*The Annals of Statistics*
1995, Vol. 23, No. 1, 255–271

# USING THE GENERALIZED LIKELIHOOD RATIO STATISTIC FOR SEQUENTIAL DETECTION OF A CHANGE-POINT

BY D. SIEGMUND[1] AND E. S. VENKATRAMAN

*Stanford University and Memorial Sloan-Kettering Cancer Center*

We study sequential detection of a change-point using the generalized likelihood ratio statistic. For the special case of detecting a change in a normal mean with known variance, we give approximations to the average run lengths and compare our procedure to standard CUSUM tests and combined CUSUM–Shewhart tests. Several examples indicating extensions to problems involving multiple parameters are discussed.

**1. Introduction.** Let $x_1, x_2, \ldots$ be independent observations. We assume that, for some unknown parameter $r$, the random variables $x_1, \ldots, x_r$ have probability density function $f_0$, while $x_{r+1}, \ldots$ have probability density function $f_1$. The density functions may be completely specified or may contain unknown parameters. The unknown parameter $r$ is called the change-point. We seek a stopping rule $N$ which allows us to observe the $x$'s sequentially, only rarely stopping the process before the change-point, but stopping it soon afterward. The canonical example is an industrial process which is in control and should be allowed to continue operating until time $r$, after which it goes out of control and should be stopped and reset as soon as possible. One formalization of these requirements is in terms of the average run lengths $E_\infty(N)$, and $\sup_r E_r(N - r|N > r)$, where $E_r$ denotes expectation under the hypothesis that the true change-point is $r$, and $E_\infty$ denotes expectation under the hypothesis of no change whatever. We require a stopping rule for which $E_\infty(N)$ is large, say, greater than a prespecified large constant, and subject to this constraint $\sup E_r(N - r|N > r)$ is as small as possible. This is still not a well-specified problem, since the indicated expectations may depend on unknown parameters, and it may not be possible to achieve our goals uniformly in those parameters.

In the special case that $f_0$ and $f_1$ are completely specified, two very good (indeed, optimal with the appropriate definitions of optimality) procedures are the Page–Lorden CUSUM test [Page (1954) and Lorden (1971)] and the quasi-Bayesian test of Shiryayev (1963) [cf. also Roberts (1966)]. The CUSUM test is defined by the stopping rule

$$N = \inf\left\{n\colon \max_{0 \le k < n} \sum_{i=k+1}^{n} \log[f_1(x_i)/f_0(x_i)] \ge a\right\}.$$

Received January 1992; revised August 1993.
[1]Research partially supported by AFOSR Grant 89-0355, National Security Agency Grant MDA 904-89-H-2040 and NSF Grant DMS-91-04432.
*AMS* 1991 *subject classification*. Primary 62L10.
*Key words and phrases*. Change-point, likelihood ratio, sequential detection.

In the more complicated case that $f_0$ is completely specified but $f_1$ contains a (one-dimensional) parameter, one frequently uses this same stopping rule with $f_1$ evaluated at some nominal value of the unknown parameter. In the special case that $f_0$ is normal with mean 0 and variance 1 while $f_1$ has its mean shifted to $\delta$, this stopping rule becomes

$$(1.1) \qquad N(\delta) = \inf\left\{n\colon \max_{0 \le k < n} \delta[S_n - S_k - \delta(n-k)/2] \ge a\right\},$$

where $S_n = x_1 + \cdots + x_n$.

For the stopping rule of a CUSUM test [in particular, for the special case (1.1)], very accurate approximations have been given for the average run lengths. $E_\infty(N)$ and $\sup E_r(N - r|N > r) = E_0(N)$ [e.g., Siegmund (1975) and (1985), Chapter 10]. Alternatively a number of authors have used the stationary Markovian structure of the process to give numerical algorithms for calculating these quantities [e.g., Brook and Evans (1972)].

The stopping rule (1.1) is useful for detecting changes in the mean $\mu$ from $\mu = 0$ to values of $\mu$ for which $\delta\mu > 0$. If one wants to detect changes to either a positive or negative value, one can use the two-sided stopping rule

$$(1.2) \qquad N = \min[N(\delta), N(-\delta)].$$

A simple renewal argument [e.g., Siegmund (1985), page 28] shows that, for $r = 0$ or $\infty$ and $N$ defined by (1.2),

$$[E_r(N)]^{-1} = [E_r(N(\delta))]^{-1} + [E_r(N(-\delta))]^{-1}.$$

Although it is not at all obvious in this case whether $\sup_r E_r(N - r|N > r)$ is attained at $r = 0$, one conventionally continues to use the surrogate $E_0(N)$ along with $E_\infty(N)$ to evaluate such two-sided stopping rules.

