma)$ approaches p. We will use this idea to define δ.

The procedure δ begins with an "exploratory" first stage and then follows with, on average, $m-1$ "conservative" stages using the MLE as an estimate for the true value of θ and a_k as an estimate for the distance from the current value of the log-likelihood ratio to the optimal boundary. If $c \in \mathcal{B}_m^+(d)$, the $(m+1)$st stage is a "critical" stage in the sense that the stopping probability is determined by $\lim_{d\to 0}(d/c)/h_m(a)$ and bounded away from 0 and 1, followed (if necessary) by geometric sampling with stopping probability approaching 1. If $c \in \mathcal{B}_m^o(d)$, no critical stage is necessary so the $(m+1)$st stage begins the geometric sampling. The stopping risk is computed after each stage and δ stops as soon as the stopping risk is no greater than d, or equivalently, when $a_k \leq 0$. The value of D is determined of course by which hypothesis has smaller posterior risk of rejection. In addition, the total sample size N has a fixed upper bound $\overline{n}$, defined below.

We first define a sub-family of tests, $\{\delta_\alpha\}_{\alpha>0}$, which we will use to define $\delta = \delta_{\alpha(d)}$ for a function $\alpha(d)$ that approaches 0 as $d \to 0$. (In practice, this limiting process can be dispensed with and δ_0 can simply be used; see Section 4.4.) After an "exploratory" first stage, δ_α essentially mimics the procedures defined in Chapters 2 and 3 by taking as large a sample as possible at each stage while keeping the sampling costs the correct order of magnitude, but while "estimating all parameters as it goes along."

Specifically, for $\alpha \geq 0$, $k = 1, 2, \ldots$ let

$$\xi_k^\alpha(\theta) = \left[1 - \frac{I(\theta)}{(1+\alpha)\overline{I}}\right]^{(1/2)^{k-1}} \left[\frac{(\theta-\theta')^2\psi''(\theta)}{I(\theta)}\right]^{1-(1/2)^{k-1}} \tag{4.12}$$

and let $\xi_k(\theta) = \xi_k^0(\theta)$. The ξ_k^α represent the units of the smallest possible (in probability) a_k and play a similar role to the κ_m in Chapters 2 and 3. Observe that the ξ_k^α satisfy

$$\xi_{k+1}^\alpha(\theta) = \sqrt{\xi_k^\alpha(\theta) \cdot \frac{(\theta - \theta')^2 \psi''(\theta)}{I(\theta)}}. \tag{4.13}$$

We will let $\xi_k^\alpha = \xi_k^\alpha(\hat{\theta}_k)$ with respect to a given procedure. Recall the constants defined in Section 2.1.3,

$$C_m^m = \prod_{i=1}^{m-1} \left[(1/2)^{m-1-i} - (1/2)^{m-1}\right]^{(1/2)^{i+1}}$$

and that Lemma 2.10 established

$$\sqrt{F_{d/c}^{(m-1)}(a)} \sim C_m^m h_m(a)$$

when $c \in \mathcal{B}_m^+(d)$. For $Q > 0$ let $z^\alpha(\theta, Q)$ be the unique solution of

$$\frac{\Phi(-z^\alpha(\theta, Q))}{\phi(z^\alpha(\theta, Q))} = \frac{QI(\theta)C_m^m}{\xi_{m+1}^\alpha(\theta)} \tag{4.14}$$

and let $z_m^\alpha(Q) = z^\alpha(\hat{\theta}_m, Q)$ with respect to a given procedure.

Now fix $0 < \alpha < 1$ and let $\delta_\alpha = (N, M, D)$. Let

$$\begin{aligned} \mu_k^* &= I(\hat{\theta}_k^*) \\ \sigma_k^{*2} &= [\hat{\theta}_k^* - (\hat{\theta}_k^*)']^2 \psi''(\hat{\theta}_k^*) \end{aligned}$$

with respect to δ_α, which we now define. Let $\overline{n} = \lceil 3a/\underline{I} \rceil$ and

$$\begin{aligned} N_1 &= \left\lceil \frac{a}{(1+\alpha)\overline{I}} \right\rceil \\ N_{k+1} &= \lceil t(\sqrt{\log(a_k/(d/c)^2 + 1)}, a_k, \mu_k^*, \sigma_k^*) \rceil 1\{a_k > 0\} \wedge (\overline{n} - N^k) \end{aligned}$$

for $1 \leq k < m$. When $c \in \mathcal{B}_m^o(d)$, let

$$N_{m+k+1} = \lceil t(z, a_{m+k}, \mu_{m+k}^*, \sigma_{m+k}^*) \rceil 1\{a_{m+k} > 0\} \wedge (\overline{n} - N^{m+k}) \tag{4.15}$$

for $k \geq 0$, where $z \to -\infty$ satisfies $h_m(a)|z| = o(d/c)$; z represents the standard normal upper quantile for geometric sampling. In this $c \in \mathcal{B}_m^o(d)$ case we then let $\delta = \delta_{\alpha(d)}$ for any function $\alpha(d) \to 0$ as $d \to 0$; e.g., $\alpha(d) = d$ suffices. If $c \in \mathcal{B}_m^+(d)$ and $Q \equiv \lim_{d \to 0}(d/c)/h_m(a) \in (0, \infty)$, then

$$N_{m+1} = \lceil t(z_m^\alpha(Q), a_m, \mu_m^*, \sigma_m^*) \rceil 1\{a_m > 0\} \wedge (\overline{n} - N^m),$$

where N_{m+1+k} is given by (4.15) for $k \geq 1$. In this boundary case also, $\delta = \delta_{\alpha(d)}$, where $\alpha(d) \to 0$ as $d \to 0$, but the function $\alpha(d)$ will be specified in the proof of Theorem 4.10. Finally, let

$$M = \inf\{k \geq 1 : a_k \leq 0 \quad \text{or} \quad N^k = \overline{n}\}.$$

Observe that $r(\lambda_M) \leq d$ under δ_α since, on $\{N < \overline{n}\}$,

$$\begin{aligned} r(\lambda_M) &= \exp(-\log r(\lambda_M)^{-1}) \\ &= \exp(a_M - a) \\ &\leq \exp(-a) \quad (\text{since } a_M \leq 0 \text{ on } \{N < \overline{n}\}) \\ &= d, \end{aligned}$$

while on $\{N = \overline{n}\}$,

$$\begin{aligned} r(\lambda_M) &= \exp(-\log r(\lambda_M)^{-1}) \\ &\le \exp(-\ell_M + O(1)) \quad \text{(by Lemma 4.2)} \\ &\le \exp(-NI(\hat{\theta}_M) + O(1)) \\ &\le \exp(-\overline{n}\underline{I} + O(1)) \\ &= \exp(-3a + O(1)) \le e^{-2a} = d^2 \end{aligned}$$

for sufficiently small d. The lemmas that follow establish further properties of sampling under δ_α.

For $\varepsilon > 0$ and $k \ge 1$ let

$$V_k(\varepsilon) = \left\{ \left| \hat{\theta}^*_k - \theta \right| \le \varepsilon \right\}$$

with respect to δ_α. Note that the dependence on θ is suppressed in notation; this should not cause confusion as its probability will always be computed under P_θ for the same value of θ. The next lemma gives a lower bound on the rate at which $P_\theta(V_k(\varepsilon)) \to 1$.

Lemma 4.5. *Let $k \ge 1$. There exists $\eta > 0$ such that*

$$P_\theta(V_k(\varepsilon)) \ge 1 - 2\exp(-\eta\varepsilon^2 a) \tag{4.16}$$

for all $0 < \varepsilon < 1$, uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$. In particular, $P_\theta(V_k(\varepsilon)) \to 1$ uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$ even if $\varepsilon \to 0$, provided $\varepsilon\sqrt{a} \to \infty$.

Proof. Let $0 < \varepsilon < 1$.

$$P_\theta(\hat{\theta}^*_k > \theta + \varepsilon) = P_\theta((\psi')^{-1}(S^k/N^k) > \theta + \varepsilon) = P_\theta(S^k > N^k\psi'(\theta + \varepsilon))$$

since ψ' is increasing. By Theorem 7.5 of [2],

$$P_\theta(S_n > x) \leq \exp[-nI((\psi')^{-1}(x/n), \theta)].$$

Using this and letting $\eta_o = [(1+\alpha)\overline{I}]^{-1}$ so that $N^k \geq N^1 \geq \eta_o a$,

$$P_\theta(\hat{\theta}_k^* > \theta + \varepsilon) \leq \exp[-\eta_o a I(\theta + \varepsilon, \theta)] \leq \exp[-\eta \varepsilon^2 a],$$

some $\eta > 0$, since $I(\theta + \varepsilon, \theta) \geq \eta_o' \varepsilon^2$, some $\eta_o' > 0$. The other tail is handled similarly and the second claim follows immediately from (4.16). $\square$

For $\varepsilon > 0$ and $k \geq 1$ let

$$U_k(\varepsilon) = \left\{ a_k > (1+\varepsilon)\xi_k^\alpha F_{d/c}^{(k-1)}(a) \right\}.$$

The next two lemmas will allow us to make precise statements about the behavior of a_k under δ_α.

Lemma 4.6. *Under δ_α there exists $\eta > 0$ such that for any $0 < \varepsilon < 1$,*

$$P_\theta \left(\left| \frac{a_1}{\xi_1^\alpha a} - 1 \right| > \varepsilon \right) = O(\Phi(-(\eta \varepsilon \sqrt{a} \wedge a^{1/7}))) \tag{4.17}$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Proof. By Lemma 4.2, $a_1 \leq a - \ell_1 + O(1)$, so

$$P_\theta(U_1(\varepsilon)) \leq P_\theta(\ell_1 < -a[(1+\varepsilon)\xi_1^\alpha - 1] + O(1)) = P_\theta\left(\frac{\ell_1 - \mu_1 N_1}{\sigma_1 \sqrt{N_1}} < \zeta \right),$$

where

$$\zeta \equiv \frac{-a[(1+\varepsilon)\xi_1^\alpha - 1] - \mu_1 N_1 + O(1)}{\sigma_1 \sqrt{N_1}}.$$

Since $0 \le N_1 - a/[(1+\alpha)\overline{I}] < 1$,

$$\begin{aligned} \zeta &\le \frac{-a\left[(1+\varepsilon)\left(1-\frac{\mu_1}{(1+\alpha)\overline{I}}\right)-1\right]-\frac{\mu_1 a}{(1+\alpha)\overline{I}}}{\sigma_1\sqrt{\frac{a}{(1+\alpha)\overline{I}}+1}}+o(1) \\ &\le \frac{-a\varepsilon\left(1-\frac{\mu_1}{(1+\alpha)\overline{I}}\right)}{\sigma_1\sqrt{\frac{2a}{(1+\alpha)\overline{I}}}}+o(1) \\ &\le -\eta\varepsilon\sqrt{a}, \end{aligned}$$

where

$$\eta \equiv \frac{\alpha}{2\sqrt{2\overline{\psi''}(1+\alpha)/\overline{I}}} > 0,$$

say. Thus $\zeta \le -\eta\varepsilon\sqrt{a} \le -(\eta\varepsilon\sqrt{a} \wedge a^{1/7})$, and

$$(\eta\varepsilon\sqrt{a} \wedge a^{1/7}) \le a^{1/7} = o(a^{1/6}) = o((N_1)^{1/6}),$$

so by large deviations,

$$P_\theta(a_1 > (1+\varepsilon)\xi_1^\alpha a) \le \Phi(-(\eta\varepsilon\sqrt{a} \wedge a^{1/7}))(1+o(1)).$$

The other tail is handled similarly to prove (4.17). □

Lemma 4.7. *If $c \in \mathcal{B}_m(d)$, then under δ_α, for $1 \le k \le m$,*

$$\frac{a_k}{\xi_k^\alpha F_{d/c}^{(k-1)}(a)} \to 1$$

in P_θ-probability as $d \to 0$, uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Proof. The $k=1$ case holds a fortiori by Lemma 4.6. Assume $2 \le k \le m$ and let F^k denote $F_{d/c}^{(k)}(a)$. Fix $0 < \varepsilon < 1$. By Lemma 4.4, $a_{k+1} \le a_k - \ell(N_{k+1}, \hat{\theta}_k) + O(\log a)$,

so

$$
\begin{aligned}
U_{k+1}(\varepsilon) &\subseteq \left\{\ell(N_{k+1},\hat{\theta}_k) < a_k - (1+\varepsilon)\xi_{k+1}^{\alpha}F^k + O(\log a)\right\} \\
&= \left\{\frac{\ell(N_{k+1},\hat{\theta}_k) - \mu_k N_{k+1}}{\sigma_k\sqrt{N_{k+1}}} < \zeta_{k+1}\right\}
\end{aligned}
$$

where

$$
\zeta_{k+1} \equiv \frac{a_k - \mu_k N_{k+1} - (1+\varepsilon)\xi_{k+1}^{\alpha}F^k + O(\log a)}{\sigma_k\sqrt{N_{k+1}}}.
$$

Let $0 < \eta \to 0$ at a rate which will be determined below. Letting primes denote complements, on $U_k'(\varepsilon/10) \cap V_k(\eta)$,

$$
\begin{aligned}
\zeta_{k+1} &= \frac{\sigma_k^*}{\sigma_k}\cdot\frac{a_k - \mu_k^* N_{k+1}}{\sigma_k^*\sqrt{N_{k+1}}} + \sqrt{N_{k+1}}\cdot\frac{\mu_k^* - \mu_k}{\sigma_k} - \frac{(1_\varepsilon)\xi_{k+1}^{\alpha}F^k + O(\log a)}{\sigma_k\sqrt{N_{k+1}}} \\
&= \frac{\sigma_k^*}{\sigma_k}\sqrt{\log a_k/(d/c)^2} + \sqrt{N_{k+1}}\cdot\frac{\mu_k^* - \mu_k}{\sigma_k} - \frac{(1_\varepsilon)\xi_{k+1}^{\alpha}F^k + O(\log a)}{\sigma_k\sqrt{N_{k+1}}} \\
&\leq \frac{\sigma_k^*}{\sigma_k}\sqrt{\log F^{k-1}/(d/c)^2} + O(\sqrt{F^{k-1}})\cdot O(|\mu_k^* - \mu_k|) - \frac{(1+\varepsilon)\xi_{k+1}^{\alpha}F^k}{\sigma_k\sqrt{(1+\varepsilon/10)\xi_k^{\alpha}F^{k-1}/\mu_k^*}} + O(1) \\
&= \frac{\sigma_k^*}{\sigma_k}\sqrt{\log F^{k-1}/(d/c)^2} + O(\sqrt{F^{k-1}}\cdot\eta) \\
&\qquad -(1+\varepsilon)\sqrt{\frac{\mu_k^*}{\sigma_k\xi_k^{\alpha}(1+\varepsilon/10)}}\cdot\xi_{k+1}^{\alpha}\sqrt{F^{k-1}/(d/c)^2} + O(1). \qquad (4.18)
\end{aligned}
$$

Let

$$
\eta = \varepsilon_1\sqrt{\frac{\log F^{k-1}/(d/c)^2}{F^{k-1}}},
$$

where $\varepsilon_1 > 0$ is small enough that the $O(\sqrt{F^{k-1}}\cdot\eta)$ term in (4.18) is less than $(\varepsilon/10)\sqrt{\log F^{k-1}/(d/c)^2}$.

$$
\begin{aligned}
\eta\sqrt{a} &= \varepsilon_1\sqrt{\frac{\log F^{k-1}/(d/c)^2}{F^{k-1}}} \\
&\geq \varepsilon_1\sqrt{\frac{\log F^{k-1}/(d/c)^2}{a}}\cdot\sqrt{a} \quad (\text{since } k \geq 2 \Rightarrow F^{k-1} \ll a) \\
&= \varepsilon_1\sqrt{\log F^{k-1}/(d/c)^2} \to \infty,
\end{aligned}
$$

so $P_\theta(V_k(\eta)) \to 1$ by Lemma 4.5. Since both

$$\frac{\sigma_k^*}{\sigma_k}, \qquad \sqrt{\frac{\mu_k^*}{\sigma_k \xi_k^\alpha}} \xi_{k+1}^\alpha \to 1$$

as $\eta \to 0$, we may assume η is small enough that

$$\frac{\sigma_k^*}{\sigma_k} \leq 1 + \varepsilon/10, \quad \xi_{k+1}^\alpha \sqrt{\frac{\mu_k^*}{\sigma_k \xi_k^\alpha}} \geq 1 - \varepsilon/10.$$

Plugging these and the above bound for the $O(\sqrt{F^{k-1}} \cdot \eta)$ term into (4.18),

$$\begin{aligned}
\zeta_{k+1} &\leq -\sqrt{\log F^{k-1}/(d/c)^2} \left[(1+\varepsilon) \frac{(1-\varepsilon/10)}{\sqrt{1+\varepsilon/10}} - (1+\varepsilon/10) - \varepsilon/10 \right] + 1 \\
&\leq -\sqrt{\log F^{k-1}/(d/c)^2} [(1+\varepsilon)(1-\varepsilon/10)(1-\varepsilon/20) - 1 - \varepsilon/5] + 1 \\
&\leq -(\varepsilon/2)\sqrt{\log F^{k-1}/(d/c)^2} + 1 \to -\infty
\end{aligned}$$

on $U_k'(\varepsilon/10) \cap V_k(\eta)$, hence $P_\theta(U_{k+1}(\varepsilon) \cap U_k'(\varepsilon/10) \cap V_k(\eta)) \to 0$. Then

$$\begin{aligned}
P_\theta(U_{k+1}(\varepsilon)) &= o(1) + P(U_{k+1}(\varepsilon) \cap (U_k'(\varepsilon/10) \cap V_k(\eta))') \\
&\leq o(1) + P_\theta(U_k(\varepsilon/10)) + P_\theta(V_k'(\eta)) = o(1)
\end{aligned}$$

using the induction hypothesis and the fact that $P_\theta(V_k(\eta)) \to 1$. The other tail is handled similarly to show

$$P_\theta\left(\left| \frac{a_{k+1}}{\xi_{k+1}^\alpha F^k} - 1 \right| > \varepsilon \right) \to 0,$$

completing the induction and the proof. □

Lemma 4.8. *If $c \in \mathcal{B}_m(d)$, then there is a function $\gamma = \gamma(d)$ such that*

$$\gamma = \begin{cases} o\left(\frac{d/c}{h_m(a)}\right)^2, & \text{if } c \in \mathcal{B}_m^o(d) \\ O(1), & \text{if } c \in \mathcal{B}_m^+(d) \end{cases} \tag{4.19}$$

and, under δ_α,

$$P_\theta(a_m > \gamma F_{d/c}^{(m-1)}(a)) = o\left(\frac{d/c}{a}\right) \tag{4.20}$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$ *as* $d \to 0$.

Proof. Let F^k denote $F_{d/c}^{(k)}(a)$ and $\tilde{U}_k(x) = \{a_k > xF^{k-1}\}$. We proceed by induction on m. For $m = 1$, since $a_1 \leq a + O(1)$ and $F^0 = a$, taking $\gamma \equiv 2$ gives

$$P_\theta(\tilde{U}_1(\gamma)) \leq P_\theta(a < O(1)) = 0$$

for sufficiently small d, which satisfies (4.19).

