The Annals of Statistics
1997, Vol. 25, No. 5, 2117–2126

A NOTE ON OPTIMAL DETECTION OF A CHANGE IN DISTRIBUTION

BY BENJAMIN YAKIR

Hebrew University

Suppose $X_1, X_2, \ldots, X_{\nu-1}$ are iid random variables with distribution F_0, and $X_\nu, X_{\nu+1}, \ldots$ are are iid with distributed F_1. The change point ν is unknown. The problem is to raise an alarm as soon as possible after the distribution changes from F_0 to F_1 (detect the change), but to avoid false alarms.

Pollak found a version of the Shiryayev–Roberts procedure to be asymptotically optimal for the problem of minimizing the average run length to detection over all stopping times which satisfy a given constraint on the rate of false alarms. Here we find that this procedure is strictly optimal for a slight reformulation of the problem he considered.

Explicit formulas are developed for the calculation of the average run length (both before and after the change) for the optimal stopping time.

1. Introduction. The traditional formulation of sequential change point detection involves a series of independent observations $X_1, X_2, \ldots$. The distribution of the observation may change at some point from the initial distribution F_0 to a different distribution F_1. Let ν be the change point. Hence, when $\nu = k$, $X_1, X_2, \ldots, X_{k-1}$ are each distributed according to F_0, and $X_k, X_{k+1}, \ldots$ are each distributed according to F_1. The change point $\nu = \infty$ corresponds to the case of "no change"—all observations are distributed according to F_0. The distribution of the sequence, given that $\nu = k$, is denoted by $P_k(\cdot)$. Denote the associated expectation by $E_k(\cdot)$.

The formulation of sequential change point detection originated in problems of quality control. When the manufacturing process is "in control" then the products are distributed according to a target distribution F_0. At an unknown point in time the process may go "out of control" and yield products that are distributed according to F_1.

The objectives are to raise an alarm as soon as possible after the change and to avoid false alarms. A detection policy, therefore, is a stopping time on the sequence of observations $X_1, X_2, \ldots$. The goal is to sample a maximal number of prechange observations and to minimize the number of postchange observations. Hence, the stopping time N should satisfy $\{N \geq \nu - 1\}$ but, at the same time, keep $N - \nu + 1$ small.

A Bayesian approach may involve some prior distribution on ν, and a loss function of the form

$$\mathbb{1}(N < \nu - 1) + c(N - \nu + 1)\mathbb{1}(N \geq \nu - 1),$$

Received August 1991; revised February 1997.
AMS 1991 *subject classifications*. Primary 62L10; secondary 62N10.
Key words and phrases. Quality control, control charts, minimax rule, Bayes rule.

for some constant c. A non-Bayesian formulation may put a constraint on the rate of false alarms and minimize a functional of the number of post-change observations. The constraint on the rate of false alarms is customarily taken to be an expectation restriction. In other words, the stopping time N is required to satisfy $E_\infty N \geq B$, for some prespecified constant B. Here we propose to use the functional

$$\sup_{1\leq k<\infty} E_k(N-k+1 \mid N \geq k-1) \tag{1}$$

as a measure of the speed of detection. Hence, a detection policy is optimal, in a non-Bayesian sense, if it minimizes (1) among all stopping rules that satisfy the given expectation constraint on the rate of false alarms.

In this paper we show that the detection policy proposed by Pollak (1985) is optimal under the non-Bayesian formulation. Let $Z_i = f_1(X_i)/f_0(X_i)$ be the likelihood ratio of the ith observation, $i = 1, 2, \ldots$, and consider the sequence of statistics R_n^*, satisfying the recursion

$$R_n^* = (R_{n-1}^* + 1)Z_n,$$

for $n = 1, 2, \ldots$, and for some initial random variable R_0^*. Given an appropriate threshold A, define the stopping rule

$$N_A^* = \inf\{n \geq 0 \mid R_n^* \geq A\}.$$

This stopping rule is a modification of the original Shiryayev–Roberts stopping rule. In the original form the recursion is initiated with $R_0^* \equiv 0$, where, as in the optimal, rule R_0^* has a nondegenerate distribution.

