The Annals of Statistics
1991, Vol. 19, No. 1, 394–416

SEQUENTIAL DETECTION OF A CHANGE IN A NORMAL MEAN WHEN THE INITIAL VALUE IS UNKNOWN

BY MOSHE POLLAK[1] AND D. SIEGMUND[2]

Hebrew University and Stanford University

Three stopping rules are proposed to detect a change in a normal mean, when the initial value of the mean is unknown but an estimate obtained from a training sample is available. Asymptotic approximations are given for the average run length when there is no change. Under certain hypotheses about the length of time before the change occurs and the magnitude of the change, we obtain asymptotic approximations for the expected delay in detection in terms of the corresponding expected delay in the much simpler case of a known initial value. The results of a Monte Carlo experiment supplement our asymptotic theory to yield some general conclusions about the relative merits of the three stopping rules and guidelines for choosing the size of the training sample.

1. Introduction. Assume $x_1, x_2, \ldots$ are independent random variables. The observations $x_1, \ldots, x_\nu$ have probability density function f_0, and $x_{\nu+1}, \ldots$ have probability density function f_1. We write P_ν and E_ν to denote probability and expectation when the change-point is ν, $\nu = 0, 1, \ldots, +\infty$. The x's are observed sequentially with the goal of detecting the change-point ν by a stopping rule T having the properties that $E_\infty(T)$, the average run length when there is no change, exceeds some large preassigned constant and the expected delay, $E_\nu(T - \nu | T > \nu)$, is small in some suitably defined sense.

The extensive literature on this problem is primarily concerned with the case that f_0 and f_1 come from a common parametric family of distributions and that f_0 is completely specified. See, for example, Page (1954), Shiryayev (1963), van Dobben de Bruyn (1968), Lorden (1971), Lucas and Crosier (1982) and Pollak (1985). The most frequently described application is to process control. As long as the observations, representing measurements on the output of a production process, come from the target distribution f_0, the process is in control and no action is required. However, after the change-point ν the process is out of control and corrective action must be taken as soon as possible.

We shall assume that f_0 is unknown but can be estimated from a training sample $x_1, \ldots, x_{\nu_0}$, where $0 \le \nu_0 \le \nu$. One possible application is to detect a change in the frequency of congenital malformations [Weatherall and Haskey (1976) and Levin and Kline (1985)]. In this case f_0 represents the unknown

Received August 1989; revised February 1990.

[1]Research supported by the U.S.–Israel Binational Science Foundation.

[2]Research supported by AFOSR Contract 89-0355, NSA Grant MDA 904-89-H-2040, NSF Grant DMS-89-05874 and the U.S.–Israel Binational Science Foundation.

AMS 1980 *subject classifications*. Primary 62L10; secondary 62N10.

Key words and phrases. Change-point problem, sequential detection.

baseline level of malformations. An application in process control is detection of a change in variability. In contrast to problems of detection of a change in the mean, there typically is no target value for the variance, which should be as small as the process permits; but the attainable value must usually be determined empirically [e.g., Wilson, Griffiths, Kemp, Nix and Rowlands (1979)].

In order to focus on the essential features of this problem, we consider only the simple case that f_0 and f_1 are normal density functions with unit variance, differing only in their means. Specifically, we assume that $x_1, \dots, x_\nu$ are independent $N(\mu_0, 1)$ and $x_{\nu+1}, \dots$ are independent $N(\mu_0 + \mu, 1)$. The parameters ν, μ_0 and μ are all unknown, but μ is assumed to be positive. We also assume that for some known value ν_0, $0 \le \nu_0 \le \nu$, we have available a training sample $x_1, \dots, x_{\nu_0}$, which is the basis for our initial knowledge about μ_0.

In Section 2 we introduce three different detection schemes. For the first we use an invariance reduction and apply the Shiryayev–Roberts procedure along the lines suggested by Pollak and Siegmund (1985) to the sequence of maximal invariant statistics. The second is a minor modification involving a mixture of likelihood ratios of maximal invariants. For the third procedure we apply a CUSUM test to the process of recursive residuals [Brown, Durbin and Evans (1975)]. For each of these procedures we show that the P_∞-expected run length is approximately the same as for an analogous procedure in the case of known μ_0. See (6), (9) and (12).

In a recent unpublished manuscript, McDonald (1989) describes an interesting nonparametric method for detecting a stochastic increase in distribution, when the original distribution is completely unknown. We make a few remarks about his method at the end of this paper.

Section 3 gives some asymptotic theory for the expected delay, $E_\nu(T - \nu | T > \nu)$. Its general flavor is that if both ν and $E_\infty(T - \nu_0)$ are large, then $E_\nu(T - \nu | T > \nu)$ is approximately the same as for an analogous procedure in the case of known μ_0, provided μ is not too small. We also find an expression for the difference between the asymptotic expected delays in the cases of known and unknown μ_0, which can be used to select an adequately large training sample. The asymptotic theory which we have developed does not suggest a strong preference for any of the three procedures.

Section 4 compares the three procedures by means of a Monte Carlo experiment. The results are consistent with the asymptotic theory for large ν_0 and suggest a definite preference for the Shiryayev–Roberts procedures when ν_0 is small or early detection of small values of μ is important.

Technical aspects of the proofs of our principal theoretical results are sketched in two appendices.

2. Three procedures. Let $x_1, \dots, x_\nu$ be independent $N(\mu_0, 1)$ and $x_{\nu+1}, \dots$ independent $N(\mu_0 + \mu, 1)$, with $\mu > 0$. Let $S_n = x_1 + \cdots + x_n$ and $S_n^* = S_n - n\mu_0$. Suppose for the moment that μ_0 is known. For any $\nu_0 \le i < n$ the likelihood ratio statistic based on $x_1, \dots, x_n$ for testing H_0: $\nu = \infty$ against

H_1: $\nu = i$, $\mu = \delta$ is

$$\exp\left[\delta\left(S_n^* - S_i^*\right) - \delta^2(n-i)/2\right].$$

Let

$$R_n^*(\delta) = \sum_{i=\nu_0}^{n-1} \exp\left[\delta\left(S_n^* - S_i^*\right) - \delta^2(n-i)/2\right].$$

The Shiryayev–Roberts procedure [Shiryayev (1963) and Roberts (1966)] is defined by the stopping rule

$$N^* = \inf\{n\colon n > \nu_0,\ R_n^*(\delta) \geq B\}. \tag{1}$$

The positive constant δ is some value of the change μ which is deemed important to detect rapidly. From the likelihood ratio structure it follows that

$$R_n^*(\delta) - (n - \nu_0)$$

is a P_∞-martingale and hence that

$$E_\infty(N^* - \nu_0) = E_\infty\left[R_{N^*}^*(\delta)\right] \geq B.$$

A more precise approximation [Pollak (1987)] is that as $B \to \infty$,

$$E_\infty(N^* - \nu_0) \sim B/h(\delta). \tag{2}$$

Here h is a special function defined by

$$h(x) = 2x^{-2}\exp\left[-2\sum_1^\infty n^{-1}\Phi(-xn^{1/2}/2)\right], \qquad x > 0,$$

and given to a good approximation by the local expansion

$$h(x) = \exp(-\rho x) + o(x^2), \qquad x \to 0, \tag{3}$$

where $\rho = 0.583\dots$ [cf. Siegmund (1985), Chapter 10].

Suppose now that μ_0 is unknown. The problem of detecting a change from μ_0 to $\mu_0 + \mu > \mu_0$ is invariant under addition of a constant c to each observation $x_1, x_2, \dots$. An invariant function of the x's is $y_1 = 0$, $y_2 = x_2 - x_1$, $y_3 = x_3 - x_1, \dots,$ whose joint distribution depends on ν and μ but not on μ_0. For any $\nu_0 \leq i < n$, the likelihood ratio of $y_2, \dots, y_n$ for testing H_0: $\nu = \infty$ against H_1: $\nu = i$, $\mu = \delta$ is easily calculated to be

$$\exp\left[\delta(iS_n/n - S_i) - \delta^2 i(1 - i/n)/2\right].$$

In analogy with (1) we define

$$R_n(\delta) = \sum_{i=\nu_0}^{n-1} \exp\left[\delta(iS_n/n - S_i) - \delta^2 i(1 - i/n)/2\right] \tag{4}$$

and

$$N = \inf\{n\colon n > \nu_0,\ R_n(\delta) \geq B\}. \tag{5}$$

Here $R_n(\delta) - (n - \nu_0)$ is a P_∞-martingale, so

$$E_\infty(N - \nu_0) = E_\infty[R_N(\delta)] \geq B;$$

and it is shown in Appendix 1 that as $B \to \infty$ we have

$$E_\infty(N - \nu_0) \sim B/h(\delta). \tag{6}$$

The procedure defined by (5) should perform well if μ is close to the hypothesized δ, but may be poor otherwise. To compensate for this weakness, it may be helpful to put

$$\hat{R}_n = \int_0^\infty R_n(\delta)\, dG(\delta), \qquad n > \nu_0, \tag{7}$$

for a suitable probability G on $(0, \infty)$ and let

$$\hat{T} = \inf\{n: n > \nu_0, \hat{R}_n \geq B\}. \tag{8}$$

Again $E_\infty(\hat{T} - \nu_0) \geq B$, and now as $B \to \infty$,

$$E_\infty(\hat{T} - \nu_0) \sim B \Big/ \int_0^\infty h(\delta)\, dG(\delta). \tag{9}$$

The corresponding modification of (1) in the case of known μ_0 was considered by Pollak and Siegmund (1985), whose calculations indicated that this modification does not offer substantial advantages over (1) unless B is extremely large. We shall see that (8) compares favorably with (5), especially if the true change μ is substantially smaller than the hypothesized δ.