Fix $m \geq 2$. We now prove by induction on k that, for $1 \leq k \leq m-1$, there are constants $C_k < \infty$ such that

$$P_\theta(\tilde{U}_k(C_k)) = o\left(\frac{d/c}{a}\right). \tag{4.21}$$

The same argument used in the $m = 1$ case shows that $C_1 \equiv 2$ suffices. Thus assume $2 \leq k+1 \leq m-1$ and that (4.21) holds; we now show it holds with k replaced by $k+1$. Since $a_{k+1} \leq a_k - \ell(N_{k+1}, \hat{\theta}_k) + O(\log a)$ by Lemma 4.4,

$$\begin{aligned} P_\theta(\tilde{U}_{k+1}(C_{k+1})) &\leq P_\theta(\ell(N_{k+1}, \hat{\theta}_k) < a_k - C_{k+1}F^k + O(\log a)) \\ &= P_\theta\left(\frac{\ell(N_{k+1}, \hat{\theta}_k) - \mu_k(\theta)N_{k+1}}{\sigma_k(\theta)\sqrt{N_{k+1}}} < \zeta\right), \end{aligned} \tag{4.22}$$

where

$$\begin{aligned} \zeta &\equiv \frac{a_k - \mu_k(\theta)N_{k+1} - C_{k+1}F^k + O(\log a)}{\sigma_k(\theta)\sqrt{N_{k+1}}} \\ &= \frac{\sigma_k^*}{\sigma_k(\theta)} \frac{a_k - \mu_k^* N_{k+1}}{\sigma_k^* \sqrt{N_{k+1}}} + \sqrt{N_{k+1}} \cdot \frac{\mu_k^* - \mu_k(\theta)}{\sigma_k(\theta)} - \frac{C_{k+1}F^k + O(\log a)}{\sigma_k(\theta)\sqrt{N_{k+1}}} \\ &= \frac{\sigma_k^*}{\sigma_k(\theta)} \sqrt{\log a_k/(d/c)^2} + \sqrt{N_{k+1}} \cdot \frac{\mu_k^* - \mu_k(\theta)}{\sigma_k(\theta)} - \frac{C_{k+1}F^k + O(\log a)}{\sigma_k(\theta)\sqrt{N_{k+1}}}. \end{aligned} \tag{4.23}$$

Let $\varepsilon > 0$ and C_k satisfy (4.21) so that on $\tilde{U}'_k(C_k) \cap V_k(\varepsilon)$,

$$\begin{aligned}
\frac{\sigma_k^*}{\sigma_k(\theta)}\sqrt{\log a_k/(d/c)^2} &= O\left(\sqrt{\log F^{k-1}/(d/c)^2}\right), && (4.24)\\
\sqrt{N_{k+1}} \cdot \frac{\mu_k^* - \mu_k(\theta)}{\sigma_k(\theta)} &\leq \sqrt{a_k/\mu_k^*} \cdot O(\varepsilon) = O(\varepsilon\sqrt{F^{k-1}}), && (4.25)\\
\frac{C_{k+1}F^k + O(\log a)}{\sigma_k(\theta)\sqrt{N_{k+1}}} &\geq \frac{F^k}{\sigma_k(\theta)\sqrt{F^{k-1}}} + o(1) \geq \eta\sqrt{\log F^{k-1}/(d/c)^2}, && (4.26)
\end{aligned}$$

some $\eta > 0$. Now let

$$\varepsilon = C\sqrt{\frac{\log F^{k-1}/(d/c)^2}{F^{k-1}}}$$

where $C < \infty$ will be determined below. By Lemma 4.5 there exists $\eta_o > 0$ such that $P_\theta(V'_k(\varepsilon)) \leq 2\exp(-\eta_o\varepsilon^2 a)$ and

$$\eta_o\varepsilon^2 a \geq \eta_o C^2 \frac{\log[F^{k-1}/(d/c)^2]}{F^{k-1}} \cdot a \geq \eta_o C^2 \log[F^{k-1}/(d/c)^2].$$

Furthermore,

$$\begin{aligned}
\frac{F^{k-1}}{(d/c)^2} &\geq \frac{F^{m-3}}{(d/c)^2} \quad (\text{since } k-1 \leq m-3)\\
&\geq \eta'_o \cdot \frac{h_{m-2}(a)^2}{(d/c)^2} \quad (\text{some } \eta'_o > 0 \text{ by Lemma 2.6})\\
&\geq \eta'_o \cdot \frac{h_{m-2}(a)^2}{h_{m-1}(a)^2} \geq a^{(1/2)^{m-1}} && (4.27)
\end{aligned}$$

for sufficiently small d, so that

$$\eta_o\varepsilon^2 a \geq \eta_o C^2 (1/2)^{m-1}\log a \geq \log a$$

by choosing C sufficiently large. Then

$$P_\theta(V'_k(\varepsilon)) \leq 2\exp(-\log a) = 2a^{-1} = o\left(\frac{d/c}{a}\right). \qquad (4.28)$$

Plugging the estimates (4.24)-(4.26) into (4.23), on $\tilde{U}'_k(C_k) \cap V_k(\varepsilon)$

$$\begin{aligned}
\zeta &\leq O\left(\sqrt{\log F^{k-1}/(d/c)^2}\right) + O\left(\sqrt{\log F^{k-1}/(d/c)^2}\right) - C_{k+1}\eta\sqrt{\log F^{k-1}/(d/c)^2} \\
&\leq -\sqrt{\log F^{k-1}/(d/c)^2} \cdot (C_{k+1}\eta - O(1)) \\
&\leq -\sqrt{\log a^2} \qquad (4.29)
\end{aligned}$$

by taking C_{k+1} sufficiently large and using (4.27). Then, using (4.21), (4.28), and a large deviations argument,

$$\begin{aligned}
&P_\theta(\tilde{U}_{k+1}(C_{k+1})) \leq P_\theta(\tilde{U}_{k+1}(C_{k+1}) \cap \tilde{U}'_k(C_k) \cap V_k(\varepsilon)) + P_\theta(\tilde{U}_k(C_k)) + P(V'_k(\varepsilon)) \\
&\leq P\left(\left\{\frac{\ell(N_{k+1}, \hat{\theta}_k) - \mu_k(\theta)N_{k+1}}{\sigma_k(\theta)\sqrt{N_{k+1}}} < -\sqrt{\log a^2}\right\} \cap \tilde{U}'_k(C_k) \cap V_k(\varepsilon)\right) + o\left(\frac{d/c}{a}\right) \\
&\leq O\left(\Phi(-\sqrt{\log a^2})\right) + o\left(\frac{d/c}{a}\right),
\end{aligned}$$

and, using Mill's Ratio,

$$\Phi(-\sqrt{\log a^2}) \sim \frac{\phi(\sqrt{\log a^2})}{\sqrt{\log a^2}} = O\left(\frac{a^{-1}}{\sqrt{\log a^2}}\right) = o\left(\frac{d/c}{a}\right), \qquad (4.30)$$

so $P_\theta(\tilde{U}_{k+1}(C_{k+1})) = o((d/c)/a)$. This completes the induction to prove (4.21), which we now use to prove (4.20).

If there exists $\beta > 0$ such that $a^\beta = o(h_{m-1}(a)/(d/c))$, which holds when $c \in \mathcal{B}_m^+(d)$, then

$$\frac{F^{m-2}}{(d/c)^2} \geq a^{2\beta}$$

for sufficiently small d by the argument used in (4.27), and (4.20) holds with γ a large constant as in the $m = 1$ case. Otherwise, by considering subsequences there exists $\varepsilon_o > 0$ such that $a^{(1/2)^{m+2}} \geq \varepsilon_o \cdot h_{m-1}(a)/(d/c)$ for sufficiently small d. Then

$$\frac{d/c}{h_m(a)} = \frac{d/c}{h_{m-1}(a)} \cdot \frac{h_{m-1}(a)}{h_m(a)} \geq \frac{\varepsilon_o}{a^{(1/2)^{m+2}}} \cdot a^{(1/2)^{m+1}} = \varepsilon_o a^{(1/2)^{m+2}}$$

and hence

$$a^{(1/2)^{m+2}} = o\left(\frac{d/c}{h_m(a)}\right)^2. \tag{4.31}$$

By the induction hypothesis, let C_{m-1} satisfy $P(\tilde{U}_{m-1}(C_{m-1})) = o((d/c)/a)$, and let

$$\gamma = K\sqrt{\frac{\log a}{\log F^{m-2}/(d/c)^2}}, \quad \varepsilon = K'\sqrt{\frac{\log a}{F^{m-2}}},$$

where K, K' will be determined below. By the argument leading to (4.22)

$$P(\tilde{U}_m(\gamma)) \leq P\left(\frac{\ell(N_m, \hat{\theta}_{m-1}) - \mu_{m-1}(\theta)N_m}{\sigma_{m-1}(\theta)\sqrt{N_m}} < \frac{a_{m-1} - \mu_{m-1}(\theta)N_m - \gamma F^{m-1} + O(\log a)}{\sigma_{m-1}(\theta)\sqrt{N_m}}\right)$$

and

$$\begin{aligned}\frac{a_{m-1} - \mu_{m-1}(\theta)N_m - \gamma F^{m-1} + O(\log a)}{\sigma_{m-1}(\theta)\sqrt{N_m}} &\leq O\left(\sqrt{\log F^{m-2}/(d/c)^2}\right) + O\left(\sqrt{\log a}\right) - K\eta'\sqrt{\log a} \\ &= -\sqrt{\log a}\cdot(K\eta' - O(1)),\end{aligned}$$

some $\eta' > 0$, by the argument leading to (4.29). Hence, taking K sufficiently large, we obtain

$$\begin{aligned}P(\tilde{U}_m(\gamma)) &\leq P(\tilde{U}_m(\gamma) \cap \tilde{U}'_{m-1}(C_{m-1}) \cap V_{m-1}(\varepsilon)) + P(\tilde{U}_{m-1}(C_{m-1}) + P(V'_{m-1}(\varepsilon)) \\ &\leq \Phi(-\sqrt{\log a^2}) + o\left(\frac{d/c}{a}\right) + P(V'_{m-1}(\varepsilon)) \\ &= o\left(\frac{d/c}{a}\right) + P(V'_{m-1}(\varepsilon)) \end{aligned} \tag{4.32}$$

by (4.30). By choosing K' sufficiently large and repeating the argument leading to (4.28),

$$P(V'_{m-1}(\varepsilon)) = o\left(\frac{d/c}{a}\right).$$

Plugging this back into (4.32) gives $P(\tilde{U}_m(\gamma) = o((d/c)/a)$, and all that remains is to verify that γ satisfies the first case of (4.19). But

$$\gamma = o(\sqrt{\log a}) = o(a^{(1/2)^{m+2}}) = o\left(\frac{d/c}{h_m(a)}\right)^2$$

by (4.31), finishing the proof. □

Next we establish the operating characteristics of δ when $c \in \mathcal{B}_m^o(d)$.

Theorem 4.9. *If $c \in \mathcal{B}_m^o(d)$, then*

$$r(\lambda_0, \delta) \leq caE_{\lambda_0} I(\theta)^{-1} + d(m+1) + o(d). \tag{4.33}$$

Proof. Let $(N, M, D) = \delta_\alpha$, $\alpha > 0$. We will prove

$$E_\theta[cN + dM + r(\lambda_M)] \leq c \log d^{-1}/I(\theta) + d(m+1) + o(d) \tag{4.34}$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Fix $\theta \in [\underline{\theta}, \overline{\theta}]$. First we show that

$$E_\theta N \leq a/I(\theta) + o(d/c). \tag{4.35}$$

Write $E_\theta N = E_\theta(N; M \leq m) + E_\theta(N; M \geq m+1)$ and consider $E_\theta(N; M = k)$ for $1 < k \leq m$. Letting $z_k = -\sqrt{\log(a_{k-1}/(d/c)^2 + 1)}$,

$$N_k \leq \frac{a_{k-1}}{\mu^*_{k-1}} - \left[\frac{z_k \sigma^*_{k-1}\sqrt{4a_{k-1}\mu^*_{k-1} + z_k^2\sigma^{*2}_{k-1}} - z_k^2\sigma^{*2}_{k-1}}{2\mu^{*2}_{k-1}}\right] \tag{4.36}$$

$$\leq \frac{a_{k-1}}{\mu^*_{k-1}} \leq \frac{a - \ell_{k-1} + O(1)}{\mu^*_{k-1}} \quad \text{(by Lemma 4.2)} \tag{4.37}$$

$$= \frac{a - N_{k-1}\mu^*_{k-1} + O(1)}{\mu^*_{k-1}} = \frac{a + O(1)}{\mu^*_{k-1}} - N^{k-1}, \tag{4.38}$$

so

$$\begin{aligned} E_\theta(N; M=k) &= E_\theta(N^{k-1} + N_k; M=k) \\ &\leq E_\theta((a+O(1))/\mu^*_{k-1}; M=k) \\ &= E_\theta(a/\mu^*_{k-1}; M=k) + O(E_\theta(\mu^{*-1}_{k-1}; M=k)) \\ &\leq E_\theta(a/\mu^*_{k-1}; M=k) + O(E_\theta(\mu^{*-1}_{k-1})) \quad \text{(since } \mu^*_{k-1} > 0\text{)} \\ &= E_\theta(a/\mu^*_{k-1}; M=k) + O(I(\theta)^{-1}) \quad \text{(by Lemma 4.1)} \\ &= E_\theta(a/\mu^*_{k-1}; M=k) + O(1) = E_\theta(a/\mu^*_{k-1}; M=k) + o(d/c) \end{aligned}$$

for $1 < k \leq m$. Also

$$E_\theta(N; M=1) = N_1 P_\theta(M=1) \leq O(a)O(\Phi(-a^{1/7})) = o(1) = o(d/c),$$

so we have

$$E_\theta(N; M \leq m) \leq \begin{cases} aE_\theta(\mu^{*-1}_{M-1}; 2 \leq M \leq m) + o(d/c), & m \geq 2 \\ o(d/c), & m = 1. \end{cases} \tag{4.39}$$

Let $z \to -\infty$ be the quantile chosen for geometric sampling, which satisfies $|z| = o((d/c)/h_m(a))$. For $k \geq m+1$, let

$$\begin{aligned} \Lambda_k &\equiv N^k - a/\mu^*_{k-1} \\ &= N^{k-1} + N_k - a/\mu^*_{k-1} \\ &= N^{k-1} + \left[\frac{a_{k-1}}{\mu^*_{k-1}} + \frac{|z|\sigma^*_{k-1}\sqrt{4a_{k-1}\mu^*_{k-1} + z^2\sigma^{*2}_{k-1}} - z^2\sigma^{*2}_{k-1}}{2\mu^{*2}_{k-1}} \right] - a/\mu^*_{k-1} \\ &= N^{k-1} + \left[\frac{a+O(1)}{\mu^*_{k-1}} - N^{k-1} + \frac{|z|\sigma^*_{k-1}\sqrt{4a_{k-1}\mu^*_{k-1} + z^2\sigma^{*2}_{k-1}} - z^2\sigma^{*2}_{k-1}}{2\mu^{*2}_{k-1}} \right] - a/\mu^*_{k-1} \\ &= \frac{|z|\sigma^*_{k-1}\sqrt{4a_{k-1}\mu^*_{k-1} + z^2\sigma^{*2}_{k-1}} - z^2\sigma^{*2}_{k-1}}{2\mu^{*2}_{k-1}} + O(1)/\mu^*_{k-1}, \end{aligned} \tag{4.40}$$

this last by the argument leading to (4.38). Then

$$\begin{aligned} E_\theta(N; M \geq m+1) &= \sum_{k \geq m+1} E_\theta(N^k; M = k) \\ &= aE_\theta(\mu^{*-1}_{M-1}; M \geq m+1) + \sum_{k \geq m+1} E_\theta(\Lambda_k; M = k), \end{aligned} \tag{4.41}$$

and we now estimate the summands in the latter term. Let F^{m-1} denote $F^{(m-1)}_{d/c}(a)$ and let γ be the function given by Lemma 4.8 such that $P_\theta(a_m > \gamma F^{m-1}) = o((d/c)/a)$ and $\gamma = o((d/c)/h_m(a))^2$. Since we may assume without loss of generality that $|z| \to \infty$ arbitrarily slowly, assume

$$|z| = o\left(\frac{d/c}{h_m(a)\sqrt{\gamma}}\right). \tag{4.42}$$

Then, using (4.40) and the crude bound $\Lambda_k \leq N^k \leq \overline{n}$,

$$\begin{aligned} E_\theta(\Lambda_{m+1}; M = m+1) &\leq E_\theta(\Lambda_{m+1}; M \geq m+1) = E_\theta(\Lambda_{m+1}; a_m > 0) \\ &= E_\theta(\Lambda_{m+1}; 0 < a_m \leq \gamma F^{m-1}) + E_\theta(\Lambda_{m+1}; a_m > \gamma F^{m-1}) \\ &\leq O(|z|\sqrt{\gamma F^{m-1}}) + \overline{n} P_\theta(a_m > \gamma F^{m-1}) \\ &= O(|z|\sqrt{\gamma} h_m(a)) + O(a) o\left(\frac{d/c}{a}\right) \quad \text{(by Lemma 2.6)} \\ &= o(d/c) + o(d/c) = o(d/c) \end{aligned}$$

by (4.42). Note that

$$E_\theta(\Lambda_{m+1} | M \geq m+1) = \frac{E_\theta(\Lambda_{m+1}; M \geq m+1)}{P_\theta(M \geq m+1)} = \frac{o(d/c)}{1 - o(1)} = o(d/c).$$

Assume that there exists $q \to 0$ such that, for $k \geq 1$,

$$P_\theta(M \geq m+1+k) \leq q^k. \tag{4.43}$$

Since Λ_{m+1+k} are stochastically decreasing in k,

$$\begin{aligned} E_\theta(\Lambda_{m+1+k}; M = m+1+k) &\leq E_\theta(\Lambda_{m+1+k}; M \geq m+1+k) \\ &= E_\theta(\Lambda_{m+1+k} | M \geq m+1+k) P_\theta(M \geq m+1+k) \\ &\leq E_\theta(\Lambda_{m+1} | M \geq m+1) P_\theta(M \geq m+1+k) \\ &\leq o(d/c) q^k \end{aligned}$$

and the $o(d/c)$ term is independent of k, so that

$$\sum_{k \geq m+1} E_\theta(\Lambda_k; M = k) \leq o(d/c) \sum_{k \geq 0} q^k = \frac{o(d/c)}{1-q} = o(d/c).$$

Plugging this back into (4.41) gives

$$E_\theta(N; M \geq m+1) \leq a E_\theta(\mu_{M-1}^{*-1}; M \geq m+1) + o(d/c)$$

and combining this with (4.39) yields

$$\begin{aligned} E_\theta N &= E_\theta(N; M \leq m) + E_\theta(N; M \geq m+1) \\ &\leq a E_\theta(\mu_{M-1}^{*-1}; M \geq 2) + o(d/c) \\ &\leq a(I(\theta)^{-1} + O(1/a)) + o(d/c) \quad \text{(by Lemma 4.1)} \\ &= a/I(\theta) + o(d/c). \end{aligned}$$

To estimate the number of stages, M, note that if (4.43) holds,

$$\begin{aligned} E_\theta M &= \sum_{k \geq 0} P_\theta(M > k) \leq m + 1 + \sum_{k \geq 1} P_\theta(M > m + k) \\ &\leq m + 1 + \sum_{k \geq 1} q^k \\ &= m + 1 + \frac{q}{1-q} = m + 1 + o(1). \end{aligned} \tag{4.44}$$

We now prove (4.43) by induction. Let $\eta > 0$, to be specified below.