This paper is concerned with the average run lengths of a CUSUM-like test using the generalized likelihood ratio statistic to detect a change-point. In the simple case of detection of a change in the mean of independent normally distributed random variables with known variance, which without loss of generality can be assumed equal to 1, the procedure was mentioned by Barnard (1959) and amounts to maximizing the expression in (1.1) as a function of $\delta$ to obtain a stopping rule of the form

$$(1.3) \qquad T = \inf\left\{n\colon \max_{0 \le k < n} |S_n - S_k|/(n-k)^{1/2} \ge b\right\}.$$

The same procedure has been suggested by a number of other authors [e.g., Basseville (1988) and references given there]. While one can easily write down such a procedure for a wide variety of parametric problems (this is presumably a reason for its attractiveness); the process does not exhibit the same simple Markovian structure as that in (1.1), and there does not seem to have been any attempt to evaluate its average run length.

In this paper we begin with the problem of a change in a normal mean with fixed, known variance and in Sections 2 and 3 give approximations for $E_\infty(T)$ and $E_0(T)$. Section 4 contains numerical results designed to show that this procedure is better in some respects than a standard CUSUM test and is competitive with the combined Shewhart–CUSUM test of Lucas (1982). It turns out that our analysis is substantially more complicated than that required to study the average run lengths of a standard CUSUM test. In Section 5 we discuss a number of related, more general models for which CUSUM tests have not been developed and which, at least heuristically, can be studied by the same methods. These include changes in a normal mean when the variance or (and) the initial value of the mean is (are) unknown, when the data are multivariate and when they are autocorrelated. Details of the proof of our main result are given in the Appendix.

**2. $E_\infty(T)$.** Throughout this section, $x_1, x_2, \ldots$ are independent $N(0,1)$ random variables and $T$ is defined by (1.3). We let $\phi$ and $\Phi$ denote the standard normal density and distribution functions, respectively. We shall also need the function $\nu(x)$ defined by

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

For purposes of numerical evaluation it is useful to know the approximation

$$\nu(x) = \exp(-\rho x) + o(x^2), \qquad x \to 0, \tag{2.1}$$

where $\rho$ is a constant whose value is approximately 0.583 [cf. Siegmund (1985), Chapter 10].

The principal result of this paper is the following theorem.

THEOREM 1. *As $b \to \infty$,*

$$E(T) \sim \frac{(2\pi)^{1/2}\exp(b^2/2)}{b\int_0^\infty x\nu^2(x)\,dx}. \tag{2.2}$$

*In fact $T$ is asymptotically exponentially distributed with expectation given by the right-hand side of* (2.2).

REMARK 1. There is some theoretical justification for using, instead of (2.2), the asymptotically equivalent approximation

$$E(T) \sim \frac{(2\pi)^{1/2}\exp(b^2/2)}{b\int_0^b x\nu^2(x)\,dx}, \tag{2.3}$$

which is substantially more accurate numerically (see Remark 3 and Section 4).

REMARK 2. A related result for Brownian motion obtained via the Poisson clumping heuristic appears in Aldous [(1989), page 212]. In principle our method permits one to give a rigorous proof of this result by careful attention to issues of uniformity and passage to a limit. Given the substantial technique involved in the proof of Theorem 1, this added layer of detail would be quite complicated. For a similar argument in a substantially simpler context see Leadbetter, Lindgren and Rootzén [(1983), Chapter 12].

The starting point in our proof of Theorem 1 is the following result, which can be obtained by the method of Siegmund (1988).

PROPOSITION 1. *Suppose* $b \to \infty$ *and* $m_0 \to \infty$ *in such a way that* $m_0 = cb^2$ *for some fixed* $0 < c < \infty$. *Then*

$$\begin{aligned} P\{T \le m_0\} &= P\left\{ \max_{0 \le i < j \le m_0} \frac{|S_j - S_i|}{(j-i)^{1/2}} > b \right\} \\ &\sim m_0 b\phi(b) \left( \int_{c^{-1/2}}^{\infty} x\nu^2(x)\,dx - c^{-1} \int_{c^{-1/2}}^{\infty} x^{-1}\,\nu^2(x)\,dx \right). \end{aligned} \tag{2.4}$$

REMARK 3. In the derivation of (2.4) the integrals arise as the limits of Riemann sums of the form

$$\sum_{k=1}^{m_0} b^4 k^{-2} \nu^2(bk^{-1/2}) b^{-2}.$$

The upper limit of $b$ appearing in the integral in (2.3) results from approximating this sum by

$$\int_{b^{-2}}^{c} y^{-2}\nu^2(y^{-1/2})\,dy = 2\int_{c^{-1/2}}^{b} x\nu^2(x)\,dx.$$

An upper limit of $b$ is also consistent with the corresponding result for Brownian motion if we put $\nu(x) = 1$ for all $x$.