Our third procedure is a CUSUM test defined in terms of recursive residuals [Brown, Durbin and Evans (1975)]. In the present context the recursive residuals are

$$Z_i = [i/(i+1)]^{1/2}(x_{i+1} - \bar{x}_i), \qquad i = 1, 2, \dots,$$

where $\bar{x}_i = S_i/i$. The Z_i are independent and normally distributed with unit variance. Also

$$E_\nu(Z_i) = \begin{cases} 0, & i < \nu, \\ \nu\mu/[i(i+1)]^{1/2}, & i \geq \nu. \end{cases} \tag{10}$$

Put $\tilde{S}_n = Z_1 + \cdots + Z_n$, and for fixed positive numbers δ, b let

$$\tau = \inf\Big\{n: n > \nu_0, \delta\Big[\tilde{S}_{n-1} - \delta(n-1)/2 - \min_{\nu_0 \leq k \leq n-1}\big(\tilde{S}_k - \delta k/2\big)\Big] \geq b\Big\}. \tag{11}$$

According to Siegmund (1985), Theorem 10.16, a good approximation to $E_\infty(\tau - \nu_0)$ is provided by

$$E_\infty(\tau - \nu_0) = 2\delta^{-2}[\exp(b + 2\rho\delta) - (b + 2\rho\delta) - 1] + o(\delta^{-1}), \qquad \delta \to 0, \tag{12}$$

where ρ is the same numerical constant that appears in (3).

For large δ a slightly better but substantially more complicated approximation was obtained by Siegmund (1975).

REMARK. It is easy to think of still more procedures. For example, one could define a Shiryayev–Roberts procedure in terms of recursive residuals or a CUSUM test in terms of the invariant process $iS_n/n - S_i$, $n = 1, 2, \dots, i = 1, \dots, n-1$. We have in fact performed some simulations for this latter process and have found that it behaves similarly to those defined above. However, the random walk theory which allows one to obtain an approximation to the P_∞-average run length in the case of known μ_0 does not apply to the invariant process.

3. Asymptotic theory for $E_\nu(T - \nu | T > \nu)$. In this section we examine the asymptotic behavior of $E_\nu(T - \nu | T > \nu)$ for large B and ν for the stopping rules (5), (8) and (11) of Section 2. Our results take the form of a comparison of the expected delay when μ_0 is unknown to the expected delay of the analogous stopping rule for the case of known μ_0. Analytic approximations for the expected delay in the simpler case of known μ_0 have been given by various authors, for example, Pollak and Siegmund (1985, 1986b), Pollak (1987) and Siegmund (1985), Chapter 10.

We begin with the almost trivial observation that as $B \to \infty$, for $\mu > \delta/2$ and $\nu_0 = \nu = 0$,

$$E_0(N^*) \sim [\delta(\mu - \delta/2)]^{-1} \log B. \tag{13}$$

In fact, by (1)

$$\log R_n^*(\delta) = \delta S_n^* - \delta^2 n/2 - \log\left[\sum_0^{n-1} \exp\{-\delta S_i^* + \delta^2 i/2\}\right].$$

Under P_0 the series converges with probability 1 and hence makes no contribution to the first-order asymptotic behavior of $\log R_n^*(\delta)$. The result (13) follows from

$$E_0\{\log R_T^*(\delta)\} \sim \log B$$

and by Wald's identity

$$E_0\left(\delta S_{N^*}^* - \delta^2 N^*/2\right) = \delta(\mu - \delta/2) E_0(N^*).$$

A slightly more complicated argument shows that (13) continues to hold if the left-hand side is replaced by $E_\nu(N^* - \nu | N^* > \nu)$.

The following theorem provides approximations for $E_\nu(N - \nu | N > \nu)$ in terms of $E_\nu(N^* - \nu | N^* > \nu)$ for $\mu > \delta/2$ and sufficiently large ν.

THEOREM 1. *Let N be defined by* (5) *and N^* by* (1). *Suppose that B and $\nu \to \infty$ in such a way that for some $1 < \eta_1, \eta_2 < \infty$,*

$$(\log B)^{\eta_1} < \nu < B\eta_2 + \nu_0. \tag{14}$$

Then for any $\mu > \delta/2$,

$$E_\nu(N-\nu|N>\nu) - E_\nu\{(N-\nu)^2/N|N>\nu\} - E_\nu(N^*-\nu|N^*>\nu) \to 0, \tag{15}$$

and hence

$$E_\nu(N-\nu|N>\nu) = E_\nu(N^*-\nu|N^*>\nu) + \nu^{-1}E_\nu^2(N^*-\nu|N^*>\nu)(1+o(1)) + o(1). \tag{16}$$

REMARKS. (i) In view of (13), we see that the left-hand inequality in condition (14) requires that the time before the change occurs should be large compared to the expected time required to detect the change after its occurrence. The right-hand inequality in (14) is purely technical. We believe it is unnecessary.

(ii) A proof of Theorem 1 is given in Appendix 2. Later in this section we sketch a proof of the related but much simpler Theorem 3.

(iii) The term $\nu^{-1}E_\nu^2(N^*-\nu|N^*>\nu)$ in (16) can be interpreted as the asymptotic cost of ignorance of μ_0 in the favorable situation that ν and μ are large enough that we do about as well as if we had known μ_0. It converges to 0 if $\nu/(\log B)^2 \to \infty$.

Theorem 2 makes a similar comparison between the stopping rule $\hat{T}$ defined in (8) and

$$T^* = \inf\left\{n: n > \nu_0, \int_0^\infty R_n^*(\delta)\, dG(\delta) \geq B\right\}. \tag{17}$$

THEOREM 2. *Suppose G has a positive, continuous density G′ such that* $\lim_{\delta\downarrow 0} G'(\delta)$ *exists and is positive. If condition* (14) *holds, then for arbitrary* $\mu > 0$, (15) *and* (16) *hold when N and* N^* *are replaced by* $\hat{T}$ *and* T^*, *respectively.*

The proof of Theorem 2 is similar to that of Theorem 1, but is technically more complicated. The details have been omitted. For asymptotic approximations to $E_0(T^*)$, see Pollak and Siegmund (1985) and Pollak (1987).

In order to make a similar comparison for $E_\nu(\tau-\nu|\tau>\nu)$, we introduce

$$\tau^* = \inf\left\{n: n>\nu_0, \delta\left[S_n^* - \delta n/2 - \min_{\nu_0 \leq i \leq n}\left(S_i^* - \delta i/2\right)\right] \geq b\right\}. \tag{18}$$

THEOREM 3. *Suppose* $b \to \infty$ *and for some* $\eta > 1$,

$$\nu \geq b^\eta. \tag{19}$$

Then for any $\mu > \delta/2$,

$$\begin{aligned} &E_\nu(\tau-\nu|\tau>\nu) \\ &\quad = E_\nu(\tau^*-\nu|\tau^*>\nu) + 1 \\ &\qquad + \mu[2\nu(\mu-\delta/2)]^{-1}E_\nu^2(\tau^*-\nu|\tau^*>\nu)(1+o(1)) + o(1). \end{aligned} \tag{20}$$

REMARK. The factor $(\mu - \delta/2)^{-1}$ appearing on the right-hand side of (20) suggests that τ may perform poorly when μ is close to $\delta/2$; and the simulations reported in the following section substantiate this expectation. It should be noted, however, that the source of this factor is the recursive residuals, not the CUSUM statistic. It would have appeared in Theorem 1 if we had used recursive residuals to define the Shiryayev–Roberts stopping rule.