$$\begin{aligned} P_\theta(M \geq m+2) &\leq 1 - P_\theta(M = m+1) \\ &\leq 1 - P_\theta(\{M = m+1\} \cap U_m(1/2) \cap V_m(\eta)) \\ &\leq 1 - P_\theta(M = m+1 | U_m(1/2) \cap V_m(\eta))(P_\theta(U_m(1/2)) - P_\theta(V_m(\eta)')). \end{aligned} \quad (4.45)$$

Let $\mu_m(\theta) = I_\theta(\hat\theta_m)$, $\sigma_m^2(\theta) = (\hat\theta_m - \hat\theta'_m)^2\psi''(\theta)$, and

$$\rho_k(\theta) = \frac{\ell(N_{k+1}, \hat\theta_k) - \mu_k(\theta)N_{k+1}}{\sigma_k(\theta)\sqrt{N_{k+1}}}.$$

Since $a_{m+1} \leq a_m - \ell(N_{m+1}, \hat\theta_m) + O(\log a)$ by Lemma 4.4, we have

$$\begin{aligned} P_\theta(M = m+1 | U_m(1/2) \cap V_m(\eta)) &= P_\theta(a_m \leq 0 | U_m(1/2) \cap V_m(\eta)) \\ &\geq P_\theta(\ell(N_{m+1}, \hat\theta_m) \geq a_m + O(\log a)) \\ &= P_\theta\left(\rho_m(\theta) \geq \frac{a_m - \mu_m(\theta)N_{m+1} + O(\log a)}{\sigma_m(\theta)\sqrt{N_{m+1}}} \middle| U_m(1/2) \cap V_m(\eta)\right). \end{aligned}$$

Then

$$\begin{aligned} \zeta_m &\equiv \frac{a_m - \mu_m(\theta)N_{m+1} + O(\log a)}{\sigma_m(\theta)\sqrt{N^{m+1}}} \\ &= \frac{\sigma_m^*}{\sigma_m(\theta)}\left[\frac{a_m - \mu_m^* N_{m+1}}{\sigma_m^*\sqrt{N^{m+1}}}\right] + \sqrt{N^{m+1}}\frac{\mu_m^* - \mu_m(\theta)}{\sigma_m(\theta)} + O\left(\frac{\log a}{\sqrt{N^{m+1}}}\right) \\ &\leq (1 + O(\eta))z + \sqrt{N^{m+1}}O(\eta) + o(1) \\ &\leq -(3/4)|z| + O(\eta\sqrt{F^{m-1}}) \end{aligned} \quad (4.46)$$

on $U_m(1/2) \cap V_m(\eta)$ for sufficiently small η, since $z \to -\infty$. Choosing

$$\eta = \varepsilon_1(|z|/\sqrt{F^{m-1}} \wedge 1),$$

where ε_1 is small enough so that the $O(\eta\sqrt{F^{m-1}})$ term in (4.46) is less than $|z|/4$, we have

$$\zeta_m \leq -(3/4)|z| + |z|/4 = -|z|/2$$

on $U_m(1/2) \cap V_m(\eta)$, and therefore

$$P_\theta(M = m+1 | U_m(1/2) \cap V_m(\eta)) \geq P_\theta(\rho_m(\theta) \geq -|z|/2 | U_m(1/2) \cap V_m(\eta)) \to 1. \quad (4.47)$$

We know $P_\theta(U_m(1/2)) \to 1$ by Lemma 4.7 and since

$$\eta\sqrt{a} = \varepsilon_1 \left(\frac{|z|}{\sqrt{F^{m-1}}} \wedge 1 \right) \sqrt{a} \geq \varepsilon_1 \left(\frac{|z|}{\sqrt{a}} \wedge 1 \right) \sqrt{a} = \varepsilon_1(|z| \wedge \sqrt{a}) \to \infty,$$

$P_\theta(V_m(\eta)') \to 0$ by Lemma 4.5. Letting

$$q = \frac{1 - P_\theta(M = m+1 | U_m(1/2) \cap V_m(\eta))(P_\theta(U_m(1/2)) - P_\theta(V_m(\eta)'))}{P_\theta(M \geq m+1)} = \frac{o(1)}{1 - o(1)} \to 0,$$

by virtue of these last estimates, we have

$$\begin{aligned} P_\theta(M \geq m+2 | M \geq m+1) = \frac{P_\theta(M \geq m+2)}{P_\theta(M \geq m+1)} & \\ \leq \quad \frac{1 - P_\theta(M = m+1 | U_m(1/2) \cap V_m(\eta))(P_\theta(U_m(1/2)) - P_\theta(V_m(\eta)'))}{P_\theta(M \geq m+1)} & \\ = \quad q, & \qquad (4.48) \end{aligned}$$

and so, a fortiori, $P_\theta(M \geq m+2) \leq q$.

Now suppose $k \geq 2$. Using the induction hypothesis,

$$\begin{aligned} P_\theta(M \geq m+1+k) &= P_\theta(M \geq m+1+k | M \geq m+k) P_\theta(M \geq m+k) \\ &\leq P_\theta(M = m+1+k | M \geq m+k) q^{k-1}, \end{aligned}$$

and the argument used in the $m = 1$ case, replacing $U_m(1/2)$ by

$$\tilde{U} = \{a_{m+1+k} \leq (3/2)\xi_m^\alpha F^{m-1}\},$$

gives $P_\theta(M = m+1+k | M \geq m+k) \leq q$, whence $P_\theta(M \geq m+1+k) \leq q^k$, proving (4.43).

Finally, we show that

$$E_\theta r(\lambda_M) = o(d). \tag{4.49}$$

Recall that $r(\lambda_M) \le d$ uniformly and $r(\lambda_M) \le d^2$ on $\{N = \overline{n}\}$. Let $\gamma_1 \to \infty$ be any function such that

$$\log a \ll \gamma_1 \ll h_m(a) \tag{4.50}$$

and define

$$\begin{aligned} W &\equiv V_m(\eta) \cap \left\{ \left| \frac{a_m}{\xi_m^\alpha F^{m-1}} - 1 \right| \le 1/2 \right\} \cap \{M = m+1\} \cap \{r(\lambda_M) \le e^{-\gamma_1} d\} \\ &\equiv V_m(\eta) \cap \bigcap_{i=1}^{3} W_i. \end{aligned}$$

Obviously $r(\lambda_M) = o(d)$ on W_3, so

$$\begin{aligned} E_\theta r(\lambda_M) &= E_\theta(r(\lambda_M); N = \overline{n}) + E_\theta(r(\lambda_M); W \cap \{N < \overline{n}\}) + E_\theta(r(\lambda_M); W' \cap \{N < \overline{n}\}) \\ &\le d^2 \cdot 1 + o(d) \cdot 1 + d \cdot P_\theta(W') = o(d) + d \cdot P_\theta(W'), \end{aligned}$$

and (4.49) will be established once we show $P_\theta(W') \to 0$. We know $P_\theta(W_1) \to 1$ by Lemma 4.7 and it was shown that $P_\theta(W_2) \to 1$. We will choose η below in such a way that $P_\theta(V_m(\eta)) \to 1$. Let $\tilde{W} = \{\ell(N_{m+1}, \hat{\theta}_m) \ge a_m + 2\gamma_1\}$. On $\tilde{W}$,

$$\begin{aligned} r(\lambda_{m+1}) &= \exp(-\log r(\lambda_{m+1})^{-1}) \le \exp(-\ell_{m+1} + O(1)) \quad \text{(by Lemma 4.2)} \\ &\le \exp(-\ell(N_{m+1}, \hat{\theta}_m) - \ell_m + O(1)) \\ &\le \exp[-(a_m + 2\gamma_1) - (\log r(\lambda_m)^{-1} + O(\log a)) + O(1)] \quad \text{(by Lemma 4.2)} \\ &= \exp[-a - 2\gamma_1 - O(\log a)] \\ &\le \exp(-a - \gamma_1) = e^{-\gamma_1} d \end{aligned}$$

by (4.50), hence $\tilde{W} \cap V_m(\eta) \cap W_1 \cap W_2 \subseteq W$. Then

$$\begin{aligned} P_\theta(W_3|V_m(\eta) \cap W_1 \cap W_2) &\geq P_\theta(\tilde{W}|V_m(\eta) \cap W_1 \cap W_2) \\ &= P_\theta\left(\rho_m(\theta) \geq \frac{a_m - \mu_m(\theta)N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} + \frac{2\gamma_1}{\sigma_m(\theta)\sqrt{N_{m+1}}} \middle| V_m(\eta) \cap W_1 \cap W_2\right) \quad (4.51) \end{aligned}$$

and on $V_m(\eta) \cap W_1 \cap W_2$,

$$\begin{aligned} \zeta &\equiv \frac{a_m - \mu_m(\theta)N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} + \frac{2\gamma_1}{\sigma_m(\theta)\sqrt{N_{m+1}}} \\ &= \frac{\sigma_m^*}{\sigma_m(\theta)}z + \sqrt{N_{m+1}}\frac{\mu_m^* - \mu_m(\theta)}{\sigma_m(\theta)} + \frac{2\gamma_1}{\sigma_m(\theta)\sqrt{N_{m+1}}} \\ &\leq (1 + O(\eta))z + O(\sqrt{F^{m-1}})O(\eta) + O(\gamma_1/\sqrt{F^{m-1}}) \\ &\leq z/2 + O(\eta\sqrt{F^{m-1}}) + O(\gamma_1/h_m(a)) \quad \text{(by Lemma 2.6)} \\ &= z/2 + O(\eta\sqrt{F^{m-1}}) + o(1) \end{aligned}$$

by (4.50). Taking $\eta = \varepsilon_1(|z|/\sqrt{F^{m-1}} \wedge 1)$ and using the same argument as above (see what follows (4.46)) we obtain $\zeta \leq z/3 \to -\infty$ and $P_\theta(V_m(\eta)) \to 1$. Plugging this back in to (4.51), we have

$$P_\theta(W_3|V_m(\eta) \cap W_1 \cap W_2) \geq P_\theta(\rho_m(\theta) \geq z/3|V_m(\eta) \cap W_1 \cap W_2) \to 1$$

and therefore $P_\theta(W) = P_\theta(W_3|V_m(\eta) \cap W_1 \cap W_2)P_\theta(V_m(\eta) \cap W_1 \cap W_2) \to 1$, establishing (4.49).

Combining (4.35), (4.44), and (4.49) we have

$$\begin{aligned} E_\theta[cN + dM + r(\lambda_M)] &\leq c[a/I(\theta) + o(d/c)] + d[m + 1 + o(1)] + o(d) \\ &= ca/I(\theta) + d(m+1) + o(d) \end{aligned}$$

uniformly in θ, and hence

$$r(\lambda_0, \delta_\alpha) \leq caE_{\lambda_0}I(\theta)^{-1} + d(m+1) + o(d).$$

This holds for all $\alpha > 0$, so by a standard asymptotic technique (e.g., [6], p. 188), there is a function $\alpha(d) \to 0$ for which it holds. Taking $\delta = \delta_{\alpha(d)}$ gives (4.34). □

Next we consider the boundary case. Let $\Delta(z) \equiv \phi(z) - \Phi(-z)z$. Let $\alpha, Q > 0$ and recall from (4.14) that $z^\alpha(\theta, Q)$ is the unique solution of

$$\frac{\Phi(-z^\alpha(\theta,Q))}{\phi(z^\alpha(\theta,Q))} = \frac{QI(\theta)C_m^m}{\xi_{m+1}^\alpha(\theta)}.$$

Let

$$u_m^\alpha(\theta,Q) \equiv m + 1 + \Phi(z^\alpha(\theta,Q)) + \frac{\Delta(z^\alpha(\theta,Q))\xi_{m+1}^\alpha(\theta)}{C_m^m I(\theta) Q}.$$

Observe that if θ is such that $I(\theta) < \overline{I}$ (which can only fail at $\theta = \underline{\theta}$ or $\overline{\theta}$), then $\xi_{m+1}^0(\theta) > 0$ and so $z^0(\theta,Q)$ and hence $u_m^0(\theta,Q)$ are well-defined. If $I(\theta) = \overline{I}$, then $\xi_{m+1}^0(\theta) = 0$ so $z^0(\theta,Q)$ and hence $u_m^0(\theta,Q)$ are not well-defined, but

$$\begin{aligned}
\lim_{\alpha\to 0} u_m^\alpha(\theta,Q) &= m + 1 + \lim_{\alpha\to 0}\left[\Phi(z^\alpha(\theta,Q)) + \Delta(z^\alpha(\theta,Q))\frac{\phi(z^\alpha(\theta,Q))}{\Phi(-z^\alpha(\theta,Q))}\right] \\
&\quad \text{(by definition of } z^\alpha(\theta,Q)) \\
&= m + 1 + \lim_{x\to-\infty}\Phi(x) + \lim_{x\to-\infty}\Delta(x)\frac{\phi(x)}{\Phi(-x)} \\
&= m + 1 + 0 + \lim_{x\to-\infty}|x|\frac{\phi(x)}{1} = m + 1 \qquad (4.52)
\end{aligned}$$

since $\Delta(x) \sim |x|$ as $x \to -\infty$. Thus, replacing $u_m^\alpha(\theta,Q)$ by its limit in this singular case, we define

$$u_m(\theta,Q) \equiv \lim_{\alpha\to 0} u_m^\alpha(\theta,Q) = \begin{cases} u_m^0(\theta,Q), & \text{for } \theta \text{ such that } I(\theta) < \overline{I} \\ m+1, & \text{for } \theta \text{ such that } I(\theta) = \overline{I} \end{cases} \qquad (4.53)$$

for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Next we establish the operating characteristics of δ in the boundary case.

Theorem 4.10. *Assume that* $c \in \mathcal{B}_m^+(d)$ *and let* $Q = \lim_{d\to 0}(d/c)/h_m(a) \in (0,\infty)$.

There is a function $\alpha(d) \to 0$ such that $\delta \equiv \delta_{\alpha(d)}$ satisfies

$$r(\lambda_0, \delta) \leq caE_{\lambda_0} I(\theta)^{-1} + d \cdot E_{\lambda_0} u_m(\theta, Q) + o(d). \tag{4.54}$$

Proof. Let $(N, M, D) = \delta_\alpha$, $\alpha > 0$. We will show that

$$E_\theta[cN + dM + r(\lambda_M)] \leq ca/I(\theta) + d \cdot u_m^\alpha(\theta, Q) + o(d)$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$.

Fix $\theta \in [\underline{\theta}, \overline{\theta}]$. First we show that

$$E_\theta N \leq a/I(\theta) + \frac{\Delta(z^\alpha(\theta, Q))\xi^\alpha_{m+1}(\theta)}{C^m_m I(\theta) Q}(d/c) + o(d/c).$$

We can write

$$E_\theta N = \sum_{k \geq 1} E_\theta(N_k; M \geq k) = \sum_{k \leq m+1} E_\theta(N_k; M \geq k) + \sum_{k > m+1} E_\theta(N_k; M \geq k)$$

and

$$\begin{aligned}
\sum_{k \leq m+1} E_\theta(N_k; M \geq k) &= \sum_{k \leq m} [E_\theta(N_k; k \leq M \leq m) + E_\theta(N_k; M \geq m+1)] \\
&\quad + E_\theta(N_{m+1}; M \geq m+1) \\
&= \sum_{k \leq m} E_\theta(N_k; M \leq m) + E_\theta(N^{m+1}; M \geq m+1) \\
&= E_\theta(N; M \leq m) + E_\theta(N^{m+1}; M \geq m+1) \\
&\leq aE_\theta(\mu^{*-1}_{(M-1 \wedge 1)}; M \leq m) + O(1) + E_\theta(N^{m+1}; M \geq m+1) \\
&\leq aE_\theta(\mu^{*-1}_{(M-1 \wedge 1)}; M \leq m) + E_\theta(N^{m+1}; M \geq m+1) + o(h_m(a))
\end{aligned} \tag{4.55}$$

by the argument leading to (4.39). Also, by the argument leading to (4.38),

$$\begin{aligned} N^{m+1} &\leq \frac{a+O(1)}{\mu_m^*} + \frac{z_m^\alpha(Q)\sigma_m^*\sqrt{4a_m\mu_m^* + z_m^\alpha(Q)^2\sigma_m^{*2}} - z_m^\alpha(Q)^2\sigma_m^{*2}}{2\mu_m^{*2}} \\ &\equiv \frac{a+O(1)}{\mu_m^*} + Y, \end{aligned} \tag{4.56}$$

say. Choose $\varepsilon > 0$. Let $\sigma^2(\theta) = (\theta - \theta')^2\psi''(\theta)$ and $\varepsilon_o = \varepsilon[2 + \overline{z^\alpha(\theta, Q)\sigma(\theta)I(\theta)^{-3/2}}]^{-1}$, recalling that $\overline{g} = \max_{[\underline{\theta},\overline{\theta}]} g(\theta)$. Let F^{m-1} denote $F_{d/c}^{(m-1)}(a)$, and let

$$U(\varepsilon_o) = \left\{ \left| \frac{a_m}{\xi_m^\alpha F^{m-1}} - 1 \right| \leq \varepsilon_o \right\}$$

and $A = U(\varepsilon_o) \cap V_m(\eta_o) \cap \{M \geq m+1\}$, where $\eta_o > 0$ will be determined below. Since $z_m^\alpha(Q), \mu_m^*, \sigma_m^*$ approach $z^\alpha(\theta, Q), I(\theta), \sigma(\theta)$ as $\eta_o \to 0$, it follows that

$$Y \leq \frac{z^\alpha(\theta, Q)\sigma(\theta)\sqrt{4a_m I(\theta) + z^\alpha(\theta, Q)^2\sigma(\theta)^2} + z^\alpha(\theta, Q)^2\sigma(\theta)^2}{2I(\theta)^2} + O(\eta_o)\sqrt{a_m}$$

on A. Since $\sqrt{a_m} \leq \sqrt{(1+\varepsilon_o)\xi_m^\alpha F^{m-1}} = O(\sqrt{F^{m-1}}) = O(h_m(a))$ on $U(\varepsilon_o)$, by taking η_o sufficiently small we may assume

$$Y \leq \frac{z^\alpha(\theta, Q)\sigma(\theta)\sqrt{4a_m I(\theta) + z^\alpha(\theta, Q)^2\sigma(\theta)^2} + z^\alpha(\theta, Q)^2\sigma(\theta)^2}{2I(\theta)^2} + (\varepsilon_o/2)h_m(a)$$

on A. Using (4.56),

$$\begin{aligned}
E_\theta(N^{m+1};A) &\le E_\theta[(a+O(1))/\mu_m^* + Y; A] \\
&\le aE_\theta(\mu_m^{*-1};A) + O(1) \\
&\quad + E_\theta\left(\frac{-z^\alpha(\theta,Q)\sigma(\theta)\sqrt{4a_m I(\theta) + z^\alpha(\theta,Q)^2\sigma(\theta)^2}}{2I(\theta)^2}; A\right) + (\varepsilon_o/2)h_m(a) \\
&\le aE_\theta(\mu_m^{*-1};A) + \frac{-z^\alpha(\theta,Q)\sigma(\theta)}{I(\theta)^{3/2}}\sqrt{(1-\mathrm{sign}(z^\alpha(\theta,Q)))\varepsilon_o\xi_m^\alpha(\theta)F^{m-1}} + \varepsilon_o h_m(a) \\
&\le aE_\theta(\mu_m^{*-1};A) + \frac{-z^\alpha(\theta,Q)\sigma_m(\theta)}{I(\theta)^{3/2}}\sqrt{\xi_m^\alpha(\theta)F^{m-1}} + \varepsilon_o[1+|z^\alpha(\theta,Q)|\sigma(\theta)I(\theta)^{-3/2}]h_m(a) \\
&\le aE_\theta(\mu_m^{*-1};A) + \frac{-z^\alpha(\theta,Q)\xi_m^\alpha(\theta)}{I(\theta)C_m^m}h_m(a) + \varepsilon_o[2+|z^\alpha(\theta,Q)|\sigma(\theta)I(\theta)^{-3/2}]h_m(a) \\
&\le aE_\theta(\mu_m^{*-1};A) + \frac{-z^\alpha(\theta,Q)\xi_m^\alpha(\theta)}{I(\theta)C_m^m}h_m(a) + \varepsilon h_m(a), \qquad (4.57)
\end{aligned}$$

by our choice of ε_o.