In Theorem 2 we show that when the distribution of R_0^* is an appropriate quasistationary distribution then N_A^* is optimal. This follows from the fact that an equalizer rule, which is a limit of Bayes rules, is minimax. The stopping rule N_A^* is an equalizer rule due to the quasistationarity. The quasistationary distribution is useful not only for the construction of the optimal rule but also for the computation of its working characteristics. In Theorem 3 formulas are derived for the average run length to false alarm and for the average run length to detection of the optimal rule. These averages are represented as functionals of the quasistationary distribution.

The first who solved a Bayesian version of the problem was Shiryayev [see Shiryayev (1978) and Pollak (1985)]. He considered the loss function

$$\mathbb{1}(N < \nu) + c(N-\nu)\mathbb{1}(N \geq \nu)$$

and the prior distribution

$$P(\nu = 0) = \pi_0, \qquad P(\nu = n) = (1-\pi_0)p(1-p)^{n-1}, \qquad n \geq 1,$$

where p and π_0 are known constants, $0 \leq p \leq 1$, $0 \leq \pi_0 \leq 1$. When $\nu = 0$ (and when $\nu = 1$) the observations are iid with distribution F_1. Note that by Shiryayev's formulation one loses for stopping after sampling $\nu-1$ observations ($N = \nu - 1$), even though in this case all the observations sampled are from the prechange distribution!

Pollak (1985) and Pollak and Siegmund (1975) extended Shiryayev's work in a non-Bayesian setting. They proposed to measure the speed of detection with the functional

$$\sup_{1\le k<\infty} E_k(N-k \mid N\ge k).$$

Observe that while this functional is very similar to (1) they are not identical, mainly since conditioning is with respect to different events.

For the Pollak–Siegmund functional, it was shown by Pollak (1985) that the (modified) Shiryayev–Roberts rule N_A^* is an asymptotic minimax rule [up to an $o(1)$ term, as $B\to\infty$]. He was able to prove only an asymptotic version of the minimax property since he used Shiryayev's formulation, which involves the artificial state $\nu=0$. Nevertheless, in our proof we rely heavily on the techniques developed in Pollak's important paper.

A different functional for measuring the speed of detection was proposed by Lorden (1971). His proposal involves conditioning on the least favorable event preceding the change:

$$\sup_{1\le k<\infty} \operatorname{ess\,sup} E_k(N-k+1 \mid X_1,\dots,X_{k-1}).$$

He also showed that Page's cusum procedure [Page (1954)] is asymptotically optimal. His result was improved by Moustakides (1986), who showed that the cusum procedure is strictly optimal. A different optimal property of the cusum procedure was proved in Ritov (1990). It was shown there that the cusum stopping rule is the statistician's minimax rule in a game against nature. Nature, in this game, can choose the time of change based on previous observations so as to make it hardest for the statistician to detect the change.

In the next section some auxiliary Bayes problems are presented. The Bayes rules for these problems are also equalizer rules. In Section 3 the main result—the minimax property of N_A^*—is proved. In Section 4 we develop exact formulas for calculating the average run length to detection and the speed of detection of the minimax rule. The numerical details are worked out for a specific example in Section 5. Some final remarks are put in Section 6.

2. A converging sequence of equalizer Bayes rules. This section is devoted to the construction of auxiliary Bayes rules. These rules are crucial for proving the minimax property of the modified Shiryayev–Roberts procedure.