We conclude this section with a heuristic proof of Theorem 3, which is substantially easier to make completely rigorous than the proof of Theorem 1 given in Appendix 2. To simplify the notation, we suppose $\nu = \nu_0$. By renewal theory applied to the excess over the boundary and Wald's identity, for τ^* defined by (18) we have

$$\begin{aligned}
&b + \text{asymptotic expected excess} + o(1)\\
&\quad = E_{\nu_0}\Big\{\delta\big[S^*_{\tau^*} - S^*_{\nu_0} - \delta(\tau^* - \nu_0)/2\big]\\
&\qquad\quad -\delta \min_{\nu_0 \le i \le \tau^*}\big[S^*_i - S^*_{\nu_0} - \delta(i - \nu_0)/2\big]\Big\}\\
&\quad = \delta(\mu - \delta/2)E_{\nu_0}(\tau^* - \nu_0) - E_0\Big\{\delta \min_{0 \le i < \infty}\big(S^*_i - \delta i/2\big)\Big\} + o(1).
\end{aligned}$$

For the recursive residual process the asymptotic expected excess will be the same. Hence by (10)

$$\begin{aligned}
&b + \text{asymptotic expected excess} + o(1)\\
&\quad = E_{\nu_0}\Big\{\delta\big[\tilde{S}_{\tau-1} - \tilde{S}_{\nu_0} - \delta(\tau - \nu_0 - 1)/2\big]\\
&\qquad\quad -\delta \min_{\nu_0 \le i \le \tau-1}\big[\tilde{S}_i - \tilde{S}_{\nu_0} - \delta(i - \nu_0)/2\big]\Big\}\\
&\quad = \delta E_{\nu_0}\left\{\mu\nu_0 \sum_{\nu_0+1}^{\tau-1}[i(i+1)]^{-1/2} - \delta(\tau - \nu_0 - 1)/2\right\}\\
&\qquad\quad - E_0\Big\{\delta \min_{0 \le i < \infty}\big(S^*_i - \delta i/2\big)\Big\} + o(1)\\
&\quad = \delta E_{\nu_0}\{\mu\nu_0 \log[1 + (\tau - \nu_0 - 1)/\nu_0] - \delta(\tau - \nu_0 - 1)/2\}\\
&\qquad\quad - E_0\big\{\delta \min\big(S^*_i - \delta i/2\big)\big\} + o(1)\\
&\quad = \delta(\mu - \delta/2)E_{\nu_0}(\tau - \nu_0 - 1) - (\delta\mu/2\nu_0)E_{\nu_0}(\tau - \nu_0)^2(1 + o(1))\\
&\qquad\quad - E_0\big\{\delta \min\big(S^*_i - \delta i/2\big)\big\} + o(1).
\end{aligned}$$

Comparing these two asymptotic equations, we see that

$$\begin{aligned}
E_{\nu_0}(\tau - \nu_0) &= E_{\nu_0}(\tau^* - \nu_0) + 1\\
&\quad + \mu[2\nu_0(\mu - \delta/2)]^{-1}E_{\nu_0}(\tau - \nu_0)^2(1 + o(1)) + o(1).
\end{aligned}$$

TABLE 1
Expected delay for $\delta = 1$

μ	ν_0 150	75	40
1.0	11.3 ± 0.2	12.3 ± 0.2	16.4 ± 0.9
	10.4 ± 0.2	11.8 ± 0.2	18.3 ± 1.0
0.5	59.4 ± 3.6	109.9 ± 7.7	202.6 ± 12.8
	75.3 ± 3.0	154.9 ± 6.6	259.7 ± 9.4
1.5	6.5 ± 0.1	6.8 ± 0.1	7.2 ± 0.1
	5.5 ± 0.1	5.7 ± 0.1	6.0 ± 0.1

Since to first order asymptotically, $E_{\nu_0}(\tau - \nu_0)^r \sim E_{\nu_0}(\tau^* - \nu_0)^r$, $r = 1, 2$, and $E_{\nu_0}(\tau^* - \nu_0)^2 \sim E_{\nu_0}^2(\tau^* - \nu_0)$, Theorem 3 follows.

4. Monte Carlo. In this section we describe the results of a Monte Carlo experiment designed to check the insights obtained from the asymptotic theory of Section 3 and to provide comparative information for small ν.

We recall that for known μ_0 Pollak and Siegmund (1985) have shown that the Shiryayev–Roberts procedure (2) and the CUSUM test (18) perform similarly. The CUSUM test is somewhat better at detecting large changes and changes which occur immediately ($\nu = 0$). The Shiryayev–Roberts procedure is better at detecting small changes and changes which occur at large values of ν. In spite of its better asymptotic performance, the mixture stopping rule (17) seems inferior to the fixed δ rules unless B is very large.

In Table 1 we compare (5) and (11) for $\delta = 1$. The values of B and b were chosen so that the P_∞-average run lengths were about 792. For example, according to (6) and (3), for the stopping rule N defined by (5) we should take $B = 442$. Similarly, by (12) $b = 4.83$. Those values were checked by simulation and found to yield the desired average run length to within the range of sampling error.

Initially we consider only the case $\nu = \nu_0$, that is, the change occurs immediately after the end of the training sample. Guided by the results for known μ_0 described above, we expect that this choice will favor the CUSUM test, which is relatively better at detecting early changes.

In each cell of Table 1 the first entry is a Monte Carlo estimate based on $n = 2500$ replications for $E_{\nu_0}(N - \nu_0)$, where N is defined by (5). The second entry concerns τ defined by (11). Each estimate is given $\pm$one estimated standard error. As in the case of known μ_0 the CUSUM test does slightly better for large μ, while the situation is reversed for small μ. As suggested in the remark following Theorem 3, the use of recursive residuals leads to comparatively poor performance for the CUSUM test when $\mu = 0.5 = \delta/2$.

It is interesting to compare the Monte Carlo estimates in Table 1 with the theoretical approximations provided by (16) and (20), but for this we must evaluate $E_{\nu_0}(N^* - \nu_0)$ and $E_{\nu_0}(\tau^* - \nu_0)$. It follows from Siegmund (1985), Theorem 10.16, and Pollak (1987) that reasonably good approximations in the case of known μ_0 can be obtained as follows. We use the appropriate expressions for a Brownian motion process [Pollak and Siegmund (1985), cf. also Appendix 3], but with the boundary level, B or b, adjusted to give the same P_∞-average run length as the discrete time process. For example, to approximate $E_{\nu_0}(N^* - \nu_0)$ for N^* defined by (2) with $\delta = 1$ and $B = 442$, so that $E_\infty(N^* - \nu_0) \cong 792$, we use Brownian motion results with $\delta = 1$ and $B = 792$. Although it is probably too optimistic to expect the resulting composite approximations suggested by (16) and (20) to be especially accurate, they should provide an indication whether our training sample is sufficiently large that we do about as well as when μ_0 is known.

The results appropriate for Table 1 of this paper can be borrowed directly from Table 1 of Pollak and Siegmund (1985) (or see Appendix 3). For example, for $\mu = 1$, $E_{\nu_0}(N^* - \nu_0) \cong 10.8$ and hence by (16) we see that $E_{\nu_0}(N - \nu_0)$ is approximately $10.8 + 0.8 = 11.6$, $10.8 + 1.6 = 12.4$ and $10.8 + 2.9 = 13.7$ for $\nu_0 = 150$, 75 and 40, respectively. For $\nu_0 = 150$ and 75 the second term is small compared to the first, so we expect that $\nu_0 = 150$ or 75 is an adequate size for the training sample. For $\nu_0 = 40$ the second term is about 25% as large as the first term, which suggests that the training sample is too small. These diagnoses are consistent with our simulations. Unfortunately a training sample adequate to detect a change of size $\mu = 1$ may not be adequate to detect a smaller change. In fact, even the sample of size $\nu_0 = 150$ is too small to detect efficiently a change of size $\mu = 0.5$, as both our simulations and diagnostics indicate.