Again using (4.56),

$$\begin{aligned}
E_\theta(N^{m+1};U(\varepsilon_o)'\cap V_m(\eta_o)\cap\{M\ge m+1\}) &\le aE_\theta(\mu_m^{*-1};U(\varepsilon_o)'\cap V_m(\eta_o)\cap\{M\ge m+1\}) \\
&\quad + O(1) + E_\theta(O(\sqrt{a_m});U(\varepsilon_o)'\cap V_m(\eta_o)\cap\{M\ge m+1\}).
\end{aligned}$$

Letting C be the constant given by Lemma 4.8 such that

$$P_\theta(a_m > C\xi_m^\alpha F^{m-1}) = o\left(\frac{d/c}{a}\right) = o(h_m(a)/a), \qquad (4.58)$$

we have

$$\begin{aligned}
&E_\theta(\sqrt{a_m};U(\varepsilon_o)'\cap V_m(\eta_o)\cap\{M\ge m+1\}) \\
&\quad = E_\theta[\sqrt{a_m};(\{a_m\le C\xi_m^\alpha F^{m-1}\}\setminus U(\varepsilon_o))\cap V_m(\eta_o)\cap\{M\ge m+1\}] \\
&\qquad + E_\theta[\sqrt{a_m};\{a_m > C\xi_m^\alpha F^{m-1}\}\cap V_m(\eta_o)\cap\{M\ge m+1\}] \\
&\quad \le O(\sqrt{F^{m-1}})P_\theta(U(\varepsilon_o)') + O(\sqrt{a})P_\theta(a_m > C\xi_m^\alpha F^{m-1}) \\
&\qquad\quad \text{(using the crude bound } a_k \le a + O(1) = O(a)) \\
&\quad = O(h_m(a)o(1) + O(\sqrt{a})o(h_m(a)/a) = o(h_m(a)), \qquad (4.59)
\end{aligned}$$

using Lemma 2.6 and (4.58), giving

$$E_\theta(N^{m+1}; U(\varepsilon_o)' \cap V_m(\eta_o) \cap \{M \geq m+1\}) \leq a E_\theta(\mu_m^{*-1}; U(\varepsilon_o)' \cap V_m(\eta_o) \cap \{M \geq m+1\}) + o(h_m(a)). \tag{4.60}$$

Also

$$E_\theta(N^{m+1}; V_m(\eta_o)' \cap \{M \geq m+1\}) \leq \overline{n} P_\theta(V_m(\eta_o)') \leq O(a)O(\Phi(-a^{1/7})) = o(1) = o(h_m(a)), \tag{4.61}$$

by Lemma 4.5. Combining (4.57), (4.60), and (4.61), we have

$$E_\theta(N^{m+1}; M \geq m+1) \leq a E_\theta(\mu_m^{*-1}; M \geq m+1) - \frac{z^\alpha(\theta, Q)\xi_m^\alpha(\theta)}{I(\theta)C_m^m} h_m(a) + (\varepsilon + o(1))h_m(a).$$

This last term may be replaced by $o(h_m(a))$ since ε is arbitrary. Doing this and plugging into (4.55),

$$\sum_{k \leq m+1} E_\theta(N_k; M \geq k) \leq a E_\theta(\mu^{*-1}_{(M-1 \wedge 1)}) - \frac{z^\alpha(\theta, Q)\xi_m^\alpha(\theta)}{I(\theta)C_m^m} h_m(a) + o(h_m(a)). \tag{4.62}$$

Next we will estimate the terms of $\sum_{k>m} E_\theta(N_k; M \geq k)$. Let $V = V_m(\eta_1) \cap V_{m+1}(\eta_1)$, where $\eta_1 > 0$ will be determined below. Given $\varepsilon > 0$, choose $0 < \varepsilon_o \leq (\varepsilon/2)[63 \cdot \overline{(\Delta(-z^\alpha(\theta, Q))\xi_{m+1}^\alpha(\theta)/(C_m^m I(\theta)))}]^{-1}$. For sufficiently small η_1,

$$\begin{aligned} & E_\theta(N_{m+2}; U(\varepsilon_o) \cap V \cap \{M \geq m+2\}) \\ \leq\ & E_\theta(a_{m+1}/\mu^*_{m+1} + O(\sqrt{a_{m+1}}); U(\varepsilon_o) \cap V \cap \{M \geq m+2\}) \\ \leq\ & \frac{(1+\varepsilon_o)}{I(\theta)} E_\theta(a_{m+1}; U(\varepsilon_o) \cap V \cap \{M \geq m+2\}) \end{aligned}$$

and

$$a_{m+1} \leq a_m - \ell(N_{m+1}, \hat{\theta}_m) + K \log a, \tag{4.63}$$

for some $K < \infty$, by Lemma 4.4. Letting $a_m^* = a_m + K \log a$, note that

$$\{M \geq m+2\} \subseteq \{a_{m+1} > 0\} \subseteq \{a_m^* - \ell(N_{m+1}, \hat{\theta}_m) > 0\} \tag{4.64}$$

by (4.63). Then, letting

$$\zeta = \frac{a_m^* - \mu_m(\theta) N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}},$$

$$\begin{aligned}
&E_\theta(N_{m+2}; U(\varepsilon_o) \cap V \cap \{M \geq m+2\}) \\
&\leq \frac{(1+\varepsilon_o)}{I(\theta)} E_\theta[(a_m^* - \ell(N_{m+1}, \hat{\theta}_m)) 1\{a_m^* - \ell(N_{m+1}, \hat{\theta}_m) \geq 0\} | U(\varepsilon_o) \cap V] \\
&\leq \frac{(1+\varepsilon_o)^2}{I(\theta)} E_\theta[\sqrt{N_{m+1}} \Delta(-\zeta) \sigma_m(\theta) | U(\varepsilon_o) \cap V] \\
&= \frac{(1+\varepsilon_o)^2}{I(\theta)} E_\theta[\sqrt{a_m/\mu_m^* + O(\sqrt{a_m})} \Delta(-\zeta) \sigma_m(\theta) | U(\varepsilon_o) \cap V] \\
&\leq \frac{(1+\varepsilon_o)^2}{I(\theta)} E_\theta[(1+\varepsilon_o)\sigma_m(\theta) \sqrt{\frac{(1+\varepsilon_o)\xi_m^\alpha F^{m-1}}{\mu_m(\theta)}} \Delta(-\zeta) | U(\varepsilon_o) \cap V] \\
&\qquad \text{(for sufficiently small } \eta_1) \\
&\leq \frac{(1+\varepsilon_o)^{7/2}}{I(\theta)} E_\theta[\xi_{m+1} \alpha \sqrt{F^{m-1}} \Delta(-\zeta) | U(\varepsilon_o) \cap V] \\
&\leq \frac{(1+\varepsilon_o)^{7/2}}{I(\theta)} E_\theta[(1+\varepsilon_o)\xi_{m+1}^\alpha(\theta) (C_m^m)^{-1} h_m(a) \Delta(-\zeta) | U(\varepsilon_o) \cap V] \\
&\qquad \text{(since } \sqrt{F^{m-1}} \sim (C_m^m)^{-1} h_m(a)) \\
&= \frac{(1+\varepsilon_o)^{11/2} \xi_{m+1}^\alpha(\theta)}{I(\theta) C_m^m h_m(a)} E_\theta[\Delta(-\zeta) | U(\varepsilon_o) \cap V] \\
&= \frac{(1+\varepsilon_o)^{11/2} \xi_{m+1}^\alpha(\theta)}{I(\theta) C_m^m} h_m(a) [\Delta(-z^\alpha(\theta, Q)) + o(1)] && (4.65) \\
&\leq \frac{\xi_{m+1}^\alpha(\theta) \Delta(-z^\alpha(\theta, Q))}{I(\theta) C_m^m} h_m(a) + \left[\frac{63\varepsilon_o \xi_{m+1}^\alpha(\theta)}{I(\theta) C_m^m} + o(1)\right] h_m(a) \\
&\leq \frac{\xi_{m+1}^\alpha(\theta) \Delta(-z^\alpha(\theta, Q))}{I(\theta) C_m^m} h_m(a) + \varepsilon h_m(a), && (4.66)
\end{aligned}$$

by our choice of ε_o, where (4.65) uses a routine argument like that of Lemma 4.1.

On V,

$$\begin{aligned} a_{m+1} &\leq a_m - \ell(N_{m+1}, \hat{\theta}_m) + O(\log a) \quad \text{(by (4.63))} \\ &\leq a_m + O(a_m) + O(\log a) \\ &= O(a_m) + o(h_m(a)), \end{aligned}$$

so

$$\begin{aligned} &E_\theta(N_{m+2}; U(\varepsilon_o)' \cap V \cap \{M \geq m+2\}) \\ &\leq E_\theta(O(a_{m+1}); U(\varepsilon_o)' \cap V \cap \{M \geq m+2\}) \\ &\leq E_\theta(O(a_m) + o(h_m(a)); U(\varepsilon_o)' \cap V \cap \{M \geq m+2\}) \\ &\leq o(h_m(a)) \end{aligned} \tag{4.67}$$

by the argument leading to (4.59). Using the crude bound $N_{m+2} \leq N \leq \overline{n}$,

$$E_\theta(N_{m+2}; V' \cap \{M \geq m+2\}) \leq \overline{n} P_\theta(V') = o(1), \tag{4.68}$$

by the argument leading to (4.61). Combining (4.66), (4.67), and (4.68),

$$E_\theta(N_{m+2}; M \geq m+2) \leq \frac{\Delta(-z^\alpha(\theta, Q))\xi^\alpha_{m+1}(\theta)}{C^m_m I(\theta)} h_m(a) + (\varepsilon + o(1)) h_m(a) \tag{4.69}$$

and we may replace this last term by $o(h_m(a))$ since ε was arbitrary.

As in the proof of Theorem 4.9, there exists $q \to 0$ such that $P_\theta(M \geq m+2+k) \leq$

q^k for $k \geq 1$, and since N_{m+2+k} are stochastically decreasing in k,

$$\begin{aligned}
&\sum_{k\geq 1} E_\theta(N_{m+2+k}; M \geq m+2+k) \\
&= \sum_{k\geq 1} E_\theta(N_{m+2+k} | M \geq m+2+k) P_\theta(M \geq m+2+k) \\
&\leq \sum_{k\geq 1} E_\theta(N_{m+2} | M \geq m+2) q^k \\
&\leq O(h_m(a)) \sum_{k\geq 1} q^k \quad \text{(by (4.69) and since } P_\theta(M \geq m+2) \text{ bounded above 0)} \\
&= O(h_m(a)) q/(1-q) = o(h_m(a)).
\end{aligned}$$

Combining this with (4.69) and (4.62),

$$\begin{aligned}
E_\theta N &= \sum_{k\leq m+1} E_\theta(N_k; M \geq k) + \sum_{k>m+1} E_\theta(N_k; M \geq k) \\
&\leq aE_\theta(\mu^{*-1}_{(M-1\wedge 1)}) - \frac{z^\alpha(\theta,Q)\xi^\alpha_{m+1}(\theta)}{I(\theta)C^m_m} h_m(a) + \frac{\Delta(-z^\alpha(\theta,Q))\xi^\alpha_{m+1}}{I(\theta)C^m_m} h_m(a) + o(h_m(a)) \\
&= aE_\theta(\mu^{*-1}_{(M-1\wedge 1)}) + \frac{\Delta(z^\alpha(\theta,Q))\xi^\alpha_{m+1}}{I(\theta)C^m_m} h_m(a) + o(h_m(a)),
\end{aligned}$$

this last since

$$-z + \Delta(-z) = -z + \phi(-z) + \Phi(z)z = -z + \phi(z) + (1-\Phi(-z))z = \Delta(z).$$

Since $h_m(a) \sim Q^{-1}(d/c)$ and $E_\theta(\mu^{*-1}_{(M-1\wedge 1)}) = I(\theta)^{-1} + O(1/a)$ by Lemma 4.1, we have

$$E_\theta N \leq a/I(\theta) + \frac{\Delta(z^\alpha(\theta,Q))\xi^\alpha_{m+1}}{I(\theta)C^m_m Q}(d/c) + o(d/c). \tag{4.70}$$

Next we estimate the number of stages, M.

$$\begin{aligned}
E_\theta M &= \sum_{k\geq 0} P_\theta(M > k) \leq m+1 + P_\theta(M \geq m+2) + \sum_{k\geq 1} P_\theta(M \geq m+2+k) \\
&\leq m+1 + P_\theta(M \geq m+2) + \sum_{k\geq 1} q^k \\
&\leq m+1 + \Phi(z^\alpha(\theta,Q)) + o(1), \qquad (4.71)
\end{aligned}$$

once we show

$$P_\theta(M \geq m+2) \leq \Phi(z^\alpha(\theta, Q)) + o(1). \tag{4.72}$$

Assume that $\eta_2 \to 0$, but slowly enough so that $P_\theta(V_m(\eta_2)) \to 1$, and note that this holds uniformly in θ by Lemma 4.5. Let

$$\rho_k(\theta) = \frac{\ell(N_{k+1}, \hat{\theta}_k) - \mu_k(\theta) N_{k+1}}{\sigma_k(\theta)\sqrt{N_{k+1}}}$$

and choose $\varepsilon > 0$.

$$\begin{aligned}
P_\theta(M \geq m+2) &\leq P_\theta(M \geq m+2 | U(\varepsilon/2) \cap V_m(\eta_2)) P_\theta(U(\varepsilon/2) \cap V_m(\eta_2)) \\
&\quad + P_\theta(U(\varepsilon/2)') + P_\theta(V_m(\eta_2)) \\
&\leq P_\theta(M \geq m+2 | U(\varepsilon/2) \cap V_m(\eta_2)) P_\theta(U(\varepsilon/2) \cap V_m(\eta_2)) + o(1) \\
&\leq P_\theta(\rho_m(\theta) \leq \zeta | U(\varepsilon/2) \cap V_m(\eta_2))
\end{aligned}$$

by (4.64). We can write

$$\begin{aligned}
\zeta &= \frac{a_m^* - \mu_m(\theta) N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} = \frac{a - \mu_m(\theta) N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} + \frac{K \log a}{\sigma_m(\theta)} \\
&= \frac{\sigma_m^*}{\sigma_m(\theta)} z_m + \sqrt{N_{m+1}} \frac{\mu_m^* - \mu_m(\theta)}{\sigma_m(\theta)} + \frac{K \log a}{\mu_m(\theta)\sqrt{N_{m+1}}} \\
&\leq O(1) + O(\eta_2 \sqrt{F^{m-1}}) + O(\log a (F^{m-1})^{-1/2}) \\
&\leq O((F^{m-1})^{1/7}) = O(N_{m+1}^{1/7})
\end{aligned}$$

uniformly on $U(\varepsilon/2) \cap V_m(\eta_2)$ if we choose $\eta_2 = (F^{m-1})^{-5/14}$, say. Note that $\eta_2 \sqrt{a} \geq a^{1/7} \to \infty$, so $P_\theta(V_m(\eta_2)) \to 1$. Hence, we can apply large deviations to get, for sufficiently small d,

$$\begin{aligned}
P_\theta(M \geq m+2) &\leq E_\theta[(1+\varepsilon/2)\Phi(\zeta) | U(\varepsilon/2) \cap V_m(\eta_2)] \qquad (4.73) \\
&= (1+\varepsilon/2)(\Phi(z^\alpha(\theta, Q)) + o(1)) \leq \Phi(z^\alpha(\theta, Q)) + \varepsilon, \qquad (4.74)
\end{aligned}$$

proving (4.72) and hence (4.71)

Finally, we show that $E_\theta r(\lambda_M) = o(d)$. Choose $\gamma_1(d), \gamma_2(d) \to \infty$ to be any functions satisfying

$$\sqrt{h_m(a)} \gg \sqrt{\gamma_1} \gg \gamma_2 \gg \log a. \tag{4.75}$$

For example, $\gamma_1 = a^{2^{-m}}$ and $\gamma_2 = a^{2^{-m-2}}$ suffice. Let

$$\begin{aligned} W_0 &= U(1/2) \cap V_{m+1}(\eta_3), \\ W_1 &= \{a_{m+1} \le -\gamma_1\}, \\ W_2 &= \{a_{m+1} \ge \gamma_1, a_{m+2} \le -\gamma_2\}, \\ \text{and} \quad W &= (W_0 \cap W_1) \cup (W_0 \cap W_2), \end{aligned}$$

where $\eta_3 > 0$ will be chosen below. On $W_0 \cap W_i$, $i = 1, 2$,

$$r(\lambda_M) = \exp[-\log r(\lambda_{m+1})^{-1}] = \exp[a_{m+1} - a] \le e^{-\gamma_i} d = o(d).$$

Then, since $r(\lambda_M) \le d^2$ on $\{N = \overline{n}\}$ and $r(\lambda_M) \le d$ a.s.,

$$\begin{aligned} E_\theta r(\lambda_M) &= E_\theta(r(\lambda_M); W \cap \{N < \overline{n}\}) + E_\theta(r(\lambda_M); N = \overline{n}) + E_\theta(r(\lambda_M); W' \cup \{N < \overline{n}\}) \\ &\le o(d) \cdot 1 + d^2 \cdot 1 + d \cdot P_\theta(W') = o(d), \end{aligned} \tag{4.76}$$

once we show that $P_\theta(W) \to 1$. Now

$$\begin{aligned} & P_\theta(W_0 \cap W_1) = P_\theta(W_0) - P_\theta(W_0 \cap W_1') \ge P_\theta(W_0) - P_\theta(W_1' | W_0) \\ &\ge P_\theta(W_0) - P_\theta(\ell(N_{m+1}, \hat{\theta}_m) \le a_m + \gamma_1 + O(\log a) | W_0) \quad \text{(by (4.63))} \\ &= P_\theta(W_0) - P_\theta\left(\rho_m(\theta) \le \frac{a_m - \mu_m(\theta) N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} + O\left(\frac{\gamma_1}{\sqrt{N_{m+1}}}\right) \middle| W_0\right) \\ &= P_\theta(W_0) - P_\theta\left(\rho_m(\theta) \le \frac{a_m - \mu_m(\theta) N_{m+1}}{\sigma_m(\theta)\sqrt{N_{m+1}}} + o(1) \middle| W_0\right) \end{aligned}$$

since $\log a \ll \gamma_1 \ll h_m(a) = O(\sqrt{N_{m+1}})$ on W_0. This last probability approaches $\Phi(z^\alpha(\theta, Q))$ by the argument leading to (4.74), hence if we choose η_3 such that