Consider the Bayesian problem $B(\pi_0, p, c)$. Suppose the prior distribution on ν is given by $P(\nu=1)=\pi_0$, and $P(\nu=n)=(1-\pi_0)p(1-p)^{n-2}$, for all $n\ge 2$ and for some $0\le p\le 1$ and $0\le \pi_0\le 1$. Let $X_1, X_2, \dots$ be a sequence of random variables. Conditional on the event $\{\nu=k\}$ all random variables in the sequence are independent. The first $k-1$ variables $X_1, X_2,\dots,X_{k-1}$ have density f_0, with respect to some σ-finite measure μ, whereas the density of the following variables $X_k, X_{k+1},\dots$ is f_1 (with respect to the same μ). If $\nu=1$, then all variables in the sequence are iid with density f_1, and if $\nu=\infty$, then they are iid with density f_0. Let $P_k(\cdot)=P(\cdot \mid \nu=k)$, $1\le k\le\infty$, be the conditional probability measures on the sequence of observations.

A rule in this problem is a stopping time on the sequence of observations. A Bayes rule, in particular, is the stopping time which minimizes the risk function

$$\rho(N, (\pi_0, p)) = P(N < \nu - 1) + cE(N - \nu + 1)^+,$$

where $c > 0$ is the relative cost of sampling a postchange observation. (Equivalently, a Bayes rule maximizes the gain

$$1 - \rho(N, (\pi_0, p)) = P(N \geq \nu - 1) - cE(N - \nu + 1)^+.)$$

Let $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and denote by $\mathscr{F}_n$ the σ-algebra generated by the first n observations, $n \geq 1$. Note that $\pi_0 = P(\nu - 1 \leq 0 \,|\, \mathscr{F}_0)$. Define, for $n \geq 1$, $\pi_n = P(\nu - 1 \leq n \,|\, \mathscr{F}_n)$ to be the posterior probability of needing to set an alarm, given the first n observations.

It can be shown that the posterior probabilities satisfy the recursion

$$\pi_n = \frac{\pi_{n-1} f_1(X_n) + (1 - \pi_{n-1}) p f_0(X_n)}{\pi_{n-1} f_1(X_n) + (1 - \pi_{n-1}) f_0(X_n)} = p \frac{(\pi_{n-1}/p) Z_n + 1 - \pi_{n-1}}{\pi_{n-1} Z_n + 1 - \pi_{n-1}},$$

with $Z_n = f_1(X_n)/f_0(X_n)$. Moreover, if one uses the same type of calculations as in Shiryayev [(1978), pages 195–196], one gets that, for any stopping time N,

$$\rho(N, (\pi_0, p)) = E\left(1 - \pi_N + c \sum_{k=0}^{N-1} \pi_k\right).$$

It follows from standard optimal stopping theory that the Bayes rule is of the form $M_{\pi_0, p, c} = \inf\{n \,|\, \pi_n \geq \delta_{p, c}\}$ for some threshold $\delta_{p, c}$. This threshold depends on p and c, but does not depend on the value of π_0.

Define $\Delta = \inf\{x \,|\, P_\infty(Z_1 \leq x) > 0\} < 1$ and fix A, $\Delta < A < \infty$. In the following theorem a sequence of Bayes problems is considered. The Bayes solution for each one of these problems has $p(A+1)/(pA+1)$ as its threshold.

THEOREM 1. *Suppose that the P_∞-distribution of Z_1 has no atoms and set $\Delta < A < \infty$. Then there exists a constant $0 < c^* < \infty$ and a sequence $\{(p_i, c(p_i)) \,|\, 1 \leq i < \infty\}$, with $p_i \to 0$ and $c(p_i) \to c^*$, such that $\delta_{p,c} = p(A+1)/(pA+1)$ is the threshold in $B(\pi_0,\ p = p_i,\ c = c(p_i))$.*

PROOF. Choose $\pi_0 = p$ and let $q = 1 - p$. Define $R^*_{q,n}$ by the recursion $R^*_{q,n} = (R^*_{q,n-1} + 1) Z_n / q$ (with $R^*_{q,0} = 0$). The proof of the theorem, given these definitions, is very similar to the relevant part of the proof of Theorem 1 in Pollak (1985) and is thus omitted. □

