The Annals of Statistics
1998, Vol. 26, No. 6, 2179–2205

QUICKEST DETECTION WITH EXPONENTIAL PENALTY FOR DELAY[1]

BY H. VINCENT POOR

Princeton University

The problem of detecting a change in the probability distribution of a random sequence is considered. Stopping times are derived that optimize the tradeoff between detection delay and false alarms within two criteria. In both cases, the detection delay is penalized exponentially rather than linearly, as has been the case in previous formulations of this problem. The first of these two criteria is to minimize a worst-case measure of the exponential detection delay within a lower-bound constraint on the mean time between false alarms. Expressions for the performance of the optimal detection rule are also developed for this case. It is seen, for example, that the classical Page CUSUM test can be arbitrarily unfavorable relative to the optimal test under exponential delay penalty. The second criterion considered is a Bayesian one, in which the unknown change point is assumed to obey a geometric prior distribution. In this case, the optimal stopping time effects an optimal trade-off between the expected exponential detection delay and the probability of false alarm. Finally, generalizations of these results to problems in which the penalties for delay may be path dependent are also considered.

1. Introduction. Quickest detection is the problem of detecting, with as little delay as possible, a change in the probability distribution of a sequence of random measurements. This problem arises in a great variety of applications, such as seismology, speech and image processing, biomedical signal processing, machinery monitoring and finance. Overviews of existing techniques for quickest detection can be found in Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1992), Carlstein, Müller and Siegmund (1994) and Kerestecioğlu (1993).

A useful formulation of the quickest detection problem is to consider a sequence $X_1, X_2, \dots$ of random observations, and to suppose that there is a change point $t \geq 1$ (possibly $t = \infty$) such that, given t, $X_1, X_2, \dots, X_{t-1}$ are drawn from one distribution and $X_t, X_{t+1}, \dots,$ are drawn from another distribution. The set of detection strategies of interest corresponds to the set of (extended) stopping times with respect to the observed sequence, with the interpretation that the stopping time T decides that the change point t has occurred at or before time k when $T = k$. We will be more specific about this model in subsequent sections.

Received March 1997; revised June 1998.
[1] Supported in part by the IDA Center for Communications Research, Princeton, New Jersey.
AMS 1991 *subject classifications*. Primary 62L10; secondary 60G40, 62L15, 94A13.
Key words and phrases. Quickest detection, change point problems, optimal stopping, exponential cost.

The design of quickest detection procedures typically involves the optimization of a trade-off between two types of performance indices, one being a measure of the delay between the time a change occurs and it is detected [i.e., $(T - t + 1)^+$, where $x^+ = \max\{0, x\}$], and the other being a measure of the frequency of false alarms (i.e., events of the type $\{T < t\}$). In essentially all such extant designs, detection delay is penalized via a linear function of delay. [An exception is found in Pelkowitz (1987), in which nonlinear delay penalties are proposed, but corresponding optimal stopping times are not derived.] This type of penalty is suitable for many applications. For example, the earliest applications of quickest detection involved the monitoring of manufacturing processes to detect possible declines in quality of the manufactured goods. In this situation, the cost of delay is accurately measured by a linear penalty, reflecting the fact that the economic cost of discarded defective goods will be proportional to the quantity produced. However, in other applications, linear cost does not capture the true cost of delayed action. Consider, for example, financial applications in which the change point may represent a time at which a fundamental shift in the performance, or expected performance, of some type of investment occurs. In this situation, the compounding of investment growth or the short shelf lives of investment opportunities point to exponential penalties as more suitable measures of the cost of delay. Similarly, in the health monitoring of components in interconnected systems (e.g., aircraft systems, communication networks, power grids, biological populations, etc.), the effects of undetected faults can exponentiate with time, again suggesting a more aggressive cost structure than is captured with a linear delay penalty.

Motivated by these types of applications, this paper considers the problem of quickest detection with exponential penalty for delay. In particular, we consider penalties on the detection delay of the form

$$\frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1}, \tag{1}$$

where $\alpha \neq 1$ is a positive constant. Note that, with $\alpha > 1$, this penalty quantifies an exponential growth of costs as a function of delay, reflecting the type of exponentiating costs mentioned in the preceding paragraph. Alternatively, with $\alpha < 1$, (1) quantifies a saturating, sublinear cost of delay. Since

$$\frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1} = \sum_{l=t}^{T} \alpha^{l-t}, \tag{2}$$

the exponential penalty in this latter case can be viewed as a discounted version of the traditional linear penalty $(T - t + 1)^+$. Of course, as $\alpha \to 1$, the quantity of (1) approaches the linear delay penalty.

For this cost function, we consider two traditional formulations of the quickest detection problem. The first of these is a minimax formulation, first proposed in the linear-delay-penalty case by Lorden (1971), in which the delay penalty is a worst-case measure of delay, and false alarms are constrained through a lower bound on the allowable mean time between false alarms. In this formulation, the worst-case delay is taken over all possible realizations of the observations leading up to the change point and over all possible values of the change point. The second formulation is a Bayesian formulation, first proposed in the linear-delay-penalty case by Kolmogorov and Shiryayev [see Shiryayev (1963)], in which the change point is endowed with a prior distribution, and the opposing performance indices are expected detection delay and false-alarm probability. We also consider modifications of these problems in which the delay penalty depends explicitly on the observed sample path of the random sequence. We develop optimal detection procedures for each of these formulations. Performance analysis is also considered in each case, with the first of the formulations offering perhaps the most interesting results in this regard. For example, among other results, it is seen that the classical Page CUSUM test (which is minimax optimal for linear delay penalty) can have infinite minimax exponential delay if the rate at which delay penalty accumulates is too large relative to the rate at which discrimination information between prechange and postchange distributions accumulates.

The remainder of this paper is organized as follows. The minimax solution is presented in Section 2, and Section 3 is devoted to performance analysis in this case. Section 4 develops the Bayesian solution. Section 5 considers the extension of the results of Sections 2 through 4 to the case of path-dependent cost of delay. Finally, Section 6 contains some concluding remarks. Appendices containing the more detailed elements of the required proofs are also included.

2. A minimax solution. We begin by considering the situation in which the change point t is a fixed, nonrandom quantity that can be either ∞ or any value in the positive integers. To model this situation, we consider a measurable space $(\Omega, \mathscr{F})$, consisting of a sample space Ω and a σ-field $\mathscr{F}$ of events. We further consider a family $\{P_i;\ t \in [1, 2, \dots, \infty]\}$ of probability measures on $(\Omega, \mathscr{F})$, such that, under P_t, $X_1, X_2, \dots, X_{t-1}$ are independent and identically distributed (i.i.d.) with a fixed marginal distribution Q_b, and $X_t, X_{t+1}, \dots$ are i.i.d. with another marginal distribution Q_a and are independent of $X_1, X_2, \dots, X_{t-1}$. For simplicity, we assume that Q_a and Q_b are mutually absolutely continuous, that the likelihood ratio $L = dQ_a/dQ_b$ has no atoms under Q_b and that $0 < D(Q_b \| Q_a) < \infty$, where $D(Q_b \| Q_a)$ denotes the Kullback–Leibler divergence of Q_a from Q_b,

$$D(Q_b \| Q_a) = -\int \log L \, dQ_b. \tag{3}$$

For technical reasons, we also assume the existence of a random variable X_0 that is uniformly distributed in $[0, 1]$ and that is independent of $X_1, X_2, \dots$ under each P_t.

We would like to consider procedures that can detect the change point, if it occurs (i.e., if $t < \infty$), as quickly as possible after it occurs. As a set of detection strategies, it is natural to consider the set $\mathcal{T}$ of all (extended) stopping times with respect to the filtration $\{\mathcal{F}_k\}_{k\geq 0}$ where $\mathcal{F}_k$ denotes the smallest σ-field with respect to which $X_0, X_1, \ldots, X_k$ are measurable. Thus, when the stopping time T takes on the value k, the interpretation is that T has detected the existence of a change point t at or prior to time k.

Following Lorden (1971), it is of interest to penalize exponential detection delay via its worst-case value

$$d(T) = \sup_{t\geq 1} d_t(T) \tag{4}$$

with

$$d_t(T) = \operatorname{ess\,sup} E_t\left\{\frac{\alpha^{(T-t+1)^+}-1}{\alpha-1}\middle|\mathcal{F}_{t-1}\right\}, \tag{5}$$

where $E_t\{\cdot\}$ denotes expectation under the distribution P_t. (Recall that the essential supremum of a random variable is the greatest lower bound of the set of constants that bound the random variable with probability one.) Note that $d_t(T)$ is the worst-case average delay under P_t, where the worst case is taken over all realizations of $X_0, X_1, \ldots, X_{t-1}$. The desire to make $d(T)$ small must be balanced with a constraint on the rate of false alarms. The rate of false alarms can be quantified by the mean time between false alarms,

$$f(T) = E_\infty\{T\}, \tag{6}$$

and a useful design criterion is then given by

$$\inf_{T\in\mathcal{T}} d(T) \quad \text{subject to } f(T) \geq \gamma, \tag{7}$$

where γ is a constant. That is, we seek a stopping time that minimizes the worst-case delay within a lower-bound constraint on the mean time between false alarms.

The solution to (7) for the linear-delay-penalty ($\alpha = 1$) case was demonstrated in Moustakides (1986). Here, we extend this solution to the case of general α. To do so, for $h \geq 0$ we define a stopping time

$$T_h = \inf\{k \geq 0 | S_k \geq h\}, \tag{8}$$

where

$$S_k = \max_{1\leq j\leq k}\left(\prod_{l=j}^{k} \alpha L(X_l)\right) = \alpha L(X_k)\max\{S_{k-1}, 1\}, \qquad k \geq 1, \tag{9}$$

and $S_0 = 0$.

We then have the following result.