$P_\theta(W_0) \to 1$,

$$P_\theta(W_0 \cap W_1) \geq 1 - \Phi(z^\alpha(\theta, Q)) + o(1). \tag{4.77}$$

Now

$$\begin{aligned} P_\theta(W_0 \cap W_2) &= P_\theta(W_2|W_0)P_\theta(W_0) = P_\theta(a_{m+1} \geq \gamma_1, a_{m+2} \leq -\gamma_2|W_0)(1+o(1)) \\ &= P_\theta(a_{m+2} \leq -\gamma_2|\{a_{m+1} \geq \gamma_1\} \cap W_0)P_\theta(a_{m+1} \geq \gamma_1|W_0)(1+o(1)) \end{aligned}$$

and $P_\theta(a_{m+1} \geq \gamma_1|W_0) \to \Phi(z^\alpha(\theta, Q))$ by replacing γ_1 by $-\gamma_1$ in the argument used on $P_\theta(W_1'|W_0)$. Using (4.63),

$$\begin{aligned} &P_\theta(a_{m+2} \geq -\gamma_2|\{a_{m+1} \geq \gamma_1\} \cap W_0) \\ &\leq P_\theta\left(\rho_{m+1}(\theta) \leq \frac{a_{m+1} - \mu_{m+1}(\theta)N_{m+2}}{\sigma_{m+1}(\theta)\sqrt{N_{m+2}}} + O\left(\frac{\gamma_2}{\sqrt{N_{m+2}}}\right) \middle| \{a_{m+1} \geq \gamma_1\} \cap W_0\right). \end{aligned} \tag{4.78}$$

Now, letting $z \to -\infty$ be the parameter of δ_α representing the geometric sampling quantile, on $\{a_{m+1} \geq \gamma_1\} \cap W_0$

$$\begin{aligned} \zeta_{m+1} &\equiv \frac{a_{m+1} - \mu_{m+1}(\theta)N_{m+2}}{\sigma_{m+1}(\theta)\sqrt{N_{m+2}}} + O\left(\frac{\gamma_2}{\sqrt{N_{m+2}}}\right) \\ &= \frac{\sigma^*_{m+1}}{\sigma_{m+1}(\theta)}z + \sqrt{N_{m+2}}\frac{\mu^*_{m+1} - \mu_{m+1}(\theta)}{\sigma_{m+1}(\theta)} + o(1) \\ &\quad \text{(by definition of } z \text{ and since } \gamma_2 \ll \sqrt{\gamma_1} = O(\sqrt{a_{m+1}}) = O(\sqrt{N_{m+2}}) \text{ on } \{a_{m+1} \geq \gamma_1\}) \\ &\leq (1+o(1))z + \sqrt{N_{m+2}}O(\eta_3) + o(1) \quad \text{(since } W_0 \subseteq V_{m+1}(\eta_3)) \\ &\leq (1+o(1))z \to -\infty \end{aligned} \tag{4.79}$$

if we choose $\eta_3 = \sqrt{|z|}/h_m(a)$, since then

$$\eta_3\sqrt{N_{m+2}} = O(\eta_3\sqrt{a_{m+1}}) = O(\eta_3\sqrt{a_m}) = O(\eta_3\sqrt{F^{m-1}}) = O(\eta_3 h_m(a)) = o(|z|)$$

on W_0. Plugging (4.79) into (4.78) yields

$$P_\theta(a_{m+2} \geq -\gamma_2|\{a_{m+1} \geq \gamma_1\} \cap W_0) \leq P_\theta(\rho_{m+1}(\theta) \leq \zeta_{m+1}|\{a_{m+1} \geq \gamma_1\} \cap W_0) \to 0$$

and thus

$$P_\theta(W_0 \cap W_2) = (1+o(1))(\Phi(z^\alpha(\theta,Q)) + o(1))(1+o(1)) = \Phi(z^\alpha(\theta,Q)) + o(1).$$

Combining this with (4.77) gives

$$\begin{aligned} P_\theta(W) &= P_\theta(W_0 \cap W_1) + P_\theta(W_0 \cap W_2) \quad (W_1, W_2 \text{ disjoint}) \\ &\geq 1 - \Phi(z^\alpha(\theta,Q)) + \Phi(z^\alpha(\theta,Q)) + o(1) = 1 + o(1), \end{aligned}$$

proving (4.76). Combining (4.70), (4.71), and (4.76),

$$\begin{aligned} E_\theta[cN + dM + r(\lambda_M)] &\leq c[a/I(\theta) + \frac{\Delta(z^\alpha(\theta,Q))\xi^\alpha_{m+1}(\theta)}{I(\theta)C^m_m Q}(d/c)] + d[m + 1 + \Phi(z^\alpha(\theta,Q))] + o(d) \\ &= ca/I(\theta) + d \cdot u^\alpha_m(\theta,Q) + o(d) \end{aligned}$$

uniformly in θ, and hence

$$r(\lambda_0, \delta_\alpha) \leq caE_{\lambda_0} I(\theta)^{-1} + d \cdot E_{\lambda_0} u^\alpha_m(\theta,Q) + o(d).$$

This holds for all $\alpha > 0$, so by a now standard asymptotic technique, there is a function $\alpha(d) \to 0$ for which it holds. Note that $u^{\alpha(d)}_m(\theta,Q) = u_m(\theta,Q) + o(1)$ uniformly in θ by (4.52), so setting $\delta = \delta_{\alpha(d)}$,

$$\begin{aligned} r(\lambda_0, \delta) &= r(\lambda_0, \delta_{\alpha(d)}) \\ &\leq caE_{\lambda_0} I(\theta)^{-1} + d \cdot E_{\lambda_0}[u_m(\theta,Q) + o(1)] + o(d) \\ &= caE_{\lambda_0} I(\theta)^{-1} + d \cdot E_{\lambda_0} u_m(\theta,Q) + o(d), \end{aligned}$$

finishing the proof. □

4.3 The Tests δ^* and $\tilde{\delta}^*$

Lemma 4.11. *there exists $K < \infty$ such that $r(\lambda_{M^*}) \leq Kd$. Conversely, if the stopping risk at the end of a stage is less than d, then δ^* will stop.*

Proof. The Bayes test δ^* stops when the stopping risk is less than or equal the smallest possible posterior expectation of the cost of continuing. One such continuation is fully sequential sampling, whose expected cost of continuation is well known to be a bounded multiple of the cost per observation, $c + d$ in this case. Thus, there exists $K < \infty$ such that $r(\lambda_{M^*}) \leq (K/2)(c + d) \leq Kd$, which is the first claim. Since any possible continuation incurs a cost of at least d, the cost of one stage, the stopping risk is less than the cost of any possible continuation when it is less than d; this is the converse claim. □

In computing the operating characteristics of a test, it is useful to have a lower bound on the size of the first stage. Lemma 4.12 establishes the existence of a test $\tilde{\delta}^*$ with such a lower bound that is "close" to the Bayes procedure in behavior and in integrated risk. The remainder of this section will largely be spent computing the operating characteristics of $\tilde{\delta}^*$; we then compare the test δ of Section 4.2 with the Bayes test, using $\tilde{\delta}^*$ as an intermediary.

Lemma 4.12. *There is a test $\tilde{\delta}^* = (\tilde{N}^*, \tilde{M}^*, \tilde{D}^*)$ satisfying*

$$\begin{aligned} \tilde{N}_1^* &\geq \varepsilon a, \quad \varepsilon > 0, \\ r(\lambda_{\tilde{M}^*}) &\leq Kd, \quad K < \infty, & (4.80) \\ r(\lambda_0, \tilde{\delta}^*) &\leq r(\lambda_0, \delta^*) + o(d). & (4.81) \end{aligned}$$

Proof. By Lemma 4.11

$$\begin{aligned} a - O(1) &\leq \log r(\lambda_{N^*})^{-1} \\ &\leq \ell(N^*, \hat{\theta}_{M^*}) + O(\log\log \ell(N^*, \hat{\theta}_{M^*})) & (4.82) \end{aligned}$$

by Lemma 4.2 since $(|S_{N^*}| \vee N^*) \to \infty$ as $d \to 0$. Now, if $a - O(1) \le x + O(\log x)$, then $x \ge a - O(\log a)$ since if $x = a - \gamma \log a$, some $\gamma \to \infty$, then

$$\begin{aligned} a - [x + O(\log x)] &= \gamma \log a - O(\log(a - \gamma \log a)) \\ &\ge \gamma \log a - O(\log a) \ne O(1), \end{aligned}$$

violating the original assumption. Hence, (4.82) implies

$$\ell(N^*, \hat{\theta}_{M^*}) \ge a - O(\log a) \tag{4.83}$$

as $d \to 0$. Then

$$N^* = I_{\hat{\theta}^*_{M^*}}(\hat{\theta}_{M^*})^{-1} \ell(N^*, \hat{\theta}_{M^*}) \ge I_{\hat{\theta}^*_{M^*}}(\hat{\theta}_{M^*})^{-1}[a - O(\log a)]. \tag{4.84}$$

Let $\varepsilon_o > 0$ be such that $\underline{\theta} - \varepsilon_o, \overline{\theta} + \varepsilon_o$ are both in the interior of the natural parameter space, and let $W = \{\hat{\theta}^*_{M^*} \in (\underline{\theta} - \varepsilon_o, \overline{\theta} + \varepsilon_o)\}$. Note that on W

$$I_{\hat{\theta}^*_{M^*}}(\hat{\theta}_{M^*}) \le [I_{\underline{\theta} - \varepsilon_o}(\underline{\theta}) \vee I_{\overline{\theta} + \varepsilon_o}(\overline{\theta})] < \infty.$$

By (4.84), there exists $\varepsilon > 0$ such that εa is an integer and

$$N^* \ge 2\varepsilon[a - O(\log a)] \ge \varepsilon a$$

on W for sufficiently small d. By Lemma 4.5 there exists $\varepsilon_1 > 0$ such that

$$P_\theta(N^* \ge \varepsilon a) \ge P_\theta(W) \ge 1 - 2\exp[-\varepsilon_1 a] \tag{4.85}$$

uniformly in θ.

Define $\tilde{\delta}^*$ as follows. Let

$$\tilde{N}_1^* = N_1^* \vee \varepsilon a,$$

for $k \geq 1$

$$\tilde{N}^*_{k+1} = \begin{cases} [N^*_{k+1} - (\tilde{N}^{*k} - N^{*k})^+]^+ & \text{if } M^* \geq k+1 \\ N^*_{(k+1)} & \text{if } M^* \leq k, \end{cases}$$

and letting $k^* = \inf\{k \geq 1 : N^{*k} = \tilde{N}^{*k}\}$, define

$$\tilde{M}^* = \begin{cases} M^* + \inf\{k \geq 0 : M^*_{(M^*+k+1)} = 0\} & \text{on } \{k^* = \infty > M^*\} \\ M^* & \text{on } \{k^* < \infty\}, \end{cases}$$

where $N^*_{(k+1)}, M^*_{(k+1)}$ are Bayes continuations after k stages of sampling under $\tilde{\delta}^*$. Note that we have assumed $M^* < \infty$ a.s. since the Bayes procedure cannot minimize the integrated risk without $EM^* < \infty$.

The test $\tilde{\delta}^*$ can be interpreted as follows. The first stage $\tilde{N}^*_1$ is at least εa and, if this is greater than N^*_1, the following stages of $\tilde{\delta}^*$ through the $(k^* \wedge M^*)$th stage are smaller than the corresponding stages of δ^*. On sample paths such that $k^* < \infty$, δ^* has "caught up" with $\tilde{\delta}^*$ after the k^*th stage in the sense that

$$\tilde{N}^{*k} = N^{*k} \quad \text{for all } k \geq k^* \tag{4.86}$$

and the two tests will coincide exactly thereafter. On sample paths such that $k^* = \infty$, δ^* stops before ever "catching up" with $\tilde{\delta}^*$ and as soon as this happens, $\tilde{\delta}^*$ begins a Bayes continuation. In either case, $\tilde{\delta}^*$ only stops when the Bayes stopping rule indicates to do so, hence (4.80) holds by Lemma 4.11.

On $\{k^* < \infty\}$, $(\tilde{N}^*, \tilde{M}^*, \tilde{D}^*) = (N^*, M^*, D^*)$ since the procedures will behave identically after the k^*th stage. On $\{k^* = \infty\}$, $\tilde{N}^*$ is no larger than the sample size of the procedure that initially samples εa and then performs a Bayes continuation.

Thus

$$\begin{aligned} E_\theta(\tilde{N}^* - N^*) &\le E_\theta(\tilde{N}^*; k^* = \infty) \\ &\le (\varepsilon a + E_\theta N^*) P_\theta(k^* = \infty) \\ &\le (\varepsilon a + E_\theta N^*) P_\theta(N^* < \varepsilon a) \quad (\text{since } \{k^* = \infty\} \subseteq \{N^* < \varepsilon a\}) \\ &\le (\varepsilon a + E_\theta N^*) \cdot 2\exp(-\varepsilon_1 a) = E_\theta N^* \cdot 2\exp(-\varepsilon_1 a) + o(1), \end{aligned}$$

by (4.85). This holds uniformly in θ and so

$$E_{\lambda_0}(c\tilde{N}^*) \le E_{\lambda_0}(cN^*)[1 + 2\exp(-\varepsilon_1 a)] + o(c) = E_{\lambda_0}(cN^*)[1 + 2\exp(-\varepsilon_1 a)] + o(d). \tag{4.87}$$

On $\{k^* = \infty\}$, $\tilde{M}^*$ is no larger than the number of stages of the procedure that performs two Bayes tests successively, so similarly

$$\begin{aligned} E_\theta(\tilde{M}^* - M^*) &\le E_\theta(\tilde{M}^*; k^* = \infty) \\ &\le 2E_\theta M^* P_\theta(k^* = \infty) \\ &\le 2E_\theta M^* \cdot 2\exp(-\varepsilon_1 a). \end{aligned}$$

This holds uniformly in θ, so

$$E_{\lambda_0}(d\tilde{M}^*) \le E_{\lambda_0}(dM^*)[1 + 4\exp(-\varepsilon_1 a)]. \tag{4.88}$$

Since the stopping risks also coincide on $\{k^* < \infty\}$,

$$\begin{aligned} E_\theta[r(\lambda_{\tilde{M}^*}) - r(\lambda_{M^*})] &\le E_\theta[r(\lambda_{\tilde{M}^*}); k^* = \infty] \\ &\le Kd \cdot P_\theta(k^* = \infty) \quad (\text{by } (4.80)) \\ &\le Kd \cdot 2\exp(-\varepsilon_1 a) = O(d) \cdot o(1) = o(d). \end{aligned}$$

This holds uniformly in θ, giving $E_{\lambda_0} r(\lambda_{\tilde{M}^*}) \le E_{\lambda_0} r(\lambda_{M^*}) + o(d)$. Combining this with (4.87) and (4.88),

$$\begin{aligned}
r(\lambda_0, \tilde{\delta}^*) &= E_{\lambda_0}[c\tilde{N}^* + d\tilde{M}^* + r(\lambda_{\tilde{M}^*})] \\
&\le E_{\lambda_0}(cN^*)[1 + 2\exp(-\varepsilon_1 a)] + E_{\lambda_0}(cM^*)[1 + 4\exp(-\varepsilon_1 a)] + E_{\lambda_0} r(\lambda_{M^*}) + o(d) \\
&\le r(\lambda_0, \delta^*) + 4\exp(-\varepsilon_1 a) E_{\lambda_0}(cN^* + dM^*) + o(d) \\
&\le r(\lambda_0, \delta^*) + 4\exp(-\varepsilon_1 a) r(\lambda_0, \delta^*) + o(d).
\end{aligned}$$

We know from Theorems (4.9) and (4.10) that $r(\lambda_0, \delta^*) \le r(\lambda_0, \delta) = O(ca)$, so

$$r(\lambda_0, \tilde{\delta}^*) \le r(\lambda_0, \delta^*) + 4\exp(-\varepsilon_1 a) \cdot O(ca) + o(d) = r(\lambda_0, \delta^*) + o(d)$$

since

$$\exp(-\varepsilon_1 a) \cdot ca = d[a\exp(-\varepsilon_1 a)](c/d) = d \cdot o(1) \cdot o(1) = o(d).$$

This establishes (4.81) and finishes the proof. □

The next lemma gives a uniform lower bound on the average sample size of $\tilde{\delta}^*$.

Lemma 4.13. $E_\theta \tilde{N}^* \ge a/I(\theta) - O(\log a)$ *uniformly for* $\theta \in [\underline{\theta}, \overline{\theta}]$.

Proof. By the argument leading to (4.83),

$$\ell(\tilde{N}^*, \hat{\theta}_{\tilde{M}^*}) \ge a - O(\log a)$$

for sufficiently small d. Then

$$\tilde{N}^* = I_{\hat{\theta}^*_{\tilde{M}^*}}(\hat{\theta}_{\tilde{M}^*})^{-1} \ell(\tilde{N}^*, \hat{\theta}_{\tilde{M}^*}) \ge I_{\hat{\theta}^*_{\tilde{M}^*}}(\hat{\theta}_{\tilde{M}^*})^{-1}[a - O(\log a)]$$

and hence

$$\begin{aligned}
E_\theta \tilde{N}^* &\ge E_\theta I_{\hat{\theta}^*_{\tilde{M}^*}}(\hat{\theta}_{\tilde{M}^*})^{-1}[a - O(\log a)] \\
&\ge [I(\theta)^{-1} - O(1/a)] \cdot [a - O(\log a)]
\end{aligned}$$

by Lemma 4.1 since $a = O(\tilde{N}_1^*)$ and $\tilde{N}^* \geq \tilde{N}_1^*$. Expanding this last proves the claim. □

For $0 < \varepsilon < 1$ and $k \geq 1$ define $\mathcal{A}_k^+(\varepsilon)$ to be the set of all (s,t) such that

$$\log\left(\frac{d}{r(\lambda_{(s,t)})}\right)^{-1} \geq (1-\varepsilon)\xi_k(\hat{\theta}(s,t))F_{d/c}^{(k-1)}(\log d^{-1}) \quad \text{and} \quad \varepsilon \leq \hat{\theta}(s,t) \leq \overline{\theta} - \varepsilon.$$

Define $\mathcal{A}_k^-(\varepsilon)$ similarly but with $\underline{\theta} + \varepsilon \leq \hat{\theta}(s,t) \leq -\varepsilon$, and let $\mathcal{A}_k(\varepsilon) = \mathcal{A}_k^+(\varepsilon) \cup \mathcal{A}_k^-(\varepsilon)$. We will sometimes abuse this notation by writing $\lambda \in \mathcal{A}_k(\varepsilon)$ to mean $\lambda_{(s,t)}$ such that $(s,t) \in \mathcal{A}_k(\varepsilon)$.