A detection procedure N is called *an equalizer rule* if, for all $k \geq 1$,

$$E_k(N - k + 1 \,|\, N \geq k - 1) = E_1 N.$$

Given a Bayes problem with $p = p_i$, $c = c(p_i)$ and threshold $\delta_{p,c} = p(A+1)/(pA+1)$, we are looking for a rule which is both equalizer and Bayes. In order to achieve that goal the definition of a Bayes problem needs to be extended. Let G be a distribution over the interval $[0, 1]$, and assume that initially, prior to taking any observation, π_0 is sampled from the distribution G. Given the observed value of π_0, the structure of the Bayes problem is that of $B(\pi_0, p, c)$. Denote this extended problem by $B(G, p, c)$, and note that the threshold is still intact, and so is the definition of π_n. The Bayes rule $M_{G,p,c}$, as before, is the first n for which π_n exceeds the threshold $\delta_{p,c} = p(A+1)/(pA+1)$. Equivalently, this rule can be represented in terms of the statistics $R^*_{q,n}$ as $M_{G,p,c} = \inf\{n \ge 0 \mid R^*_{q,n} \ge A\}$, where

$$R^*_{q,0} = \frac{\pi_0 q}{(1-\pi_0)p} - 1.$$

Define

$$\begin{aligned} F_n(x) &= P_\infty\Big(R^*_{q,n} \le x \;\Big|\; \max_{m\le n} R^*_{q,m} < A\Big),\\ \varphi(t,x) &= P_\infty\Big(R^*_{q,n+1} \le x \;\Big|\; R^*_{q,n} = t, \max_{m\le n+1} R^*_{q,m} < A\Big),\\ \psi(t) &= P_\infty\Big(\max_{m\le n+1} R^*_{q,m} < A \;\Big|\; R^*_{q,n} = t, \max_{m\le n} R^*_{q,m} < A\Big). \end{aligned}$$

Both $\varphi(t,x)$ and $\psi(t)$ are continuous in t and x for $0 \le t, x < A$, since the P_∞-distribution of Z_1 has no atoms. Note that both φ and ψ do not depend on G nor on n.

Consider the transformation T, defined by

$$(T \circ F)(x) = \frac{\int_0^A \varphi(t,x)\psi(t)\,dF(t)}{\int_0^A \psi(t)\,dF(t)}. \tag{2}$$

This transformation maps the set of probability measures, supported by the interval $[0, A]$, into itself. Considerations similar to those given in Pollak [(1985), Lemma 11] lead to the existence of a distribution ϕ $(= \phi_p)$ which solves the equation $T \circ \phi = \phi$. Let R^* and Z be random variables. The distribution of R^* is ϕ, and the distribution of Z is the P_∞-distribution of Z_1. Let G be the distribution of

$$\frac{(R^*+1)Z/q + 1}{(R^*+1)Z/q + 1 + p/q}.$$

It follows that for all k, $k \ge 1$, the conditional P_k-distribution of π_{k-1}, given the event $\{M_{G,p,c} \ge k-1\}$, is again G. It can be concluded that the Bayes rule $M_{G,p,c}$ is also an equalizer rule. Note that $P(\nu = 1) = E\pi_0$. It follows that the gain from this rule is

$$
\begin{aligned}
&P(M_{G,\,p,\,c} \geq \nu - 1) - cE(M_{G,\,p,\,c} - \nu + 1)^+ \\
&\quad = \sum_{k=1}^{\infty} (1 - cE_1 M_{G,\,p,\,c}) P_k(M_{G,\,p,\,c} \geq k - 1) P(\nu = k) \\
&\quad = (1 - cE_1 M_{G,\,p,\,c}) \\
&\qquad \times \left[E\pi_0 + (1 - E\pi_0) p \sum_{m=1}^{\infty} P_\infty(M_{G,\,p,\,c} \geq m)(1-p)^{m-1} \right], \qquad (3) \\
&\quad = p(1 - cE_1 M_{G,\,p,\,c}) \left[E\left(\frac{\pi_0}{p}\right) + \frac{(1 - E\pi_0) P(R^*_{q,0} < A)}{p + (1-p) P(R^*_{q,0} \geq A)} \right],
\end{aligned}
$$

since the P_∞-distribution of $M_{G,\,p,\,c} + 1$ is geometric.