In Table 2 we compare all three stopping rules: (5), (8) and (11). For variety and to simplify the simulations slightly, we put $\delta = 2$ and consider stopping rules which have been modified to detect two-sided changes. For (5) this means we replace $\exp[\delta(iS_n/n - S_i)]$ by $\cosh[\delta(iS_n/n - S_i)]$ in the definition (4) of $R_n(\delta)$. For (8) we take a distribution G on $(-\infty, \infty)$; in particular for Table 2, G

TABLE 2
Expected delay for $\delta = 2$ (two-sided)

	ν_0		
μ	150	75	40
2.0	3.5 ± 0.0	3.5 ± 0.0	3.7 ± 0.0
	3.2 ± 0.0	3.3 ± 0.0	3.4 ± 0.0
	4.6 ± 0.0	4.7 ± 0.0	4.9 ± 0.0
1.0	15.1 ± 0.4	24.4 ± 1.4	47.9 ± 2.8
	17.2 ± 0.5	33.9 ± 2.1	72.3 ± 3.4
	13.3 ± 0.2	14.8 ± 0.3	17.8 ± 0.3
3.0	2.0 ± 0.0	2.0 ± 0.0	2.0 ± 0.0
	1.8 ± 0.0	1.9 ± 0.0	1.9 ± 0.0
	2.7 ± 0.0	2.7 ± 0.0	2.8 ± 0.0

is the standard normal distribution. For (11) two CUSUM tests are run simultaneously. One is designed to detect a change of $\mu = \delta > 0$, the other to detect a change of $\mu = -\delta$; and the composite test stops as soon as at least one of these one-sided tests stops [cf. Siegmund (1985), page 27ff.]. The values of B and b were chosen to make the P_∞-average run lengths about equal to $0.5 \times 792 = 396$. For the two-sided version of (5) with $\delta = 2$, $B = 123$, while $B = 265$ for (8) with G the standard normal distribution, and $b = 5.04$ for the two-sided CUSUM test. The three entries in each cell of Table 2 are estimated average delays ($n = 2500$) for these two-sided modifications of (5), (11) and (8), respectively.

The comparison between (5) and (11) in Table 2 is much the same as in Table 1. The most striking feature of Table 2 is that (8), which for large $|\mu|$ is somewhat inferior to (5) and (11), is substantially superior for $\mu = \delta/2$ and small ν_0. Another experiment, not reported here, gives similar results for $\delta = 1$. This is even more remarkable in view of the fact that (8) does not depend on the somewhat arbitrarily chosen value of δ. It appears that if early detection of a change as small as one-half the targeted δ is important, the stopping rule (8) may be preferable to both (5) and (11).

Up to now we have considered only the case $\nu = \nu_0$, where a change occurs immediately after the end of the training period. Consider for a moment the other extreme, where ν is much larger than ν_0. The training period is now effectively of length ν. In addition, if the procedure has not yet stopped, at time ν the process is not in its fixed initial state but is in a (random) quasistationary state. The effect of a longer training period can be inferred from the results already presented, which included different values of ν_0. The effect of starting from a quasistationary state is probably much the same as in the case of known μ_0. This situation was studied by Pollak and Siegmund (1985), who found that the decrease in expected delay when ν is large is greater for a Shiryayev–Roberts procedure than for a comparable CUSUM test. The upshot is that the advantage enjoyed by the CUSUM test in detecting changes of magnitude greater than or equal to δ when ν is small more or less disappears when ν is large. For a specific example, suppose that in Table 1 we have $\nu_0 = 0$ and $\nu = 150$. Monte Carlo estimates of $E_\nu(N - \nu | N > \nu)$ and $E_\nu(\tau - \nu | \tau > \nu)$ decrease to 9.7 and 10.1, respectively, when $\mu = 1$ and to the common value 5.3 for both stopping rules when $\mu = 1.5$.

McDonald (1989) has recently proposed a nonparametric CUSUM test to detect a stochastic increase from an initially unknown distribution. He forms a CUSUM statistic from the process of sequential ranks, which, as long as there is no change of distribution, are independent and uniformly distributed on their sets of possible values. His numerical results indicate that his procedure does extraordinarily well in comparison with the parametric competitor (18) for the range of conditions he explores. A simple computation of expectations indicates that McDonald's statistic can be expected to behave qualitatively like the recursive residual CUSUM test defined in (11). In particular, its performance may deteriorate dramatically when the time and magnitude of the change are small. McDonald's standard numerical example has a change occurring after 500 observations, which is quite large in comparison with the cases we have considered and may conceal this possible defect in his procedure.

APPENDIX 1. The limit of $E_\infty(N)/B$ as $B \to \infty$.

THEOREM 4. *For N defined by* (5), *uniformly in* ν_0,

$$\lim E_\infty(N - \nu_0)/B = 1/h(\delta).$$

PROOF. We begin by recalling the result of Pollak (1987) for the related stopping rule N^* defined by (1). Without loss of generality we can assume $\mu_0 = 0$, so $S_n^* = S_n$. For simplicity we write R_n^* and R_n for $R_n^*(\delta)$ and $R_n(\delta)$. Since $R_n - (n - \nu_0)$ is a P_∞-martingale having mean 0, we have

$$E_\infty(N^* - \nu_0) = BE_\infty(R_{N^*}^*/B).$$

Pollak (1987), under the unessential restriction that $\nu_0 = 0$, has shown that $R_{N^*}^*/B$ converges in law and is uniformly integrable, and has identified $\lim_{B\to\infty} E_\infty\{R_{N^*}^*/B\}$ as $1/h(\delta)$ [cf. (2)].

Since $R_n - (n - \nu_0)$ is also a P_∞-martingale with mean 0, we have

$$E_\infty(N - \nu_0) = BE_\infty(R_N/B). \tag{21}$$

To complete the proof of Theorem 4, we show that R_n is sufficiently close to R_n^* for all large n for which R_n or R_n^* is large, so that with overwhelming probability $N = N^*$ and $R_N/R_{N^*}^* \cong 1$. The desired result then follows from Pollak's theorem.

More precisely, let $\varepsilon > 0$ and put

$$A_\varepsilon = \{N = N^* > B\varepsilon + \nu_0, 1 - \varepsilon < R_N/R_{N^*}^* < 1 + \varepsilon\}.$$

Obviously,

$$E_\infty(R_N) = E_\infty(R_N; A_\varepsilon) + E_\infty(R_N; A_\varepsilon^c),$$

and hence by the Schwarz inequality

$$\begin{aligned}(1 - \varepsilon)E_\infty(R_{N^*}^*; A_\varepsilon) &\le E_\infty(R_N)\\ &\le (1 + \varepsilon)E_\infty(R_{N^*}^*; A_\varepsilon) + \left[E_\infty R_N^2 P_\infty(A_\varepsilon^c)\right]^{1/2}.\end{aligned} \tag{22}$$

By Lemma 2 below $E_\infty R_N^2 \le$ const. B^2; and by Lemmas 3 and 4

$$\limsup_{B\to\infty} P_\infty(A_\varepsilon^c) \to 0$$

as $\varepsilon \to 0$. The theorem now follows from (21) and Pollak's result by first letting $B \to \infty$, then $\varepsilon \to 0$. □

LEMMA 1. *Let* $\alpha > 1$. *There exists* $C_1 > 0$ *such that for all* $0 \le y \le \alpha\delta$,

$$E_\infty\{\exp[y(x_1 - a)] | x_1 \ge a\} \le \begin{cases} C_1, & \text{if } a \ge 0,\\ 2\exp(\alpha^2\delta^2/2 - \alpha\delta a), & \text{if } a < 0.\end{cases}$$

PROOF. The proof is a straightforward calculation, which is omitted. □

LEMMA 2. *Let $\alpha > 1$. There exists $C_2 > 0$ such that for all B and ν_0,*

$$E_\infty(R_N^\alpha) \le C_2 B^\alpha.$$

PROOF. Observe that R_n increases in x_n, and

$$\begin{aligned} R_{n+1} = {} & \exp\left[\delta n(n+1)^{-1}(x_{n+1} - \bar{x}_n) - \delta^2 n(n+1)^{-1}/2\right] \\ & + \sum_{k=\nu_0}^{n-1} \exp\left[\delta(k\bar{x}_n - S_k) - \delta^2 k(1 - k/n)/2\right] \\ & \qquad \times \exp\left\{\delta k(n+1)^{-1}(x_{n+1} - \bar{x}_n) - \delta^2 k^2/[2n(n+1)]\right\}, \end{aligned} \tag{23}$$

where $\bar{x}_k = S_k/k$. Letting F_{N-1} denote the σ-field generated by N, $x_1, \dots, x_{N-1}$, we have from (23)

$$\begin{aligned} E_\infty\left(R_N^\alpha | F_{N-1}\right) & \le E_\infty\left(1_{\{x_N > \bar{x}_{N-1}\}} R_N^\alpha | F_{N-1}\right) + (1+B)^\alpha \\ & \le E_\infty\left(R_N^\alpha | F_{N-1}, x_N > \bar{x}_{N-1}\right) + (1+B)^\alpha. \end{aligned}$$

