Taylor & Francis
Taylor & Francis Group

Optimal Multistage Sampling in a Boundary-Crossing Problem

Jay Bartroff

Department of Statistics, Stanford University, Stanford, California, USA

Abstract: Brownian motion with known positive drift is sampled in stages until it crosses a positive boundary a. A family of multistage samplers that control the expected overshoot over the boundary by varying the stage size at each stage is shown to be optimal for large a, minimizing a linear combination of overshoot and number of stages. Applications to hypothesis testing are discussed.

Keywords: Asymptotic; Brownian motion; Group sequential; Multistage; Optimality.

Subject Classifications: 62L10; 91A20.

1. INTRODUCTION AND SUMMARY

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 Brownian motion $X(t)$, with known drift $\mu > 0$ and variance per unit time 1, is sampled in stages until $X(t) \geq a > 0$ 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 sampler is the sum of these costs upon first crossing the boundary. In this paper we describe a family of samplers and show they are first-order optimal as $a \to \infty$.

Many aspects of the boundary-crossing or "first-exit" problem are well studied, though without the multistage aspects considered here. The powerful methods of

renewal theory address successive exits and the time between such events (see Feller, 1971, pp. 358–388). Lorden (1970) obtained sharp, uniform bounds for the excess over the boundary of random walks. Siegmund (1985) discusses further applications in sequential analysis.

Schmitz (1993), Cressie and Morgan (1993), and Morgan and Cressie (1997) have proved general existence results for a large class of multistage sampling problems. In particular, the theorems of Schmitz show that an optimal sampler 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) and group-sequential (n-at-a-time) sampling, there have been few investigations of the performance of procedures that vary their sample size from stage to stage.

The families of samplers constructed below, $\delta_{m,h}^{o}$ and $\delta_{m,z}^{+}$, are shown to be first-order asymptotically optimal in Theorem 2.3. They have variable stage sizes that decrease roughly as successive iterations of the function $x \mapsto \sqrt{x \log x}$, while the average number of stages required is determined by the ratio of the cost per stage to the cost per unit time in relation to a family of critical functions, h_m. These critical functions define "critical bands"—i.e., regions of the first quadrant that are closely related to how close any efficient procedure can be to the boundary after each stage of sampling; Lemma 2.4 gives a precise "in-probability" lower bound on this distance. Theorem 2.4 then provides a converse statement to the optimality of $\delta_{m,h}^{o}$, $\delta_{m,z}^{+}$, showing that any competing sampler must use at least as many stages and follow the same "schedule" of $\delta_{m,h}^{o}$ and $\delta_{m,z}^{+}$, described in Lemma 2.4.

2. MULTISTAGE SAMPLERS

Define a *multistage sampling rule T* to be a sequence of nonnegative random variables $(T_1, T_2, \dots)$ such that, for $k \geq 1$

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

where $\mathscr{E}_t$ 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 $M \equiv \inf\{m \geq 1 : X(T^m) \geq a\}$, the total number of stages required to cross the boundary a. We will then describe a *multistage sampler* by the pair $\delta(a) = (T, M)$, where the argument is the initial distance to the boundary. When there is no confusion as to which sampler is being used, the shorthand $X_k \equiv X(T^k)$, $X_0 \equiv 0$ will be employed.

Let $c, d > 0$ denote the cost per unit time and cost per stage, respectively, and consider the problem of finding the multistage sampler $\delta(a) = (T, M)$ 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 sampler 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 "first-order" asymptotic theory.

To describe a sampler 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*

$$h_m(x) \equiv x^{(1/2)^m} (\log x)^{1/2 - (1/2)^m}$$

for $m = 1, 2, \ldots$ and $x \geq 1$, with $h_0(x) \equiv x$. 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

$$h \sim Q h_m(a)$$

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 $\mathscr{B}_m^o$ be the class of positive functions h such that $\{(a, h(a))\}$ is in the mth critical band (for every sequence of a's approaching ∞) and let $\mathscr{B}_m^+$ be the class of positive functions h such that $\{(a, h(a))\}$ is on the boundary between critical bands m and $m + 1$ (for every sequence of a's approaching ∞). That is,

$$\mathscr{B}_m^o \equiv \{h : (0, \infty) \to (0, \infty) \mid h_m(x) \ll h(x) \ll h_{m-1}(x) \text{ as } x \to \infty\},$$

$$\mathscr{B}_m^+ \equiv \left\{h : (0, \infty) \to (0, \infty) \,\bigg|\, \lim_{x \to \infty} \frac{h(x)}{h_m(x)} \in (0, \infty)\right\},$$

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

$$m \qquad \text{if } h \in \mathscr{B}_m^o$$
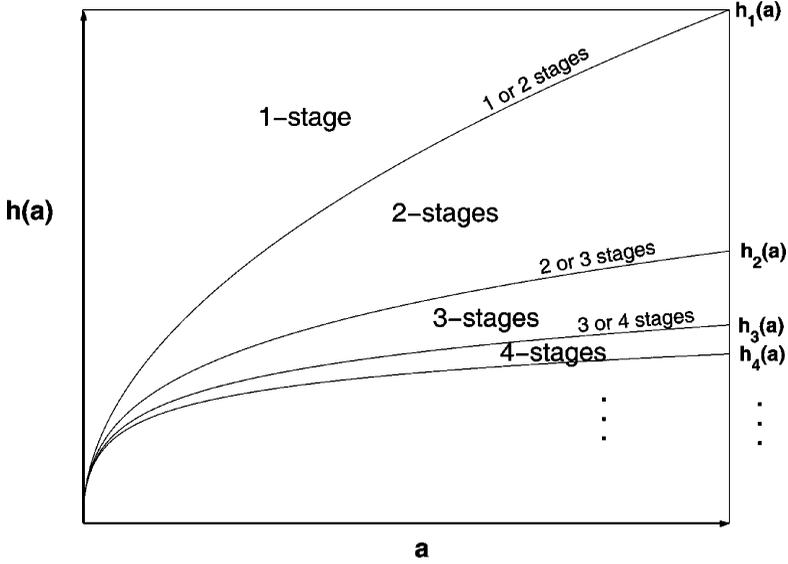$$m + \eta \quad \text{if } h \in \mathscr{B}_m^+,$$

Figure 1. The critical bands and the average number of stages used by an efficient procedure.

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

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

for a given $h \in \mathcal{B}_m$, some $m \geq 1$. Define the optimal sampler $\delta^*(a) = (T^*, M^*)$ to be one that achieves $R^*(a) \equiv \inf_\delta R(\delta(a))$. Note that, 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.

A convenient way of parametrizing stage sizes is by the upper standard normal quantile of the probability of stopping at the end of the stage. Thus, for $x > 0$ and $z \in \mathbb{R}$ let $t(x, z)$ be the unique solution of

$$\frac{x - \mu t(x, z)}{\sqrt{t(x, z)}} = z. \tag{2.7}$$

A simple computation gives

$$t(x, z) = x/\mu - \frac{z\sqrt{4x\mu + z^2} - z^2}{2\mu^2}.$$

Letting z_p denote the upper p-quantile of standard normal curve, the probability of being across a boundary x units away at the end of a stage of size $t(x, z_p)$ is p. An important asymptotic property of $t(x, z)$ is that

$$t(x, z) = x/\mu + O(z\sqrt{x})$$

as $x, |z| \to \infty$ if $|z| \ll \sqrt{x}$.

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 expected overshoot or undershoot. For example, letting $E(Y; A)$ denote $E(Y \cdot 1_A)$, if $Y \sim N(\lambda, \sigma^2)$ then

$$E(Y; Y \geq y) = \int_y^\infty P(Y > w)dw + y \cdot P(Y \geq y) \quad \text{integration by parts}$$

$$= \int_y^\infty \Phi\left(-\frac{w-\lambda}{\sigma}\right)dw + y \cdot \Phi\left(-\frac{y-\lambda}{\sigma}\right)$$

$$= \sigma \int_{(y-\lambda)/\sigma}^\infty \Phi(-v)dv + y \cdot \Phi\left(-\frac{y-\lambda}{\sigma}\right) \quad \text{change of variables}$$

$$= \sigma \cdot \Delta\left(\frac{y-\lambda}{\sigma}\right) + y \cdot \Phi\left(-\frac{y-\lambda}{\sigma}\right). \tag{2.8}$$

In particular, by taking $y = 0$, $\lambda = \mu t(x, z) - x = -z\sqrt{t(x, z)}$, and $\sigma = \sqrt{t(x, z)}$ in (2.8) we have

$$E[X(t(x, z)) - x; X(t(x, z)) \geq x] = \sqrt{t(x, z)} \cdot \Delta(z). \tag{2.9}$$

Useful asymptotic properties of Δ are

$$\Delta(z) \sim \begin{cases} |z|, & \text{as } z \to -\infty \\ \phi(z)/z^2, & \text{as } z \to +\infty. \end{cases} \tag{2.10}$$

The former is trivial while the latter follows from the classical expansion

$$\Phi(-z) = \frac{\phi(z)}{z}[1 - z^{-2} + O(z^{-4})]. \tag{2.11}$$

2.1. Geometric Sampling

For $z \in \mathbb{R}$, let $\dot{\delta}_z(a) = (T, M)$ be the sampler such that the probability of stopping at the end of each stage is constant at $p \equiv \Phi(-z)$ across the stages, i.e.,

$$T_k \equiv t(a - X_{k-1}, z) \cdot 1\{X_{k-1} < a\}.$$

Then M is a geometric random variable with mean $1/p$, and we thus refer to $\dot{\delta}_z(a)$ as *geometric sampling*. Although p is constant across the stages, we do allow it to vary with a, the initial distance to the boundary.