3. A minimax rule. In this section we show that the (modified) Shiryayev–Roberts procedure N_A^*, proposed by Pollak (1985), is optimal. Define the (modified) Shiryayev–Roberts statistics by the recursion $R_n^* = (R_{n-1}^* + 1) Z_n$, where R_0^* is random. Let

$$
\begin{aligned}
F_n(x) &= P_\infty\Big(R_n^* \leq x \mid \max_{m \leq n} R_m^* < A \Big), \\
\varphi(t, x) &= P_\infty\Big(R_{n+1}^* \leq x \mid R_n^* = t,\ \max_{m \leq n+1} R_m^* < A \Big), \\
\psi(t) &= P_\infty\Big(\max_{m \leq n+1} R_m^* < A \mid R_n^* = t,\ \max_{m \leq n} R_m^* < A \Big).
\end{aligned}
$$

The transformation T is defined in (2), with the current φ and ψ. Let ϕ_0 be the probability measure on the interval $[0, A]$ which is invariant under the transformation T. Let the distribution of R^* be ϕ_0, and let the distribution of Z be the P_∞-distribution of Z_1. Both R^* and Z are independent of the sequence of observations $X_1, X_2, \dots$. Then $R_0^* = (R^* + 1)Z$. The (modified) Shiryayev–Roberts procedure is $N_A^* = \inf\{n \geq 0 \mid R_n^* \geq A\}$.

For the given B, one can find A for which $E_\infty N_A^* = B$. The stopping rule N_A^* is also an equalizer rule. Let N be any stopping time. In the proof it is shown that if $E_\infty N \geq B$ but $E_k(N - k + 1 \mid N \geq k - 1) < E_1 N_A^*$, for all $1 \leq k < \infty$, then the risk of N is smaller than the Bayes risk in the problem $B(G, p, c)$, for some $0 < p < 1$. This contradiction establishes the minimax property of N_A^*.

THEOREM 2. *Suppose that the P_∞-distribution of Z_1 has no atoms. For all $\Delta < A < \infty$, let N_A^* be the stopping time defined above.*

(i) *For every $1 < B < \infty$ there exists $\Delta < A < \infty$ such that $E_\infty N_A^* = B$.*
(ii) *This N_A^* minimizes*

$$
\sup_{1 \leq k < \infty} E_\nu(N - k + 1 \mid N \geq k - 1)
$$

among all stopping times N that satisfy the constraint $E_\infty N \geq B$.

PROOF. The proof of (i) is similar to the proofs of Lemmas 15, 16 and 17 in Pollak (1985) and is thus omitted.

It can be shown that $\phi_p \to \phi_0$ in distribution (for some sequence $p \to 0$). It follows that $\pi_0/p \to R_0^* + 1$ in distribution and that $M_{G,\,p,\,c} \to N_A^*$ in P_1-distribution. Moreover, by Lebesgue's dominated convergence theorem, $E\pi_0/p \to ER_0^* + 1$ and $E_1 M_{G,\,p,\,c} \to E_1 N_A^*$. (Note that all relevant stopping times are dominated by the Shiryayev–Roberts stopping time with $R_0^* = 0$.) The P_∞-distribution of $N_A^* + 1$ is geometric. Therefore,

$$B = E_\infty N_A^* = \frac{P(R_0^* < A)}{P(R_0^* \geq A)}.$$

It follows from (3) that

$$\lim_{p \to 0} \frac{P(M_{G,\,p,\,c} \geq \nu - 1) - cE(M_{G,\,p,\,c} - \nu + 1)^+}{p} = (1 - c^* E_1 N_A^*)[ER_0^* + 1 + B].$$