REMARK 4. Both Theorem 1 and Proposition 1 are concerned with the maximum of the random field $Z_{i,j} = |S_j - S_i|/(j-i)^{1/2}$, $1 \le i < j \le m$. In Proposition 1 the value of $m = m_0$ is of order $b^2$ and the indicated probability goes to 0 exponentially fast. In Theorem 1, $m$ is of order $\exp(b^2/2)/b$, and the same probability converges to a limit in $(0, 1)$. To make the transition from small to large $m$, we must analyze the behavior of $Z_{i,j}$, $1 \le i < j \le m$, for values $(i, j)$ with $(j - i) \le cb^2$ and $(j - i) > cb^2$. It turns out that the range $(j - i) > cb^2$ makes a negligible contribution as $c \to \infty$, but showing this is where much of the work lies.

REMARK 5. By setting the right-hand side of (2.2) equal to $\exp(x)$ and solving asymptotically for $b$ as a function of $m$ and $x$, one can easily reformulate the limit theorem for $T$ in the statement of Theorem 1 as a double exponential limit in distribution for

$$\frac{\left[\max_{0 \le i < j \le m} |S_j - S_i|/(j-i)^{1/2} - a_m\right]}{b_m},$$

for suitable sequences $a_m$, $b_m$. As Aldous [(1989), Section A10] points out, this reformulation will rarely provide a useful approximation, but it may nevertheless be interesting to compare it to the classical one-dimensional result of Darling and Erdös (1956). In particular, the integral appearing on the right-hand side of (2.2) is present in the final approximation, showing that our result is distribution dependent, whereas that of Darling and Erdös is not.

We begin with a comparatively trivial bound.

LEMMA 1. *Assume* $m = t/[b\phi(b)\int_0^\infty x\nu^2(x)\,dx]$, *and let* $0 < \varepsilon < 1$. *Then for all large* $b$, *uniformly in* $t$ *bounded away from* 0,

$$P\{T > m\} \le \exp[-t(1-\varepsilon)].$$

*In particular*, $\{T/[b^{-1}\exp(b^2/2)],\ b > 10^6\}$ *is uniformly integrable.*

PROOF. Let $m_0 = cb^2$, as above. Then

$$\begin{aligned} P\{T > m\} &\le P\left\{\max_{0\le k<m/m_0}\ \max_{km_0\le i<j<(k+1)m_0} \frac{|S_j - S_i|}{(j-i)^{1/2}} < b\right\} \\ &= \prod_{0\le k<m/m_0} P\left\{\max_{km_0\le i<j<(k+1)m_0} \frac{|S_j - S_i|}{(j-i)^{1/2}} < b\right\} \qquad (2.5) \\ &= \left(1 - P\left\{\max_{0\le i<j<m_0} \frac{|S_j - S_i|}{(j-i)^{1/2}} \ge b\right\}\right)^{m/m_0}. \end{aligned}$$

By Proposition 1 we can choose $c$ large enough that

$$P\left\{\max_{0\le i<j<m_0} \frac{|S_j - S_k|}{(j-i)^{1/2}} \ge b\right\} \ge m_0 b\phi(b)\int_0^\infty x\nu^2(x)\,dx(1-\varepsilon),$$

for all large $b$. The lemma now follows from the elementary inequality $1 - x \le \exp(-x)$. □

PROOF OF THEOREM 1. Let $m = t/[b\phi(b)\int_0^\infty x\nu^2(x)\,dx]$ and $m_0 = cb^2$. By Lemma 1 we know that

$$\liminf P\{T \le m\} \ge 1 - \exp(-t)$$

and $\{T/[b^{-1}\exp(b^2/2)],\ b > 10^6\}$ is uniformly integrable. To complete the proof it suffices to show

$$\limsup P\{T \le m\} \le 1 - \exp(-t).$$

Obviously,

$$\begin{aligned} &P\{T \le m\} \\ &\quad\le P\left\{\max_{0\le k<m/m_0}\ \max_{km_0\le i<j\le(k+1)m_0} \frac{|S_j - S_i|}{(j-i)^{1/2}} > b\right\} \qquad (2.6) \\ &\qquad + P\left\{\max_{0\le k<m/m_0}\ \max_{km_0\le i<(k+1)m_0,(k+1)m_0\le j\le m} \frac{|S_j - S_i|}{(j-i)^{1/2}} > b\right\}. \end{aligned}$$