Lemma 4.14. *Assume $c \in \mathcal{B}_m(d)$ and let $\lambda_k = \lambda_{(S^k, \tilde{N}^{*k})}$. Given $\varepsilon > 0$ and $1 \leq k \leq m$, there exists $\eta > 0$ such that*

$$P_{\lambda_0}(\lambda_k \in \mathcal{A}_k(\eta)) \geq 1 - \varepsilon. \tag{4.89}$$

Proof. Let $A_k^{(\cdot)}(\eta) = \{\lambda_k \in \mathcal{A}_k^{(\cdot)}(\eta)\}$. First we handle the $k = 1$ case. Assume that

$$\limsup_{d \to 0} \frac{\tilde{N}_1^*}{a} \leq \overline{I}^{-1}. \tag{4.90}$$

Suppose $\varepsilon > 0$. Choose $\varepsilon_o > 0$ such that

$$\lambda_0(\underline{\theta} + \varepsilon_o, -\varepsilon_o) + \lambda_0(\varepsilon_o, \overline{\theta} - \varepsilon_o) \geq 1 - \varepsilon/2,$$

and let $\eta = \varepsilon_o/2$. We can write

$$P_{\lambda_0}(A_1(\eta)) \geq \int_{\underline{\theta}+\varepsilon_o}^{-\varepsilon_o} P_\theta(A_1^-(\eta))\lambda_0(\theta)d\theta + \int_{\varepsilon_o}^{\overline{\theta}-\varepsilon_o} P_\theta(A_1^+(\eta))\lambda_0(\theta)d\theta. \tag{4.91}$$

Let $V_k(\varepsilon) \equiv \{|\hat{\theta}_k - \theta| \leq \varepsilon\}$. Let $0 < \varepsilon_1 \leq \varepsilon_o/2$, where ε_1 will be determined below. On $V_1(\varepsilon_1)$ for $\theta \in [\varepsilon_o, \overline{\theta} - \varepsilon_o]$,

$$\hat{\theta}_1 \geq \theta - \varepsilon_1 \geq \varepsilon_o - \varepsilon_1 \geq \varepsilon_o/2 = \eta$$

and

$$\hat{\theta}_1 \le \theta + \varepsilon_1 \le \overline{\theta} - \varepsilon_o + \varepsilon_1 \le \overline{\theta} - \varepsilon_o/2 = \overline{\theta} - \eta.$$

Similarly, $\underline{\theta} + \eta \le \hat{\theta}_1 \le -\eta$ for $\theta \in [\underline{\theta} + \varepsilon_o, -\varepsilon_o]$. Let

$$\begin{aligned} \mu_k(\theta) &= I_\theta(\hat{\theta}_k) \\ \sigma_k^2(\theta) &= (\hat{\theta}_k - \hat{\theta}_k')^2 \psi''(\theta) \\ \rho_k(\theta) &= \frac{\ell(\tilde{N}_k^*, \hat{\theta}_k) - \mu_k(\theta)\tilde{N}_k^*}{\sigma_k(\theta)\sqrt{\tilde{N}_k^*}}. \end{aligned}$$

Now, $a_1 = a - \log r(\lambda_1)^{-1} \ge a - \ell_1 + O(1)$ by Lemma 4.2, so that

$$\begin{aligned} A_1^+ \cap V_1(\varepsilon_1) &\supseteq \{\ell_1 \le a[1 - (1-\eta)(1 - I(\hat{\theta}_1)/\overline{I})] + O(1)\} \cap V_1(\varepsilon_1) \\ &\supseteq \left\{ \rho_1(\theta) \le \frac{a[1 - (1-\eta)(1 - I(\hat{\theta}_1)/\overline{I})] - \mu_1(\theta)\tilde{N}_1^*}{\sigma_1(\theta)\sqrt{\tilde{N}_1^*}} + O(\tilde{N}_1^{*-1/2}) \right\} \cap V_1(\varepsilon_1). \end{aligned}$$

Let

$$\eta_o = 1 - \frac{I(\underline{\theta} + \varepsilon_o/2) \vee I(\overline{\theta} - \varepsilon_o/2)}{\overline{I}} > 0$$

and note that $1 - I(\hat{\theta}_1)/\overline{I} \ge \eta_o$ on $V_1(\varepsilon_1)$. Also, using (4.90) and the fact that $\tilde{N}_1^* \ge \varepsilon a \to \infty$,

$$\begin{aligned} &\frac{a[1 - (1-\eta)(1 - I(\hat{\theta}_1)/\overline{I})] - \mu_1(\theta)\tilde{N}_1^*}{\sigma_1(\theta)\sqrt{\tilde{N}_1^*}} + O(\tilde{N}_1^{*-1/2}) \\ &\quad \ge \frac{a[1 - (1-\eta)(1 - I(\hat{\theta}_1)/\overline{I})] - \mu_1(\theta)a(1+\eta\eta_o/4)/\overline{I}}{\sigma_1(\theta)\sqrt{a(1+\eta\eta_o/4)/\overline{I}}} + o(1) \\ &\quad = \sqrt{\frac{a\overline{I}}{\sigma_1(\theta)^2(1+\eta\eta_o/4)}} \left[1 - (1-\eta)(1 - I(\hat{\theta}_1)/\overline{I}) - \mu_1(\theta)(1+\eta\eta_o/4)/\overline{I}\right] \end{aligned}$$

for sufficiently small d. As $\varepsilon_1 \to 0$, the expression in brackets approaches

$$\begin{aligned} \eta[1 - I(\theta)/\overline{I} - (I(\theta)/\overline{I})\eta_o/4)] &\ge \eta[1 - I(\theta)/\overline{I} - \eta_o/4] \\ &\ge \eta[3\eta_o/4] > 0, \end{aligned}$$

and therefore

$$\frac{a[1-(1-\eta)(1-I(\hat{\theta}_1)/\overline{I})]-\mu_1(\theta)\tilde{N}_1^*}{\sigma_1(\theta)\sqrt{\tilde{N}_1^*}}+O(\tilde{N}_1^{*-1/2})\geq \eta'\sqrt{a},$$

some $\eta' > 0$, for sufficiently small ε_1. Thus, for $\theta \in [\varepsilon_o, \overline{\theta}-\varepsilon_o]$,

$$\begin{aligned} P_\theta(A_1^+(\eta)) &\geq P_\theta(A_1^+(\eta)|V_1(\varepsilon_1))P_\theta(V_1(\varepsilon_1)) \\ &\geq P_\theta(\rho_1(\theta)\leq \eta'\sqrt{a}|V_1(\varepsilon_1))P_\theta(V_1(\varepsilon_1)) \to 1 \end{aligned}$$

uniformly since $\eta'\sqrt{a}\to\infty$ and we choose $\varepsilon_1 \to 0$ so that $P_\theta(V_1(\varepsilon_1)) \to 1$ by a now routine argument. Similarly, $P_\theta(A_1^-(\eta)) \to 1$ uniformly for $\theta \in [\underline{\theta}+\varepsilon_o, -\varepsilon_o]$, and plugging into (4.91) gives

$$\begin{aligned} P_{\lambda_0}(A_1(\eta)) &\geq (1-o(1))\lambda_0(\underline{\theta}+\varepsilon_o,-\varepsilon_o)+(1-o(1))\lambda_0(\varepsilon_o,\overline{\theta}-\varepsilon_o) \\ &\geq (1-o(1))(1-\varepsilon/2)\geq 1-\varepsilon \end{aligned}$$

by the time the $o(1)$ term is less than $\varepsilon/2$.

All that remains for the $k=1$ case is to verify (4.90). Suppose that, contrary to (4.90), $\limsup \tilde{N}_1^*/a > \overline{I}^{-1}$. Then there exists $\eta > 0$ and a sequence of d's approaching 0 on which $\tilde{N}_1^* \geq (\overline{I}^{-1}+2\eta)a$. Assume $\overline{I}=I(\overline{\theta})$; the other case, $\overline{I}=I(\underline{\theta})$, is handled similarly. By continuity there exists $\theta_2 < \overline{\theta}$ such that $I(\theta)^{-1} \leq \overline{I}^{-1}+\eta$ for all $\theta\in[\theta_2,\overline{\theta}]$, and hence

$$\tilde{N}_1^* \geq (I(\theta)^{-1}+\eta)a \tag{4.92}$$

for all $\theta \in [\theta_2, \overline{\theta}]$. Since

$$E_\theta \tilde{N}^* \geq I(\theta)^{-1}a - O(\log a) \tag{4.93}$$

uniformly for $\theta \in [\underline{\theta}, \overline{\theta}]$ by Lemma 4.13, it follows that

$$\begin{aligned}
r(\lambda_0, \tilde{\delta}^*) &\geq E_{\lambda_0}(c\tilde{N}^*) \\
&= c\int_{\underline{\theta}}^{\theta_2} E_\theta \tilde{N}^* \lambda_0(\theta) d\theta + c\int_{\theta_2}^{\overline{\theta}} E_\theta \tilde{N}^* \lambda_0(\theta) d\theta \\
&\geq ca\int_{\underline{\theta}}^{\theta_2} I(\theta)^{-1} \lambda_0(\theta) d\theta + ca\int_{\theta_2}^{\overline{\theta}} (I(\theta)^{-1} + \eta) \lambda_0(\theta) d\theta - O(c\log a) \\
&\qquad \text{(by (4.92) and (4.93))} \\
&\geq ca[E_{\lambda_0} I(\theta)^{-1} + \eta'] - O(c\log a),
\end{aligned}$$

where $\eta' = \eta\lambda_0(\theta_2, \overline{\theta}) > 0$. We know from Theorems 4.9 and 4.10 that

$$r(\lambda_0, \delta) \leq caE_{\lambda_0} I(\theta)^{-1} + O(d),$$

which leads to

$$\begin{aligned}
r(\lambda_0, \delta^*) - r(\lambda_0, \delta) &= [r(\lambda_0, \delta^*) - r(\lambda_0, \tilde{\delta}^*)] + [r(\lambda_0, \tilde{\delta}^*) - r(\lambda_0, \delta)] \\
&\geq -o(d) + [\eta' ca - O(c\log a) - O(d)] \quad \text{(by Lemma 4.12)} \\
&= \eta' ca - o(ca) > 0
\end{aligned}$$

for sufficiently small d, which contradicts the optimality of δ^*. This proves (4.90) and completes the $k = 1$ case.

To handle $2 \leq k \leq m$, we will first prove that for sufficiently small $\eta > 0$,

$$P_{\lambda_0}(A_k^{\pm}(3\eta/4) | A_{k-1}^{\pm}(\eta)) \to 1 \tag{4.94}$$

as $d \to 0$. Let $\lambda_{k-1} \in \mathcal{A}_{k-1}^{\pm}(\eta)$ and $\varepsilon_1 > 0$, which will be chosen below. Consider $P_\theta(A_k^{\pm}(3\eta/4))$ for $|\theta - \hat{\theta}_{k-1}| \leq \varepsilon_1$. If $\varepsilon_1 \leq \eta/8$, then on $A_{k-1}^{+}(\eta) \cap V_k(\varepsilon_1)$,

$$\hat{\theta}_k \geq \theta - \varepsilon_1 \geq \hat{\theta}_{k-1} - 2\varepsilon_1 \geq \eta - 2\varepsilon_1 \geq 3\eta/4$$

and

$$\hat{\theta}_k \le \theta + \varepsilon_1 \le \hat{\theta}_{k-1} + 2\varepsilon_1 \le \overline{\theta} - \eta + 2\varepsilon_1 \le \overline{\theta} - 3\eta/4.$$

Similarly, on $A^-_{k-1}(\eta) \cap V_k(\varepsilon_1)$,

$$\underline{\theta} + 3\eta/4 \le \hat{\theta}_k \le -3\eta/4$$

so in either case, the requirements of $\hat{\theta}_k$ on $A^\pm_k(3\eta/4)$ are satisfied on $V_k(\varepsilon_1)$ if $\varepsilon_1 \le \eta/8$, which we assume for the remainder of the proof. Let

$$\zeta = \frac{a_{k-1} - \mu_k(\theta)\tilde{N}^*_k}{\sigma_k(\theta)\sqrt{\tilde{N}^*_k}}$$

and let F^{k-1} denote $F^{(k-1)}_{d/c}(a)$. By Lemma 4.4, $a_k \ge a_{k-1} - \ell(\tilde{N}^*_k, \hat{\theta}_k) - O(\log a)$, so that

$$\begin{aligned} A^\pm_k(3\eta/4) \cap V_k(\varepsilon_1) &\supseteq \{\ell(\tilde{N}^*_k, \hat{\theta}_k) \le a_{k-1} - (1-3\eta/4)\xi_k F^{k-1} + O(\log a)\} \cap V_k(\varepsilon_1) \\ &\supseteq \left\{\rho_k(\theta) \le \zeta - \frac{(1-3\eta/4)\xi_k F^{k-1} + O(\log a)}{\sigma_k(\theta)\sqrt{\tilde{N}^*_k}}\right\} \cap V_k(\varepsilon_1). \end{aligned}$$

Solving for $\sqrt{\tilde{N}^*_k}$ we obtain

$$\zeta - \frac{(1-3\eta/4)\xi_k F^{k-1} + O(\log a)}{\sigma_k(\theta)\sqrt{\tilde{N}^*_k}} = \zeta - \frac{(1-3\eta/4)\xi_k F^{k-1} + O(\log a)}{\sigma_k(\theta)\mu_k(\theta)^{-1}(\sqrt{a_{k-1}\mu_k(\theta) + \zeta^2\sigma_k(\theta)^2/4} - \zeta\sigma_k(\theta)/2)}.$$

This last is increasing in ζ, so letting $U = \{\zeta \ge \sqrt{\log[F^{k-2}/(d/c)^2]} - 1\}$, on $V_k(\varepsilon_1) \cap$

$U \cap A_{k-1}^{\pm}(\eta)$,

$$\begin{aligned}
&\zeta - \frac{(1-3\eta/4)\xi_k F^{k-1} + O(\log a)}{\sigma_k(\theta)\sqrt{\tilde{N}_k^*}} \\
&\geq \sqrt{\log F^{k-2}/(d/c)^2} - 1 - \frac{(1-3\eta/4)\xi_k F^{k-1}}{\sigma_k(\theta)\mu_k(\theta)^{-1}\sqrt{a_{k-1}\mu_k(\theta)}}(1+o(1)) \\
&\geq \sqrt{\log F^{k-2}/(d/c)^2} - \frac{(1-3\eta/4)\xi_k F^{k-1}}{\sigma_k(\theta)\mu_k(\theta)^{-1/2}\sqrt{(1-\eta)\xi_{k-1}F^{k-2}}}(1+o(1)) \\
&\geq \sqrt{\log F^{k-2}/(d/c)^2} - \frac{(1-3\eta/4)}{\sqrt{1-\eta}}\sqrt{\frac{\mu_k(\theta)}{\sigma_k(\theta)\xi_{k-1}}}\xi_k \frac{F^{k-1}}{\sqrt{F^{k-2}}}(1+o(1)) \\
&\geq \sqrt{\log F^{k-2}/(d/c)^2} - (1-\eta/12)(1+\eta/24)\sqrt{\log F^{k-2}/(d/c)^2} \\
&\geq (\eta/24)\sqrt{\log F^{k-2}/(d/c)^2} \to \infty,
\end{aligned}$$

this last since

$$\frac{(1-3\eta/4)}{\sqrt{1-\eta}} \leq 1-\eta/12 \quad \text{and} \quad \sqrt{\frac{\mu_k(\theta)}{\sigma_k(\theta)\xi_{k-1}}} \cdot \xi_k(1+o(1)) \leq 1+\eta/24$$

for sufficiently small ε_1. Thus,

$$\begin{aligned}
P_\theta(A_k^{\pm}(3\eta/4)) &\geq P_\theta(A_k^{\pm}(3\eta/4)|V_k(\varepsilon_1)\cap U)P_\theta(V_k(\varepsilon_1)\cap U) \\
&\geq P_\theta(\rho_k(\theta) \leq (\eta/24)\sqrt{\log F^{k-2}/(d/c)^2}|V_k(\varepsilon_1)\cap U)P_\theta(V_k(\varepsilon_1)\cap U) \\
&= (1+o(1))P_\theta(V_k(\varepsilon_1)\cap U) \sim P_\theta(U), \qquad (4.95)
\end{aligned}$$

since $P_\theta(V_k(\varepsilon_1)) \to 1$ by a routine argument.

Now, letting

$$\tilde{\lambda}_0(\theta) = \frac{P_\theta(A_{k-1}^{\pm}(\eta))\lambda_0(\theta)}{P_{\lambda_0}(A_{k-1}^{\pm}(\eta))}$$

and using ϖ_θ to denote the distribution function of $(S_{k-1}, \tilde{N}^{k-1})$ given the true parameter value θ, we write

$$
\begin{aligned}
&P_{\lambda_0}(A_k^{\pm}(3\eta/4)|A_{k-1}^{\pm}(\eta)) = E_{\lambda_0}(P_{\lambda_{k-1}}(A_k^{\pm}(3\eta/4))|A_{k-1}^{\pm}(\eta)) \\
&= \int_{[\underline{\theta},\bar{\theta}]} E_{\tilde{\theta}}[P_{\lambda_{k-1}}(A_k^{\pm}(3\eta/4))|A_{k-1}^{\pm}(\eta)]\tilde{\lambda}_0(\tilde{\theta})d\tilde{\theta} \\
&= \int_{[\underline{\theta},\bar{\theta}]} \int_{\mathcal{A}_{k-1}^{\pm}(\eta)} P_{\lambda_{k-1}}(A_k^{\pm}(3\eta/4))\frac{d\varpi_{\tilde{\theta}}(s,t)}{P_{\tilde{\theta}}(A_{k-1}^{\pm}(\eta)}\tilde{\lambda}_0(\tilde{\theta})d\tilde{\theta} \\
&= \int_{[\underline{\theta},\bar{\theta}]} \int_{\mathcal{A}_{k-1}^{\pm}(\eta)} \int_{[\underline{\theta},\bar{\theta}]} P_{\theta}(A_k^{\pm}(3\eta/4))\lambda_{(s,t)}(\theta)d\theta\frac{d\varpi_{\tilde{\theta}}(s,t)}{P_{\tilde{\theta}}(A_{k-1}^{\pm}(\eta)}\tilde{\lambda}_0(\tilde{\theta})d\tilde{\theta} \\
&\gtrsim \int_{[\underline{\theta},\bar{\theta}]} \int_{\mathcal{A}_{k-1}^{\pm}(\eta)} \int_{[\underline{\theta},\bar{\theta}]} P_{\theta}(U)\lambda_{(s,t)}(\theta)d\theta\frac{d\varpi_{\tilde{\theta}}(s,t)}{P_{\tilde{\theta}}(A_{k-1}^{\pm}(\eta)}\tilde{\lambda}_0(\tilde{\theta})d\tilde{\theta}, \qquad (4.96)
\end{aligned}
$$

this last by (4.95). Thus if (4.94) were to fail there would be a sequence of d's approaching zero on which the right hand side of (4.96) is bounded below 1. Letting

$$
\nu(J_1 \times J_2 \times J_3) \equiv \int_{J_1}\int_{J_2}\int_{J_3} \lambda_{(s,t)}(\theta)d\theta\frac{d\varpi_{\tilde{\theta}}(s,t)}{P_{\tilde{\theta}}(A_{k-1}^{\pm}(\eta))}\tilde{\lambda}_0(\tilde{\theta})d\tilde{\theta},
$$

this would imply that there exists $\varepsilon_2 > 0$ and $J \subseteq [\underline{\theta},\overline{\theta}] \times \mathcal{A}_{k-1}^{\pm}(\eta) \times [\underline{\theta},\overline{\theta}]$ such that $\nu(J) \geq \varepsilon_2$ and $P_\theta(U') \geq \varepsilon_2$ on this sequence. Let

$$
J_1 = \{x : (x,y,z) \in J\}, \quad J_2(\theta) = \{y : (\theta,y,z) \in J\}, \quad J_3(s,t) = \{z : (x,(s,t),z) \in J\}.
$$

For $\theta \in J_1$, using Wald's equation

$$
E_\theta \tilde{N}^* = [E_\theta I_\theta(\hat{\theta}_{\tilde{M}^*})]^{-1} E_\theta \ell_{\tilde{M}^*}.
$$

By Theorem 6.1.1 of [16],

$$
[E_\theta I_\theta(\hat{\theta}_{\tilde{M}^*})]^{-1} = [I(\theta) + O(1/a)]^{-1} = I(\theta)^{-1} + O(1/a),
$$

since $a = O(\tilde{N}_1^*)$ by Lemma 4.12. Let $\varepsilon_3 > 0$ and

$$\begin{aligned} W_o(\theta) &= \{\lambda_{k-1} \in J_2(\theta)\} \cap \{\theta \in J_3(S_{k-1}, \tilde{N}^{*k-1})\} \cap U' \\ W(\theta) &= W_o(\theta) \cap \{\hat{\theta}_{k-1}, \hat{\theta}_k \in (\theta - \varepsilon_3, \theta + \varepsilon_3)\}. \end{aligned}$$