In order to prove (2.2) assume the contrary holds. There exist, thus, a stopping time N and some $\varepsilon > 0$ such that $E_\infty N \geq B$ but

$$\sup_{1 \leq k < \infty} E_k(N - k + 1 \mid N \geq k - 1) \leq E_1 N_A^* - \varepsilon.$$

Reasoning similar to (3) leads to the conclusion that

$$\frac{P(N \geq \nu - 1) - cE(N - \nu + 1)^+}{p} \geq (1 - cE_1 N_A^* + c\varepsilon)\left[E\frac{\pi_0}{p} + (1 - E\pi_0)\sum_{m=1}^{\infty} P_\infty(N \geq m)(1-p)^{m-1}\right].$$

The right-hand side of the above inequality converges, as $p \to 0$, to

$$(1 - c^* E_1 N_A^* + c^*\varepsilon)[ER_0^* + 1 + E_\infty N] \geq (1 - c^* E_1 N_A^* + c^*\varepsilon)[ER_0^* + 1 + B].$$

It follows that, for some small (but positive) p, the stopping rule N is better than the Bayes rule of the problem $B(G, p, c)$. This is a contradiction, which establishes the proof of (2.2). □

4. Average run length. When applying detection schemes in real problems, it is important to be able to compute their performance characteristics. This is, in particular, true for the average run length, both before and after the change, of the optimal policy described in Theorem 2. Here we present formulas for these average run lengths when the distribution ϕ_0 can be attained. Approximate expressions (asymptotic, as $B \to \infty$) for the average run lengths are derived in Pollak (1987). Our results are exact.

THEOREM 3. *Define*

$$p_0 = P(R_0^* \geq A) = 1 - E_\infty \phi_0(A/Z_1 - 1),$$
$$\mu_0 = E(R_0^* \mid R_0^* < A) = \int_0^A x\, d\phi_0(x),$$

and let N_A^* be the minimax detection scheme associated with A. Then the following hold:

(i) $E_\infty N_A^* = (1 - p_0)/p_0$;
(ii) $E_1 N_A^* = (\mu_0 + 1)(1 - p_0)/(p_0(\mu_0 + 1) + 1)$.

PROOF. The P_∞-distribution of $N_A^* + 1$ is geometric. Relation (i) thus follows.

Given $p > 0$, consider the equalizer Bayes rule $M_{G,\,p,\,c}$. From the proof of Theorem 2 it follows that

$$E(M_{G,\,p,\,c} - \nu + 1)^+/p \to E_1 N_A^*(ER_0^* + 1 + E_\infty N_A^*). \tag{4}$$

On the other hand, if one uses the same calculations as in Shiryayev [(1978), pages 195–196], one gets

$$E(M_{G,\,p,\,c} - \nu + 1)^+ = E\Bigg(\sum_{k=0}^{M_{G,\,p,\,c}-1} \pi_k\Bigg) = E\Bigg(\sum_{k=0}^{\infty} \pi_k \mathbb{1}(M_{G,\,p,\,c} \geq k+1)\Bigg),$$

hence

$$E(M_{G,\,p,\,c} - \nu + 1)^+/p = E\Bigg(\sum_{k=0}^{\infty} (\pi_k/p) \mathbb{1}(M_{G,\,p,\,c} \geq k+1)\Bigg). \tag{5}$$

Consider, for any fixed k, the limit of $E(\pi_k/p)\mathbb{1}(M_{G,\,p,\,c} \geq k+1)$ as $p \to 0$. The joint density of $X_1, \ldots, X_k$, given π_0, is

$$\pi_0 \prod_{i=1}^k f_1(x_i) + (1-\pi_0) \sum_{m=2}^k pq^{m-2} \prod_{i=1}^{m-1} f_0(x_i) \prod_{i=m}^k f_1(x_i) + (1-\pi_0) q^{k-1} \prod_{i=1}^k f_0(x_i).$$