Let $\varepsilon > 0$. The first term on the right-hand side of (2.6) can be treated as in the proof of Lemma 1 and shown to have limit superior less than or equal to $1 - \exp(-t(1 + \varepsilon))$ provided $c$ is large enough. Hence, to complete the proof it suffices to show that for $c$ large enough the second term is less than $\varepsilon$ for all large $b$. This is the content of Lemmas 2–9 in the Appendix. □

**3. $E_0(T)$.** When there is a change, we are interested in the expected delay until its detection, $E_r(N - r|N > r)$, which is a function of the two variables $r$ and $\mu$. To simplify the process of studying this function, it is customary to consider $E_0(N)$, which only depends on $\mu$. For many one-sided stopping rules, for example, for stopping rule (1.2) or (1.3) with the absolute values removed, which might be used to detect a change from mean value 0 to some positive value, it is easy to see that $\sup E_r(N - r|N > r) = E_0(N)$. It is not obvious that this relation remains true for two-sided stopping rules. However, since $E_0(T)$ is certainly of interest and is reasonably easy to approximate, we consider it as a surrogate for the more complex function $E_r(T - r|T > r)$.

Throughout this section $x_1, x_2, \ldots$ are independently and normally distributed with mean $\mu$ and unit variance. It will be convenient to change our notation for the rest of this section and to write $P_\mu$ and $E_\mu$ to denote dependence of probability and expectation on the value of $\mu$. (Note the inconsistency with our previous notation, where a subscript denoted the value of $r$ and the dependence on the underlying distribution was suppressed. Now we take $r = 0$ and the subscript emphasizes dependence on $\mu$.) Our approximation for $E_\mu(T)$ involves a synthesis of results of Pollak and Siegmund (1975), Lai and Siegmund (1979) and Siegmund (1979). [See also Siegmund (1985), Chapters 9 and 10.]

A direct consequence of the ideas in the first two of these papers is that for fixed $\mu > 0$, as $b \to \infty$,

$$(3.1)\quad E_\mu(T) = \mu^{-2}(b^2 - 1) + \mu^{-1}\left[2E_{\mu/2}\left(\min_{n\ge 0} S_n\right) + \frac{E_{\mu/2}\left(S_{\tau_0}^2\right)}{E_{\mu/2}\left(S_{\tau_0}\right)}\right] + o(1),$$

where $\tau_0 = \inf\{n: S_n > 0\}$. In order to evaluate the second and third terms on the right-hand side of (3.1), we suppose that $\mu \to 0$. According to Siegmund (1979), we have

$$(3.2)\qquad E_{\mu/2}\left(S_{\tau_0}^2\right)/E_{\mu/2}\left(S_{\tau_0}\right) = 2\rho + \mu/4 + o(\mu^2),$$

where $\rho$ is the constant introduced in (2.1). Also,

$$(3.3)\qquad -E_{\mu/2}\left(\min_{n\ge 0} S_n\right) = \mu^{-1} - \rho + \mu/8 + o(\mu).$$

Substitution of (3.2) and (3.3) into (3.1) leads to the approximation

$$(3.4)\qquad E_\mu(T) \approx (b^2 - 3)/\mu^2 + 4\rho/\mu.$$

We do not know a precise mathematical interpretation of approximation (3.4). By comparing the derivation of (3.1) with the much more precise

TABLE 1
$E_\infty(T)$

| $b$ | Monte Carlo | (2.3) |
|---|---|---|
| 3.30 | 288 ± 6 | 256 |
| 3.45 | 431 ± 9 | 399 |
| 3.60 | 685 ± 15 | 638 |
| 3.75 | 1108 ± 24 | 1047 |
| 3.90 | 1876 ± 42 | 1764 |
| 4.05 | 3244 ± 70 | 3048 |
| 4.20 | 5651 ± 113 | 5399 |

approximations available for the stopping rule (1.1) [e.g., Siegmund (1985), Chapter 10], we see that (3.1) cannot be expected to provide a good approximation for values of $\mu$ close to 0, and hence we should not expect (3.4) to provide good approximations for $\mu$ close to 0. Since we have used the small-$\mu$ asymptotics of (3.2) and (3.3), we should not be surprised if it is not especially good for large values of $\mu$ also. Nevertheless, we shall see in the next section that (3.4) provides reasonable approximations for an interesting range of parameter values.

**4. Numerical examples.** The purpose of this section is to give some indication of the accuracy of the approximations (2.3) and (3.4) and to show that the procedure defined by (1.3) compares favorably with others in the literature. The point of the examples is not to demonstrate the superiority of (1.3), but only its reasonableness. In the following section we show heuristically that similar procedures can be developed for a number of more difficult problems, where there are few competitors.