Not only is geometric sampling an interesting random process on its own, but it also has been conjectured that optimal multistage procedures share its stationarity property. While Theorem 2.4 will show this is not true, geometric sampling will prove to be a useful tool for designing the final stages of our optimal samplers in the next section.

Lemma 2.1 establishes a fundamental upper bound on the behavior of geometric sampling, and Lemma 2.2 gives an asymptotic bound on the overshoot of X under geometric sampling when the probability of stopping at each stage approaches 1. With the exception of the the main results, Theorems 2.3 and 2.4, the proofs of all theorems and lemmas can be found in the appendix.

Lemma 2.1. *Let* $z \in \mathbb{R}$, $q = \Phi(z)$, *and*

$$g(x) \equiv \frac{\Delta(-z)}{qz} \cdot (x - \mu t(x, z)) = \frac{\Delta(-z)}{2\mu q}\left(\sqrt{4x\mu + z^2} - z\right). \qquad (2.12)$$

Then $\dot{\delta}_z(a) = (T, M)$ *satisfies*

$$ET - a/\mu \leq \begin{cases} \dfrac{q\Delta(z)}{\mu\Delta(-z)} \cdot g(a) + \mu^{-1}\displaystyle\sum_{k\geq 2} g^{(k)}(a)q^k, & \text{if } z \geq 0 \\[2ex] \dfrac{q\Delta(z)}{\mu\Delta(-z)} \displaystyle\sum_{k\geq 1} g^{(k)}(a)q^{k-1}, & \text{if } z \leq 0, \end{cases} \qquad (2.13)$$

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

Lemma 2.2. *Let* $z(x)$ *be a nonpositive function such that* $\lim_{x\to\infty} z(x) = -\infty$ *and* $|z(x)| = o(\sqrt{x})$ *as* $x \to \infty$. *Then* $\dot{\delta}_{z(a)}(a) = (T, M)$ *satisfies*

$$EX_M - a = O\big(|z(a)|\sqrt{a}\big)$$

as $a \to \infty$.

2.2. The Samplers $\delta^o_{m,h}$ and $\delta^+_{m,z}$

In this section we define two families of samplers that will later be shown to be first-order optimal under different conditions. Namely, the sampler $\delta^o_{m,h}$ will be optimal when $h \in \mathscr{B}^o_m$ and $\delta^+_{m,z}$ will be optimal when $h \in \mathscr{B}^+_m$.

Let $f(x) = (6/\sqrt{\mu})\sqrt{x\log(x+1)}$ and note that f^{-1} is well defined since f is increasing. The family of samplers $\delta^o_{m,h}(a)$ are indexed by a positive integer m and a positive function h, and the argument a is the initial distance to the boundary. We define the family of samplers inductively on m as follows. Letting $\wedge$ denote min,

$$\delta^o_{1,h}(a) \equiv \dot{\delta}_{\zeta(a)}(a), \quad \text{where } \zeta(x) = -\left(\sqrt{h(x)}/x^{1/4} \wedge x^{1/7}\right).$$

$$\delta^o_{m+1,h}(a) \equiv \text{1st stage } t\left(a, \sqrt{\log(a/h(a)^2 + 1)}\right),$$

$$\text{followed (if necessary) by } \delta^o_{m,h\circ f^{-1}}(a - X_1).$$

The family of samplers $\delta^+_{m,z}(a)$ are indexed by a positive integer m and a number $z \in \mathbb{R}$. They are defined inductively on m as follows:

$$\delta^+_{1,z}(a) \equiv \text{1st stage } t(a, z), \text{ followed (if necessary) by } \dot{\delta}_{\nu(a-X_1)}(a - X_1),$$

$$\text{where } \nu(x) = -\sqrt{\log(x+1)}.$$

$$\delta^+_{m+1,z}(a) \equiv \text{1st stage } t(a, \sqrt{(1 - 2^{-m})\log(a + 1)}),$$

$$\text{followed (if necessary) by } \delta^+_{m,z}(a - X_1).$$

The main theorems in this section establish the operating characteristics of $\delta_{m,h}^o$ and $\delta_{m,z}^+$. Theorem 2.1 will show that $\delta_{m,h}^o$ uses m stages almost always and has overshoot bounded by h, asymptotically. Theorem 2.2 will show that $\delta_{m,z}^+$ uses m or $m+1$ stages almost always, and that the probability of using m stages has upper quantile approaching z.

Theorem 2.1. *If $h \in \mathcal{B}_m^o$ then $\delta_{m,h}^o(a) = (T, M)$ satisfies*

$$ET - a/\mu = o(h(a)) \tag{2.14}$$

$$EM \to m \tag{2.15}$$

as $a \to \infty$.

Before establishing the operating characteristics of $\delta_{m,z}^+$ in Theorem 2.2, we introduce a family of positive constants that appear in the first-order overshoot of $\delta_{m,z}^+$. For $m \geq 1$ define

$$\kappa_m = \kappa_m(\mu) = \mu^{-2+(1/2)^m} \prod_{i=1}^{m-1} \left[(1/2)^{m-1-i} - (1/2)^{m-1} \right]^{(1/2)^{i+1}}. \tag{2.16}$$

The next lemma contains the key property of the κ_m that we will need.

Lemma 2.3. *For $m \geq 1$, as $x \to \infty$*

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

We adopt the notation $F \lesssim G$ to denote $F \leq (1 + o(1)) \cdot G$. Theorem 2.2 establishes the operating characteristics of $\delta_{m,z}^+$.

Theorem 2.2. *Let $m \in \mathbb{N}$ and $z \in \mathbb{R}$. Then $\delta_{m,z}^+(a) = (T, M)$ satisfies*

$$ET - a/\mu \lesssim \Delta(z)\kappa_m h_m(a) \tag{2.17}$$

$$EM \to m + \Phi(z) \tag{2.18}$$

as $a \to \infty$.

2.3. Optimality of $\delta_{m,h}^o$ and $\delta_{m,z}^+$

In this section we state our main optimality results of the paper, Theorems 2.3 and 2.4. Theorem 2.3 shows that $\delta_{m,h}^o$ is first-order optimal when $h \in \mathcal{B}_m^o$ and that δ_{m,z^*}^+ is first-order optimal when $h \in \mathcal{B}_m^+$, where the value z^* is determined by $\lim_{x \to \infty} h(x)/h_m(x)$. Theorem 2.4 provides a converse to Theorem 2.3, showing that any efficient sampler must behave like $\delta_{m,h}^o$ and δ_{m,z^*}^+.

Theorem 2.3. *If $h \in \mathcal{B}_m^o$, then*

$$R(\delta_{m,h}^o(a)) \sim m \cdot h(a) \sim R^*(a) \tag{2.19}$$

as $a \to \infty$. If $h \in \mathscr{B}_m^+$, then

$$R(\delta_{m,z^*}^+(a)) \sim \left[m + \Phi(z^*) + \Delta(z^*) \cdot \frac{\phi(z^*)}{1 - \Phi(z^*)} \right] h(a) \sim R^*(a) \qquad (2.20)$$

as $a \to \infty$, where z^* is the unique solution of

$$\frac{\phi(z^*)}{1 - \Phi(z^*)} = \lim_{x \to \infty} \frac{\kappa_m h_m(x)}{h(x)}. \qquad (2.21)$$

Before proving Theorem 2.3 we introduce a family of iterated functions $F_h^{(k)}$ that are the order of magnitude of the best-possible undershoot after k stages, $a - X_k$. Lemma 2.4 makes this explicit and Lemma 2.5 establishes a link between the $F_h^{(k)}$ and the critical functions h_k.