Suppose we are given $x_1, \dots, x_n$, $N > n$ and $x_{n+1} > \bar{x}_n$. There exists a unique $\Delta = \Delta(n, x_1, \dots, x_n)$ such that $N = n + 1$ if and only if $x_{n+1} - \bar{x}_n \ge \Delta$. There are two cases to consider: $\Delta \ge 0$ or $\Delta < 0$. We first suppose $\Delta \ge 0$, which is the more difficult case. If $x_{n+1} - \bar{x}_n$ were replaced by Δ in (23), the resulting expression would equal B, and hence if $x_{n+1} - \bar{x}_n \ge \Delta$,

$$R_{n+1} \le \exp[\delta(x_{n+1} - \bar{x}_n - \Delta)] B.$$

Hence by Lemma 1, when $\Delta \ge 0$,

$$\begin{aligned} & E\left(R_{n+1}^\alpha | x_1, \dots, x_n, N > n, x_{n+1} \ge \bar{x}_n + \Delta\right) \\ & \quad \le \begin{cases} C_1 B^\alpha, & \text{if } \bar{x}_n + \Delta \ge 0, \\ 2\exp\left[\alpha^2\delta^2/2 - \alpha\delta(\bar{x}_n + \Delta)\right] B^\alpha, & \text{if } \bar{x}_n + \Delta < 0, \end{cases} \\ & \quad \le B^\alpha \left\{ C_1 + 2\exp\left[\alpha^2\delta^2/2 + \alpha\delta \max_{i \ge 1} |\bar{x}_i|\right] \right\}. \end{aligned}$$

If $\Delta < 0$, the condition $N = n + 1$ is implied by $x_{n+1} > \bar{x}_n$, and hence by (23)

$$\begin{aligned} & E_\infty\left(R_{n+1}^\alpha | x_1, \dots, x_n, N = n + 1, x_{n+1} > \bar{x}_n\right) \\ & \quad \le E_\infty\left\{\exp[\alpha\delta(x_{n+1} - \bar{x}_n)](1+B)^\alpha | x_1, \dots, x_n, x_{n+1} > \bar{x}_n\right\}. \end{aligned}$$

This in turn, by Lemma 1, is bounded by

$$\begin{aligned} & \begin{cases} C_1(1+B)^\alpha, & \text{if } \bar{x}_n \ge 0, \\ 2\exp\left[\alpha^2\delta^2/2 - \alpha\delta\bar{x}_n\right](1+B)^\alpha, & \text{if } \bar{x}_n < 0, \end{cases} \\ & \quad \le (1+B)^\alpha \left\{ C_1 + 2\exp\left[\alpha^2\delta^2/2 + \alpha\delta \max_{i \ge 1} |\bar{x}_i|\right] \right\}. \end{aligned}$$

The final inequality follows by taking expectations. □

LEMMA 3. *For any* $0 < \varepsilon < 1$,

$$P_\infty\{N - \nu_0 \le \varepsilon B\} \le \varepsilon \quad \text{and} \quad P_\infty\{N^* - \nu_0 \le \varepsilon B\} \le \varepsilon.$$

PROOF. Since R_n is a nonnegative submartingale and $E_\infty R_n = n - \nu_0$, for any integer $m \ge 1$,

$$P_\infty\{N - \nu_0 \le m\} \le B^{-1}E_\infty\{R_N; N - \nu_0 \le m\} \le B^{-1}E_\infty\{R_{m+\nu_0}\} = B^{-1}m.$$

The proof for R_n^* is the same. □

LEMMA 4. *For any* $\varepsilon > 0$ *there exists* $\eta(\varepsilon) \to 0$ *as* $\varepsilon \to 0$ *such that for all large* B,

$$P_\infty\{N = N^* > \varepsilon B + \nu_0, 1 - \varepsilon \le R_N/R_{N^*}^* \le 1 + \varepsilon\} \ge 1 - \eta.$$

PROOF. Let $0 < \beta < \frac{1}{2}$ and observe that

$$\frac{R_n}{R_n^*} = \frac{Q_n + (H_n - Q_n) + U_n + \Sigma_{1,n}}{V_n + \Sigma_{2,n}}, \tag{24}$$

where

$$H_n = \sum_{k=\nu_0}^{n^\beta} \exp\{\delta(kS_n/n - S_k) - \delta^2 k(1 - k/n)/2\},$$

$$Q_n = \sum_{k=\nu_0}^{n^\beta} \exp(-\delta S_k - \delta^2 k/2),$$

$$U_n = \sum_{k=n^\beta \vee \nu_0}^{n-n^\beta} \exp\{\delta(kS_n/n - S_k) - \delta^2 k(1 - k/n)/2\},$$

$$V_n = \sum_{k=\nu_0}^{n-n^\beta} \exp\{\delta(S_n - S_k) - \delta^2(n - k)/2\},$$

$$\sum_{1,n} = \sum_{k=(n-n^\beta)\vee\nu_0}^{n-1} \exp\{\delta(S_n - S_k) - \delta^2(n - k)/2\} \times \exp\{\delta(n - k)[\delta(1 - k/n)/2 - S_n/n]\}$$

and

$$\sum_{2,n} = \sum_{k=(n-n^\beta)\vee\nu_0}^{n-1} \exp\{\delta(S_n - S_k) - \delta^2(n - k)/2\}.$$

In these expressions a sum is understood to equal 0 if the lower index of summation exceeds the upper.

It is easy to see that (uniformly in ν_0)

$$H_n - Q_n = \sum_{k=\nu_0}^{n^\beta} \exp\!\left(-\delta S_k - \delta^2 k/2\right) \times \left[\exp\{\delta k S_n/n + \delta^2 k^2/(2n)\} - 1\right] \to 0 \tag{25}$$

with probability 1, while Q_n is bounded with probability 1, that is,

$$Q_n \le \sum_0^\infty \exp\left(-\delta S_k - \delta^2 k/2\right). \tag{26}$$

Since

$$\sum_{k=n^\beta \vee \nu_0}^{n-n^\beta} \exp\left[-\delta^2 k(1-k/n)/4\right] \le n\exp(-\delta^2 n^\beta/8) \to 0,$$

we have

$$P_\infty\{U_n \ge n\exp(-\delta^2 n^\beta/8)\} \le \sum_{k=n^\beta \vee \nu_0}^{n-n^\beta} P_\infty\{kS_n/n - S_k \ge \delta k(1-k/n)/4\}$$
$$\le 9(2\pi)^{-1/2}\delta^{-1} n\exp(-\delta^2 n^\beta/40)$$

and hence by Borel–Cantelli,

$$U_n \to 0 \quad \text{with probability } 1. \tag{27}$$

Similarly,

$$V_n \to 0 \quad \text{with probability } 1. \tag{28}$$

Finally, it is easy to see, for example, by the law of the iterated logarithm, that

$$\sum_{1,n} \Big/ \sum_{2,n} \to 1 \tag{29}$$

with probability 1.

It follows from (24), (25), (26), (27), (28) and (29) that with probability close to 1 if either R_n or R_n^* is large, say $\ge B/2$, then the ratio R_n/R_n^* is close to 1. According to Pollak (1987), $R_{N^*}^*/B$ has a continuous limiting distribution as $B \to \infty$ and hence except for the unlikely event that $R_{N^*}^*/B$ is close to 1, $N = N^*$ and R_N/R_{N^*} is close to 1. This together with Lemma 3 completes the proof. □

APPENDIX 2. Asymptotic behavior of $E_\nu(N - \nu | N > \nu)$.

In this appendix we give the essential ingredients of the rather technical proof of Theorem 1. The crucial tool is Lemma 12, which provides a precise coupling of $\log R_n$ and $\log R_n^*$. Without loss of generality we assume $\mu_0 = 0$.