THEOREM 2.1. *Suppose $h \geq 0$ and*

$$P_\infty(\alpha L(X_1) > 1) > 0. \tag{10}$$

Then T_h solves (7) *with $\gamma = f(T_h)$. That is,*

$$f(T) \geq f(T_h) \Rightarrow d(T) \geq d(T_h). \tag{11}$$

REMARK 2.2. Note that condition (10) is trivially satisfied if $\alpha \geq 1$.

PROOF. In proving this result, we will make heavy use of Moustakides' method of proof for the linear-delay case [Moustakides (1986)]. For the case in which $\log \alpha < D(Q_b \| Q_a)$, the extension to exponential cost is rather straightforward. For larger α, some variations are needed. These considerations will arise in the proof of Lemma 2.4 below. (Since the case $h = 0$ is trivial, we consider only $h > 0$.)

It is easily seen that, in seeking solutions to (7), we can restrict attention to stopping times that achieve the constraint on $f(T)$ with equality. This follows since, if $\gamma < f(T) < \infty$, then we can produce a stopping time that achieves the constraint with equality without increasing the worst-case exponential delay, simply by randomizing between T and the stopping time that is identically zero. (Such randomized stopping times are in $\mathscr{T}$ by virtue of the inclusion of $\mathscr{F}_0$ in the filtration.) Stopping times for which $f(T) = \infty$ can be eliminated from consideration, since for this case we can choose sufficiently large n so that $f(\min\{T, n\}) \geq \gamma$, and we always have $d(\min\{T, n\}) \leq d(T)$. [That $f(T_h) < \infty$ follows from Theorem 3.1 below.]

We now state the following two intermediate results, whose proofs are given in Appendix A and from which the theorem follows.

LEMMA 2.3. *Suppose $T \in \mathscr{T}$ is such that $0 < f(T) < \infty$. Then*

$$d(T) \geq \bar{d}(T) \triangleq \frac{E_\infty\left\{\sum_{k=0}^{T-1} \max\{S_k, 1\}\right\}}{E_\infty\left\{\sum_{k=0}^{T-1} (1 - S_k)^+\right\}} \tag{12}$$

with equality if $T = T_h$.

LEMMA 2.4. *Suppose* (10) *holds. Then T_h solves the following maximization problem for all continuous nonincreasing functions $g\colon [0, \infty) \to \mathbb{R}$:*

$$\sup_{T \in \mathscr{T}} \left\{ \sum_{k=0}^{T-1} g(S_k) \right\} \quad \text{subject to } f(T) = \gamma. \tag{13}$$

Taking $g(x) = -\max\{x, 1\}$ and $g(x) = (1 - x)^+$, respectively, Lemma 2.4 asserts that T_h simultaneously minimizes the numerator and maximizes the denominator of $\bar{d}(T)$ within the constraint $f(T) = \gamma$. Since $d(T_h) = \bar{d}(T_h)$, the theorem follows. □

Theorem 2.1 asserts the optimality of the stopping time based on the first exit of S_k from the interval $[0, h)$. We henceforth assume that $h \geq 1$, in which case the stopping time T_h can be written equivalently as

$$T_h = \inf\{k \geq 0 | m_k \geq \log h\}, \tag{14}$$

where

$$m_k = \log \max\{S_k, 1\}, \qquad k = 0, 1, \ldots. \tag{15}$$

It is easily seen that the sequence $\{m_k\}$ can be computed recursively via

$$m_k = \max\{m_{k-1} + \log L(X_k) + \log \alpha, 0\}, \qquad k = 1, 2, \dots, \tag{16}$$

with $m_0 = 0$. That is, the test based on T_h accumulates the adjusted log-likelihoods, $\log L(X_k) + \log \alpha$, resetting the accumulation to zero whenever it goes negative. The alarm is sounded when this accumulation crosses the upper threshold $\log h$. With $\alpha = 1$ the stopping time T_h thus reduces to the classical CUSUM test of Page (1954). For $\alpha \neq 1$, the test is more or less aggressive in sounding alarms than is Page's test, depending on whether $\alpha > 1$ or $\alpha < 1$. This is, of course, completely consistent with intuition, since larger values of α correspond to greater penalties on delay.

3. Performance analysis. In the preceding section, we showed that the stopping time T_h is optimal in the sense of Theorem 2.1. In this section, we consider the performance of this stopping time by determining the quantities $d(T_h)$ and $f(T_h)$.

We begin with the following result, which gives exact expressions for these two quantities. (Here, and throughout this paper, 1_A denotes the indicator function of the event A.)

THEOREM 3.1. *Suppose* $h > 1$, *and* (10) *holds. Then*

$$f(T_h) = \frac{E_\infty\{N\}}{1 - P_\infty(F_0)} < \infty \tag{17}$$

and

$$d(T_h) = \frac{E_1\{\alpha^N\} - 1}{(1-\alpha)\left(1 - E_1\left\{\alpha^N 1_{F_0}\right\}\right)}, \tag{18}$$

where N is the stopping time

$$N = \min\left\{ n \geq 1 \,\middle|\, \sum_{l=1}^{n} \left[\log L(X_l) + \log \alpha\right] \notin (0, \log h) \right\}, \tag{19}$$

and where F_0 denotes the event

$$\left\{ \sum_{l=1}^{N} \left[\log L(X_l) + \log \alpha\right] \leq 0 \right\}. \tag{20}$$

REMARK 3.2. The proof of this result relies on the renewal properties of the accumulated sum m_k of (16), arising from the resetting of this sum each time it crosses zero. This analysis is similar to the classical analysis of Page's CUSUM [e.g., Basseville and Nikiforov (1993), pages 195–197 or Siegmund (1985), Section II.6]. However, a distinction between this result and that for Page's CUSUM arises in the treatment of the exponential delay, $d(T_h)$, and so a proof is included in Appendix B.

Since N of (19) is the first exit time of a random walk from an interval, its statistical behavior can be analyzed via the classical methods of Wald approximation, diffusion approximation, and so on [cf. James, James and Siegmund (1988), Khan (1978) or Siegmund (1985)]. However, even without Theorem 3.1, $f(T_h)$ and $d(T_h)$ can be estimated directly by approximating the behavior of T_h with that of the stopping time

$$\tilde{T}_h = \inf\left\{u \geq 0 | Z_u - \min_{0 \leq s \leq u} Z_s \geq \log h\right\}, \tag{21}$$

where $\{Z_u;\ u \geq 0\}$ is a Brownian motion approximating the random walk

$$\sum_{l=1}^{n} [\log L(X_l) + \log \alpha], \qquad n = 1, 2, \ldots. \tag{22}$$

Approximation of this type for linear delay penalty and the classical Page test has been considered by Reynolds (1975). Analogous results can be obtained for the exponential-penalty case, as we now develop.

Consider the case in which the observations are in $\mathbb{R}^n$, and Q_b and Q_a are $\mathscr{N}(\mu_0, \Sigma)$ and $\mathscr{N}(\mu_1, \Sigma)$ distributions, respectively, with Σ positive definite. Then, under P_∞ we may use the model

$$Z_u = \sqrt{2D}\, B_u + D(\beta - 1)u, \qquad u \geq 0, \tag{23}$$

and under P_1 we may use the model

$$Z_u = \sqrt{2D}\, B_u + D(\beta + 1)u, \qquad u \geq 0, \tag{24}$$

where $\{B_u;\ u \geq 0\}$ is a standard Brownian motion, $D = D(Q_b \| Q_a)$, with $D(Q_b \| Q_a)$ given by (3), and the constant β is defined as

$$\beta = \frac{\log \alpha}{D(Q_b \| Q_a)}. \tag{25}$$

Here, of course, $D = (\mu_1 - \mu_0)' \Sigma^{-1} (\mu_1 - \mu_0)/2$. [Note that the model (23) and (24) can also be used in the local testing case if the two distributions Q_a and Q_b are sufficiently close to one another that the asymptotes $\mathrm{Var}_\infty(\log L(X_1)) \sim \mathrm{Var}_1(\log L(X_1)) \sim 2D$ and $E_1\{\log L(X_1)\} \sim -E_\infty\{\log L(X_1)\} \equiv D$ give accurate approximations.)

The statistics of stopping times of the form (21) have been analyzed in several works, including Kennedy (1976), Lehoczky (1977) and Taylor (1975). We can conclude from this analysis [cf. equations (3.1.104) and (3.1.105) of Basseville and Nikiforov (1993)] that, under the model (23) and (24), we have

$$D \times f(\tilde{T}_h) = \begin{cases} (\log h)^2/2, & \beta = 1, \\ \left(\dfrac{h^{1-\beta} - 1}{1 - \beta} - \log h \right) \Big/ (1 - \beta), & \beta \neq 1 \end{cases} \tag{26}$$

and

$$(27)\qquad (e^{\beta D}-1)\times d(\tilde{T}_h) = \begin{cases} (h-1-\log h)/(1+\log h), & \beta = 1, \\ ((1-\beta)h^{1+\beta}-h+\beta h^{\beta})/(h-\beta h^{\beta}), & \beta \neq 1. \end{cases}$$

(Recall that $e^{\beta D}=\alpha$.) As $\beta \to 0$ (i.e., $\alpha \to 1$) these expressions reduce to those of Reynolds (1975) for the diffusion approximation to the performance of Page's test under linear delay penalty.

Note that, for large γ (and consequently large h), (26) implies

$$(28)\qquad D \times f(\tilde{T}_h) \sim \begin{cases} h^{1-\beta}/(1-\beta)^2, & \beta < 1, \\ \log h/(\beta-1), & \beta > 1 \end{cases}$$

and so

$$(29)\qquad d(\tilde{T}_h) \sim \begin{cases} \mathscr{O}(\gamma^{\beta/(1-\beta)}), & 0 < \beta < 1, \\ \mathscr{O}(\exp^{(\gamma D(\beta-1))}), & \beta > 1. \end{cases}$$

Thus, the asymptotic behavior of the expected delay penalty is fundamentally different depending on whether the rate of delay-penalty increase is greater or less than the rate at which discrimination information between prechange and postchange distributions accumulates.