For $x \geq y^2 > 0$ define $F_y(x) = \sqrt{x \log(x/y^2)}$. For a positive function h and x such that $h^2(x) \leq x$ define

$$F_h^{(k)}(x) \equiv F_y^{(k)}(x)\big|_{y=h(x)}.$$

Note that $h(\cdot)$ is not iterated, e.g.,

$$F_h^{(2)}(x) = \left[x \log(x/h^2(x)) \right]^{1/4} \sqrt{\log \left(\frac{\sqrt{x \log(x/h^2(x))}}{h^2(x)} \right)}$$

$$\neq \left[x \log \left(x/h^2 \left(\sqrt{x \log(x/h^2(x))} \right) \right) \right]^{1/4} \sqrt{\log \left(\frac{\sqrt{x \log(x/h^2(x))}}{h^2(\sqrt{x \log(x/h^2(x))})} \right)}.$$

Lemma 2.4 establishes an "in probability" lower bound of the order $F_h^{(k)}$ on how close to the boundary any efficient sampler can be after each of the first $m - 1$ stages when $h \in \mathscr{B}_m$.

Lemma 2.4. *If $h \in \mathscr{B}_m$ and δ is any sampler such that $R(\delta(a)) = O(h(a))$, then for any $\varepsilon > 0$ and $0 \leq k \leq m - 1$*

$$P\left(a - X_k \geq (1 - \varepsilon) \cdot (1/\mu)^{1-2^{-k}} F_h^{(k)}(a) \right) \to 1 \qquad (2.22)$$

as $a \to \infty$.

Lemma 2.5 shows that, when $h \in \mathscr{B}_m$, square roots of the iterates $F_h^{(k-1)}(a)$ are roughly constant multiples of the critical functions h_k. The constants themselves are given by the solutions of the following recurrence relation. For $1 \leq k \leq m$, define C_k^m to be the unique solution of

$$C_{k+1}^m = \sqrt{C_k^m} \cdot \left[(1/2)^{k-1} - (1/2)^{m-1} \right]^{1/4}; \quad C_1^m = 1. \qquad (2.23)$$

After taking logarithms, solving (2.23) amounts to solving a difference equation. This computation gives

$$C_k^m = \prod_{i=1}^{k-1} \left[(1/2)^{k-1-i} - (1/2)^{m-1} \right]^{(1/2)^{i+1}}, \qquad (2.24)$$

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

$$\kappa_m = (1/\mu)^{2-(1/2)^m} C_m^m. \tag{2.25}$$

Lemma 2.5. *If* $h \in \mathscr{B}_m^+$, *then*

$$\sqrt{F_h^{(k-1)}(x)} \sim C_k^m h_k(x) \quad \text{as } x \to \infty, \text{ for } 1 \le k \le m. \tag{2.26}$$

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

$$C_k^{m-1} \lesssim \frac{\sqrt{F_h^{(k-1)}(a)}}{h_k(x)} \lesssim C_k^m \quad \text{as } x \to \infty, \text{ for } 1 \le k < m. \tag{2.27}$$

Proof of Theorem 2.3. Assume that $h \in \mathscr{B}_m^o$. The left-hand side of (2.19) holds by Theorem 2.1. Now

$$R^*(a) \le R(\delta_{m,h}^o(a)) \quad \text{the Bayes property}$$
$$= O(h(a))$$

by Theorem 2.1, so Lemma 2.4 applies to $\delta^*(a)$. Letting X_k^* denote the δ^*-sampled process,

$$R^*(a) \ge h(a)EM^*$$
$$\ge h(a)mP(M^* \ge m)$$
$$= h(a)mP(a - X_{m-1}^* > 0)$$
$$\ge h(a)mP\big(a - X_{m-1}^* \ge (1/2)(1/\mu)^{1-2^{-(m-1)}} F_h^{(m-1)}(a)\big)$$
$$\sim h(a)m \cdot 1$$

by Lemma 2.4, proving (2.19).

If $h \in \mathscr{B}_m^+$ then $\lim_{x \to \infty} h(x)/h_m(x)$ is positive and finite, hence

$$\lim_{x \to \infty} \frac{\kappa_m h_m(x)}{h(x)} = \kappa_m \left[\lim_{x \to \infty} \frac{h(x)}{h_m(x)} \right]^{-1}$$

is positive and finite as well. The function

$$z \mapsto \frac{\phi(z)}{1 - \Phi(z)}$$

increases from 0 to ∞ as z ranges from $-\infty$ to ∞, so equation (2.21) has a unique solution, z^*. Theorem 2.2 shows that

$$R(\delta_{m,z^*}^+(a)) \lesssim \Delta(z^*)\kappa_m h_m(a) + h(a)[m + \Phi(z^*)]$$
$$\sim \Delta(z^*)\frac{\phi(z^*)}{1 - \Phi(z^*)}h(a) + h(a)[m + \Phi(z^*)] \quad \text{by definition of } z^*$$
$$= \left[m + \Phi(z^*) + \Delta(z^*)\frac{\phi(z^*)}{1 - \Phi(z^*)} \right]h(a).$$

Suppose $\varepsilon > 0$. Again $R^*(a) \leq R(\delta^+_{m,z^*}(a)) = O(h(a))$, so by Lemma 2.4

$$P\big(a - X^*_{m-1} \geq (1 - \varepsilon)(1/\mu)^{1-2^{-(m-1)}} F_h^{(m-1)}(a)\big) \to 1. \tag{2.28}$$

Let $(T^{*(m)}, M^{*(m)})$ denote the continuation of δ^* after the $(m-1)$st stage, i.e.,

$$M^{*(m)} = [M^* - (m-1)]^+,$$
$$T^{*(m)} = T^* - (T^*_1 + \cdots + T^*_{m-1}),$$

and for $y > 0$ define

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

We will show below that $\varphi(y)$ is nondecreasing in y. Let

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

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

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

we have

$$\begin{aligned}
\mu^{-1} &E(X_{M^{*(m)}} - (a - X^*_{m-1}) \mid a - X^*_{m-1} = \gamma) \\
&\geq \mu^{-1} E\big[(X_{M^{*(m)}} - (a - X^*_{m-1}))1\{M^{*(m)} = 1\} \mid a - X^*_{m-1} = \gamma\big] \\
&= \mu^{-1}\Delta(z_p)\sqrt{t(\gamma, z_p)} \\
&\sim \mu^{-1}\Delta(z_p)\sqrt{\gamma/\mu} \\
&= \mu^{-1}\Delta(z_p)\sqrt{(1 - \varepsilon)(1/\mu)^{2-2^{-m+1}} F_h^{(m-1)}(a)} \\
&\sim \Delta(z_p)\sqrt{1 - \varepsilon} \cdot (1/\mu)^{2-2^{-m}} C_m^m h_m(a) \quad \text{by Lemma 2.8} \\
&= \Delta(z_p)\sqrt{1 - \varepsilon} \cdot \kappa_m h_m(a) \quad \text{by (2.25)} \\
&\sim \Delta(z_p)\sqrt{1 - \varepsilon} \cdot \frac{\phi(z^*)}{1 - \Phi(z^*)} h(a), \tag{2.29}
\end{aligned}$$

by definition of z^*. Also, $E(M^{*(m)} \mid a - X^*_{m-1} = \gamma) \geq 2 - p$, and combining this with (2.29) gives, for sufficiently large a,

$$\begin{aligned}
\varphi(\gamma) &\geq \left[\Delta(z_p)\sqrt{1 - \varepsilon} \cdot \frac{\phi(z^*)}{1 - \Phi(z^*)} + 2 - p\right]h(a) \cdot (1 - \varepsilon) \\
&\geq \left[\Delta(z_p)\frac{\phi(z^*)}{1 - \Phi(z^*)} + 2 - p\right]h(a) \cdot (1 - 2\varepsilon). \tag{2.30}
\end{aligned}$$

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

$$\begin{aligned}
R^*(a) &= E[\mu^{-1}(X_{M^*} - a) + h(a)M^*] \\
&\geq E[\mu^{-1}(X_{M^{*(m)}} - Y) + h(a)(m - 1 + M^{*(m)}); V]
\end{aligned}$$

$$\geq E[\varphi(Y) + (m-1)h(a); V]$$

$$\geq [\varphi(\gamma) + (m-1)h(a)]P(V) \quad (\varphi \text{ nondecreasing})$$

$$\geq \left[\Delta(z_p)\frac{\phi(z^*)}{1-\Phi(z^*)} + m + 1 - p\right]h(a) \cdot (1-2\varepsilon) \cdot P(V) \quad \text{by (2.30)}$$

$$\geq \left[\Delta(z_p)\frac{\phi(z^*)}{1-\Phi(z^*)} + m + 1 - p\right]h(a) \cdot (1-3\varepsilon) \tag{2.31}$$

for sufficiently large a since $P(V) \to 1$ by (2.28). Using basic calculus, it can be shown that the expression in brackets in (2.31) achieves its unique minimum when $p = 1 - \Phi(z^*)$, hence

$$R^*(a) \geq \left[\Delta(z^*)\frac{\phi(z^*)}{1-\Phi(z^*)} + m + \Phi(z^*)\right]h(a) \cdot (1-3\varepsilon)$$

for sufficiently large a. Since ε was arbitrary, this completes the proof of (2.20) and hence the theorem once we verify that $\varphi(\cdot)$ is nondecreasing.