Table 1 compares the results of a 2000-repetition Monte Carlo experiment with approximation (2.3) for various values of $b$. In Table 2 we consider a single value of $b$, chosen to give $E_\infty(N) \approx 400$, and compare three different procedures. The first is the stopping rule (1.3) (GLR). The second is the

TABLE 2
*Comparison of three stopping rules*

| $\mu$ | LR $b = 3.45$ | | CS $\delta = 1$, $a = 4.83$ | CSCS $\delta = 1$, $a = 5$, $a_1 = 3.5$ |
|---|---|---|---|---|
| 0.00 | 431. | 399. * | 396. | 391. |
| 0.25 | 106. | | 129. | 131. |
| 0.50 | 34. | | 36. | 37. |
| 1.00 | 10.9 | 11.2* | 10.0 | 10.2 |
| 1.50 | 5.6 | 5.5* | 5.5 | 5.6 |
| 2.00 | 3.5 | 3.4* | 3.8 | 3.8 |
| 3.00 | 1.9 | 1.8* | 2.3 | 2.1 |
| 4.00 | 1.3 | 1.1* | 1.7 | 1.3 |

two-sided version, (1.2), of the standard CUSUM test (CS). Since the CUSUM test has been compared unfavorably to the Shewhart chart for detecting large changes, the third test is the combined Shewhart–CUSUM (CSCS), studied by Lucas (1982). The stopping rule of this test is the minimum of (1.2) and

$$N_1 = \inf\{n: |x_n| > a_1\}. \tag{4.1}$$

For the standard CUSUM test, the entries are obtained from the approximation given by Siegmund [(1985), Chapter 10], which is known to be very accurate. The combined Shewhart–CUSUM entries are taken from Lucas' paper, where they were obtained by numerical calculation. For the generalized likelihood ratio procedure, the entries marked with an asterisk were obtained from (2.3) or (3.4). The unmarked entries are the result of a 2000-repetition Monte Carlo experiment. In those cells where there is no analytic approximation, formula (3.2) gives a very poor approximation, as we suspected it would.

The generalized likelihood ratio procedure is better than the standard CUSUM test for detection of large and small changes, and slightly inferior for detecting changes of the nominal size $\delta = 1$. There are only minor differences between the generalized likelihood ratio test and the combined CUSUM–Shewhart test.

Pollak and Siegmund (1985) have suggested using, as an alternative measure of the average delay until detection, the limiting value of $E_r(N - r|N > r)$ as $r \to \infty$. This is smaller than the value of $E_0(N)$, although the difference is not large. For example, the entry for the CUSUM test in Table 2 in the cells for $\mu = 1.5$, 1.0 and 0.5 would decrease to about 5.1, 9.2 and 34, respectively [Pollak and Siegmund (1985, 1986)]. For the likelihood ratio test defined by (1.3) the difference is almost completely negligible, unless $\mu$ is close to 0. Consequently, if one uses this criterion to measure expected delay, the region where the CUSUM test dominates the likelihood ratio test becomes somewhat broader. The modified criterion will not be noticeable in the behavior of the combined Shewhart–CUSUM test when $\mu$ is large; when $\mu$ is small, its effect will be much the same as with the unmodified CUSUM test.

**5. Related problems.** The problem studied in the preceding sections has been widely discussed in the literature. Compared to other stopping rules, the structure of (1.3) makes it difficult to analyse; and although it seems quite good, it is not overwhelmingly better than the competition. However, at least on a heuristic level the same method can be applied to a large number of more difficult related problems which do not appear to have received a satisfactory treatment previously. In this section we briefly discuss a number of examples.

EXAMPLE 1. Suppose the $x$'s are $p$-variate normal with a known covariance matrix, which without loss of generality can be taken to be the identity.

For detecting a change in the mean vector from 0 to an arbitrary value $\mu \neq 0$, the analogues of (1.2) and (1.3) are, respectively

$$N = \inf\left\{j\colon \max_{0 \le i < j} \delta\left[\|S_j - S_i\| - \delta(j-i)/2\right] > a\right\} \tag{5.1}$$

and

$$T = \inf\left\{j\colon \max_{0 \le i < j} \|S_j - S_i\|/(j-i)^{1/2} > b\right\}. \tag{5.2}$$

The stopping rule (5.1) does not have the relatively simple Markovian structure which permits detailed analysis of (1.2). However, one can study (5.1) or (5.2) by the method of Theorem 1. For example, as $b \to \infty$,

$$E_\infty(T) \sim \frac{\Gamma(p/2)2^{p/2}\exp(b^2/2)}{b^p \int_0^b x\nu^2(x)\,dx}.$$