It is interesting to compare (27) with the performance of Page's test under exponential delay penalty. To do so, let us denote by $\tilde{T}_P$ the stopping time (21) where the Brownian motion $\{Z_u;\ u \geq 0\}$ approximates the random walk

$$(30)\qquad \sum_{l=1}^{n} \log L(X_l), \qquad n = 1, 2, \ldots.$$

Again assuming Gaussian observations as above, this Brownian motion behaves statistically as (23) with $\beta = 0$ and as (24) with $\beta = 0$, respectively, under prechange and postchange conditions. The statistical behavior of $\tilde{T}_P$ thus approximates the statistical behavior of Page's test. Applying (26) with $\beta = 0$ yields

$$(31)\qquad f(\tilde{T}_P) = \frac{h-1-\log h}{D}.$$

To analyze $d(\tilde{T}_P)$ for $\alpha \neq 1$, we consider two cases, $\beta < 1/4$ and $\beta > 1/4$. For $\beta < 1/4$, equation (3.1.104) of Basseville and Nikiforov (1993) straightforwardly yields

$$(32)\qquad d(\tilde{T}_P) = \frac{2\xi(\sqrt{h})^{1+\xi} - (\xi+1)(\sqrt{h})^{\xi^2} - \xi + 1}{(e^{\beta D}-1)\left[(\xi+1)(\sqrt{h})^{\xi^2} + \xi - 1\right]},$$

with $\xi = \sqrt{1-4\beta}$. It follows that, asymptotically in the false-alarm constraint γ, we have

$$(33)\qquad d(\tilde{T}_P) = \mathscr{O}(\gamma^{(\xi+4\beta)/2}), \qquad 0 < \beta < 1/4.$$

Alternatively, for $\beta > 1/4$, we can show that $d(\tilde{T}_P) = \infty$. In particular, we note that, for any $x \in (0, \log h)$, $\tilde{T}_P$ is no smaller than the first exit time of $\{Z_u;\ u \geq 0\}$ from the interval $(0, \log h)$ after the first time that $Z_u = x$. Statistically this latter time is no smaller than the first exit of $\{Z_u + x;\ u \geq 0\}$ from $(0, \log h)$, a stopping time that we denote by $\tilde{T}^x$. It follows from page 258 of Dvoretsky, Kiefer and Wolfowitz (1953) [see also equation (A:194) of Wald (1947)] that, under the postchange conditions, $\tilde{T}^x$ has a probability density that is a mixture of densities of the form

$$\frac{\lambda_i}{2\Gamma(1/2)t^{3/2}}\exp\left\{-\frac{\lambda_i^2}{4t}-\frac{Dt}{4}\right\}, \qquad t \geq 0, \tag{34}$$

where the λ_i's are positive constants. Clearly, then, we have

$$E_1\{\alpha^{\tilde{T}_P}\} = \infty, \tag{35}$$

if $\log \alpha > D/4$.

So we see that the delay penalty incurred by the continuous-time version of Page's test in this case is infinite if the rate of penalty increase is greater than one-fourth the rate at which discrimination information between prechange and postchange distributions accumulates. Even for smaller β, (29) and (33) imply that

$$\frac{d(\tilde{T}_P)}{d(\tilde{T}_h)} = \mathscr{O}(\gamma^{\varepsilon}), \qquad 0 < \beta < 1/4, \tag{36}$$

with $\varepsilon = \xi/2 + \beta(1-2\beta)/(1-\beta)$. Since ε is strictly positive, the optimization problem posed in the preceding section clearly yields a significantly better test under exponential penalty than does the classical linear-penalty formulation.

4. A Bayesian solution: the exponential disorder problem. We now turn to a Bayesian version of the quickest detection problem, in which the change point t is assumed to be a random variable with a known prior distribution on the nonnegative integers. In particular, we consider the general set-up of Section 2, with an additional probability distribution P on $(\Omega, \mathscr{F})$ under which t has the given prior distribution and the P_k's considered previously are the conditional distributions given the events $\{t = k\}$. Here we do not need the assumptions that $L(X_k)$ is nonatomic under Q_b and that $D(Q_b \| Q_a)$ is finite, and so we drop them. Additionally, as randomization will not be needed here either, we can replace the uniform random variable X_0 with a constant.

In this situation, for a stopping time T, as a measure of delay we can adopt the expected exponential delay,

$$E\left\{\frac{\alpha^{(T-t+1)^+}-1}{\alpha-1}\right\}, \tag{37}$$

where $E\{\cdot\}$ denotes expectation under the measure P. Similarly, as a measure of false-alarm rate we can adopt the false-alarm probability,

$$P(T < t). \tag{38}$$

Analogously with the case of minimax design, we would like to determine stopping times T that effect optimal trade-offs between the two objectives of small detection delay and small false-alarm rate. A convenient way of implementing such a trade-off is to seek $T \in \mathcal{T}$ to solve the optimization problem

$$\inf_{T\in\mathcal{T}}\left[P(T < t) + cE\left\{\frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1}\right\}\right], \tag{39}$$

where $c > 0$ is a constant controlling the relative importance of the two performance indices.

Note that, if we replace $(\alpha^{(T-t+1)^+} - 1)/(\alpha - 1)$ with its $\alpha \to 1$ limit $(T - t + 1)^+$, then the criterion (39) reduces to the classical Kolmogorov–Shiryayev criterion for detection of a "disorder" [see Shiryayev (1963)], with the exception that Shiryayev (1963) uses $(T - t)^+$ in place of $(T - t + 1)^+$. We could have equivalently considered the delay $(T - t)^+$ rather than $(T - t + 1)^+$. In particular, since T and t are integer valued, it is easy to see that (39) is equivalent to

$$\inf_{T\in\mathcal{T}}\left[(1 - c)P(T < t) + \alpha c\, E\left\{\frac{\alpha^{(T-t)^+} - 1}{\alpha - 1}\right\}\right]. \tag{40}$$

Interestingly, (40) implies that, with $c \geq 1$, the optimal stopping time for (39) is $T \equiv 0$.

It is also noteworthy that, for $\alpha < 1$, a delay penalty of the "opportunity-loss" form,

$$1_{\{T < t\}} - \alpha^{(T-t+1)^+}, \tag{41}$$

is easily treated via (39) by appropriate adjustment of the constant c. In particular, the problem

$$\inf_{T\in\mathcal{T}}\left[P(T < t) + c''E\left\{1_{\{T < t\}} - \alpha^{(T-t+1)^+}\right\}\right], \tag{42}$$

with $c'' > 0$, is equivalent to (39) with $c = c''(1 - \alpha)/(1 + c'') < 1$.

As in previous analyses of the Bayesian disorder problem, we will assume a prior distribution on the change point t of the form

$$P(t = 0) = \pi \quad \text{and} \quad P(t = k|t \geq k) = \rho, \tag{43}$$

where π and ρ are two constants lying in the interval $(0, 1)$. That is, there is a probability π that a change has already occurred when we start observing the sequence; and there is a conditional probability ρ that the sequence will transition to the postchange state at any time, given that it has not done so prior to that time. This model gives rise to a geometric prior distribution

$$P(t = k) = \begin{cases} \pi, & \text{if } k = 0, \\ (1 - \pi)\rho(1 - \rho)^{k-1}, & \text{if } k = 1, 2, \ldots. \end{cases} \tag{44}$$

The solution to problem (39) with the geometric prior (44) is summarized in the following result.

THEOREM 4.1. *For appropriate chosen threshold* $R^* \geq 0$, *the stopping time*

$$T_B = \inf\{k \geq 0 | R_k \geq R^*\} \tag{45}$$

with

$$R_k = \frac{\alpha L(X_k)}{1-\rho}(R_{k-1}+\rho), \qquad k = 1,2,\ldots, \; R_0 = \frac{\alpha\pi}{1-\pi}, \tag{46}$$

is Bayes optimal [*i.e., it solves* (39) *with the geometric prior* (44)]. *Moreover, if* $c \geq 1$, *then* $R^* = 0$.

REMARK 4.2. The stopping time T_B can be written equivalently as

$$T_B = \inf\{k \geq 0 | r_k \geq R^*/(1+R^*)\}, \tag{47}$$

where $r_k = R_k/(1+R_k)$ satisfies the recursion

$$r_k = \frac{\alpha L(X_k)[r_{k-1}+\rho(1-r_{k-1})]}{\alpha L(X_k)[r_{k-1}+\rho(1-r_{k-1})]+(1-\rho)(1-r_{k-1})}, \qquad k = 1,2,\ldots, \tag{48}$$

with

$$r_0 = \frac{\alpha\pi}{1-\pi+\alpha\pi}. \tag{49}$$

With $\alpha = 1$ it is easily seen that $r_k \equiv \pi_k \triangleq P(t \leq k|\mathscr{F}_k)$, and the result of Theorem 4.1 reduces to that of Shiryayev (1963, 1978).

REMARK 4.3. A proof of Theorem 4.1 is given in Appendix C. The basic idea of this proof is to first convert (39) to a standard optimal stopping problem by rewriting the objective of (39) as $E\{Y_T\}$, where

$$Y_k = E\left\{1_{\{k<t\}} + c\frac{\alpha^{(k-t+1)^+}-1}{\alpha-1}\middle| \mathscr{F}_k\right\}, \qquad k = 0,1,\ldots,\infty. \tag{50}$$

This can be done because of the nonnegativity of the Y_k's and the monotone convergence theorem. The sequence $\{Y_k\}$ is given explicitly by

$$Y_k = (1-\pi_k)l(R_k) - \frac{c}{\alpha-1}, \qquad k = 0,1,\ldots, \tag{51}$$

where l is the line

$$l(R) = \frac{\alpha-1+c+cR}{\alpha-1} \tag{52}$$

and where π_k is defined as in Remark 4.2; that is, $\pi_k = P(t \leq k|\mathscr{F}_k)$. [We also have $Y_\infty = \infty$ if $\alpha > 1$ and $Y_\infty = c/(1-\alpha)$ if $\alpha < 1$.] Since (π_k, R_k) forms a homogeneous Markov process, (39) thus reduces to a Markov optimal stopping problem, which can be solved by standard methods (i.e., dynamic programming). In particular, the pay-off resulting from the initial condition $(\pi_0, R_0) = (\pi, R)$ can be shown to be

$$\inf_{T\in\mathscr{T}} E\{Y_T\} = (1-\pi)s(R) - \frac{c}{\alpha-1}, \tag{53}$$

where s is a function satisfying the condition

$$s(R) = l(R) \Leftrightarrow R \ge R^* \quad \text{where } R^* = \inf\{R \ge 0 | s(R) = l(R)\}. \tag{54}$$

Markov optimal stopping theory then implies that the optimal stopping time is given by

$$T_{\text{opt}} = \inf\{k = 0, 1, \dots | l(R_k) = s(R_k)\} = \inf\{k = 0, 1, \dots | R_k \ge R^*\}, \tag{55}$$

which equals T_B.