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 probability of being over the boundary at the end of each stage as $(T^{*(m)}, M^{*(m)})$ when starting from $a - X^*_{m-1} = y'$. Then

$$E\big(M'^{(m)} \mid a - X^*_{m-1} = y\big) = E\big(M^{*(m)} \mid a - X^*_{m-1} = y'\big) \tag{2.32}$$

and, letting

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

we have

$$E\big[(X_{M'^{(m)}} - y)1\{M'^{(m)} = 1\} \mid a - X^*_{m-1} = y\big]$$

$$= \Delta(z_{p_1})\sqrt{t(y, z_{p_1})}$$

$$\leq \Delta(z_{p_1})\sqrt{t(y', z_{p_1})} \quad \text{since } y \leq y'$$

$$= E\big[(X_{M^{*(m)}} - y')1\{M^{*(m)} = 1\} \mid a - X^*_{m-1} = y'\big].$$

Similar arguments inductively give

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

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

and these last two bounds show

$$E\big(X_{M'^{(m)}} - y \mid a - X^*_{m-1} = y\big) \leq E\big(X_{M^{*(m)}} - y' \mid a - X^*_{m-1} = y'\big). \tag{2.33}$$

Then

$$\varphi(y) \leq E\big[\mu^{-1}(X_{M'^{(m)}} - y) + h(a)M'^{(m)} \mid a - X^*_{m-1} = y\big] \quad \text{(optimality of } (T^{*(m)}, M^{*(m)}))$$

$$\leq E[\mu^{-1}(X_{M^{*(m)}} - y') + h(a)M^{*(m)} \mid a - X^*_{m-1} = y'] \quad \text{(by (2.32) and (2.33))}$$

$$= \varphi(y'),$$

finishing the proof. $\qquad \square$

The final result of this section is a type of converse to Theorem 2.3, showing that good samplers must behave like $\delta_{m,h}^o$, $\delta_{m,z}^+$ in not only the sense that m stages are necessary when $h \in \mathscr{B}_m$, but also that any efficient sampler must follow the same "schedule" that $\delta_{m,h}^o$ and $\delta_{m,z}^+$ follow for the first $m-1$ stages, described in Lemma 2.4.

Theorem 2.4. *Assume that $h \in \mathscr{B}_m$ and let*

$$\delta(a) = \begin{cases} \delta_{m,h}^o(a) & \text{if } h \in \mathscr{B}_m^o \\ \delta_{m,z^*}^+(a) & \text{if } h \in \mathscr{B}_m^+, \end{cases}$$

where z^* is as in (2.21). If $\delta'(a) = (T, M)$ is a sampler such that there is a sequence $a_i \to \infty$ with

$$P\left(a_i - X_k \geq (1-\varepsilon)(1/\mu)^{1-2^{-k}} F_h^{(k)}(a_i)\right) \quad \text{bounded below } 1 \tag{2.34}$$

for some $1 \leq k < m$ and $\varepsilon > 0$, then

$$\frac{R(\delta'(a_i)) - R(\delta(a_i))}{h(a_i)} \to +\infty \tag{2.35}$$

as $i \to \infty$. In particular, (2.35) holds if $P(M \geq m) \not\to 1$.

Proof.   If (2.34) holds, then Lemma 2.4 implies that

$$\frac{R(\delta'(a_i))}{h(a_i)} \to \infty$$

as $i \to \infty$. We know that $R(\delta(a_i)) = O(h(a_i))$ by Theorem 2.3, so

$$\frac{R(\delta'(a_i)) - R(\delta(a_i))}{h(a_i)} = \frac{R(\delta'(a_i))}{h(a_i)} - O(1) \to +\infty$$

as $i \to \infty$, proving the first claim.

   If

$$P\left(a - X_k \geq (1-\varepsilon)(1/\mu)^{1-2^{-k}} F_h^{(k)}(a)\right) \to 1$$

for all $1 \leq k < m$, then

$$P(M \geq m) = P(a - X_{m-1} > 0)$$
$$\geq P\left(a - X_{m-1} \geq (1-\varepsilon)(1/\mu)^{1-2^{-(m-1)}} F_h^{(m-1)}(a)\right) \to 1,$$

proving the second assertion. $\square$

3.   AN APPLICATION TO HYPOTHESIS TESTING

In this section we discuss how the above multistage samplers can be used to construct efficient multistage hypothesis tests and give a numerical example of the

performance of such a test. Let $Y_1, Y_2, \ldots$ be independent and identically distributed with density f and consider testing the simple hypotheses

$$H_0 : f = f_0 \quad \text{vs.} \quad H_1 : f = f_1 \tag{3.1}$$

in stages. We can describe multistage tests of these hypotheses by triples (N, M, D), where N is the total number of observations, M is the total number of stages used, and D is the decision variable, taking values in $\{0, 1\}$. One measure of the performance of (N, M, D) in testing H_0, H_1 is the *integrated risk*, which we define as

$$r(N, M, D) = \sum_{i=0}^{1} [cE_i N + dE_i M + w_i P_i (D \neq i)] \pi_i, \tag{3.2}$$

where $0 < c, d < 1$ represent the cost per observation and cost per stage, π_i is the prior distribution on $\{f_0, f_1\}$, and $w_i > 0$ represent the penalty for a wrong decision. Here E_i and P_i denote expectation and probability under f_i.

The multistage samplers discussed above, with some simple modifications (like making sure each stage size is an integer), tell us how to sample the Y_j in stages by observing some random process until it crosses a boundary. The relevant random process here, taking the place of the Brownian motion above, is the log-likelihood process, which we now define.

Assume that

$$E_i \log^2 \left(\frac{f_i(Y_1)}{f_{1-i}(Y_1)} \right) < \infty \quad \text{for } i = 0, 1$$

and let

$$\sigma_i^2 = \text{Var}_i \log \left(\frac{f_i(Y_1)}{f_{1-i}(Y_1)} \right), \quad \mu_i = \sigma_i^{-1} E_i \log \left(\frac{f_i(Y_1)}{f_{1-i}(Y_1)} \right).$$

Define the log-likelihood process

$$X_i(n) = \sigma_i^{-1} \sum_{j=1}^{n} \log \left(\frac{f_i(Y_j)}{f_{1-i}(Y_j)} \right).$$

Note that, like the Brownian motion above, $E_i X_i(n) = n\mu_i > 0$ and $\text{Var}_i X_i(n) = n$. Thus, we can use a given multistage sampler δ' with boundary a to sample the Y_j by treating f_i as the true, underlying density and sampling according to δ' until $X_i(n) \geq a$ at the end of a stage. If a is large, this is compelling evidence that f_i is indeed the true density, and thus provides us with a "one-decision" test of hypotheses (3.1). We can construct an ordinary "two-decision" test from two samplers, δ_0 and δ_1, by somehow choosing a first-stage size, computing the maximum likelihood estimated of the true hypothesis $\hat{\imath} \in \{0, 1\}$ from the data observed in that first stage, and then continuing with δ_i and a slightly modified stopping rule (e.g., "stop when $|X_i(n)| \geq a$") to protect against an error in $\hat{\imath}$. A natural choice for the size of the first stage is the minimum of the two stage sizes dictated by δ_0, δ_1.

Let δ be the test constructed in this manner from the two samplers $\delta^o_{m_i^*, d/c}(a_i)$, $i = 0, 1$, where $a_i \equiv \sigma_i^{-1} \log d^{-1}$ and

$$m_i^* \equiv \inf\{m \geq 1 : \kappa_m(\mu_i) h_m(a_i) \leq d/c \leq \kappa_{m+1}(\mu_i) h_{m+1}(a_i)\}.$$

Table 1. Results for testing $\mu = .25$ vs. $\mu = -.25$ about a Gaussian mean, with $d = 0.001$, $\pi_i = 1/2$, $w_i = 1$

Procedure	EN	EM	r	$r(\delta)/r$ (%)
		$d/c = 1$		
δ	62.2	5.2	0.068	100
$\delta_g(1)$	57.5	57.5	0.115	59.1
$\delta_g(15)$	64.9	4.6	0.073	92.9
$\delta_g(30)$	76.7	2.6	0.080	85.0
		$d/c = 5$		
δ	68.3	2.9	0.017	100
$\delta_g(1)$	57.5	57.5	0.070	24.0
$\delta_g(22)$	72.7	3.3	0.018	92.8
$\delta_g(44)$	83.6	1.9	0.019	88.4
		$d/c = 10$		
δ	76.6	1.9	0.0097	100
$\delta_g(1)$	57.5	57.5	0.0644	15.1
$\delta_g(37)$	80.5	2.2	0.0104	93.3
$\delta_g(74)$	97.6	1.3	0.0112	86.6

Bartroff (2004, 2006) shows that, under general conditions on the f_i, δ minimizes the integrated risk (3.2) to second order as $c, d \to 0$ at specified rates. Bartroff (2004) and a third paper by the author in this series will extend these results to composite hypotheses about the parameter of an exponential family. See these references for more details on the testing problem.