EXAMPLE 2. A more interesting case occurs when there are unknown nuisance parameters, for example, if the variance $\sigma^2$ of the $x$'s or the mean value $\mu_0$ of $x_1, \dots, x_r$, or perhaps both, are unknown. The case of unknown $\mu_0$ was discussed by Pollak and Siegmund (1991), who suggested a stopping rule of Shiryayev–Roberts form [Shiryayev (1963) and Roberts (1966)], which in the case of known $\mu_0$ has been shown [Pollak and Siegmund (1985)] to behave similarly to the CUSUM test (1.2). In the formulation of Pollak and Siegmund (1991) there is a training sample of size $0 \le r_0 \le r$ which provides an initial estimator of $\mu_0$. Their method can in principle be used in the case of unknown $\sigma^2$, although the statistic is quite complicated. The analogue of (1.3) is easily obtained in all cases. For $\mu_0$ unknown and $\sigma^2 = 1$ it is

$$T_1 = \inf\left\{j\colon j > r_0,\ \max_{r_0 \le i < j} \frac{|iS_j/j - S_i|}{[i(j-i)/j]^{1/2}} > b\right\}; \tag{5.3}$$

for $\mu_0 = 0$ and $\sigma^2$ unknown it is

$$T_2 = \inf\left\{j\colon j > r_0,\ \max_{r_0 \le i < j} -j\log\left(1 - \frac{(S_j - S_i)^2}{[(j-i)\Sigma_1^j x_n^2]}\right) > b^2\right\}; \tag{5.4}$$

and for both $\mu_0$ and $\sigma^2$ unknown it involves an appropriate combination of (5.3) and (5.4). The argument of Theorem 1 suggests that in all cases as $b \to \infty$, $E_\infty(T_i) \sim E_\infty(T)$, where $T$ is defined by (1.3). To see why this should be so, consider $T_1$ defined by (5.3). We put $m_0 = cb^2$ as before and divide the range $[0, m]$ into the intervals $[km_0, (k+1)m_0)$. For different values of $k$ the random fields $\{iS_j/j - S_i,\ km_0 \le i < j < (k+1)m_0\}$ are easily seen by a computation of covariances to be stochastically independent, but not, as in Theorem 1, identically distributed. The method of Siegmund (1988) yields

$$P\left\{\max_{km_0 \le i < j < (k+1)m_0} |iS_j/j - S_i|/[i(j-i)/j]^{1/2} > b\right\}$$

$$\sim 2^{-1}b^3\phi(b)\int_{kc<s<t<(k+1)c} (t-s)^{-2}\nu\left([s/t(t-s)]^{1/2}\right) \times \nu\left([t/s(t-s)]^{1/2}\right) ds\,dt. \tag{5.5}$$

Since $k$ ranges from 0 to $m/m_0$, and $m$ is exponentially large compared to $m_0$, except for a negligibly small fraction of $k$-values close to 0 the ratio $s/t$ in (5.5) is uniformly close to 1. Hence, for essentially all $k$, the double integral in (5.5) essentially equals

$$\int_{kc<s<t<(k+1)c} (t-s)^{-2}\nu^2\left[(t-s)^{-1/2}\right] ds\,dt.$$

Substitution of this approximation and some manipulation show that the right-hand side of (5.5) asymptotically equals the right-hand side of (2.4).

One can also consider multivariate generalizations of these examples.

EXAMPLE 3. The preceding examples all exhibit invariance with respect to some group of transformations, and hence an adaptation of the Shiryayev–Roberts procedure along the lines suggested by Pollak and Siegmund (1985, 1991) is in principle possible, although the required calculations can be complicated. A problem for which that method does not seem applicable is detection of a change in the mean of a first-order autoregressive process. Suppose $x_0 = 0$ and, for $n \ge 1$, $x_n - \mu 1\{n > r\} = \rho(x_{n-1} - \mu 1\{n > r + 1\}) + \varepsilon_n$, where $\varepsilon_1, \varepsilon_2, \dots$ are independent $N(0, \sigma^2)$, $|\rho| < 1$, and for simplicity we assume $\sigma^2$ is known and equals 1.