This density converges to $\prod_{i=1}^k f_0(x_i)$, since $\pi_0 \to 0$ as $p \to 0$. Moreover, $\pi_k/p \to R_k^* + 1$ and $\mathbb{1}(M_{G,\,p,\,c} \geq k+1) \to \mathbb{1}(N_A^* \geq k+1)$ (in distribution), as $p \to 0$. By definition $\pi_k/p < (A+1)/(pA+1)$ on the event $\{N_p \geq k+1\}$, and $R_k^* < A$ on the event $\{N_A^* \geq k+1\}$. Hence, by Lebesgue's dominated convergence theorem,

$$E(\pi_k/p)\mathbb{1}(M_{G,\,p,\,c} \geq k+1) \to E_\infty(R_k^* + 1)\mathbb{1}(N_A^* \geq k+1). \tag{6}$$

The conditional P_∞-distribution of R_k^*, given the event $\{N_A^* \geq k+1\}$, is ϕ_0. Therefore,

$$E_\infty R_k^* \mathbb{1}(N_A^* \geq k+1) = \mu_0 P_\infty(N_A^* \geq k+1). \tag{7}$$

The stopping times $N_{G,p,c}$ are uniformally integrable. It follows, thereby, from (5), (6) and (7) that

$$\lim_{p\to 0} E(N_p - \nu + 1)^+/p = (\mu_0 + 1)E_\infty N_A^*. \tag{8}$$

Combine (i), (4) and (8) in order to conclude the proof of (ii). $\square$

5. An example. Let $f_0(x) = \exp\{-x\}\mathbb{1}(x > 0)$ and $f_1(x) = 2\exp\{-2x\} \times \mathbb{1}(x > 0)$, and consider A, $0 < A < 2$. This example is considered in Pollak (1985). It is shown there that the quasistationary distribution ϕ_0 is given by

$$\phi_0 = \begin{cases} 0, & \text{if } x \le 0, \\ x/A, & \text{if } 0 < x \le A, \\ 1, & \text{if } A < x. \end{cases}$$

Hence, $\mu_0 = A/2$. The P_∞-distribution of Z_1 is uniform on the interval $[0, 2]$ since

$$P_\infty\big(2\exp\{(1-2)X_1\} \le z\big) = P_\infty(X_1 \ge -\log(z/2)) = z/2,$$

for $0 \le z \le 2$. It follows that

$$p_0 = 1 - E_\infty \phi_0(A/Z_1 - 1) = 1 - (\log A)/2.$$

6. Concluding remarks. In the problem of change point detection, the optimal policy is to stop when R_n^* is large. The statistic R_n^* can be thought of as a measure of the information on the likelihood of the process being "out of control." The statistic R_0^* can be interpreted as a prior belief regarding the likelihood of a change when surveillance is initiated. The distribution of prior beliefs, for the optimal policy, is the quasistationary distribution when the process is "in control."

In practice it seems unlikely that one would apply a procedure which initiates with a random R_0^*. We believe, however, that the results presented here have theoretical significance: they fill a gap in the present optimal change point detection theory. It is known from Pollak's work that the Shiryayev–Roberts procedure is optimal when the average run length to false alarm is large. Here we showed that it is actually optimal for any such average run length.

Our results suggest a method for the calculation of the working characteristics of the optimal policy: One starts by solving the operator equation $T \circ \phi = \phi$ and then applying Theorem 3. Unfortunately, in most examples, solving the equation is an analytically intractable task. Evaluation of ϕ_0 may require numerical methods.

Acknowledgment. The author would like to thank Moshe Pollak for his suggestions and support.

REFERENCES

LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

MOUSTAKIDES, G. V. (1986). Optimal stopping time for detecting a change in distribution. *Ann. Statist.* **14** 1379–1387.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrics* **41** 100–115.

POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227.

POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779.

POLLAK, M. and SIEGMUND, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3** 1267–1282.

RITOV, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18** 1464–1469.

ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430.

SHEWHART, W. A. (1931). *The Economic Control of the Quality of Manufactured Product.* Macmillan, New York.

SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules.* Springer, New York.

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM 91905
ISRAEL
E-MAIL: msby@olive.mscc.huji.ac.il