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. $\mu = -.25$, about the mean of Gaussian random variables with unit variance. $\delta_g(k)$ denotes group-sequential testing with constant stage-size k, which samples until

$$\left| \sum_j \log \left(\frac{f_1(Y_j)}{f_0(Y_j)} \right) \right| \geq \log d^{-1}$$

at the end of a stage, which is equivalent to the stopping rule of δ. For each value of d/c, the operating characteristics of $\delta_g(k)$ are given for three values of $k : k = 1$, the best possible k (determined by simulation), and two times the best possible k. The results show significant improvement in the integrated risk of the variable stage-size test δ upon δ_g. This improvement decreases for large values of d/c, but this is to be expected since the number of stages of any reasonable test will approach 1 in this limit.

Here we constructed tests from the samplers $\delta_{m,h}^o$. In practice, tests constructed from the samplers $\delta_{m,z}^+$ also perform well and behave almost identically as those constructed from $\delta_{m,h}^o$. The choice here was made merely to simplify presentation.

APPENDIX A: PROOFS OF THEOREMS AND LEMMAS

Proof of Lemma 2.1.   First we will prove

$$E(a - X_k \mid M > k) \leq g^{(k)}(a) \quad \text{for all } k \geq 0. \tag{A.1}$$

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

$$E(a - X_{k+1} \mid M > k + 1, X_k) = E[(a - X_k) - (X_{k+1} - X_k) \mid M > k + 1, X_k]$$
$$= \Delta(-z)\sqrt{t(a - X_k, z)}/q$$
$$= \Delta(-z)\frac{(a - X_k) - \mu t(a - X_k, z)}{qz}$$
$$= g(a - X_k).$$

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

$$E(a - X_{k+1} \mid M > k + 1) = E(g(a - X_k) \mid M > k + 1)$$
$$\leq g(E(a - X_k \mid M > k + 1))$$
$$= g(E(a - X_k \mid M > k))$$
$$\leq g(g^{(k)}(a)) = g^{(k+1)}(a), \tag{A.2}$$

proving (A.1). In (A.2) we use that $E(a - X_k \mid M > k + 1) = E(a - X_k \mid M > k)$; 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 (2.13). Let $p = 1 - q = \Phi(-z)$. Assume first that $z \geq 0$ so that $p \leq 1/2$. $E(T_1 \mid M \geq 1) = t(a, z)$ and for $k \geq 2$,

$$E(T_k \mid M \geq k) = E(t(a - X_{k-1}, z) \mid M > k - 1)$$
$$\leq \mu^{-1}E(a - X_{k-1} \mid M > k - 1) \quad \text{since } p \leq 1/2$$
$$\leq \mu^{-1}g^{(k-1)}(a)$$

by (A.1). Using these two relations

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

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

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

using the relation between g and $t(\cdot, z)$ in (2.12).

Now assume $z \leq 0$. Consequently, $t(\cdot, z)$ is concave, so using Jensen's inequality and (A.1),

$$E(T_k \mid M \geq k) = E[t(a - X_{k-1}, z) \mid M > k - 1]$$
$$\leq t(E[a - X_{k-1} \mid M > k - 1], z) \leq t(g^{(k-1)}(a), z)$$

and, as computed above,

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

again using (2.12) for the final step. $\qquad\qquad\square$

Proof of Lemma 2.2. Let $z = z(a)$ and g be as in Lemma 2.1 with $p = \Phi(-z)$ and $q = 1 - p$. A simple computation shows that $g(a)$ has a unique positive fixed point

$$x^* = x^*(a) = \frac{\Delta(-z)\phi(z)}{\mu q^2}$$

so that $g(a) \leq (a \vee x^*)$, where $\vee$ denotes max. Then

$$EX_M - a = \mu(ET - a/\mu) \quad \text{Wald's equation}$$

$$\leq \frac{q\Delta(z)}{\Delta(-z)} \sum_{k=1}^{\infty} g^{(k)}(a) q^{k-1} \quad \text{by Lemma 2.1}$$

$$\leq \frac{q\Delta(z)}{\Delta(-z)} \cdot \begin{cases} \sum_{k=1}^{\infty} g(a) q^{k-1} = g(a)/p \leq 2g(a) & \text{when } a > x^*, \\[2mm] \sum_{k=1}^{\infty} x^* q^{k-1} = x^*/p \leq 2x^* & \text{when } a \leq x^*, \end{cases}$$

$$\text{since } g(a) \leq (a \vee x^*)$$

$$\leq 2 \cdot \frac{q\Delta(z)}{\Delta(-z)} \cdot (g(a) \vee x^*). \qquad\qquad (A.3)$$

Now

$$\frac{q\Delta(z)}{\Delta(-z)} \cdot g(a) = \frac{\Delta(z)}{2\mu}\left(\sqrt{4a\mu + z^2} - z\right) \sim \frac{|z|}{2\mu} \cdot O(\sqrt{a}) = O(|z|\sqrt{a}) \qquad (A.4)$$

by (2.10) and since $|z| = o(\sqrt{a})$. Also,

$$\frac{q\Delta(z)}{\Delta(-z)} \cdot x^* = \frac{q\Delta(z)}{\Delta(-z)} \cdot \frac{\Delta(-z)\phi(z)}{\mu q^2} = \frac{\Delta(z)\phi(z)}{\mu q}$$

$$\sim \frac{|z|\phi(z)}{\mu\Phi(z)} \quad \text{by (2.10) and since } q = \Phi(z)$$

$$= O(z^2) \quad \text{by (2.11)}$$

$$= O(|z|\sqrt{a}) \qquad\qquad (A.5)$$

since $|z| = o(\sqrt{a})$. Plugging (A.4) and (A.5) into (A.3) gives the claim. $\qquad\square$

Proof of Theorem 2.1. We proceed by induction on m. For $m = 1$, assume $h \in \mathscr{B}_1^o$ and let $\delta_{1,h}^o(a) = (T, M)$. Note that

$$|\zeta(a)| \leq \frac{\sqrt{h(a)}}{a^{1/4}} = \frac{\sqrt{h(a)/a}}{a^{1/4}} \cdot \sqrt{a} = o(1) \cdot \sqrt{a} = o(\sqrt{a}),$$

so Lemma 2.2 applies. Then

$$
\begin{aligned}
ET - a/\mu = \mu^{-1}(EX_M - a) \quad &\text{by Wald's equation} \\
= O(|\zeta(a)|\sqrt{a}) \quad &\text{by Lemma 2.2} \\
= O\left(\frac{\sqrt{h(a)}}{a^{1/4}} \cdot \sqrt{a}\right) \quad &\text{by definition of } \zeta \\
= O\left(\sqrt{\frac{a^{1/2}}{h(a)}} \cdot h(a)\right) \\
= o(1) \cdot h(a) \quad &\text{by virtue of } h \in \mathscr{B}_1^o \\
= o(h(a)),
\end{aligned}
$$

showing that the theorem holds for $m = 1$.