If we momentarily also assume that $\rho$ is known, a straightforward calculation shows that the analogue of (1.3), to which it reduces when $\rho = 0$, is

$$M = \inf\left\{j: \max_{0 \le i < j} \frac{|x_{i+1} - \rho x_i + (1-\rho)\Sigma_{i+2}^{j}(x_n - \rho x_{n-1})|}{\left[1 + (j-i-1)(1-\rho)^2\right]^{1/2}} > b\right\}. \tag{5.6}$$

When $\rho$ is unknown, (5.6) should be modified by substituting for $\rho$ its maximum likelihood estimator $\hat{\rho}_{i,j}$ based on $j$ observations when it is assumed that the change occurs at $r = i$. This estimator is given implicitly in the solution of a pair of equations which also involve the corresponding maximum likelihood estimator $\hat{\mu}_{i,j}$ of $\mu$. As in Example 2, one might also add a training sample of size $r_0$, to provide a preliminary estimator of $\rho$.

It is easy to see that under $P_\infty$ the process defining stopping rule (5.6) has exactly the same distribution as does that defining (1.3) under the assumptions of Section 2. In view of the heuristic calculation given in Example 2, it seems plausible to conjecture that for large $j$ and values of $i$ in the critical range close to $j$, the value of $\rho$ can be estimated so accurately from the overwhelming number of observations prior to the $i$th that the same result holds asymptotically for the $P_\infty$ average run length when $\rho$ is unknown. However, a rigorous or even convincing heuristic analysis of this example appears to be much more complicated than the earlier ones.

It would be easy to add to this list of examples. The natural setting of Siegmund (1988) is exponential families of distributions, and it appears possible to set the theory discussed in this paper at that level of generality. Detecting a change from a known initial value in a one-parameter exponential family seems reasonably straightforward, but problems involving nuisance parameters appear to be substantially more complicated. In particular, one cannot expect the $P_\infty$ average run length to be asymptotically independent of the value of unknown nuisance parameters, but it should be roughly independent over broad ranges of values of the parameters.

Although the heuristic argument presented above suggests that in a wide range of examples, asymptotically the $P_\infty$ average run length does not depend on unknown nuisance parameters, the practical significance of this result requires two qualifications. The first is merely to emphasize that the conjectured results, if correct, are asymptotic. We have done extensive simulations of stopping rules (5.3) and (5.4). From these results it is clear that $E_\infty(T_1)$ is virtually equal to $E_\infty(T)$, regardless of the size of the training sample $r_0$, but $E_\infty(T_2)$ is sensitive to the size of the training sample. Unless $b$ is fairly large it can be somewhat smaller than $E_\infty(T)$ unless $r_0$ is about 20. For example, for $b = 3.45$ and $r_0 = 0$, 10 and 25, Monte Carlo experiments with 2000 repetitions yielded estimates for $E_\infty(T_2)$ of 345, 392 and 400, respectively. The second point to keep in mind is that, although we may appear to pay little in the $P_\infty$ average run length for lack of knowledge of these nuisance parameters, we pay heavily in the expected delay after a change. In fact, it is no longer meaningful to look only at the expected delay under the assumption that the change occurs immediately, since the more information we have to estimate the unknown nuisance parameter, the faster we can detect a change. For the case of an unknown initial value of the mean, our simulations show that stopping rule (5.3) compares favorably with the best stopping rules studied by Pollak and Siegmund (1991). In the case of unknown variance, the average time to detection after a change-point is substantially increased when $r$ is small and the change itself is large. One can see that this must be true by observing that instead of the leading term of $(b/\mu)^2$ in (3.4), if a change occurs immediately and there is no training sample to estimate $\sigma^2$, the leading term in approximating the average delay to detection is $b^2/\log(1 + \theta^2)$, where $\theta = \mu/\sigma$. A training sample of about 10 observations to provide an initial estimator of $\sigma^2$ seems to close most of the gap between the cases of known and unknown variance.

## APPENDIX

The following lemmas provide a suitable upper bound for the second term on the right-hand side of (2.6) and hence allow us to complete the proof of Theorem 1.

Let $x_1, x_2, \dots$ and $x'_1, x'_2, \dots$ be independent $N(0,1)$ random variables and let $S_i$ and $S'_j$ be their partial sums. Let $m, b \to \infty$ such that $mb\phi(b) \to \lambda \in (0, \infty)$. Let $c \gg 0$, $m_0 = cb^2$, $m_1 = c^{1/2} m_0$ and $m_2 = m_0^2$.