REMARK 4.4. It follows from the proof of Theorem 4.1 given in Appendix C that the function s appearing in the payoff (53) is the pointwise monotone limit from above of the sequence of functions $\{\mathscr{Q}^n l;\ n = 0, 1, \dots\}$ where l is the line (52) and where the operator $\mathscr{Q}$ is defined in (125). It can be shown (see Appendix C) that $\{\mathscr{Q}^n l;\ n = 0, 1, \dots\}$ is a monotone nonincreasing sequence of continuous functions, from which it follows that the sequence $\{R_n^*\}$ defined by

$$R_n^* = \inf\{R \ge 0 | \mathscr{Q}^n l(R) = l(R)\}, \qquad n = 0, 1, \dots \tag{56}$$

converges upward to the decision threshold R^*. Thus, computation of the threshold and optimal cost can be performed iteratively.

5. Quickest detection with path-dependent exponential costs. Thus far, we have considered delay penalties that depend on the sample path of observations only through the stopping time T. It is also of interest to allow for delay penalties that depend on the sample path in more direct ways. For example, we might wish to replace the exponential penalty $\alpha^{(T-t+1)^+}$ with

$$\prod_{k=t}^{T} \phi_k, \tag{57}$$

where $\{\phi_k\}$ is a nonnegative sequence adapted to the observations. (We adopt the conventions $\prod_a^b = 1$ if $b < a$, and $\sum_a^b = 0$ if $b < a$.) Such a penalty might arise, for example, in the detection of changes in financial time series, where the quantity ϕ_k is the return that an investment would have generated during time period k had it been in force then. [For a related problem, see Beibel and Lerche (1997).]

It is straightforward to show that the replacement of the linear penalty on $(T-t+1)^+$ with path-dependent penalties of the form $\sum_{k=t}^T \phi(X_k)$, where ϕ is a real-valued, measurable function satisfying $0 < \int \phi(x) Q_a(dx) < \infty$, does not materially change the form of the solutions of the Lorden and Shiryayev problems. In this section, we provide results analogous to those of Sections 2 through 4 for certain problems of this type in which the path-dependent costs are exponential.

We first consider the cost structure of (57) in the Bayesian case. In particular, we generalize Theorem 4.1 as follows.

THEOREM 5.1. *Consider the model of Section* 4 *and the problem*

$$\inf_{T\in\mathscr{T}}\left[P(T<t)+cE\left\{\frac{\prod_{k=t}^{T}\phi(X_k)-1}{\alpha-1}\right\}\right], \tag{58}$$

where $c>0$, *and* ϕ *is a real-valued, nonnegative function satisfying*

$$\alpha \triangleq \int \phi(x)Q_a(dx)<\infty \tag{59}$$

and $\alpha\notin\{0,1\}$. *Then, for appropriately chosen threshold* $\hat{R}^*\geq 0$, *the stopping time*

$$\hat{T}_B=\inf\left\{k\geq 0 \,|\, \hat{R}_k\geq \hat{R}^*\right\} \tag{60}$$

with

$$\hat{R}_k=\frac{\phi(X_k)L(X_k)}{1-\rho}\left(\hat{R}_{k-1}+\rho\right),\qquad k=1,2,\dots,\ \hat{R}_0=\frac{\phi(X_0)\pi}{1-\pi}, \tag{61}$$

is Bayes optimal [*i.e., it solves* (58) *with the geometric prior* (44)]. *Moreover, if* $c\geq 1$, *then* $\hat{R}^*=0$.

PROOF. The key to this theorem is the following result, a proof of which is found in Appendix D.

LEMMA 5.2. *Consider the model of Section* 4 *with the constant* $\alpha=\int\phi\, dQ_a\neq 1$ *and the post-change distribution* Q_a *replaced with* $\hat{Q}_a$ *given by*

$$\hat{Q}_a(dx)=\frac{\phi(x)Q_a(dx)}{\alpha}. \tag{62}$$

Let $\hat{P}$ *denote the new probability measure on* $(\Omega,\mathscr{F})$ *defined by this change. Then* (58) *is solved by the solution to the following problem*:

$$\inf_{T\in\mathscr{T}}\left[\hat{P}(T<t)+c\hat{E}\left\{\frac{\alpha^{(T-t+1)^+}-1}{\alpha-1}\right\}\right], \tag{63}$$

where $\hat{E}\{\cdot\}$ *denotes expectation under* $\hat{P}$.

Since $\hat{T}_B$ is the optimal stopping time for (63) (via Theorem 4.1), the theorem follows. □

We now turn to the analogous problem in the minimax setting of Section 2. We have not been able to generalize Theorem 2.1 to the case of a direct substitution of (57) for $\alpha^{(T-t+1)^+}$ in (5). However, from (2) we see that an alternative substitution of interest is to replace (1) with

$$\sum_{l=t}^{T} \prod_{j=t}^{l-1} \phi(X_j), \tag{64}$$

where ϕ is a positive, real-valued measurable function on the range of the observation X_k such that $\int \phi(x) Q_a(dx) < \infty$ and such that $\phi(X_k)L(X_k)$ has no atoms under Q_b. That is, we can consider the following problem:

$$\inf_{T \in \mathscr{T}} \hat{d}(T) \quad \text{subject to } f(T) \geq \gamma, \tag{65}$$

where

$$\hat{d}(T) = \sup_{t \geq 1} \operatorname{ess\,sup} E_t \left\{ \sum_{l=t}^{T} \prod_{j=t}^{l-1} \phi(X_j) \,\middle|\, \mathscr{F}_{t-1} \right\}. \tag{66}$$

Of course (65) reduces to (7) when $\phi \equiv \alpha$.

In this connection, we consider stopping times of the form

$$\hat{T}_h = \inf\{t = 0, 1, 2, \ldots \mid \hat{S}_k \geq h\}, \tag{67}$$

with

$$\hat{S}_k = \max_{1 \leq j \leq k} \left(\prod_{l=j}^{k} \phi(X_l) L(X_l) \right) = \phi(X_k) L(X_k) \max\{\hat{S}_{k-1}, 1\}, \qquad k \geq 1, \tag{68}$$

where $\hat{S}_0 = 0$ and $L = dQ_a/dQ_b$. We then have the following result.

THEOREM 5.3. *Consider the probability model of Section* 2 *and suppose that* $h \geq 0$ *and*

$$P_\infty(\phi(X_1)L(X_1) > 1) > 0. \tag{69}$$

Then $\hat{T}_h$ *solves* (65) *with* $\gamma = f(\hat{T}_h)$. *Moreover, if* $h > 1$, *the quantities* $f(\hat{T}_h)$ *and* $\hat{d}(\hat{T}_h)$ *are given by* (17) *and* (18), *respectively, with* N *replaced with the random variable*

$$\hat{N} = \min\left\{ n \geq 1 \,\middle|\, \sum_{l=1}^{n} [\log L(X_l) + \log \phi(X_l)] \notin (0, \log h) \right\}. \tag{70}$$

PROOF. Analogously to the situation with Theorem 5.1, Theorem 5.3 follows as a corollary to Theorems 2.1 and 3.1, after we note the following result, which is proved in Appendix D.

LEMMA 5.4. *Consider the model of Section* 2 *with* α *chosen as in Lemma* 5.2 *and with* Q_a *replaced with* $\hat{Q}_a$ *as in Lemma* 5.2. *Suppose* $T \in \mathscr{T}$ *is such that* $0 < f(T) < \infty$. *Then*

$$\hat{d}(T) = \sup_{t \geq 1} \operatorname{ess\,sup} \hat{E}_t \left\{ \frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1} \,\middle|\, \mathscr{F}_{t-1} \right\}, \tag{71}$$

where $\hat{E}_t\{\cdot\}$ *denotes expectation under the measure* $\hat{P}_t$ *on* $(\Omega, \mathscr{F})$ *corresponding to a change point at* t *and a postchange distribution* $\hat{Q}_a$.

The theorem follows immediately. □

6. Conclusion. We have considered the quickest detection problems of Lorden and Shiryayev when the linear penalty on detection delay is replaced with a possibly path-dependent exponential delay penalty. We have seen that each of these problems is solved by replacing, in the corresponding linear-penalty optimal stopping rules, the likelihood ratio between pre- and postchange distributions with a scaled version of itself. We have also explored the issue of performance analysis in each of these problems. In the minimax problem this involves separate analysis of the exit statistics of the optimal stopping rule; whereas in the Bayesian problem this involves the computation of the payoff function, which is an adjunct to the determination of the optimal stopping rule.

In each of the problems considered, we have examined the tradeoff of opposing performance indices: $d(T)$ and $f(T)$ in the minimax formulation and $P(T < t)$ and $E\{(\alpha^{(T-t+1)^+} - 1)/\alpha - 1\}$ in the Bayesian formulation. In the minimax case, the tradeoff was effected by minimizing one of these indices with a constraint on the other, and in the Bayesian formulation we opted for the minimization of a linear combination of the two indices. These optimization criteria were chosen because they are the traditional ones for their respective change-detection formulations. However, we could of course have considered alternatives such as a linear combination of the performance indices in the minimax case, or a false-alarm constrained minimization in the Bayesian case. As with their linear-delay counterparts, these alternative problems should have essentially the same solutions (aside from the choice of threshold) as the problems presented here [cf. Theorem 4.10 of Shiryayev (1978)]. Similarly, one might also introduce other combinations of performance indices, such as trading false-alarm probability against minimax delay [e.g., Yakir (1996)].