Now assume that $h \in \mathscr{B}_{m+1}^o$ and let $\delta_{m+1,h}^o(a) = (T, M)$. Let $\xi = \sqrt{\log(a/h(a)^2 + 1)}$ and $t = t(a, \xi)$, the size of the first stage of $\delta_{m+1,h}^o(a)$. Note that

$$\frac{a}{h(a)^2} = \left(\frac{h_m(a)}{h(a)}\right)^2 \cdot \frac{a}{h_m(a)^2} \geq \left(\frac{h_m(a)}{h(a)}\right)^2 \cdot \frac{a}{h_1(a)^2} = \left(\frac{h_m(a)}{h(a)}\right)^2 \cdot 1 \to \infty$$

so that $\xi \to \infty$, which is important in the definition of the first stage of $\delta_{m+1,h}^o$. Using the definition of f, as $x \to \infty$

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

or equivalently, $h_m = o(h \circ f^{-1})$. A similar argument shows $h \circ f^{-1} = o(h_{m-1})$, which shows that

$$h \circ f^{-1} \in \mathscr{B}_m^o. \tag{A.6}$$

For $y > 0$ let $\delta_{m, h \circ f^{-1}}^o(y) = (T', M')$ and

$$\psi(y) = EM'$$
$$\varphi(y) = EX_{M'} - y.$$

Suppose $\varepsilon > 0$. By the induction hypothesis there is a constant y_o such that

$$|\psi(y) - m| \leq \varepsilon \tag{A.7}$$
$$\varphi(y) \leq \varepsilon \cdot h(f^{-1}(y)) \tag{A.8}$$

for all $y \geq y_o$, where (A.8) uses (2.14) and Wald's equation.

$$EM - 1 = E(M - 1; 0 < a - X_1 < y_o) + E(M - 1; a - X_1 \geq y_o)$$
$$= E(\psi(a - X_1); 0 < a - X_1 < y_o) + E(\psi(a - X_1); X_1 \geq y_o). \quad \text{(A.9)}$$

Letting $A = \sup_{0 < y < y_o} \psi(y) < \infty$,

$$E(\psi(a - X_1); 0 < a - X_1 < y_o) \leq A \cdot P(0 < a - X_1 < y_o)$$
$$\leq A \cdot P(X_1 > a - y_o)$$
$$= A \cdot P\left(\frac{X_1 - \mu t}{\sqrt{t}} > \frac{a - \mu t}{\sqrt{t}} - \frac{y_o}{\sqrt{t}}\right)$$
$$= A \cdot P\left(\frac{X_1 - \mu t}{\sqrt{t}} > \xi + o(1)\right)$$
$$= A \cdot o(1) = o(1), \quad \text{(A.10)}$$

by, say, Chebyshev's inequality since $\xi \to \infty$. Plugging this into (A.9),

$$|EM - (m + 1)| \leq o(1) + |E(\psi(a - X_1) - m; a - X_1 \geq y_o)| + m \cdot P(a - X_1 \leq y_o)$$
$$\leq o(1) + E(|\psi(a - X_1) - m|; a - X_1 \geq y_o) + m \cdot o(1)$$
$$\leq o(1) + \varepsilon \cdot P(a - X_1 \geq y_o) + o(1) \quad \text{by (A.7)}$$
$$\leq \varepsilon + \varepsilon + \varepsilon = 3\varepsilon$$

for sufficiently large a. This shows that $EM \to m + 1$.

To show that (2.15) holds for $m + 1$, first write

$$\mu E(T - a/\mu) = EX_M - a = E(X_M - a; M = 1) + E(X_M - a; 0 < a - X_1 < y_o)$$
$$+ E(X_M - a; y_o \leq a - X_1 \leq (1/2)f(a))$$
$$+ E(X_M - a; a - X_1 > (1/2)f(a))$$
$$\equiv I + II + III + IV. \quad \text{(A.11)}$$

First,

$$I = E(X_1 - a; X_1 \geq a) = \sqrt{t} \cdot \Delta(\xi) \quad \text{by (2.9)}$$
$$\sim \sqrt{t} \cdot \frac{\phi(\xi)}{\xi^2} \quad \text{by (2.10)}$$
$$= O(\sqrt{a}) \cdot \frac{O(h(a)/\sqrt{a})}{\xi^2}$$
$$\text{(since } t \sim a/\mu \text{ and } \phi(\xi) \propto h/\sqrt{a})$$
$$= o(h(a)). \quad \text{(A.12)}$$

Next we have

$$II = E(\varphi(a - X_1); 0 < a - X_1 < y_o) = o(1) \quad \text{(A.13)}$$

by the same argument leading to (A.10).

Before considering *III*, note that we may assume without loss of generality that h is nondecreasing. Otherwise, we could replace h by $\underline{h}(x) \equiv \inf_{y \geq x} h(y)$ throughout the proof, since $\underline{h}$ is nondecreasing and bounded above by h. Since f^{-1} is also nondecreasing, $h \circ f^{-1}$ is thus non decreasing. Now

$$
\begin{aligned}
III &= E(\varphi(a - X_1); y_o \leq a - X_1 \leq (1/2)f(a)) \\
&\leq E(\varepsilon \cdot h(f^{-1}(a - X_1)); y_o \leq a - X_1 \leq (1/2)f(a)) \quad \text{by (A.8)} \\
&\leq \varepsilon \cdot h(f^{-1}((1/2)f(a))) \quad h \circ f^{-1} \text{ nondecreasing} \\
&\leq \varepsilon \cdot h(f^{-1}(f(a))) \quad f^{-1} \text{ nondecreasing} \\
&= \varepsilon \cdot h(a).
\end{aligned}
\tag{A.14}
$$

We know that $h(f^{-1}(x)) = o(h_{m-1}(x))$ and hence $h(f^{-1}(x)) = o(x)$ since $m = 1$ gives the largest case, asymptotically. Thus we assume that a is large enough so that

$$
h(f^{-1}(y)) \leq y \quad \text{for all } y \geq (1/2)f(a).
\tag{A.15}
$$

Then

$$
\begin{aligned}
IV &= E(\varphi(a - X_1); a - X_1 > (1/2)f(a)) \\
&\leq E(\varepsilon \cdot h(f^{-1}(a - X_1)); a - X_1 > (1/2)f(a)) \quad \text{by (A.8)} \\
&\leq \varepsilon E(a - X_1; a - X_1 > (1/2)f(a)) \quad \text{by (A.15)} \\
&= \varepsilon \left[\sqrt{t} \cdot \Delta\left(\frac{f(a)}{2\sqrt{t}} - \xi \right) + (f(a)/2) \cdot \Phi\left(\xi - \frac{f(a)}{2\sqrt{t}} \right) \right],
\end{aligned}
\tag{A.16}
$$

using (2.8) for this last step. Now

$$
\begin{aligned}
\frac{f(a)}{2\sqrt{t}} - \xi &= -\xi + \frac{f(a)}{2\sqrt{a/\mu}} \cdot (1 + o(1)) \quad \text{since } t \sim a/\mu \\
&= -\sqrt{\log(a/h(a)^2)} + o(1) + 3\sqrt{\log(a+1)} \cdot (1 + o(1)) \\
&\geq -\sqrt{\log(a/h_{m+1}(a)^2)} + 2\sqrt{\log a} \quad \text{since } h \in \mathcal{B}_m^o \\
&\geq -\sqrt{\log(a/a^{(1/2)^m})} + 2\sqrt{\log a} \\
&= -\sqrt{(1 - 2^{-m})\log a} + 2\sqrt{\log a} \geq \sqrt{\log a}.
\end{aligned}
$$

Plugging this back into (A.16) gives

$$
\begin{aligned}
IV &\leq \varepsilon \left[\sqrt{t} \cdot \Delta(\sqrt{\log a}) + (f(a)/2) \cdot \Phi(-\sqrt{\log a}) \right] \\
&= O(\sqrt{a}) \cdot \frac{\phi(\sqrt{\log a})}{\log a} + O(\sqrt{a \log a}) \cdot \frac{\phi(\sqrt{\log a})}{\sqrt{\log a}} \quad \text{by (2.10) and Mills's ratio} \\
&= O(\sqrt{a}) \cdot \phi(\sqrt{\log a}) = O(1) = o(h(a)).
\end{aligned}
\tag{A.17}
$$

Plugging (A.12), (A.13), (A.14), and (A.17) into (A.11) gives

$$
\mu ET - a/\mu \leq o(h(a)) + o(1) + \varepsilon h(a) + o(h(a)) \leq 4\varepsilon h(a)
$$

for sufficiently large a. This shows that $ET - a/\mu = o(h(a))$, finishing the proof. $\square$

Proof of Lemma 2.3.