Note that

$$\exp\{\delta(n-\nu)[S_n-(n-\nu)\mu]/n\}(R_n/R_n^*) \tag{30}$$
$$= \frac{\Sigma_{\nu_0}^{n-1}\exp[\delta(S_\nu - S_k) - \delta^2(\nu-k)/2] \times \exp\left\{\delta n^{-1}\left[\frac{1}{2}\delta(n-k)^2 - (\nu-k)S_n - \mu(n-\nu)^2\right]\right\}}{\Sigma_{\nu_0}^{n-1}\exp[\delta(S_\nu - S_k) - \delta^2(\nu-k)/2]}.$$

This is an average of $\exp\{\delta n^{-1}[\frac{1}{2}\delta(n-k)^2 - (\nu-k)S_n - \mu(n-\nu)^2]\}$ weighted by $\exp\{\delta(S_\nu - S_k) - \delta^2(\nu-k)/2\}$. We shall show that only terms with k close to ν contribute significantly to this average. Observe also that

$$\begin{aligned} &\delta n^{-1}\left[\tfrac{1}{2}\delta(n-k)^2 - (\nu-k)S_n - \mu(n-\nu)^2\right] \\ &\quad = -\delta(\mu-\delta/2)(n-\nu)^2/n \\ &\qquad +(\nu-k)\{\delta^2(n-k)/(2n) - \delta(\mu-\delta/2)(n-\nu)/n \\ &\qquad\qquad -\delta[S_n-(n-\nu)\mu]/n\}. \end{aligned} \tag{31}$$

Because of the lower bound imposed on ν by (14), the important values of $n-\nu$ are $O(\log B)$ (cf. the discussion of Theorem 3 given above), and hence for k close to ν all but the first term on the right-hand side of (31) can be neglected. The result, stated more precisely in Lemma 13, is that outside an event of small probability, uniformly for n in an interval of values which contains N with overwhelming probability

$$\begin{aligned} \log R_n = \log R_n^* &- \delta(\mu-\delta/2)(n-\nu)^2/n \\ &- \delta(n-\nu)[S_n-(n-\nu)\mu]/n + o(1). \end{aligned} \tag{32}$$

Integrating (32) on a suitable event and arguing along established lines of nonlinear renewal theory [e.g., Siegmund (1985), Chapter 9, Siegmund (1986) and Hogan (1984)], we have

$$\begin{aligned} &\log B + \text{asymptotic expected excess} + o(1) \\ &\quad = E_\nu\{\log R_N^* | N>\nu\} - \delta(\mu-\delta/2)E_\nu[(N-\nu)^2/N | N>\nu] \\ &\qquad - \delta E_\nu\{(N-\nu)[S_N-(N-\nu)\mu]/N | N>\nu\}. \end{aligned}$$

By the definition of R_n^* and Wald's lemma we have

$$\begin{aligned} E_\nu\{\log R_N^* | N>\nu\} &= \delta(\mu-\delta/2)E_\nu(N-\nu | N>\nu) \\ &\quad + E_\nu\left[\log\left\{\sum_0^{\nu-1}\exp\left[\delta(S_\nu-S_k) - \tfrac{1}{2}\delta^2(\nu-k)\right]\right.\right. \\ &\qquad \left.\left.+\sum_\nu^{N-1}\exp\left[-\delta(S_k-S_\nu)+\tfrac{1}{2}\delta^2(k-\nu)\right]\right\}\middle| N>\nu\right]. \end{aligned}$$

Under P_ν the argument of the logarithmic term on the right-hand side of this equation converges to $Q_1 + Q_2$, where Q_1 and Q_2 are independent, Q_1 has the P_∞-distribution of $\sum_1^\infty \exp[\delta S_k - \frac{1}{2}\delta^2 k]$ and Q_2 has the P_0-distribution of $\sum_0^\infty \exp[-\delta S_k + \frac{1}{2}\delta^2 k]$. It follows from the results of Appendix 1, a uniform integrability argument and Pollak and Siegmund (1986a) that

$$E_\nu\{\log R_N^* | N > \nu\} = \delta(\mu - \delta/2) E_\nu(N - \nu | N > \nu) + E\{\log(Q_1 + Q_2)\} + o(1).$$

From this point the argument is much like that sketched for Theorem 3, except for the crucial observation that

$$E_\nu\{(N - \nu)[S_N - (N - \nu)\mu]/N | N > \nu\} = 0. \tag{33}$$

This result follows from the observation that

$$(n - \nu)[S_n - (n - \nu)\mu]/n = \{S_n - S_\nu - (n - \nu)\mu\} + \{S_\nu - \nu[S_n - (n - \nu)\mu]/n\},$$
$$n = \nu, \nu + 1, \dots,$$

is the sum of two mean zero martingales, the first relative to the σ-fields $\mathscr{F}_n$ generated by the x's and the second relative to the smaller σ-fields $\mathscr{G}_n = \sigma(x_2 - x_1, \dots, x_n - x_1)$. In fact,

$$\nu[S_n - (n - \nu)\mu]/n - S_\nu = E_\nu\{S_n - S_\nu - (n - \nu)\mu | \mathscr{G}_n\}$$

with probability 1 for $n \geq \nu$. Since N is a $\mathscr{G}_n$-stopping time, it is also an $\mathscr{F}_n$-stopping time, and (33) follows.

REMARK. The first inequality in (14) requires that sufficient data to estimate μ_0 accurately are available by time ν. The second inequality is purely technical and probably can be eliminated. Under P_∞ the distribution of $N - \nu_0$ is approximately exponential with mean $B/h(\delta)$ when B is large. This follows from the proof of Theorem 4, where it is shown that $N = N^*$ with arbitrarily large P_∞-probability. Since R_n^* is Markovian and has a limiting distribution, the waiting time for it to enter a set of states having small probability is well-known to be approximately exponential. Hence the right-hand inequality in (14) guarantees that $P_\nu\{N > \nu\} = P_\infty\{N > \nu\}$ is bounded away from 0, so we do not have to condition on an event with vanishingly small probability. In practice the condition is not restrictive, since if ν is large compared to B, the procedure will most likely terminate before the change occurs. However, for Theorem 3, where the process is Markovian and has a P_∞-quasistationary distribution, no a priori upper bound on ν need be assumed.

We now present some of the details useful in turning the preceding argument into a rigorous proof. To simplify the notation, we take $\nu_0 = 0$.

Our first result is an easy consequence of standard arguments. Its proof is omitted.

LEMMA 5. *Let $\xi, \zeta > 0$. There exists $\gamma = \gamma(\xi, \zeta) > 0$ such that for every m,*

$$P_\infty\{|S_n| \geq \zeta n^{1/2+\xi} \text{ for some } n \geq m\} \leq \exp(-\gamma m^{2\xi})$$

and

$$P_\infty\left\{\max_{n \leq m}|S_n| \geq \zeta m^{1/2+\xi}\right\} \leq \exp(-\gamma m^{2\xi}).$$

In what follows c denotes a numerical constant, the value of which does not concern us and indeed may change from one appearance to the next. Let $N_n(k)$, $D(k)$ denote the term indexed by k in the numerator and denominator of (30), respectively. Note that the denominator is bounded away from 0, since $D(\nu) = 1$.

Let $0 < \kappa_0 < 1$, $\kappa_1 > 0$, $n_0 = \nu + (1 - \kappa_0)(\log B)/[\delta(\mu - \delta/2)]$ and $n_1 = \nu + (1 + \kappa_1)(\log B)/[\delta(\mu - \delta/2)]$. Also put $b = \log B$.

LEMMA 6. *Let $0 < \alpha, \beta < 1$. There exists a constant $c > 0$ such that for all large B,*

$$P_\nu\left\{\sum_{k=0}^{\infty} D(k) - \sum_{k=\nu-\nu^\beta}^{\nu+(n_0-\nu)^\alpha} D(k) \geq \exp(-b^c)\right\} \leq \exp(-b^c).$$

PROOF. Observe that

$$\log D(k) = \begin{cases} (\nu - k)\delta[(S_\nu - S_k)/(\nu - k) - \delta/2], & \text{for } k < \nu, \\ -(k - \nu)\delta[\{S_k - S_\nu - (k - \nu)\mu\}/(k - \nu) + \mu - \delta/2], & \\ & \text{for } k \geq \nu. \end{cases}$$

Any easy application of Lemma 5 with $\xi = 1/2$ completes the proof. □

LEMMA 7. *Let $0 < \rho < 1$. There exists $c > 0$ such that for all large B,*

$$P_\nu\left\{\sum_{k=0}^{(1-\rho)\nu} N_n(k) \geq B^{-c} \text{ for some } n \geq n_0\right\} \leq B^{-c}.$$

PROOF. For $0 \leq k \leq (1 - \rho)\nu$,

$$\begin{aligned} \log N_n(k) &= -\delta S_k + \delta^2(n - \nu)/2 - \delta(S_n - S_\nu) - \delta(\delta/2 - S_n/n)k \\ &\quad + \delta^2 k^2/(2n) + \delta(n - \nu)[S_n - (n - \nu)\mu]/n \\ &\leq -\delta(n - \nu)[\mu - \delta/2 - \{S_n - S_\nu - (n - \nu)\mu\}/(n - \nu)] - \delta S_k \\ &\quad + \delta k[\mu(n - \nu)/n + \delta(1 - \rho)\nu/(2n) - \delta/2 \\ &\qquad - \{S_n - (n - \nu)\mu\}/n] + \delta(n - \nu)[S_n - (n - \nu)\mu]/n. \end{aligned}$$