Several other problems of interest are suggested by this work. For example, it may be interesting to explore formal connections between the minimax and Bayesian formulations of this problem, as has been done in the case of linear delay penalty in Beibel (1996), Bojdecki and Hosza (1984) and Ritov (1990). Also, continuous-time versions are of interest; in the minimax case, they formalize the connection between results such as (26) and (27) and optimal stopping solutions [cf. Beibel (1996)]; and in the Bayesian case, they can give rise to closed-form solutions for the payoff [e.g., Theorem 4.9 of Shiryayev (1978)]. Moreover, continuous-time solutions are sometimes particularly simple when viewed as "generalized parking" problems [see, e.g., Beibel (1994) or Beibel and Lerche (1997)]. Finally, alternative ways of invoking exponential penalties may be of interest. For example, it is common to consider problems of optimal stopping in which the rewards are exponentially discounted by a deflator [see, e.g., Dubins and Teicher (1967)]. A similar formulation in which the deflator changes at an unknown time could be used to model an exponential cost of delay in quickest detection problems.

APPENDIX

A. Proofs for Section 2.

PROOF OF LEMMA 2.3. This result is the exponential-delay analog of Lemma 3 of Moustakides (1986), the proof of which can be adapted straightforwardly to this case. In particular, we define

$$b_t(T) = E_t\left\{ \frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1} \,\middle|\, \mathscr{F}_{t-1} \right\}. \tag{72}$$

On applying the identity (2) we can write

$$b_t(T) = \sum_{k=t}^{\infty} \sum_{l=0}^{k-t} \alpha^l P_t(T = k \mid \mathscr{F}_{t-1}) + b_\infty P_t(T = \infty \mid \mathscr{F}_{t-1}), \tag{73}$$

with $b_\infty = 1/(1-\alpha)$ if $\alpha < 1$, and $b_\infty = \infty$ if $\alpha \geq 1$. We can then write

$$\begin{aligned} b_t(T) &= \sum_{k=t}^{\infty} \alpha^{k-t} P_t(T \geq k \mid \mathscr{F}_{t-1}) \\ &= \sum_{k=t}^{\infty} \alpha^{k-t} E_\infty\left\{ \prod_{l=1}^{k-1} L(X_l) 1_{\{T \geq k\}} \,\middle|\, \mathscr{F}_{t-1} \right\} \\ &= E_\infty\left\{ \left[\sum_{k=t}^{T} \sum_{l=t}^{k-1} \alpha L(X_l) \right] \,\middle|\, \mathscr{F}_{t-1} \right\}. \end{aligned} \tag{74}$$

Also, analogously to Lemma 1 of Moustakides (1986), the sequence $U_n = \max\{S_n, 1\}$ has the properties that, for any $n > m \geq 1$ and for fixed $\{X_{m+1}, \dots, X_n\}$, S_n is a nondecreasing function of U_m, and that

$$U_n = \sum_{l=1}^{n+1} (1 - S_{l-1})^+ \prod_{j=l}^{n} \alpha L(X_l). \tag{75}$$

With these modifications, Lemma 2.3 follows by an argument identical to that used in Lemma 3 of Moustakides (1986) [namely, equations (11)–(14) of Moustakides (1986)]. □

PROOF OF LEMMA 2.4. In proving this result, we distinguish three cases: $\beta < 1$, $\beta = 1$ and $\beta > 1$, where

$$\beta = \frac{\log \alpha}{D(Q_b \| Q_a)}, \tag{76}$$

with $D(Q_b \| Q_a)$ from (3).

Case 1: $\beta < 1$. Lemma 2.4 is the exponential-delay analog of Theorem 1 of Moustakides (1986). The proof of Theorem 1 of Moustakides applies almost exactly to Lemma 2.4 in this case, with the exception that Lemma 4 of Moustakides must be generalized to assert that the moments of the stopping time

$$\nu_1 = \inf\{n > 0 | S_n \leq 1\} \tag{77}$$

are finite under P_∞. To see that this is true, we first note that $\beta < 1$ implies, via Lemma 2.8 of Chow, Robbins and Siegmund (1971), that $P_\infty(\nu_1 < \infty) = 1$. For each integer $r \geq 0$, we can write

$$E_\infty\{(\nu_1)^r\} \leq \sum_{k=1}^{\infty} k^r P_\infty\left(\prod_{l=1}^{k-1} \alpha L(X_l) > 1\right), \tag{78}$$

where we have used the fact the

$$\{\nu_1 = k\} \subset \{S_1 > 1, \dots, S_{k-1} > 1\} \subset \left\{\prod_{l=1}^{k-1} \alpha L(X_l) > 1\right\}.$$

The probability in the summand in the right-hand term in (78) is the probability that a random walk with finite, negative mean exceeds zero. Since $E_\infty\{\alpha L\} = \alpha < \infty \Rightarrow E_\infty\{[(\log \alpha L)^+]^r\} < \infty$, for all integers $r \geq 0$, Theorem 1[(i) $\Leftrightarrow$ (x)] of Janson (1986) implies that the right-hand side of (78) is finite. Thus, we conclude that all moments of ν_1 are finite under P_∞, as required.

With this generalization of Lemma 4 of Moustakides (1986) it is easily checked that the remainder of the proof of Theorem 1 of Moustakides goes through exactly for the exponential delay case with the substitution of $\alpha L(X_k)$ for $L(X_k)$.

Case 2: $\beta = 1$. Fix the distributions Q_b and Q_a and consider variation in β due to variation in α only. For $\beta \leq 1$, let $\{S_k^\beta\}$ denote the sequence $\{S_k\}$ defined by the corresponding choice of α. Note that S_k^β is strictly increasing in β for each $k \geq 1$, and thus we have (recall that g is nonincreasing)

$$\sup_{T \in \bar{\mathscr{T}}_\gamma} E_\infty\left\{\sum_{k=1}^{T-1} g(S_k^1)\right\} \leq \sup_{T \in \bar{\mathscr{T}}_\gamma} E_\infty\left\{\sum_{k=1}^{T-1} g(S_k^\beta)\right\} \qquad \forall \beta \leq 1, \tag{79}$$

where $\bar{\mathscr{T}}_\gamma$ denotes the set of all elements of $\mathscr{T}$ satisfying the constraint $f(T) = \gamma$.

For each $\beta \in [0, 1]$, let T^β denote the element of $\bar{\mathscr{T}}_\gamma$ defined by T_h for the corresponding value of α. Since $L(X_k)$ has no atoms under P_∞ and since g is continuous, we have that $T^\beta \to T^1$ and $g(S_k^\beta) \to g(S_k^1)$ almost surely under P_∞, as $\beta \uparrow 1$. Thus, we conclude that

$$\lim_{\beta \uparrow 1} \sum_{k=1}^{T^\beta - 1} g(S_k^\beta) = \sum_{k=1}^{T^1 - 1} g(S_k^1) \qquad \text{a.s. } [P_\infty]. \tag{80}$$

Let h_β denote the threshold used in T^β. Then, for $\beta \in [0,1]$, we have

$$\begin{aligned}(81)\qquad \left|\sum_{k=1}^{T^\beta - 1} g(S_k^\beta)\right| &\le T^\beta \max\{|g(0)|, |g(h_\beta)|\}\\ &\le T^\beta \max\{|g(0)|, |g(h_0)|, |g(h_1)|\},\end{aligned}$$

where we have used the fact that h_β is nondecreasing in β. We can conclude further that $T^\beta \le T_1^\beta \le T_1^0$ a.s. $[P_\infty]$ where, for $x \in \{\beta, 1\}$, T_1^x denotes the first time S_k^x crosses h_1. Since $E_\infty\{T_1^0\} < \infty$, the left-hand side of (81) is thus bounded by a fixed integrable random variable for all $\beta \in [0,1]$. The dominated convergence theorem and (80) then imply that

$$(82)\qquad \lim_{\beta \uparrow 1} E_\infty\left\{\sum_{k=1}^{T^\beta - 1} g(S_k^\beta)\right\} = E_\infty\left\{\sum_{k=1}^{T^1 - 1} g(S_k^1)\right\}.$$

Combining (79) with (82), we conclude that T^1 is optimal in $\bar{\bar{\mathscr{T}}}_\gamma$ for $\beta = 1$, and thus Lemma 2.4 follows for this case.

Case 3: $\beta > 1$. The moments of ν_1 of (77) are infinite in this case [cf. Doob (1953), page 308, for the case $\beta = 1$. The case $\beta > 1$ then follows since ν_1 is nondecreasing in α]. So, it is more difficult to adapt the proof of Theorem 1 of Moustakides (1986) to Lemma 2.4. However, it is straightforward to prove Lemma 2.4 directly in this case.

It suffices to show that there is a $\lambda \in \mathbb{R}$ such that T_h solves

$$(83)\qquad \sup_{T \in \mathscr{T}} E_\infty\left\{\sum_{k=0}^{T-1} [g(S_k) - \lambda]\right\}.$$

It is easily seen that T_0 solves this problem when $\lambda \ge g(0)$, and T_∞ solves it when $\lambda \le g(\infty) \triangleq \lim_{s\to\infty} g(s) = \inf_{s \ge 0} g(s)$. So, let us consider the situation $\lambda \in (g(\infty), g(0))$.