$$\log(\kappa_{m+1}/\kappa_m) = -(1/2)^{m+1}\log\mu + \sum_{i=1}^{m}(1/2)^{i+1}\log\left[(1/2)^{m-i} - (1/2)^m\right]$$

$$-\sum_{i=1}^{m-1}(1/2)^{i+1}\log\left[(1/2)^{m-1-i} - (1/2)^{m-1}\right]$$

$$= -(1/2)^{m+1}\log\mu + (1/2)^{m+1}\log\left[1 - 2^{-m}\right] - \sum_{i=1}^{m-1}(1/2)^{i+1}\log 2$$

$$= (1/2)^{m+1}\log(1 - 2^{-m})/\mu + (1/2 - (1/2)^m)\log(1/2).$$

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

$$\log(h_m(y)/h_{m+1}(x)) = (1/2)^m\log y + (1/2 - (1/2)^m)\log\log y$$

$$- (1/2)^{m+1}\log x - (1/2 - (1/2)^{m+1})\log\log x$$

$$= (1/2)^{m+1}[\log(1 - (1/2)^m)/\mu + \log x + \log\log x]$$

$$+ (1/2 - (1/2)^m)[\log(1/2) + \log\log x + o(1)]$$

$$- (1/2)^{m+1}\log x - (1/2 - (1/2)^{m+1})\log\log x$$

$$= (1/2)^{m+1}\log(1 - 2^{-m})/\mu + (1/2 - (1/2)^m)\log(1/2) + o(1)$$

$$= \log(\kappa_{m+1}/\kappa_m) + o(1)$$

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

Proof of Theorem 2.2.   We proceed by induction on m. Let $\delta_{1,z}^+(a) = (T, M)$. Using Wald's equation,

$$\mu E(T - a/\mu) = E(X_M - a; M = 1) + E(X_M - a; M > 1).$$

Let $t = t(a, z)$, the size of the first stage of $\delta_{1,z}^+(a)$. Using (2.10),

$$E(X_M - a; M = 1) = E(X_1 - a; X_1 \geq a)$$

$$= \sqrt{t} \cdot \Delta(z) \quad \text{by (2.9)}$$

$$\sim \sqrt{a/\mu} \cdot \Delta(z) \tag{A.18}$$

since $t \sim a/\mu$. For $y > 0$ let

$$\dot\delta_{v(y)}(y) = (T', M'),$$

$$\varphi(y) = EX_{M'} - y,$$

$$\psi(y) = EM',$$

where $v(y) = -\sqrt{\log(y + 1)}$ as in the definition of $\delta_{1,z}^+$. Suppose $\varepsilon > 0$. By Lemma 2.2 and since $EM' = 1/\Phi(|v(y)|) \to 1$ as $y \to \infty$ there are constants y_o, C such that for all $y \geq y_o$,

$$\varphi(y) \leq C\sqrt{y\log(y + 1)} \tag{A.19}$$

$$|\psi(y) - 1| \leq \varepsilon. \tag{A.20}$$

Let $Y = a - X_1$. Note that

$$P(Y \geq y_o) = \Phi(z - y_o/\sqrt{t}) \to \Phi(z) \tag{A.21}$$

by continuity, and

$$P(Y < 0) = 1 - \Phi(z) \tag{A.22}$$

so that $P(0 < Y < y_o) \to 0$.

Now

$$
\begin{aligned}
E(X_M - a; M > 1) &= E(\varphi(Y); Y > 0) \\
&= E(\varphi(Y); 0 < Y < y_o) + E(\varphi(Y); Y \geq y_o) \\
&= o(1) + E\left(C\sqrt{Y \log(Y+1)}; Y \geq y_o\right) \quad \text{by (A.22)} \\
&\leq o(1) + C\sqrt{E(Y \mid Y \geq y_o) \log[E(Y \mid Y \geq y_o) + 1]},
\end{aligned}
$$

where this last uses concavity of $y \mapsto y\log(y+1)$ with Jensen's inequality. Now

$$
\begin{aligned}
E(Y \mid Y \geq y_o) \\
&= P(Y \geq y_o)^{-1} E(Y; Y \geq y_o) \\
&= [\Phi(z) + o(1)]^{-1} \cdot \left[\sqrt{t} \cdot \Delta(-z + y_o/\sqrt{t}) + y_o \cdot \Phi(z - y_o/\sqrt{t})\right] \quad \text{using (2.8)} \\
&= O(1)[O(\sqrt{a}) \cdot O(1) + y_o \cdot O(1)] = O(\sqrt{a}).
\end{aligned}
$$

Thus

$$E(X_M - a; M > 1) \leq o(1) + C\sqrt{O(\sqrt{a}) \log O(\sqrt{a})} = o(\sqrt{a}),$$

and combining this with (A.18) gives

$$ET - a/\mu \leq \Delta(z)\mu^{-3/2}\sqrt{a} + o(\sqrt{a}) = (1 + o(1))\Delta(z)\kappa_1 h_1(a),$$

establishing (2.17) for $m = 1$.

$$
\begin{aligned}
|EM - (1 + \Phi(z))| &= |1 + E(\psi(Y); Y > 0) - (1 + \Phi(z))| \\
&\leq |E(\psi(Y); 0 < Y < y_o)| + E(|\psi(Y) - 1|; Y \geq y_o) \\
&\quad + |P(Y \geq y_o) - \Phi(z)| \\
&\leq o(1) + \varepsilon + o(1) \quad \text{by (A.20) and (A.21)} \\
&\leq 3\varepsilon
\end{aligned}
$$

for sufficiently large a, proving that $EM \to 1 + \Phi(z)$.

Next let $\delta^+_{m+1,z}(a) = (T, M)$ and $\delta^+_{m,z}(y) = (T', M')$ for $y > 0$. Let

$$\varphi(y) = EX_{M'} - y,$$
$$\psi(y) = EM',$$

and suppose $\varepsilon > 0$. By the induction hypothesis there is a constant y_1 such that, for all $y \geq y_1$,

$$\varphi(y) \leq (1 + \varepsilon)\mu\Delta(z)\kappa_m h_m(y) \qquad (A.23)$$

$$|\psi(y) - (m + \Phi(z))| \leq \varepsilon. \qquad (A.24)$$

Let $t = t(a, \sqrt{(1 - 2^{-m})\log(a + 1)})$, the size of the first stage of $\delta^+_{m+1,z}(a)$. Again letting $Y = a - X_1$,

$$\mu E(T - a/\mu) = E(X_M - a; M = 1) + E(X_M - a; M > 1)$$
$$= E(X_1 - a; X_1 \geq a) + E(\varphi(Y); 0 < Y < y_1) + E(\varphi(Y); Y \geq y_2). \qquad (A.25)$$

Now

$$E(X_1 - a; X_1 \geq a) = \sqrt{t} \cdot \Delta\left(\sqrt{(1 - 2^{-m})\log(a + 1)}\right) \quad \text{by (2.9)}$$
$$\sim \sqrt{a/\mu} \cdot \frac{\phi\left(\sqrt{(1 - 2^{-m})\log(a + 1)}\right)}{(1 - 2^{-m})\log(a + 1)} \quad \text{by (2.10)}$$
$$= \sqrt{a/\mu} \cdot \frac{O(a^{-1/2 + (1/2)^{m+1}})}{(1 - 2^{-m})\log(a + 1)}$$
$$= o(a^{(1/2)^{m+1}}) = o(h_{m+1}(a)). \qquad (A.26)$$

$E(\varphi(Y); 0 < Y < y_1) = o(1)$ by a now routine argument and using (A.23),

$$\frac{E(\varphi(Y); Y \geq y_1)}{(1 + \varepsilon)\mu\Delta(z)\kappa_m} \leq E(h_m(Y); Y \geq y_1)$$
$$\leq h_m(E(Y; Y \geq y_1)P(Y \geq y_1)^{-1})$$

by Jensen's inequality since h_m is concave. Now

$$E(Y; Y \geq y_1) \sim \sqrt{t} \cdot \Delta\left(-\sqrt{(1 - 2^{-m})\log(a + 1)}\right) \sim \sqrt{a/\mu} \cdot \sqrt{(1 - 2^{-m})\log(a + 1)}$$

by a routine application of (2.8). Using that $P(Y \geq y_1) \to 1$, for sufficiently large a we have

$$E(\varphi(Y); Y \geq y_1) \leq (1 + \varepsilon)^2 \mu\Delta(z)\kappa_m h_m(\sqrt{a/\mu})$$
$$\leq (1 + \varepsilon)^3 \mu\Delta(z)\kappa_{m+1} h_{m+1}(a), \qquad (A.27)$$

using Lemma 2.3. Plugging (A.26) and (A.27) into (A.25) gives

$$ET - a/\mu \leq o(h_{m+1}(a)) + o(1) + (1 + \varepsilon)^3 \Delta(z)\kappa_{m+1} h_{m+1}(a)$$
$$\leq [\varepsilon + \varepsilon + (1 + \varepsilon)^3]\Delta(z)\kappa_{m+1} h_{m+1}(a)$$
$$\leq [1 + 9\varepsilon]\Delta(z)\kappa_{m+1} h_{m+1}(a)$$

for sufficiently large a, which shows that (2.17) holds for $m + 1$. As for the number of stages used,

$$
\begin{aligned}
|EM - (m + 1 + \Phi(z))| &= |1 + E(\psi(Y); Y > 0) - (m + 1 + \Phi(z))| \\
&\leq |E(\psi(Y); 0 < Y < y_1)| + E(|\psi(Y) - (m + \Phi(z))|; Y \geq y_1) \\
&\quad + (m + \Phi(z))P(Y < y_1) \\
&\leq o(1) + \varepsilon \cdot P(Y \geq y_1) + (m + \Phi(z)) \cdot o(1) \\
&\quad \text{by (A.24)} \quad \text{and since } P(Y < y_1) \to 0 \\
&\leq o(1) + \varepsilon + o(1) \leq 3\varepsilon
\end{aligned}
$$

for sufficiently large a. This shows that $EM \to m + 1 + \Phi(z)$ and completes the proof. $\qquad\square$