By Lemma 5, for any $\zeta > 0$ there exists $c > 0$ such that for all large B,

$$P_\nu\{|S_n - S_\nu - (n-\nu)\mu| \ge \zeta(n-\nu) \text{ for some } n \ge n_0\} \le B^{-c}, \tag{34}$$

$$P_\nu\{|S_n - (n-\nu)\mu| \ge \zeta n \text{ for some } n \ge n_0\} \le B^{-c}, \tag{35}$$

$$P_\nu\{|S_k| \ge \zeta(n_0 - \nu) \text{ for some } k \le n_0 - \nu\} \le B^{-c} \tag{36}$$

and

$$P_\nu\{|S_k| \ge \zeta k \text{ for some } n_0 - \nu \le k \le (1-\rho)\nu\} \le B^{-c}. \tag{37}$$

Choosing ζ to be sufficiently small, we obtain the lemma by bounding $\log N_n(k)$ from above on the intersection of the complements of the events in (34)–(37) and summing over $k \le (1-\rho)\nu$. □

LEMMA 8. *Let* $\frac{1}{2} < \beta < 1$, $0 < \rho < 1$. *There exists* $c > 0$ *such that for all large* B,

$$P_\nu\left\{ \sum_{k=(1-\rho)\nu}^{\nu-\nu^\beta} N_n(k) \ge \exp(-b^c) \text{ for some } \nu < n < n_1 \right\} \le \exp(-b^c).$$

PROOF. From the equality

$$\begin{aligned} \log N_n(k) = {} & \delta(S_\nu - S_k) - \delta(\mu - \delta/2)(n-\nu) \\ & - \delta(\nu - k)[S_n - (n-\nu)\mu]/n \\ & + \delta k[(\mu - \delta/2)(n-\nu)/n - \delta(\nu-k)/(2n)], \end{aligned} \tag{38}$$

we obtain for all $(1-\rho)\nu \le k \le \nu - \nu^\beta$,

$$\log N_n(k) \le \delta(S_\nu - S_k) - \delta(\nu-k)[S_n - (n-\nu)\mu]/n - \delta^2\nu^\beta k/(2n).$$

Let $\zeta < \delta(1-\rho)/4$. If

$$|S_\nu - S_k| < \zeta\nu^\beta \quad \text{for all } (1-\rho)\nu \le k \le \nu - \nu^\beta \tag{39}$$

and

$$|S_n - (n-\nu)\mu| < \zeta\nu^\beta \quad \text{for all } \nu < n < n_1, \tag{40}$$

then $\sum_{k=(1-\rho)\nu}^{\nu-\nu^\beta} N_n(k) \le \exp(-b^c)$ for some $c > 0$, and an application of Lemma 5 to the events in (39) and (40) completes the proof. □

LEMMA 9. *Let* $0 < \beta_2 < 1$ *and* $\max\{\beta_2/2, \beta_2 - \frac{1}{2}\} < \beta_1 < \beta_2$. *Then for some* $c > 0$ *and all large* B,

$$P_\nu\left\{ \sum_{k=\nu-\nu^{\beta_2}}^{\nu-\nu^{\beta_1}} N_n(k) \ge \exp(-b^{-c}) \text{ for some } \nu < n < n_1 \right\} \le \exp(-b^c).$$

PROOF. Let $0 < \varepsilon < \beta_1 - \beta_2 + \frac{1}{2}$. From (38) we see that for all $\nu - \nu^{\beta_2} \le k \le \nu - \nu^{\beta_1}$,

$$\log N_n(k) \le \delta(S_\nu - S_k) - \delta(\nu-k)[S_n - (n-\nu)\mu]/n - \delta^2 k\nu^{\beta_1}/(2n).$$

If for all $\nu - \nu^{\beta_2} \le k \le \nu - \nu^{\beta_1}$, $n_0 < n < n_1$ we have

$$|S_\nu - S_k| < \nu^{\beta_1/2+\beta_2/4} \quad \text{and} \quad |S_n - (n-\nu)\mu| < \nu^{1/2+\varepsilon},$$

then for some $c > 0$, $\sum_{k=\nu-\nu^{\beta_2}}^{\nu-\nu^{\beta_1}} N_n(k) \le \exp(-b^c)$ for all $\nu < n < n_1$. Another appeal to Lemma 5 completes the proof. □

Lemma 10. *Let $0 < \beta < 1$. There exists $c > 0$ such that for all large B,*

$$P_\nu\left\{\sum_{k=0}^{\nu-\nu^\beta} N_n(k) \ge \exp(-b^c) \text{ for some } n_0 < n < n_1\right\} \le \exp(-b^c).$$

Proof. We apply Lemma 9 repeatedly until $\beta_1 < \beta$ and combine the result with Lemmas 7 and 8. □

Lemma 11. *Let $0 < \alpha < 1$. There exists $c > 0$ such that for all large B,*

$$P_\nu\left\{\sum_{k=\nu+(n_0-\nu)^\alpha}^{n-1} N_n(k) \ge \exp(-b^c) \text{ for some } \nu < n < n_1\right\} \le \exp(-b^c).$$

Proof. Let $0 < \varepsilon < \frac{1}{2}$. We write

$$\begin{aligned}\log N_n(k) = &-(k-\nu)\left[\delta\mu - \delta^2 k/(2n) + \delta\{S_k - S_\nu - (k-\nu)\mu\}/(k-\nu)\right]\\ &-\delta(\mu - \delta/2)(n-\nu)(n-k)/n\\ &-\delta(\nu-k)[S_n - (n-\nu)\mu]/n.\end{aligned}$$

If for all $\nu + (n_0 - \nu)^\alpha \le k < n < n_1$ we have

$$|S_k - S_\nu - (k-\nu)\mu| \le (k-\nu)^{3/4}$$

and

$$|S_n - (n-\nu)\mu| < \nu^{1/2+\varepsilon},$$

then for some $c > 0$ and all $\nu < n < n_1$,

$$\sum_{k=\nu+(n_0-\nu)^\alpha}^{n-1} N_n(k) \le \exp(-b^c).$$

Yet another application of Lemma 5 completes the proof. □

Lemma 12. *For some $c > 0$,*

$$P_\nu\Big\{\Big|\log(R_n/R_n^*) + \delta(\mu - \delta/2)(n-\nu)^2/n + \delta(n-\nu)[S_n - (n-\nu)\mu]/n\Big| \ge \exp(-b^c) \text{ for some } n_0 < n < n_1\Big\} \le \exp(-b^c).$$

Proof. Recall (30) and the lines following. Let $0 < \alpha < \min\{\frac{1}{4}, \eta_1 - 1\}$, $0 < \beta < \min\{\frac{1}{2}, (\eta_1 - 1)/\eta_1\}$, and consider (31) for $\nu - \nu^\beta \le k \le \nu + (n_0 -$

$\nu)^{\alpha}$. For any $\varepsilon < \min\{\frac{1}{2} - \alpha/\eta_1, \frac{1}{2} - \beta\}$, if $|S_n - (n-\nu)\mu| < \nu^{1/2+\varepsilon}$ for all $n_0 < n < n_1$ the final term on the right-hand side of (31) is negligible. Hence from Lemmas 5, 10 and 11 we obtain Lemma 12. □

LEMMA 13. *For large enough* κ_1, $E_\nu[(N-\nu)1\{N > n_1\}|N > \nu] \to 0$.