Fix $\lambda \in (g(\infty), g(0))$ and, for all $s \ge 0$, define the sequence

$$(84)\qquad Y_n^s = \sum_{k=0}^{n-1} g(\bar{S}_k^s) - n\lambda, \qquad n = 0, 1, \dots,$$

where

$$(85)\qquad \bar{S}_k^s = \max\{\bar{S}_{k-1}^s, 1\}\alpha L(X_k), \qquad k = 1, 2, \dots, \ \bar{S}_0^s = s.$$

Note that (83) can be rewritten as

$$(86)\qquad \sup_{T \in \mathscr{T}} E_\infty\{Y_T^s\},$$

with $s = 0$. For each $s \ge 0$, we can write

$$\begin{aligned}(87)\qquad (Y_n^s)^+ &\le \sum_{k=0}^{n-1} \left(g(\bar{S}_k^s) - \lambda\right)^+\\ &\le (g(s) - \lambda)^+ + \sum_{k=1}^{n-1}\left(g\left(\max\{s,1\}\alpha^k L_1^k\right) - \lambda\right)^+,\end{aligned}$$

with $L_1^k = \prod_{j=1}^k L(X_j)$, where we use the facts that $\bar{S}_k^s \geq \max\{s, 1\}\alpha^k L_1^k$ and that g is nonincreasing. We then have

$$\begin{aligned} E_\infty\left\{\sup_n (Y_n^s)^+\right\} &\leq g(0) - \lambda + E_\infty\left\{\sum_{k=1}^\infty \left(g\left(\max\{s,1\}\alpha^k L_1^k\right) - \lambda\right)^+\right\} \\ &\leq [g(0) - \lambda] \sum_{k=0}^\infty P_\infty\left(\alpha^k L_1^k < g^{-1}(\lambda)/\max\{s,1\}\right), \end{aligned} \tag{88}$$

where $g^{-1}(\lambda) = \inf\{s \geq 0 | g(s) \geq \lambda)\} < \infty$. Since $\beta > 1$, the summand in the rightmost term of (88) is the probability that a random walk with finite positive mean falls below a real threshold. Analogously with the situation in (78), this probability is summable, and thus we can conclude that

$$E_\infty\left\{\sup_n (Y_n^s)^+\right\} < \infty. \tag{89}$$

It follows from (89) and Theorem 4.5′ of Chow, Robbins and Siegmund (1971) that (86) is solved by the stopping time

$$T_{\mathrm{opt}}^s = \inf\left\{k \geq 0 | Y_k^s = \bar{\gamma}_k^s\right\}, \tag{90}$$

where $\{\bar{\gamma}_k^s\}$ is the Snell envelope of $\{Y_k^s\}$,

$$\bar{\gamma}_k^s = \operatorname*{ess\,sup}_{T \in \mathscr{T}_k} E_\infty\left\{Y_T^s \mid \mathscr{F}_k\right\}, \qquad k = 0, 1, \dots, \tag{91}$$

and where $\mathscr{T}_k$ denotes the subset of $\mathscr{T}$ satisfying $P_\infty(T \geq k) = 1$. (Recall that the essential supremum of a family of random variables is the smallest random variable that almost surely dominates all members of the family.)

The homogeneous Markovity of $\{\bar{S}_k^s\}$ allows us to represent $\bar{\gamma}_k^s$ as

$$\bar{\gamma}_k^s = Y_k + v\left(\bar{S}_k^s, \lambda\right), \tag{92}$$

where

$$v(s, \lambda) = \sup_{T \in \mathscr{T}} E_\infty\left\{Y_T^s\right\}, \qquad s \geq 0. \tag{93}$$

We can now use an argument similar to that used on page 1386 of Moustakides (1986) to show that there is a λ such that $T_{\mathrm{opt}}^0 = T_h$. In particular, we look for a root of the function

$$b(\lambda) = g(h) - \lambda + \mathscr{V} v(h, \lambda), \tag{94}$$

with

$$\mathscr{V} v(s, \lambda) = \int v(\max\{s, 1\}\alpha L(x), \lambda) Q_b(dx), \qquad s \geq 0. \tag{95}$$

Since $v(s, \lambda)$ is the maximum of a family of decreasing, linear functions in λ, it is a nonincreasing convex function of λ. It follows that b is convex (and hence continuous) and nonincreasing on $(g(\infty), \infty)$. Since $v(s, g(0)) \equiv 0$ and $g(h) \leq g(0)$ we have that $b(g(0)) \leq 0$. Moreover, the nonnegativity of v implies that b has a root in $(g(\infty), g(0)]$ if there is a $\lambda \in (g(\infty), g(0))$ for which $b(\lambda) > 0$. This is trivially true if $g(h) > g(\infty)$. If, on the other hand $g(h) =$

$g(\infty)$, then this will be true if there is a $\lambda \in (g(\infty), g(0))$ such that $\mathscr{V}v(h, \lambda) > 0$. The latter condition requires only that there be a $\lambda \in (g(\infty), g(0))$ such that $P_\infty(\max\{h, 1\}L(X_1) > g^{-1}(\lambda)) > 0$, a condition that can always be met because $g^{-1}(\lambda)$ can be made arbitrarily small.

Since $\overline{S}_k^s$ is nondecreasing in s, and g is nonincreasing, it follows that $v(s, \lambda)$ is a nonincreasing function of s. Moreover, v satisfies the integral equation [see Theorem II.16 of Shiryayev (1973)]

$$v(s, \lambda) = \max\{0, g(s) - \lambda + \mathscr{V}v(s, \lambda)\}, \qquad s \geq 0, \tag{96}$$

where the operator $\mathscr{V}$ is as above. With λ chosen as a root of b, it is clear from (96) that T_{opt}^0 will equal T_h.

Thus, Lemma 2.4 follows for $\beta > 1$ and hence for all β. □

B. Proof of Theorem 3.1. We consider first $d(T_h)$. Under the distribution P_∞ the observations $X_1, X_2, \dots$ are i.i.d. with marginal distribution Q_b. It is clear that T_h arises from a renewal process, with renewals occurring whenever the accumulated sum m_k of (16) is reset to zero, and with a termination when m_k exits from $[0, \log h)$. It follows that we can write

$$T_h = \sum_{j=1}^{J} N_j \qquad \text{a.s.}\ [P_\infty], \tag{97}$$

where $N_1, N_2, \dots$ are i.i.d. repetitions (under P_∞) of the random variable N of (19) and where J denotes the number of repetitions of N that occur before the sum exits at the upper boundary $\log h$. Let M_j denote the indicator of the event that the jth repetition of N results in an exit at the upper boundary. Then J is a stopping time with respect to the sequence $(N_1, M_1), (N_2, M_2), \dots,$ which is i.i.d. under P_∞. Since $E\{T_h\}$ clearly exists, the generalized Wald identity [see, e.g., Robbins and Samuel (1966)] allows us to write

$$E_\infty\{T_h\} = E_\infty\{J\}E_\infty\{N\}. \tag{98}$$

It is easy to see that, under P_∞, J is a geometric random variable with

$$P_\infty(J = j) = [1 - P_\infty(F_0)][P_\infty(F_0)]^{j-1}, \qquad j = 1, 2, \dots, \tag{99}$$

and the equality in (17) thus follows. To prove the inequality in (17), we note that (10) implies $P_\infty(F_0) < 1$, from which it follows that $E_\infty\{J\} < \infty$. Furthermore, a lemma of Stein's [cf. Siegmund (1985), Proposition 2.19] implies that $E_\infty\{N\} < \infty$, and so the inequality in (17) follows as well.

To analyze $d(T_h)$ it is useful to note first that the worst-case prechange sample paths are those that lead to a resetting of m_k just before the change point. Consequently, the stopping time T_h is an equalizer rule, and

$$d(T_h) = d_1(T_h) = E_1\left\{\frac{\alpha^{T_h} - 1}{\alpha - 1}\right\}. \tag{100}$$

Since, under the measure P_1, $X_1, X_2, \dots$ is an i.i.d. sequence drawn from Q_a, the analysis of this quantity can proceed in much the same fashion as that of

$f(T_h)$. In particular, using (97) we can write

$$E_1\{\alpha^{T_h}\} = E_1\{\alpha^{N_1} \cdots \alpha^{N_J}\} = \sum_{j=1}^{\infty} E_1\{\alpha^{N_1} \cdots \alpha^{N_j} | J = j\} P_1(J = j), \tag{101}$$

where the second equality follows since $\alpha^{T_h} > 0$.

Consider the summand $E_1\{\alpha^{N_1} \cdots \alpha^{N_j} | J = j\} P_1(J = j)$. As in the preceding analysis, we can write

$$P_1(J = j) = [1 - P_1(F_0)][P_1(F_0)]^{j-1}, \qquad j = 1, 2, \ldots. \tag{102}$$

We can further write

$$E_1\left\{\prod_{l=1}^{j} \alpha^{N_l} \,\middle|\, J = j\right\} = E_1\left\{\prod_{l=1}^{j} \alpha^{N_l} \,\middle|\, M_1 = \cdots = M_{j-1} = 0, M_j = 1\right\}, \tag{103}$$

where $M_1, M_2, \ldots$ are defined as above. The random variables $N_1, N_2, \ldots$ are conditionally independent given the random variables $M_1, M_2, \ldots$. Thus, (103) becomes

$$E_1\{\alpha^{N_1} \cdots \alpha^{N_j} | J = j\} = E_1\{\alpha^N | F_0^c\}[E_1\{\alpha^N | F_0\}]^{j-1}. \tag{104}$$

Since N is the first exit time of the random walk $\sum_{l=1}^{n}[\log L(X_l) + \log \alpha]$ from an interval, it follows straightforwardly from Proposition IV-4-19 of Neveu (1975) that

$$E_1\left\{\left[\prod_{l=1}^{N} L(X_l)\right]^{-1}\right\} = 1. \tag{105}$$

Moreover, (10) implies that $P_1(F_0^c) > 0$ and, since $L(X_k)$ is almost surely positive under Q_a, it further follows that

$$E_1\left\{\left[\prod_{l=1}^{N} L(X_l)\right]^{-1} 1_{F_0}\right\} < 1. \tag{106}$$

On F_0 we have that $\alpha^N \prod_{l=1}^{N} L(X_l) \le 1$, and thus (106) implies that

$$E_1\{\alpha^N 1_{F_0}\} < 1. \tag{107}$$

On combining (101), (102), (104) and (107), we can write

$$E_1\{\alpha^{T_h}\} = \frac{E_1\{\alpha^N 1_{F_0^c}\}}{1 - E_1\{\alpha^N 1_{F_0}\}}, \tag{108}$$

and (18) follows.