Proof of Lemma 2.4. Let $G_k = (1/\mu)^{1-2^{-k}} F_h^{(k)}(a)$. Suppose $\varepsilon > 0$. Let

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

The $k = 0$ case is trivial since (2.22) is equivalent to $P(a \geq a) \to 1$. Fix $1 \leq k < m$ and assume that $P(V_{k-1}) \to 1$. Let $\delta(a) = (T, M)$ and

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

Note that

$$
\begin{aligned}
h(a)^2 &= o(h_{m-1}(a)^2) \quad \text{since } h \in \mathscr{B}_m \\
&= o(F_h^{(m-2)}(a)) \quad \text{by Lemma 2.5} \\
&= o(G_{m-2}) \\
&= o(G_{k-1})
\end{aligned}
$$

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

$$
P\big(\zeta_k \geq \sqrt{\log(G_{k-1}/h^2(a))} - 1 \mid V_{k-1}\big) \to 1. \tag{A.28}
$$

Let $\xi = \sqrt{\log(G_{k-1}/h^2(a))} - 1$ and $U = \{\zeta_k < \xi\}$. If (A.28) were to fail there would be a constant $\eta > 0$ and a sequence of a's approaching ∞ on which $P(U \mid V_{k-1}) > \eta$. Then

$$
\begin{aligned}
\mu R(\delta(a)) \geq \mu E(T - a/\mu) = E(X_M - a) &\geq E\big[(X_M - a)1\{M = k\}; U \cap V_{k-1}\big] \\
&= E\big[\Delta(\zeta_k)\sqrt{t(a - X_{k-1}, \zeta_k)}; U \cap V_{k-1}\big]. \tag{A.29}
\end{aligned}
$$

The function inside the expectation in (A.29) is decreasing in both ζ_k and X_{k-1}, hence

$$
\mu R(\delta(a)) \geq \Delta(\xi)\sqrt{t((1 - \varepsilon)G_{k-1}, \xi)} \cdot P(U \cap V_{k-1}). \tag{A.30}
$$

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

$$P(U \cap V_{k-1}) \geq \eta/2, \qquad (A.31)$$

say, for large enough a. Also

$$\Delta(\xi)\sqrt{t((1-\varepsilon)G_{k-1}, \xi)} \sim \frac{\phi(\xi)}{\xi^2}\sqrt{(1-\varepsilon)G_{k-1}/\mu}$$

$$\geq \varepsilon' h(a) \frac{\exp\left(\sqrt{\log(G_{k-1}/h^2(a))} - 1/2\right)}{\left(\sqrt{\log(G_{k-1}/h^2(a))} - 1\right)^2} \quad \text{some } \varepsilon' > 0$$

$$= h(a)/o(1). \qquad (A.32)$$

Plugging (A.31) and (A.32) into (A.30) gives $h(a) = o(R(\delta(a)))$, which contradicts our assumption that $R(\delta(a)) = O(h(a))$. Hence, (A.28) must hold. Then

$$P(V_k \mid U' \cap V_{k-1})$$
$$= P(a - X_k \geq (1 - \varepsilon)G_k \mid U' \cap V_{k-1})$$
$$= P\left(\frac{(X_k - X_{k-1}) - \mu T_k}{\sqrt{T_k}} \leq \frac{a - X_{k-1} - (1 - \varepsilon)G_k - \mu T_k}{\sqrt{T_k}} \middle| U' \cap V_{k-1}\right). \qquad (A.33)$$

On V_{k-1},

$$\frac{a - X_{k-1} - (1 - \varepsilon)G_k - \mu T_k}{\sqrt{T_k}} = \zeta_k - \frac{2\mu(1 - \varepsilon)G_k}{\sqrt{4\mu(a - X_{k-1}) + \zeta_k^2} - \zeta_k}$$

$$\geq \zeta_k - \frac{2\mu(1 - \varepsilon)G_k}{\sqrt{4\mu(1 - \varepsilon)G_{k-1} + \zeta_k^2} - \zeta_k},$$

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

$$\zeta_k - \frac{2\mu(1 - \varepsilon)G_k}{\sqrt{4\mu(1 - \varepsilon)G_{k-1} + \zeta_k^2} - \zeta_k} \geq \xi - \frac{2\mu(1 - \varepsilon)G_k}{\sqrt{4\mu(1 - \varepsilon)G_{k-1} + \xi^2} - \xi}$$

$$= \xi - \frac{2\mu(1 - \varepsilon)G_k}{\sqrt{4\mu(1 - \varepsilon)G_{k-1}}}(1 + o(1))$$

$$= \xi - \sqrt{(1 - \varepsilon)\log(F_h^{(k-1)}(a)/h^2(a))}(1 + o(1))$$

$$\sim (1 - \sqrt{1 - \varepsilon})\sqrt{\log(G_{k-1}/h^2(a))} \equiv \gamma \to \infty.$$

Substituting this back into (A.33) gives

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

by Chebyshev's inequality. Thus $P(V_k) \geq P(V_k \mid U' \cap V_{k-1})P(U' \cap V_{k-1}) \to 1$ since $P(U' \cap V_{k-1}) \to 1$ by the induction hypothesis and (A.28), finishing the induction and proving the lemma. □

Proof of Lemma 2.5.   Let F^k denote $F_h^{(k)}(x)$. First we prove (2.26) by induction on k. For $k = 1$,

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

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

$$\log\left(\frac{F^{k-1}}{h(x)^2}\right) \sim \log\left(\frac{(C_k^m h_k(x))^2}{(Qh_m(x))^2}\right)$$

$$\sim \log\left(\frac{h_k(x)^2}{h_m(x)^2}\right)$$

$$\sim \log\left(\frac{x^{(1/2)^{k-1}}(\log x)^{1-(1/2)^{k-1}}}{x^{(1/2)^{m-1}}(\log x)^{1-(1/2)^{m-1}}}\right)$$

$$\sim \left[(1/2)^{k-1} - (1/2)^{m-1}\right]\log x, \qquad (A.34)$$

so

$$\sqrt{F^k} = \left\{F^{k-1}\log(F^{k-1}/h(x)^2)\right\}^{1/4}$$

$$\sim \left\{(C_k^m h_k(x))^2\left[(1/2)^{k-1} - (1/2)^{m-1}\right]\log x\right\}^{1/4}$$

$$= \sqrt{C_k^m} \cdot x^{(1/2)^{k+1}}(\log x)^{1/4-(1/2)^{k+1}}\left[(1/2)^{k-1} - (1/2)^{m-1}\right]^{1/4}(\log x)^{1/4}$$

$$= \sqrt{C_k^m} \cdot \left[(1/2)^{k-1} - (1/2)^{m-1}\right]^{1/4}h_{k+1}(x)$$

$$= C_{k+1}^m h_{k+1}(x), \qquad (A.35)$$

by (2.23).

Next we prove (2.27) 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 $m \ge 2$. Now assume $2 \le k + 1 < m$ and that (2.27) holds for k. Then, since $h_m \ll h \ll h_{m-1}$,

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

by the same argument leading to (A.34). Then, by repeating the argument leading to (A.35) with $\lesssim$ in place of $\sim$,

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

by (2.23). The other bound is similar:

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

and so

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

by replacing m by $m-1$ in (A.35) and (2.23). $\square$

ACKNOWLEDGMENTS

REFERENCES

Bartroff, J. (2004). Asymptotically Optimal Multistage Hypothesis Tests, Ph.D. diss., California Institute of Technology–Pasadena.

Bartroff, J. (2006). Asymptotically Optimal Multistage Tests of Simple Hypotheses, *Annals of Statistics*, to appear.

Cressie, N. and Morgan, P. B. (1993). The VPRT: A Sequential Testing Procedure Dominating the SPRT, *Econometric Theory* 9: 431–450.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, Second edition, New York: Wiley.

Lorden, G. (1970). On Excess over the Boundary, *Annals of Mathematical Statistics* 41: 520–527.

Morgan, P. B. and Cressie, N. (1997). A Comparison of the Cost-Efficiencies of the Sequential, Group-Sequential, and Variable-Sample-Size-Sequential Probability Ratio Tests, *Scandinavian Journal of Statistics* 24: 181–200.

Schmitz, N. (1993). *Optimal Sequentially Planned Decision Procedures*, Lecture Notes in Statistics, Vol. 79, New York: Springer-Verlag.

Siegmund, D. (1985). *Sequential Analysis*, New York: Springer-Verlag.