By Lemma 4.2

$$\begin{aligned} \ell_{\tilde{M}^*} &\geq \log r(\lambda_{\tilde{M}^*})^{-1} - O(\log a) \\ &\geq a - O(1) - O(\log a) \quad \text{(by Lemma 4.11)} \\ &= a - O(\log a), \end{aligned}$$

and therefore

$$\begin{aligned} E_\theta \ell_{m^*} - a + O(\log a) &\geq E_\theta[\ell_{\tilde{M}^*} - a + O(\log a)|W]P_\theta(W) \\ &\geq E_\theta[\ell_{\tilde{M}^*} - a + O(\log a)|W]P_\theta(W) \\ &\geq E_\theta[(\ell_{\tilde{M}^*} - a + O(\log a))1\{\tilde{M}^* = k\}|W]P_\theta(W) \\ &\geq E_\theta[(\ell(\tilde{N}_k^*, \hat{\theta}_{k-1}) - a_{k-1} + O(\log a))1\{\tilde{M}^* = k\}|W]P_\theta(W) \qquad (4.97) \end{aligned}$$

since, on $\{\tilde{M}^* = k\}$,

$$\begin{aligned} \ell_{\tilde{M}^*} &= \ell(\tilde{N}^{*k}, \hat{\theta}_k) \geq \ell(\tilde{N}^{*k}, \hat{\theta}_{k-1}) \\ &= \ell(\tilde{N}_k^*, \hat{\theta}_{k-1}) + \ell(\tilde{N}^{*k-1}, \hat{\theta}_{k-1}) \\ &= \ell(\tilde{N}_k^*, \hat{\theta}_{k-1}) + \log r(\lambda_{k-1})^{-1} + O(\log a) \quad \text{(by Lemma 4.2)} \\ &= \ell(\tilde{N}_k^*, \hat{\theta}_{k-1}) + a - a_{k-1} + O(\log a). \end{aligned}$$

Letting $\varepsilon_3 \to 0$ in such a way that $\varepsilon\sqrt{a} \ll \zeta$ yields $P_\theta(\hat{\theta}_{k-1}, \hat{\theta}_k \in (\theta - \varepsilon_3, \theta + \varepsilon_3)) \to 1$

and $\varepsilon_3\sqrt{\tilde{N}_k^*}$ on W. Let

$$\begin{aligned}\zeta' &\equiv \frac{a_{k-1} - O(\log a) - \mu_{k-1}(\theta)\tilde{N}_k^*}{\sigma_{k-1}(\theta)\sqrt{\tilde{N}_k^*}} \\ &= \frac{\sigma_k(\theta)}{\sigma_{k-1}(\theta)}\zeta + \sqrt{\tilde{N}_k^*}\cdot O(\mu_k(\theta) - \mu_{k-1}(\theta)) + o(1) \\ &= (1 + o(1))\zeta + O(\varepsilon_3\sqrt{\tilde{N}_k^*}) \\ &\sim \zeta = o((\tilde{N}_k^*)^{1/6})\end{aligned}$$

on $W \subseteq A_{k-1}(\eta)$ Then by Lemma 2.10

$$\begin{aligned}&E_\theta[(\ell(\tilde{N}_k^*, \hat{\theta}_{k-1}) - a_{k-1} + O(\log a))1\{\tilde{M}^* = k\}|W] \sim E_\theta[\Delta(\zeta)\sqrt{\tilde{N}_k^*}|W] \\ &\gtrsim E_\theta[\Delta(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)\sqrt{a_{k-1}/I(\theta)}|W] \quad (\text{on } U') \\ &\gtrsim E_\theta\left[\frac{\phi(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)}{(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)^2}\cdot\sqrt{\eta' F^{k-2}}\,\middle|\, W\right] \\ &\qquad (\text{some } \eta' > 0 \text{ on } A_{k-1}(\eta), \text{ and since } \Delta(z) \sim \phi(z)/z^2) \\ &\propto \frac{\exp[-(1/2)(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)^2]}{(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)^2}\sqrt{F^{k-2}} \\ &= \frac{d/c}{\sqrt{F^{k-2}}}\cdot\frac{\exp[\sqrt{\log[F^{k-2}/(d/c)^2]} - 1/2]}{(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)^2}\sqrt{F^{k-2}} \\ &= (d/c)\frac{\exp[\sqrt{\log[F^{k-2}/(d/c)^2]} - 1/2]}{(\sqrt{\log[F^{k-2}/(d/c)^2]} - 1)^2} \equiv (d/c)\cdot\gamma \gg d/c. \qquad (4.98)\end{aligned}$$

Also, since

$$\int_{J_1} P_\theta(W_o(\theta))\tilde{\lambda}_0(\theta)d\theta = \nu(J) \geq \varepsilon_2 > 0,$$

there exists $\tilde{J}_1 \subseteq J_1$ such that $\tilde{\lambda}_0(\tilde{J}_1) \geq \tilde{\varepsilon}_2 > 0$ and $P_\theta(W_o(\theta)) \geq \tilde{\varepsilon}_2$ for all $\theta \in \tilde{J}_1$. Since $P_\theta(\hat{\theta}_{k-1}, \hat{\theta}_k \in (\theta - \varepsilon_3, \theta + \varepsilon_3)) \to 1$, this last implies $P_\theta(W(\theta)) \geq \tilde{\varepsilon}_2/2 > 0$, say, for all $\theta \in \tilde{J}_1$ and sufficiently small d and also implies that

$$\nu(\tilde{J}_1 \times J_2 \times J_3) \geq \int_{\tilde{J}_1} P_\theta(W_o(\theta))\tilde{\lambda}_0(\theta)d\theta \geq \tilde{\varepsilon}_2^2 > 0. \qquad (4.99)$$

Plugging this and (4.98) into (4.97), we have

$$E_\theta \tilde{N}^* \geq [I(\theta)^{-1} + O(1/a)][(d/c)\cdot\gamma\cdot\tilde{\varepsilon}_2/2 + a - O(\log a)] = a/I(\theta) + \tilde{\gamma},$$

where $\tilde{\gamma} \gg d/c$, and this holds uniformly for $\theta \in \tilde{J}_1$; we will use this lower bound for $\theta \in \tilde{J}_1$ and the uniform lower bound provided by Lemma 4.13 for $\theta \notin \tilde{J}_1$. Now, since

$$\begin{aligned}\lambda_0(\tilde{J}_1) &= \int_{\tilde{J}_1} \lambda_0(\theta)d\theta = \int_{\tilde{J}_1} \frac{P_{\lambda_0}(A^{\pm}_{k-1}(\eta))\tilde{\lambda}_0(\theta)d\theta}{P_\theta(A^{\pm}_{k-1}(\eta))} \\ &\geq P_{\lambda_0}(A^{\pm}_{k-1}(\eta))\tilde{\lambda}_0(\tilde{J}_1) \geq P_{\lambda_0}(A^{\pm}_{k-1}(\eta))\nu(\tilde{J}_1 \times J_2 \times J_3) \\ &\geq \tilde{\varepsilon}_3 > 0\end{aligned}$$

by the induction hypothesis and (4.99),

$$\begin{aligned}r(\lambda_0, \delta^*) &\geq E_{\lambda_0}(c\tilde{N}^*) = cE_{\lambda_0}(\tilde{N}^*; \theta \in \tilde{J}_1) + cE_{\lambda_0}(\tilde{N}^*; \theta \notin \tilde{J}_1) \\ &\geq cE_{\lambda_o}(a/I(\theta) + \tilde{\gamma}; \theta \in \tilde{J}_1) + cE_{\lambda_0}(a/I(\theta) - O(\log a); \theta \notin \tilde{J}_1) \quad \text{(by Lemma 4.13)} \\ &\geq caE_{\lambda_0}I(\theta)^{-1} + c\tilde{\gamma}\tilde{\varepsilon}_3 - c\cdot O(\log a) \\ &= caE_{\lambda_0}I(\theta)^{-1} + c\tilde{\gamma}\tilde{\varepsilon}_3 - o(c\tilde{\gamma})\end{aligned}$$

since $\tilde{\gamma} \gg d/c \gg \log a$. But we know from Theorems 4.9 and 4.10 that

$$r(\lambda_0, \delta) \leq caE_{\lambda_0}I(\theta)^{-1} + O(d) = caE_{\lambda_0}I(\theta)^{-1} + o(c\tilde{\gamma}),$$

which implies

$$\begin{aligned}r(\lambda_0, \delta^*) - r(\lambda_0, \delta) &= [r(\lambda_0, \delta^*) - r(\lambda_0, \tilde{\delta}^*)] + [r(\lambda_0, \tilde{\delta}^*) - r(\lambda_0, \delta)] \\ &\geq -o(d) + [c\tilde{\gamma}\tilde{\varepsilon}_3 - o(c\tilde{\gamma})] \quad \text{(by Lemma 4.12)} \\ &= c\tilde{\gamma}\tilde{\varepsilon}_3 - o(c\tilde{\gamma}) > 0 \qquad (4.100)\end{aligned}$$

for sufficiently small d, a contradiction. We have thus established (4.94).

With this in hand, we now finish the induction by proving the k case of (4.89).

Given $\varepsilon > 0$, let $\eta > 0$ be such that $P_{\lambda_0}(A_1(\eta)) \geq 1-\varepsilon/2$ via the $k = 1$ case of (4.89). Then

$$\begin{aligned} P_{\lambda_0}(A_k^{\pm}((3/4)^{k-1}\eta)) &\geq P_{\lambda_0}(A_1^{\pm}(\eta)) \prod_{i=2}^{k} P_{\lambda_0}(A_i^{\pm}((3/4)^{i-1}\eta) | A_{i-1}^{\pm}((3/4)^{i-2}\eta)) \\ &= P_{\lambda_0}(A_1^{\pm}(\eta)) \prod_{i=2}^{k} (1 - o(1)) \quad \text{(by (4.94))} \\ &= P_{\lambda_0}(A_1^{\pm}(\eta)) \cdot (1 - o(1)). \end{aligned}$$

Assuming d is sufficiently small that this last $o(1)$ term is less than $\varepsilon/2$ (for both the $+$ and $-$ cases), we have

$$\begin{aligned} P_{\lambda_0}(A_k((3/4)^{k-1}\eta)) &= P_{\lambda_0}(A_k^{+}((3/4)^{k-1}\eta)) + P_{\lambda_0}(A_k^{-}((3/4)^{k-1}\eta)) \\ &\geq P_{\lambda_0}(A_1^{+}(\eta)) \cdot (1-\varepsilon/2) + P_{\lambda_0}(A_1^{-}(\eta)) \cdot (1-\varepsilon/2) \\ &= P_{\lambda_0}(A_1(\eta))(1-\varepsilon/2) \\ &\geq (1-\varepsilon/2)(1-\varepsilon/2) \geq 1-\varepsilon, \end{aligned}$$

finishing the proof. □

Lemma 4.15. *Assume that $c \in \mathcal{B}_m^+(d)$ and let $Q = \lim_{d\to 0}(d/c)/h_m(a) \in (0, \infty)$. For every $\varepsilon > 0$ there exists $\eta > 0$ such that*

$$r(\lambda, \tilde{\delta}^*) \geq c \log(d/r(\lambda))^{-1} E_\lambda I(\theta)^{-1} + d[E_\lambda u_m(\theta, Q) - m] - \varepsilon d$$

uniformly for $\lambda \in \mathcal{A}_m(\eta)$.

Proof. Let $\eta, \eta_1, \eta_2 > 0$, to be chosen below. Let $(s,t) \in \mathcal{A}_m^{\pm}(\eta)$ and let $\tilde{\delta}^* = (\tilde{N}^*, \tilde{M}^*, \tilde{D}^*)$ denote the continuation from (s,t); also let $\hat{\theta} = \hat{\theta}(s,t)$, $\lambda = \lambda_{(s,t)}$, and $\hat{\theta}_{\tilde{M}^*} = \hat{\theta}(s + S^{\tilde{M}^*}, t + \tilde{N}^*)$. Write

$$\begin{aligned} r(\lambda, \tilde{\delta}^*) &\geq E_\lambda(c\tilde{N}^* + d\tilde{M}^*) \\ &= \int_{|\theta-\hat{\theta}|\leq\eta_1} E_\theta(c\tilde{N}^* + d\tilde{M}^*)\lambda(\theta)d\theta + \int_{|\theta-\hat{\theta}|>\eta_1} E_\theta(c\tilde{N}^*)\lambda(\theta)d\theta. \end{aligned}$$

We first consider $E_\theta \tilde{N}^*$ for $|\theta - \hat{\theta}| \le \eta_1$. Let $V = \{|\hat{\theta}_{\tilde{M}^*} - \hat{\theta}| \le \eta_2\}$. By Wald's equation, $E_\theta(\ell(\tilde{N}^*, \hat{\theta})|V) = \tilde{\mu}(\theta) E_\theta(\tilde{N}^*|V)$, where $\tilde{\mu}(\theta) = I_{E_\theta(X_1|V)}(\hat{\theta})$, so that

$$E_\theta \tilde{N}^* \ge E_\theta(\tilde{N}^*|V) P_\theta(V) = \tilde{\mu}(\theta)^{-1} E_\theta(\ell(B^*, \hat{\theta})|V) P_\theta(V).$$

Note that for sufficiently small η_2, Lemma 4.3 applies on V so that

$$\begin{aligned}
\ell(\tilde{N}^*, \hat{\theta}) &= \log r(\lambda_{\tilde{M}^*})^{-1} - \log r(\lambda)^{-1} + o(1) \\
&\ge \log d^{-1} - O(1) - \log r(\lambda)^{-1} + o(1) \quad \text{(by Lemma 4.12)} \\
&\ge \log(d/r(\lambda))^{-1} - K
\end{aligned}$$

for some $K < \infty$. Letting $\tilde{\sigma}(\theta)^2 = (\hat{\theta} - \hat{\theta}')\mathrm{Var}(X_1|V)$ and $q(\theta) = P_\theta(\tilde{M}^* = 1|V)$,

$$\begin{aligned}
E_\theta(\ell(\tilde{N}^*, \hat{\theta})|V) &= E_\theta[\ell(\tilde{N}^*, \hat{\theta}) - (\log(d/r(\lambda))^{-1} - K)|V] + \log(d/r(\lambda))^{-1} - K \\
&\ge E_\theta[(\ell(\tilde{N}^*, \hat{\theta}) - (\log(d/r(\lambda))^{-1} - K))1\{\tilde{M}^* = 1\}|V] + \log(d/r(\lambda))^{-1} - K \\
&= \Delta(z_{q(\theta)})\tilde{\sigma}(\theta)\sqrt{\tilde{N}_1^*} \cdot (1 + o(1)) + \log(d/r(\lambda))^{-1} - K.
\end{aligned}$$

Assume that

$$\frac{\log(d/r(\lambda))^{-1} - K - \tilde{\mu}(\theta)\tilde{N}_1^*}{\tilde{\sigma}(\theta)\sqrt{\tilde{N}_1^*}} = O(1) \tag{4.101}$$

as $d \to 0$; if this were to fail then a contradiction to the optimality of δ^* could be reached by an argument like that leading to (4.100). Then, letting F^{m-1} denote $F_{d/c}^{(m-1)}(a)$, it follows from (4.101) that

$$\begin{aligned}
\Delta(z_{q(\theta)})\tilde{\sigma}(\theta)\sqrt{\tilde{N}_1^*} &= \Delta(z_{q(\theta)})\tilde{\sigma}(\theta)[\sqrt{\log(d/r(\lambda))^{-1}/\tilde{\mu}(\theta)} + O(1)] \\
&\ge \Delta(z_{q(\theta)})\tilde{\sigma}(\theta)\sqrt{(1-\eta)\xi_m(\hat{\theta})F^{m-1}/\tilde{\mu}(\theta)}(1 + o(1)),
\end{aligned}$$

by virtue of $\lambda \in \mathcal{A}_m^{\pm}(\eta)$. Hence

$$(E_\theta \tilde{N}^*)P_\theta(V)^{-1} \ge \tilde{\mu}(\theta)^{-1}\log(d/r(\lambda))^{-1} + \Delta(z_{q(\theta)})\tilde{\mu}(\theta)^{-3/2}\tilde{\sigma}(\theta)\sqrt{(1-\eta)\xi_m(\hat{\theta})}\frac{h_m(a)}{C_m^m}(1 + o(1)),$$

using Lemma 2.6. Now, by an argument like that of Lemma 4.1,

$$\tilde{\mu}(\theta)^{-1} = I_\theta(\hat{\theta})^{-1} + O(1/a) \geq I(\theta)^{-1} + O(1/a),$$

since $I_\theta(\hat{\theta}) = I(\theta) - I(\theta, \hat{\theta}) \leq I(\theta)$, and similarly $\tilde{\sigma}(\theta)^{-3/2} = \mu(\theta)^{-3/2} + o(1)$. Also, for sufficiently small η_1,

$$I(\theta)^{-3/2}\sigma(\theta)\sqrt{\xi_m(\hat{\theta})} \geq I(\theta)^{-1}\sqrt{\frac{\sigma(\theta)^2\xi_m(\theta)}{I(\theta)}}(1-\eta) = I(\theta)^{-1}\xi_{m+1}(\theta)(1-\eta).$$

Combining these estimates, we obtain for sufficiently small d

$$\begin{aligned}(E_\theta\tilde{N}^*)P_\theta(V)^{-1} &\geq I(\theta)^{-1}\log(d/r(\lambda))^{-1} + \frac{\Delta(z_{q(\theta)})\xi_{m+1}(\theta)h_m(a)}{I(\theta)C_m^m}(1-\eta)^2 \\ &\geq I(\theta)^{-1}\log(d/r(\lambda))^{-1} + \frac{\Delta(z_{q(\theta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(d/c)(1-\eta)^3.\end{aligned}$$

For the remainder of the proof assume that $\eta_1 \leq \eta_2/2$, which implies that $V \supseteq \{|\hat{\theta}_{\tilde{M}^*} - \theta| \leq \eta_2/2\}$ and hence

$$\begin{aligned}\log(d/r(\lambda))^{-1}P_\theta(V') &\leq O(\log a)P_\theta(|\hat{\theta}_{\tilde{M}^*} - \theta| > \eta_2/2) \\ &= O(\log a)O(\Phi(-\eta_2' a^{1/7})) = o(1),\end{aligned}$$

for some $\eta_2' > 0$, by the argument of Lemma 4.5. Thus

$$E_\theta\tilde{N}^* \geq I(\theta)^{-1}\log(d/r(\lambda))^{-1} + \frac{\Delta(z_{q(\theta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(d/c)(1-\eta)^3.$$

Also $E_\theta\tilde{M}^* \geq (2 - q(\theta))(1 - o(1))$ so that for $|\theta - \hat{\theta}| \leq \eta_1$,

$$E_\theta(c\tilde{N}^* + d\tilde{M}^*) \geq I(\theta)^{-1}c\log(d/r(\lambda))^{-1} + \left[\frac{\Delta(z_{q(\theta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3 + 2 - q(\theta)\right]d - o(d).$$