This completes the proof of Theorem 3.1. □

C. Proof of Theorem 4.1. In this proof all statements concerning random variables and sequences of random variables are taken to hold almost surely under the measure P.

In view of Remark 4.2, in proving Theorem 4.1 we need only consider the case $\alpha \neq 1$, which we henceforth assume. Moreover, in view of the discussion following (40), the situation for $c \geq 1$ is trivial. Thus, we may also restrict attention in the remainder of the proof the case $c < 1$.

We first note that (51) follows straightforwardly from Bayes' formula, which implies

$$(109)\quad E\{1_{\{k<t\}}|\mathscr{F}_k\} \equiv 1 - \pi_k = \frac{1}{1+W_k} \quad \text{and} \quad E\{\alpha^{(k-t+1)^+}|\mathscr{F}_k\} = \frac{R_k+1}{1+W_k},$$

where

$$(110)\qquad W_k = \frac{L(X_k)}{1-\rho}(W_{k-1}+\rho), \qquad k = 1,2,\dots,\ W_0 = \pi,$$

and where R_k is given by (46). [The values of Y_∞ follow directly from the definition and from the asymptotic properties of the likelihood ratio; cf. Chung (1968), Theorem 8.2.5.]

Since $\{Y_k\}$ is adapted and is bounded from below, the optimal stopping time [see Theorem 4.5′ of Chow, Robbins and Siegmund (1971)] is

$$(111)\qquad T_{\mathrm{opt}} = \inf\{k \geq 0 | Y_k = \gamma_k\},$$

where $\{\gamma_k\}$ is the Snell envelope of $\{Y_k\}$,

$$(112)\qquad \gamma_k = \operatorname*{ess\,inf}_{T\in\mathscr{T}_k} E\{Y_T|\mathscr{F}_k\}$$

and where $\mathscr{T}_k$ is the subset of $\mathscr{T}$ that satisfies $P(T \geq k) = 1$.

In order to determine the Snell envelope, we first prove the following result.

LEMMA C.1. *For each integer $n \geq 0$, define a sequence of random variables*

$$(113)\qquad \gamma_k^n = \operatorname*{ess\,inf}_{T\in\mathscr{T}_k^n} E\{Y_T|\mathscr{F}_k\}, \qquad k = 0,1,\dots,n,$$

where $\mathscr{T}_k^n$ is the subset of $\mathscr{T}_k$ that satisfies $P(T \leq n) = 1$. Then, for all $k \geq 0$,

$$(114)\qquad \lim_{n\to\infty} \gamma_k^n = \gamma_k \qquad a.s.\,[P].$$

PROOF. Note first that, for $\alpha < 1$, Y_k is bounded from above as well as from below, and Lemma C.1 follows immediately from the "triple limit theorem" [cf. Theorem 4.8(b) of Chow, Robbins and Siegmund (1971)]. So we restrict attention to the case $\alpha > 1$.

Fix k and n with $k < n < \infty$. Since Y_k is bounded from below by zero, it follows from Theorem 4.5′ of Chow, Robbins and Siegmund (1971) [applied under the conditional probability measure $P(\cdot|\mathscr{F}_k)$] that there is a stopping time $T_k \in \mathscr{T}_k$ such that $P(T_k < \infty|\mathscr{F}_k) = 1$ and $\gamma_k = E\{Y_{T_k}|\mathscr{F}_k\}$. Since $\gamma_k \leq Y_k < \infty$ and T_k is almost surely finite conditioned on $\mathscr{F}_k$, we can write

$$(115)\quad \gamma_k = E\{Y_{n\wedge T_k}|\mathscr{F}_k\} + \sum_{l=n+1}^{\infty} E\{Y_l 1_{\{T_k=l\}}|\mathscr{F}_k\} - E\{Y_n 1_{\{T_k>n\}}|\mathscr{F}_k\},$$

where $a \wedge b = \min\{a, b\}$.

Now, again using the fact that $\gamma_k < \infty$, we have that

$$\lim_{n\to\infty} \sum_{l=n+1}^{\infty} E\{Y_l 1_{\{T_k=l\}} | \mathscr{F}_k\} = 0. \tag{116}$$

On using the facts that $\mathscr{F}_k \subset \mathscr{F}_n$ and $\{T_k > n\} \in \mathscr{F}_n$, we can write

$$\begin{aligned} E\{Y_n 1_{\{T_k>n\}} | \mathscr{F}_k\} &= E\left\{ E\left\{ 1_{\{n<t\}} + c\frac{\alpha^{(n-t+1)^+} - 1}{\alpha - 1} \middle| \mathscr{F}_n \right\} 1_{\{T_k>n\}} \middle| \mathscr{F}_k \right\} \\ &= E\left\{ \left(1_{\{n<t\}} + c\frac{\alpha^{(n-t+1)^+} - 1}{\alpha - 1} \right) 1_{\{T_k>n\}} \middle| \mathscr{F}_k \right\} \\ &\le P(T_k > n | \mathscr{F}_k) + c \sum_{l=n+1}^{\infty} E\left\{ \frac{\alpha^{(n-t+1)^+} - 1}{\alpha - 1} 1_{\{T_k=l\}} \middle| \mathscr{F}_k \right\}. \end{aligned} \tag{117}$$

Since T_k is almost surely finite conditioned on $\mathscr{F}_k$, we have that

$$\lim_{n\to\infty} P(T_k > n | \mathscr{F}_k) = 0. \tag{118}$$

Consider the summand in the second term to the right of the inequality in (117) for a fixed $l > n$. Because $\mathscr{F}_k \subset \mathscr{F}_l$ and $\{T_k = l\} \in \mathscr{F}_l$ we have

$$\begin{aligned} E\left\{ \frac{\alpha^{(n-t+1)^+} - 1}{\alpha - 1} 1_{\{T_k=l\}} \middle| \mathscr{F}_k \right\} &= E\left\{ E\left\{ \frac{\alpha^{(n-t+1)^+} - 1}{\alpha - 1} \middle| \mathscr{F}_l \right\} 1_{\{T_k=l\}} \middle| \mathscr{F}_k \right\} \\ &\le E\left\{ E\left\{ \frac{\alpha^{(l-t+1)^+} - 1}{\alpha - 1} \middle| \mathscr{F}_l \right\} 1_{\{T_k=l\}} \middle| \mathscr{F}_k \right\}. \end{aligned} \tag{119}$$

We can conclude from (116) that the sum

$$\sum_{l=n+1}^{\infty} E\left\{ E\left\{ \frac{\alpha^{(l-t+1)^+} - 1}{\alpha - 1} \middle| \mathscr{F}_l \right\} 1_{\{T_k=l\}} \middle| \mathscr{F}_k \right\} \tag{120}$$

converges almost surely to zero conditioned on $\mathscr{F}_k$, and it thus follows that

$$\lim_{n\to\infty} E\{Y_n 1_{\{T_k>n\}} | \mathscr{F}_k\} = 0. \tag{121}$$

Now, since $n \wedge T_k \in \mathscr{T}_k^n$, it follows from the definition of γ_k^n that

$$E\{Y_{n\wedge T_k} | \mathscr{F}_k\} \ge \gamma_k^n, \tag{122}$$

from which we conclude [via (115), (116), (121) and (122)] that

$$\gamma_k \ge \limsup_{n\to\infty} \gamma_k^n. \tag{123}$$

Since we clearly have $\gamma_k \le \gamma_k^n$, the lemma follows. □

Lemma C.1 allows us to find the Snell envelope by first considering the finite-horizon problem, $\inf_{T\in\mathscr{T}_0^n} E\{Y_T\}$ and then passing to the limit. In view of the recursive nature of the sequences π_k and R_k, the finite-horizon problem can be solved using dynamic programming [see, e.g., Theorem 3.2 of Chow, Robbins and Siegmund (1971)]. This approach leads straightforwardly to the representation

$$\gamma_k = (1-\pi_k)\lim_{n\to\infty}\mathcal{Q}^n l(R_k) - \frac{c}{\alpha-1}, \tag{124}$$

where the operator $\mathcal{Q}$ is defined as

$$\mathcal{Q}h(R) = \min\left\{l(R), (1-\rho)\int h\left(\frac{\alpha L(x)}{1-\rho}(R+\rho)\right)Q_b(dx)\right\}, \quad R \geq 0. \tag{125}$$

[It is easily seen that the integral in (125) exists when h is any member of the sequence $\{\mathcal{Q}^n l\}$.]