PROOF. For $n \geq \nu$,

$$\log R_n \geq \delta(\nu S_n/n - S_\nu) - \delta^2\nu(n-\nu)/(2n).$$

Since S_ν is independent of $\{N > \nu\}$,

$$P_\nu\{N > n|N > \nu\} \leq \Phi\{[b - \delta(\mu - \delta/2)\nu(1-\nu/n)]/\delta[\nu(1-\nu/n)]^{1/2}\}.$$

Hence for large enough κ_1,

$$BP_\nu\{N > n_1|N > \nu\} \to 0. \tag{41}$$

Now let $\nu < m < n$ and write

$$\begin{aligned} R_n = \sum_0^{m-1} &\exp\left[\delta(kS_n/n - S_k) - \tfrac{1}{2}\delta^2k(n-k)/n\right] \\ &+ \sum_m^{n-1} \exp\{\delta\left[kn^{-1}(S_n - S_m) + km^{-1}S_m - (S_k - S_m) + S_m\right] \\ &\qquad\qquad - \tfrac{1}{2}\delta^2k(n-k)/n\}. \end{aligned}$$

Let $\tilde{R}_n$ denote R_n but with μ subtracted from each x_k for $k > m$, and observe that $\tilde{R}_n \leq R_n$. Let $\mathscr{F}_m$ be the σ-field generated by $x_1, \dots, x_m$ and note that conditional on $\mathscr{F}_m$ the P_ν joint distribution of $\tilde{R}_n$, $n = m+1, \dots,$ equals the P_∞ joint distribution of R_n, $n = m+1, \dots$. Let $m = n_1$. Then on $\{N > n_1\}$,

$$\begin{aligned} E_\nu(N - \nu|\mathscr{F}_{n_1}) &= (n_1 - \nu) + \sum_{n_1}^{\infty} P_\nu\{N > n|\mathscr{F}_{n_1}\} \\ &= (n_1 - \nu) + \sum_{n_1}^{\infty} P_\nu\{R_k < B \text{ for all } n_1 < k \leq n|\mathscr{F}_{n_1}\} \\ &\leq (n_1 - \nu) + \sum_{n_1}^{\infty} P_\nu\{\tilde{R}_k < B \text{ for all } n_1 < k \leq n|\mathscr{F}_{n_1}\} \\ &= (n_1 - \nu) + \sum_{n_1}^{\infty} P_\infty\{R_k < B \text{ for all } n_1 < k \leq n|\mathscr{F}_{n_1}\} \\ &= (n_1 - \nu) + E_\infty(N - n_1|\mathscr{F}_{n_1}) \leq (n_1 - \nu) + E_\infty(R_N|\mathscr{F}_{n_1}). \end{aligned}$$

An application of (41) and the proof of Lemma 2 complete the proof of Lemma 13. □

LEMMA 14. *For any* $\varepsilon > 0$ *there exists* $c > 0$ *such that*

$$P_\nu\{\log R_n \geq \log R_n^* + \varepsilon b \text{ for some } \nu \leq n \leq n_0\} \leq \exp(-b^c).$$

PROOF. Since

$$R_n \leq R_n^* \left\{ \sum_{k=0}^{\nu-\nu^\beta} N_n(k) + \sum_{k=\nu+(n_0-\nu)^\alpha}^{n-1} N_n(k) + \frac{\sum_{k=\nu-\nu^\beta}^{\nu+(n_0-\nu)^\alpha \wedge (n-1)} N_n(k)}{\sum_{k=0}^{n-1} D_k} \right\}$$
$$\times \exp\{-\delta(n-\nu)[S_n - (n-\nu)\mu]/n\},$$

the result follows from appropriate modifications of Lemmas 7 and 10 to cover the range $\nu \leq n \leq n_0$, Lemmas 8, 9, 11 and the proof of Lemma 12. □

LEMMA 15. *For any* $\varepsilon > 0$ *there exists* $c > 0$ *such that for all large* B,

$$P_\nu\{R_\nu^* > B^\varepsilon | N > \nu\} \leq cB^{-\varepsilon}.$$

PROOF. Observe that

$$P_\nu\{R_\nu^* > B^\varepsilon | N > \nu\} \leq P_\infty\{R_\nu^* > B^\varepsilon\}/P_\infty\{N > \nu\}.$$

From the upper bound imposed on ν by condition (14) and the results of Appendix 1, we see that

$$P_\infty\{N > \nu\} \geq P_\infty\{N > B\eta_2\} \sim P_\infty\{N^* > B\eta_2\} \to \exp(-h(\delta)\eta_2) \quad \text{as } B \to \infty.$$

Also, under P_∞, R_ν^* has the same distribution as

$$\sum_0^{\nu-1} \exp(\delta S_k - \delta^2 k/2),$$

which is dominated by

$$\sum_0^{\infty} \exp(\delta S_k - \delta^2 k/2). \tag{42}$$

According to Kesten (1973), Theorem 5, the tail of distribution of (42) goes to 0 at the rate $1/x$. □

APPENDIX 3. A remark about Pollak and Siegmund (1985).

In several places we have referred to Pollak and Siegmund (1985), who were concerned with detecting a change in the drift of Brownian motion when the initial drift μ_0 is known (say $\mu_0 = 0$). Two of their results are a double integral giving the expected delay for a Shiryayev–Roberts procedure when $\nu = 0$ (Proposition 2) and a precise asymptotic expression for the expected delay when ν is large (Theorem 1). Here we observe that as $B \to \infty$, the difference between these expected delays converges to an infinite series, which is often easily evaluated. Hence in many cases the double integral given by Pollak and Siegmund (1985) can be approximately evaluated by hand.

Let $W(t)$, $0 \le t < \infty$, denote Brownian motion with drift 0 for $t \le \nu$ and drift $\mu > 0$ for $t > \nu$. Let $\delta, B > 0$ and define

$$T = \inf\left\{t: \int_0^t \exp\left[\delta\{W(t) - W(s)\} - \delta^2(t-s)/2\right] ds \ge B\right\}.$$

THEOREM. *For any $\mu > 0$, as $\nu, B \to \infty$,*

$$E_0(T) - E_\nu(T - \nu | T > \nu) \to 2\delta^{-2} \sum_1^\infty [k(2\alpha - 1 + k)]^{-1}, \tag{43}$$

where $\alpha = \mu/\delta$.

REMARK. Special cases of interest are $\alpha = 1, 1/2, 3/2, 2$. The series on the right-hand side of (43) is 1, $\Sigma_1^\infty k^{-2} = \pi^2/6$, 3/4, and 11/18, respectively.

PROOF. From the argument given in Appendix B of Pollak and Siegmund (1985), we see that as $\nu, B \to \infty$,

$$E_\nu(T - \nu | T > \nu) + o(1)$$

$$= \begin{cases} 2\delta^{-2} \int_0^A \{[1 - (x/A)^{2\alpha - 1}]/(2\alpha - 1)\} \exp(-1/x)\, dx/x, & \alpha \ne 1/2, \\ 2\delta^{-2} \int_0^A \log(A/x) \exp(-1/x)\, dx/x, & \alpha = 1/2, \end{cases} \tag{44}$$

where $A = \delta^2 B/2$. The expression for $E_0(T)$ given in Proposition 2 together with some routine calculus yields

$$E_0(T) = 2\delta^{-2} \int_{A^{-1}}^\infty \int_{u^{-1}}^A z^{-2\alpha} \exp(1/z)\, dz\, u^{-2\alpha} \exp(-u)\, du.$$

We now expand $\exp(1/z) = \Sigma_{k=0}^\infty z^{-k}/k!$ and integrate term by term. The $k = 0$ term is exactly the right-hand side of (44); and the remaining terms converge as $A \to \infty$ to the series appearing on the right-hand side of (43). □

REFERENCES

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

VAN DOBBEN DE BRUYN, C. S. (1968). *Cumulative Sum Tests*. Griffin, London.

HOGAN, M. (1984). Problems in boundary crossings for random walks. Ph.D dissertation, Stanford Univ.

KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248.

LEVIN, B. and KLINE, J. (1985). The CUSUM test of homogeneity with an application to spontaneous abortion epidemiology. *Statist. Med.* **4** 469–488.

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

LUCAS, J. M. and CROSIER, R. B. (1982). Fast initial response for CUSUM quality control schemes: Give your CUSUM a head start. *Technometrics* **24** 199–205.

MCDONALD, D. (1989). A CUSUM procedure based on sequential ranks. Unpublished manuscript.

PAGE, E. A. (1954). Continuous inspection schemes. *Biometrika* **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. (1985). A diffusion process and its application to detecting a change in the drift of Brownian motion. *Biometrika* **72** 267–280.

POLLAK, M. and SIEGMUND, D. (1986a). Convergence of quasi-stationary to stationary distributions for stochastically monotone Markov processes. *J. Appl. Probab.* **23** 215–220.

POLLAK, M. and SIEGMUND, D. (1986b). Approximations to the average run lengths of CUSUM tests. Technical Report, Stanford Univ.

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

SHIRYAYEV, A. N. (1963). On optimal method in earliest detection problems. *Theory Probab. Appl.* **8** 26–51.

SIEGMUND, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser. B* **37** 394–401.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.

WEATHERALL, J. A. C. and HASKEY, J. C. (1976). Surveillance of malformations. *British Med. Bull.* **32** 39–44.

WILSON, D. W., GRIFFITHS, K., KEMP, V. W., NIX, A. B. J. and ROWLANDS, R. J. (1979). Interval quality control of radioimmunoassays: Monitoring of error. *J. Endocrinology* **80** 365–372.

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM
ISRAEL

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305