Using some calculus,

$$\begin{aligned}\frac{\Delta(z_{q(\theta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3+2-q(\theta) &\geq \inf_{p\in(0,1)}\left[\frac{\Delta(z_p)\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3+2-p\right] \\ &= \frac{\Delta(z_{p^*(\theta,\eta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3+2-p^*(\theta,\eta),\end{aligned}$$

where $p^*(\theta,\eta)$ is the unique solution of

$$\frac{p^*(\theta,\eta)}{\phi(z_{p^*(\theta,\eta)})}=\frac{I(\theta)C_m^m Q}{\xi_{m+1}(\theta)(1-\eta)^3}.$$

Now

$$\frac{\Delta(z_{p^*(\theta,\eta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3+2-p^*(\theta,\eta)\to u_m(\theta,Q)-m$$

as $\eta\to 0$, so that

$$\frac{\Delta(z_{p^*(\theta,\eta)})\xi_{m+1}(\theta)}{I(\theta)C_m^m Q}(1-\eta)^3+2-p^*(\theta,\eta)\geq u_m(\theta,Q)-m-\varepsilon/2$$

for sufficiently small η, uniformly for $|\theta-\hat{\theta}|\leq\eta_1$. Thus

$$E_\theta(c\tilde{N}^*+d\tilde{M}^*)\geq I(\theta)^{-1}c\log(d/r(\lambda))^{-1}+d(u_m(\theta,Q)-m)-(\varepsilon/2+o(1))d,$$

giving

$$\begin{aligned}\int_{|\theta-\hat{\theta}|\leq\eta_1}E_\theta(c\tilde{N}^*+d\tilde{M}^*)\lambda(\theta)d\theta\geq &c\log(d/r(\lambda))^{-1}E_\lambda(I(\theta)^{-1};|\theta-\hat{\theta}|\leq\eta_1)\\ &+d[E_\lambda(u_m(\theta,Q)-m;|\theta-\hat{\theta}|\leq\eta_1)-\varepsilon/2-o(1)].\end{aligned} \quad (4.102)$$

To handle $|\theta-\hat{\theta}|>\eta_1$, we use the uniform bound

$$\begin{aligned}E_\theta\tilde{N}^* &\geq I(\theta)^{-1}\log(d/r(\lambda))^{-1}-O(\log\log(d/r(\lambda))^{-1}) \quad \text{(Lemma 4.13)}\\ &\geq I(\theta)^{-1}\log(d/r(\lambda))^{-1}-O(\log a),\end{aligned}$$

since $\log(d/r(\lambda))^{-1} \leq a + O(1)$, and therefore

$$\begin{aligned}\int_{|\theta-\hat{\theta}|>\eta_1} E_\theta(c\tilde{N}^*)\lambda(\theta)d\theta &\geq c\log(d/r(\lambda))^{-1}E_\lambda(I(\theta)^{-1}; |\theta-\hat{\theta}|>\eta_1) - cO(\log a) \\ &\geq c\log(d/r(\lambda))^{-1}E_\lambda(I(\theta)^{-1}; |\theta-\hat{\theta}|>\eta_1) - o(d).\end{aligned}$$

Combining this with (4.102) gives

$$\begin{aligned}r(\lambda,\tilde{\delta}^*) &\geq c\log(d/r(\lambda))^{-1} + dE_\lambda(u_m(\theta,Q)-m; |\theta-\hat{\theta}|\leq\eta_1) - (\varepsilon/2+o(1))d \\ &\geq c\log(d/r(\lambda))^{-1} + dE_\lambda(u_m(\theta,Q)-m) - (\varepsilon/2+o(1))d \qquad (4.103)\end{aligned}$$

since

$$E_\lambda(u_m(\theta,Q)-m; |\theta-\hat{\theta}|>\eta_1) \leq 2\cdot P_\lambda(|\theta-\hat{\theta}|>\eta_1) = o(1).$$

Assuming d is small enough so that the $o(1)$ term in (4.103) is less than $\varepsilon/2$, this relation establishes the claim. □

The final theorem gives a lower bound on the integrated risk of the Bayes procedure and thereby shows that δ is second-order optimal.

Theorem 4.16. *Let $m \geq 1$ and $u_m(\theta,Q)$ be as in (4.53). Then, as $d \to 0$,*

$$r(\lambda_0,\delta^*) \geq \begin{cases} caE_{\lambda_0}I(\theta)^{-1} + d(m+1) - o(d), & \text{if } c \in \mathcal{B}_m^o(d) \\ caE_{\lambda_0}I(\theta)^{-1} + d\cdot E_{\lambda_0}u_m(\theta,Q) - o(d), & \text{if } c \in \mathcal{B}_m^+(d),\ Q = \lim \frac{(d/c)}{h_m(a)}. \end{cases} \qquad (4.104)$$

Therefore, as $d \to 0$, δ minimizes the stopping risk to second-order in the sense that

$$r(\lambda_0,\delta) - r(\lambda_0,\delta^*) = o(d), \qquad (4.105)$$

provided $c \in \mathcal{B}_m(d)$ for some $m \geq 1$.

Proof. We prove that the lower bounds (4.104) hold for $\tilde{\delta}^*$ and then use Lemma 4.12 to compare the integrated risks of $\tilde{\delta}^*$ and δ^*.

Assume that $c \in \mathcal{B}_m^o(d)$ and choose $\varepsilon > 0$. Since $\log a = o(h_m(a)) = o(d/c)$, by Lemma 4.13, $E_\theta \tilde{N}^* \geq a/I(\theta) - o(d/c)$ uniformly in θ and hence

$$E_{\lambda_0}(c\tilde{N}^*) \geq caE_{\lambda_0} I(\theta)^{-1} - o(d). \tag{4.106}$$

Let $A_m(\eta) = \{\lambda_m \in \mathcal{A}_m(\eta)\}$ and choose $\eta > 0$ such that

$$P(A_m(\eta)) \geq 1 - \frac{\varepsilon}{2(m+1)} \tag{4.107}$$

by virtue of Lemma 4.14. Since $\tilde{M}^* \geq m+1$ on A_m,

$$E(d\tilde{M}^*) \geq dE(\tilde{M}^*; A_m(\eta)) \geq d(m+1)P(A_m(\eta)) \geq d(m+1) - (\varepsilon/2)d,$$

by our choice of η. Combining this with (4.106) and assuming d is small ehough so that the $o(d)$ term in (4.106) is less than $(\varepsilon/2)d$,

$$\begin{aligned} r(\lambda_0, \tilde{\delta}^*) &\geq E(c\tilde{N}^* + d\tilde{M}^*) \\ &\geq caEI(\theta)^{-1} - (\varepsilon/2)d + d(m+1) - (\varepsilon/2)d \\ &= caEI(\theta)^{-1} + d(m+1) - \varepsilon d, \end{aligned}$$

proving that

$$r(\lambda_0, \tilde{\delta}^*) \geq caEI(\theta)^{-1} + d(m+1) - o(d).$$

The first case of (4.104) follows since $r(\lambda_0, \delta^*) \geq r(\lambda_0, \tilde{\delta}^*) - o(d)$ by Lemma 4.12.

Next we consider the boundary case. Assume that $c \in \mathcal{B}_m^+(d)$ and let $Q = \lim_{d\to 0}(d/c)/h_m(a) \in (0, \infty)$. Choose $\varepsilon > 0$. By Lemmas 4.14 and 4.15 there exists $\eta > 0$ such that $P_{\lambda_0}(A'_m(\eta)) \leq \varepsilon/[6(m+2)]$ and the conclusion of Lemma 4.15

holds with ε replaced by $\varepsilon/6$; one additional restriction is imposed on η below. Then

$$\begin{aligned} E_{\lambda_0}[c\tilde{N}^* + d\tilde{M}^* + r(\lambda_{\tilde{M}^*}); A_m(\eta)] &= E_{\lambda_0}[c\tilde{N}^{*m} + dm + r(\lambda_m, \tilde{\delta}^*); A_m(\eta)] \\ &\geq E_{\lambda_0}[c\tilde{N}^{*m} + dm + c\log(d/r(\lambda_m))^{-1}E_{\lambda_m}I(\theta)^{-1} + E_{\lambda_m}(u_m(\theta, Q) - m) - (\varepsilon/6)d; A_m(\eta)] \\ &= c\log d^{-1}E_{\lambda_0}(E_{\lambda_m}I(\theta)^{-1}; A_m(\eta)) + dE_{\lambda_0}(E_{\lambda_m}u_m(\theta, Q); A_m(\eta)) \\ &\quad + cE_{\lambda_0}(\tilde{N}^{*m} - \log r(\lambda_m)^{-1}E_{\lambda_m}I(\theta)^{-1}; A_m(\eta)) - (\varepsilon/6)d. \end{aligned}$$

Thus,

$$\begin{aligned} E_{\lambda_0}[c\tilde{N}^* + d\tilde{M}^* + r(\lambda_{0,\tilde{M}^*}); A_m(\eta)] - c\log d^{-1}E_{\lambda_0}(E_{\lambda_m}I(\theta)^{-1}; A_m(\eta)) - dE_{\lambda_0}u_m(\theta, Q) \\ \geq -dE_{\lambda_0}(E_{\lambda_m}u_m(\theta, Q); A'_m(\eta)) + cE_{\lambda_0}(\tilde{N}^{*m} - \log r(\lambda_m)^{-1}E_{\lambda_m}I(\theta)^{-1}; A_m(\eta)) - (\varepsilon/6)d \\ \geq -d(m+2)P_{\lambda_0}(A'_m(\eta)) - c \cdot o(d/c) - (\varepsilon/6)d \\ \geq -(\varepsilon/3 + o(1))d, \end{aligned} \tag{4.108}$$

by our choice of η. Also,

$$\begin{aligned} E_{\lambda_0}(c\tilde{N}^*; A'_m(\eta)) &\geq c[\log d^{-1}E_{\lambda_0}(I(\theta)^{-1}|A'_m(\eta)) - O(d/c)]P_{\lambda_0}(A'_m(\eta)) \\ &= c\log d^{-1}E_{\lambda_0}(I(\theta)^{-1}; A'_m(\eta)) - O(d)P_{\lambda_0}(A'_m(\eta)) \\ &\geq c\log d^{-1}E_{\lambda_0}(I(\theta)^{-1}; A'_m(\eta)) - (\varepsilon/6)d, \end{aligned}$$

assuming η is sufficiently small. Combining this with (4.108),

$$\begin{aligned} r(\lambda_0, \tilde{\delta}^*) &\geq E_{\lambda_0}(c\tilde{N}^* + d\tilde{M}^* + r(\lambda_m, \tilde{\delta}^*); A_m(\eta)) + E_{\lambda_0}(c\tilde{N}^*; A'_m(\eta)) \\ &\geq c\log d^{-1}[E_{\lambda_0}(E_{\lambda_m}I(\theta)^{-1}; A_m(\eta)) + E_{\lambda_0}(I(\theta)^{-1}; A'_m(\eta))] + dE_{\lambda_0}u_m(\theta, Q) - (\varepsilon/2 + o(1))d \\ &\geq c\log d^{-1}E_{\lambda_0}I(\theta)^{-1} + dE_{\lambda_0}u_m(\theta, Q) - \varepsilon d \end{aligned}$$

by the time the last $o(1)$ term is less than $\varepsilon/2$. This shows

$$r(\lambda_0, \tilde{\delta}^*) \geq c\log d^{-1}E_{\lambda_0}I(\theta)^{-1} + dE_{\lambda_0}u_m(\theta, Q) - o(d)$$

and consequently that the same bound holds for δ^* by Lemma 4.12. This finishes the boundary case and hence proves (4.104). Comparing this with the integrated risk of δ from Theorems 4.9 and 4.10 establishes (4.105). □

4.4 A Numerical Example

As discussed in Section 3.3 for simple hypotheses, there are many possibilities for small sample procedures that are asymptotically equivalent to the test δ, defined and proved asymptotically optimal above. In this section we describe one natural choice and give the results of a numerical experiment comparing it with group-sequential sampling.

Recall that the "exploratory" first stage of δ_α does not depend on m, where m is such that $c \in \mathcal{B}_m(d)$. Thus, a small sample version of δ_α may use the data of the first stage to determine its choice of m. Using this idea, let δ denote the test $\delta_{\alpha=0}$ with parameter m^* chosen to be the smallest k such that

$$C_k^k \sqrt{\xi_k(\hat{\theta}_1)} \cdot h_k(a/\sigma_1) \leq d/c \leq C_{k-1}^{k-1} \sqrt{\xi_{k-1}(\hat{\theta}_1)} \cdot h_{k-1}(a/\sigma_1). \tag{4.109}$$

It is immediate from (4.109) that $m^* = m$ for sufficiently small d when $c \in \mathcal{B}_m^o(d)$, whence δ is asymptotically optimal by Theorem 4.16.

Table 2 contains the results of a numerical experiment comparing δ with group-sequential testing of the hypotheses

$$-1 \leq \mu \leq -.25 \quad \text{vs.} \quad .25 \leq \mu \leq 1$$

about the mean of normally distributed random variables with unit variance, with a "flat" prior, $\lambda_0(\mu) = (1/2) \cdot 1\{|\mu| \leq 1\}$, and 0-1 loss function $w(\mu) = 1\{.25 \leq |\mu| \leq 1\}$. $\delta_g(k)$ denotes group-sequential testing with constant stage-size k, which samples until the stopping risk is less than d, the same stopping rule employed by δ.

For each value of d/c, the operating characteristics of $\delta_g(k)$ are given for $k = 1$,

Table 2
Numerical Results for Testing Normal Mean
$-1 \le \mu \le -.25$ vs. $.25 \le \mu \le 1$ ($d = 10^{-4}$)

Test	EN	EM	int. risk (d)	2nd-order risk (d)
		$d/c = 1$		
δ	61.7	4.13	65.8	8.9
$\delta_g(1)$	55.9	55.9	111.8	55.1
$\delta_g(12)$	64.4	5.47	69.8	12.9
$\delta_g(20)^\dagger$	77.1	3.85	81.0	24.1
		$d/c = 5$		
δ	73.8	2.61	17.4	5.2
$\delta_g(1)$	55.9	55.9	67.1	54.9
$\delta_g(12)$	64.4	5.47	18.4	6.2
$\delta_g(32)$	77.2	2.57	18.0	5.8
		$d/c = 10$		
δ	81.0	2.47	10.6	3.9
$\delta_g(1)$	55.9	55.9	61.5	54.8
$\delta_g(12)$	64.4	5.47	11.9	5.3
$\delta_g(40)$	92.2	2.30	11.5	4.9

$k = 12$ (the size of the first stage of δ for the values of the parameters considered), and the best possible k (determined by simulation). Since both δ and δ_g must sample until the stopping risk is less than d, the cost of the number of observations required for this and the first stage represents a "fixed cost" which all procedures will incur. Thus, we obtain a more accurate comparison of the efficiency due to sampling by considering the *2nd-order risk* of the procedures, defined as

$$\text{integrated risk } -(cEN^{(1)} + d),$$

where $N^{(1)}$ is the number of observations of $\delta_g(1)$.

The results show significant improvement in the risk and 2nd-order risk of δ over δ_g. As we noted in Section 3.3, the size of the smallest possible 2nd-order risk is not known, so it is difficult to say how much further improvement is possible without backward induction type calculations, which remain prohibitively large in this general

†In the $d/c = 1$ case, $k = 12$ is the best possible sample size so we report $k = 20$ as the third group size.

setting. We would expect the difference between δ and the best group sequential test to decrease for larger values of d/c, since $EM^* \to 1$ in this limit.

The differences in risk between δ and group sequential tests here is roughly comparable to that seen in the simple hypotheses setting. One would expect that a procedure that uses estimates of the true state of nature to design future stages would be more robust over a range of parameter values, and hence show more pronounced improvement over constant stage-size sampling in this composite hypotheses setting. This indicates that a higher level of refinement is necessary to indicate how to achieve higher efficiencies in practical use.

Bibliography

[1] Bechhofer, R. E., Dunnett, C. W., & Sobel, M. (1954), A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41** 170-176.

[2] Brown, L. D. (1986), *Fundamentals of Statistical Exponential Families.* Institute of Mathematical Statistics, Hayward, CA.

[3] Chernoff, H. (1959), Sequential design of experiments. *Ann. Math. Stat.* **30** 755-770.

[4] Chernoff, H. (1972), *Sequential Analysis and Optimal Design.* Society for Industrial and Applied Mathematics, Philadelphia.

[5] Chow, Y.S., Robbins, H., & Siegmund, D. (1971), *Great Expectations: The Theory of Optimal Stopping.* Houghton, Mifflin, & Co., New York.

[6] Chung, K. L. (1968), *A Course in Probability Theory.* Harcourt, New York.

[7] Cressie, N. & Morgan, P. B. (1993), The VPRT: A sequential testing procedure dominating the SPRT, *Econometric Theory* **9**, 431-450.

[8] DeMets, D. L. & Ware, J. H. (1980), Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* **67** 651-660.

[9] DeMets, D. L. & Lan, K. K. (1983), Discrete sequential boundaries for clinical trials. *Biometrika* **70** 659-663.

[10] Feller, W. (1971), *An Introduction to Probability Theory and its Applications, vol. II.* Wiley, New York.

[11] Govindarajulu, Z. (1975), *Sequential Statistical Procedures* Academic Press, New York.

[12] Keifer, J. & Sacks, J. (1963), Asymptotically optimal sequential inference and design. *Ann. Math. Stat.* **34** 705-750.

[13] Hall, P. (1981), Asymptotic theory of triple sampling for sequential estimation of a mean, *Ann. Statist.* **9**, 1229-1238.

[14] Kopka, H. & Daly, P. W. (1999), *A Guide to LaTeX.* Addison-Wesley, Harlow, England.

[15] Lehmann, E. L. (1986), *Testing Statistical Hypotheses.* Springer, New York.

[16] Lehmann, E. L. & Casella, G. (1998), *Theory of Point Estimation.* Springer, New York.

[17] Lorden, G. (1967), Integrated risk of asymptotically Bayes sequential tests, *Ann. Math. Stat.* **38**, 1399-1422.

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

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

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

[21] Lorden, G. (1980), Structure of sequential tests minimizing an expected sample size, *Z. Wahrscheinlichkeistheorie verw. Gebiete* **51**, 291-302.

[22] Lorden, G. (1983), Asymptotic efficiency of three-stage hypothesis tests. *Ann. Stat.* **11** 129-140.

[23] Lorden, G., Nearly-optimal sequential tests for exponential families, to appear.

[24] Morgan, P. B. & Cressie, N. (1997), A comparison of the cost-efficiencies of the sequential, group-sequential, and variable-sample-size-sequential probability ratio tests, *Scand. J. of Stat.* **24**, 181-200.

[25] Pocock, S. J. (1984) *Clinical Trials: A Practical Approach.* John Wiley & Sons, Chichester.

[26] Rudin, W. (1985), *Principles of Mathematical Analysis.* McGraw-Hill, New York.

[27] Schmitz, N. (1993) *Optimal Sequentially Planned Decision Procedures.* Lecture Notes in Statistics, **79**. Springer-Verlag, New York.

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

[29] Schwarz, G. (1969), A second-order approximation to optimal stopping regions. *Ann. Math. Stat.* **40** 313-315.

[30] Siegmund, D. (1985), *Sequential Analysis*. Springer-Verlag, New York.

[31] Stein, C. (1945), A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.* **16** 243-258.

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

[33] Wald, A. (1951), Asymptotic minimax solutions of sequential estimation problems. *Proc. Second Berkeley Symp. Math. Stat. Prob.* 1-11. Univ. of California Press.

[34] Wald, W. and Wolfowitz, J. (1948), Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* **19** 326-339.

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