A straightforward inductive argument shows that each of the functions $\mathcal{Q}^n l$ is continuous and concave and moreover that $\{\mathcal{Q}^n l\}$ is a pointwise monotone nonincreasing sequence of functions. So, this sequence has a pointwise limit s, which, as the monotone pointwise limit from above of a sequence of concave functions, must also be concave. Further, by the monotone convergence theorem, s must satisfy the nonlinear integral equation

$$s(R) = \min\{l(R), \mathcal{Q}s(R)\}, \qquad R \geq 0, \tag{126}$$

which implies that it is continuous. Thus, we can write

$$\gamma_k = (1-\pi_k)s(R_k) - \frac{c}{\alpha-1}, \qquad k = 0, 1, \dots, \tag{127}$$

where s is a continuous concave function bounded from above by the line l of (52). It thus follows from the above properties that the optimal stopping time (111) is given by

$$T_{\mathrm{opt}} = \inf\{k = 0, 1, \dots \mid l(R_k) = s(R_k)\}. \tag{128}$$

A further inductive argument shows that the difference $l - \mathcal{Q}^n l (\geq 0)$ is a monotone nonincreasing function for each n, from which we can conclude that $l - s$ is nonnegative and nonincreasing. This implies that s satisfies (54), and in turn that

$$T_{\mathrm{opt}} = \inf\{k = 0, 1, \dots \mid R_k \geq R^*\}, \tag{129}$$

where R^* is the constant (54). (Note that, for $\alpha < 1$, it is possible that $R^* = \infty$, in which case the optimal stopping time is $T \equiv \infty$.) Since T_{opt} is the same as T_B, this completes the proof of the theorem. □

D. Proofs for Section 5.

Proof of Lemma 5.2. Similarly to the situation in Theorem 4.1, the objective of the optimization (58) can be written as $E\{Z_T\}$, where

$$\begin{aligned} Z_k &= E\left\{1_{\{k<t\}} + c\frac{\prod_{j=t}^{T}\phi(X_j) - 1}{\alpha-1}\,\middle|\,\mathcal{F}_k\right\} \\ &= (1-\pi_k)l(\hat{R}_k) - \frac{c}{\alpha-1}, \end{aligned} \tag{130}$$

with $l(R)$ as before and where $Z_\infty = c/(1-\alpha)$ for $\alpha < 1$ and $Z_\infty = \infty$ for $\alpha > 1$. (In this proof, statements concerning random variables are taken to hold almost surely under P.)

We can write

$$\begin{aligned} E\{Z_T\} &= \sum_{k=0}^{\infty} E\{Z_k 1_{\{T=k\}}\} + Z_\infty P(T=\infty) \\ &= \sum_{k=0}^{\infty} E\Big\{(1-\pi_k)\, l\big(\hat{R}_k\big) 1_{\{T=k\}}\Big\} \\ &\quad + \left(Z_\infty + \frac{c}{\alpha-1}\right) P(T=\infty) - \frac{c}{\alpha-1}. \end{aligned} \tag{131}$$

The summand in the rightmost term of this equation is given by

$$E\Big\{(1-\pi_k)\, l\big(\hat{R}_k\big) 1_{\{T=k\}}\Big\} = \hat{E}\Big\{\lambda_k (1-\pi_k)\, l\big(\hat{R}_k\big) 1_{\{T=k\}}\Big\}, \tag{132}$$

where λ_k is the Radon–Nikodym derivative of the restriction of P to $\mathscr{F}_k$, with respect to the restriction of $\hat{P}$ to $\mathscr{F}_k$. It is easily seen that $\lambda_k(1-\pi_k) = 1 - \hat{\pi}_k$ with $\hat{\pi}_k = \hat{P}(t \le k | \mathscr{F}_k)$, and thus we can write

$$\begin{aligned} E\{Z_T\} &= \sum_{k=0}^{\infty} \hat{E}\Big\{(1-\hat{\pi}_k)\, l\big(\hat{R}_k\big) 1_{\{T=k\}}\Big\} \\ &\quad + \left(Z_\infty + \frac{c}{\alpha-1}\right) P(T=\infty) - \frac{c}{\alpha-1}. \end{aligned} \tag{133}$$

Now, for $\alpha < 1$ we have $Z_\infty + c/(\alpha - 1) = 0$, and so

$$E\{Z_T\} = \sum_{k=0}^{\infty} \hat{E}\Big\{\hat{Y}_k 1_{\{T=k\}}\Big\} + \hat{Y}_\infty \hat{P}(T=\infty), \tag{134}$$

with

$$\hat{Y}_k = (1-\hat{\pi}_k)\, l\big(\hat{R}_k\big) - \frac{c}{\alpha-1}, \qquad k = 0, 1, \dots, \tag{135}$$

and $\hat{Y}_\infty = c/(1-\alpha)$. From Lemma 4.3, it follows that

$$E\{Z_T\} = \hat{P}(T < t) + c\hat{E}\left\{\frac{\alpha^{(T-t+1)^+} - 1}{\alpha - 1}\right\}, \tag{136}$$

and the lemma follows for $\alpha < 1$.

Now suppose $\alpha > 1$. We know from Theorem 4.1 that (71) is solved by $\hat{T}_B$ from (60) with $\hat{R}^* < \infty$. Since

$$R_k \ge \prod_{j=1}^{k} \phi(X_j) L(X_j) \to \infty, \tag{137}$$

we see that $P(\hat{T}_B < \infty) = 1$. So, we can conclude from (133) that $\hat{T}_B$ also solves (58), and the lemma follows for $\alpha > 1$ as well. □

PROOF OF LEMMA 5.4. Here, we need only note that [with b_∞ as in (73)]

$$
\begin{aligned}
&E_t\left\{\sum_{l=t}^{T}\prod_{j=t}^{l-1}\phi(X_j)\,\middle|\,\mathscr{F}_{t-1}\right\}\\
&\qquad = E_t\left\{\sum_{k=t}^{\infty}\sum_{l=t}^{k}\prod_{j=t}^{l-1}\phi(X_j)1_{\{T=k\}}\,\middle|\,\mathscr{F}_{t-1}\right\} + b_\infty P_t(T=\infty)\\
&\qquad = E_t\left\{\sum_{k=t}^{\infty}\prod_{l=t}^{k-1}\phi(X_l)1_{\{T\geq k\}}\,\middle|\,\mathscr{F}_{t-1}\right\} \qquad (138)\\
&\qquad = \sum_{k=t}^{\infty}E_\infty\left\{\prod_{l=t}^{k-1}\phi(X_l)L(X_l)1_{\{T\geq k\}}\,\middle|\,\mathscr{F}_{t-1}\right\}\\
&\qquad = E_\infty\left\{\prod_{l=t}^{T-1}\phi(X_l)L(X_l)\,\middle|\,\mathscr{F}_{t-1}\right\}.
\end{aligned}
$$

Thus we see that the constant α and measure Q_a enter the problem of Section 2 only through the product $\alpha\, dQ_a/dQ_b(x)$, which, by the final expression in the above equation, can be replaced by

$$\phi(x)L(x) = \int\phi(y)Q_a(dy)\times\frac{d\hat{Q}_a}{dQ_b}(x). \qquad (139)$$

The lemma follows. □

REFERENCES

BASSEVILLE, M. and NIKIFOROV, I. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

BEIBEL, M. (1994). Bayes problems in change-point models for the Wiener process. In *Change Point Problems* (E. Carlstein, H.-G. Müller and D. Siegmund, eds.) 1–6. IMS, Hayward, CA.

BEIBEL, M. (1996). A note on Ritov's approach to the minimax property of the cusum procedure. *Ann. Statist.* **24** 1804–1812.

BEIBEL, M. and LERCHE, H. R. (1997). A new look at optimal stopping problems related to mathematical finance. *Statist. Sinica* **7** 93–108.

BOJDECKI, T. and HOSZA, J. (1984). On a generalized disorder problem. *Stochastic Process. Appl.* **18** 349–359.

BRODSKY, B. E. and DARKHOVSKY, B. S. (1992). *Nonparametric Methods in Change-point Problems*. Kluwer, Boston.

CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change Point Problems*. IMS, Hayward, CA.

CHOW, Y.S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton-Mifflin, Boston.

CHUNG, K. L. (1968). *A Course in Probability Theory*. Academic Press, New York.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

DUBINS, L. E. and TEICHER, H. (1967). Optimal stopping when the future is discounted. *Ann. Math. Statist.* **38** 601–605.

DVORETSKY, A., KIEFER, J. and WOLFOWITZ, J. (1953). Sequential decision problems for processes with continuous time parameter. Testing hypotheses. *Ann. Math. Statist.* **24** 254–264.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1988). Conditional boundary crossing probabilities with applications to change-point problems. *Ann. Probab.* **16** 825–839.

JANSON, S. (1986). Moments for first-passage and last-exit times, the minimum, and related quantities for random walks with positive drift. *Adv. in Appl. Probab.* **18** 865–879.
KHAN, R. A. (1978). Wald's approximations to the average run length in cusum procedures. *J. Statist. Plann. Inference* **2** 63–77.
KENNEDY, D. P. (1976). Martingales related to cumulative sum tests and single-server queues. *Stochastic Process. Appl.* **4** 261–269.
KERESTECIOǦLU, F. (1993). *Change Detection and Input Design in Dynamical Systems*. Research Studies Press, Taunton, UK.
LEHOCZKY, J. P. (1977). Formulas for stopped diffusion processes with stopping times based on the maximum. *Ann. Probab.* **5** 601–607.
LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.
MOUSTAKIDES, G. B. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.
NEVEU, J. (1975). *Discrete Parameter Martingales*. North-Holland, Amsterdam.
PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.
PELKOWITZ, L. (1987). The general discrete time disorder problem. *Stochastics* **20** 89–110.
REYNOLDS, M. R., JR. (1975). Approximations to the average run length in cumulative sum control charts. *Technometrics* **17** 65–71.
RITOV, Y. (1990). Decision theoretic optimality of the cusum procedure. *Ann. Statist.* **18** 1464–1469.
ROBBINS, H. E. and SAMUEL, E. (1966). An extension of Wald's Lemma. *J. Appl. Probab.* **3** 272–273.
SHIRYAYEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8** 22–46.
SHIRYAYEV, A. N. (1973). *Statistical Sequential Analysis*. Amer. Math. Soc., Providence, RI.
SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York.
SIEGMUND, D. (1985). *Sequential Analysis*. Springer, New York.
TAYLOR, H. M. (1975). A stopped Brownian motion formula. *Ann. Probab.* **3** 234–246.
WALD, A. (1947). *Sequential Analysis*. Wiley, New York.
YAKIR, B. (1996). Dynamic sampling policy for detecting a change in distribution, with a probability bound on false alarm. *Ann. Statist.* **24** 2199–2214.

DEPARTMENT OF ELECTRICAL ENGINEERING
PRINCETON UNIVERSITY
B-210 ENGINEERING QUADRANGLE OLDEN STREET
PRINCETON, NEW JERSEY 08544
E-MAIL: poor@princeton.edu