LEMMA 2. *There exists $\delta(c) \to 0$ as $c \to \infty$ such that, for all large $b$,*

$$P\left\{\max_{k} \max_{i,j} (S_j - S_i)/(j-i)^{1/2} > b\right\} \le \delta(c). \tag{A.1}$$

*where the maximum is taken over $0 < k \le m/m_0$, $(k-1)m_0 \le i < km_0$ and $km_0 \le j \le m$.*

PROOF. The maximum over $(k-1)m_0 \le i < km_0$, $km_0 < j \le km_0 + m_1$ can be bounded by appealing again to Siegmund (1988). For $(i,j)$ such that $(k-1)m_0 \le i < km_0$ and $km_0 + m_1 < j \le m$ we have $j - i = j - km_0 + km_0 - i$, and $S_j - S_i$ equals the sum of the independent random walks $S_j - S_{km_0}$ and $S_{km_0} - S_i$. Hence in order to complete the proof of Lemma 2 it suffices to show the following result stated as Lemma 3. □

LEMMA 3. *There exists $\delta(c) \to 0$ as $c \to \infty$ such that, for all large $b$,*

$$\left(\frac{m}{m_0}\right) P\left\{\max_{0<i\le m_0,\, m_1<j\le m} \frac{(S_i + S'_j)}{(i+j)^{1/2}} > b\right\} \le \delta(c). \tag{A.2}$$

PROOF. Let $A_1 = \{i\colon 0 < i \le m_0\}$, $A_2 = \{j\colon m_1 < j \le m_2\}$ and $A_3 = \{j\colon m_2 < j \le m\}$. Then observe that

$$\begin{aligned} &\left\{\max_{0<i\le m_0,\, m_1<j\le m} \frac{(S_i + S'_j)}{(i+j)^{1/2}} > b\right\} \\ &\quad = \left\{\max_{i\in A_1,\, j\in A_3} \frac{(S_i + S'_j)}{(i+j)^{1/2}} > b\right\} \\ &\qquad \cup \left\{\max_{i\in A_1,\, j\in A_2} \frac{(S_i + S'_j)}{(i+j)^{1/2}} > b,\ \max_{i\in A_1,\, j\in A_3} \frac{(S_i + S'_j)}{(i+j)^{1/2}} \le b\right\}. \end{aligned} \tag{A.3}$$

The proof of the lemma follows from Lemmas 4 and 5, which give bounds for the probabilities of the two events on the right-hand side of (A.3). □

LEMMA 4. *There exists a numerical constant $K$ such that, for all large $b$,*

$$P\left\{\max_{0<i\le m_0,\, m_2<j\le m} \frac{(S_i + S'_j)}{(i+j)^{1/2}} > b\right\} \le Kb\phi(b)\log(m)\exp\left(\frac{b^2 m_0}{2m_2}\right).$$

PROOF. Let $B(t)$ be standard Brownian motion. Observe that

$$P\left\{\max_{0<i\le m_0,\, m_2<j\le m}\frac{(S_i+S'_j)}{(i+j)^{1/2}}>b\right\}$$

$$\le P\left\{\max_{0<i\le m_0} S_i+S'_j>bj^{1/2}\text{ for some } m_2<j\le m\right\}$$

$$\le P\left\{\max_{0<t\le m_0} B(t)>bm_0^{1/2}\right\} \tag{A.4}$$

$$+\int_0^{bm_0^{1/2}} P\left\{\max_{0<t\le m_0} B(t)\in dy\right\}$$

$$\times P\left\{\max_{m_2<j\le m}\left(S'_j-\left(bj^{1/2}-y\right)\right)>0\right\}.$$

The first term equals $2(1-\Phi(b))$ and is clearly small enough. By a simple boundary crossing argument for standard Brownian motion [cf. Itô and McKean (1965), page 34], we can see that

$$P\left\{\max_{m_2<j\le m}\left(S'_j-\left(bj^{1/2}-y\right)\right)>0\right\}$$

$$\le\int_{m_2}^{m}\left(bt^{1/2}-y\right)t^{-3/2}\phi\left(\frac{bt^{1/2}-y}{t^{1/2}}\right)dt+2\left[1-\Phi\left(b-\frac{y}{m_2^{1/2}}\right)\right]. \tag{A.5}$$

If we substitute the right-hand side of equation (A.5) into the integral in equation (A.4), we get two terms, the larger of which is given by

$$\int_0^{b\sqrt{m_0}} P\left(\max_{0\le t\le m_0} B(t)\in dy\right)\int_{m_2}^{m}\left(bt^{1/2}-y\right)t^{-3/2}\phi\left(\left(bt^{1/2}-y\right)/t^{1/2}\right)dt$$

$$=2\int_{m_2}^{m}\int_0^{bm_0^{1/2}}\phi\left(y/m_0^{1/2}\right)\left(bt^{1/2}-y\right)t^{-3/2}\phi\left(\left(bt^{1/2}-y\right)/t^{1/2}\right)dy\,dt/m_0^{1/2}$$

$$\le 2\int_{m_2}^{m}(b/t)\int_{-\infty}^{\infty}\phi\left(\left(bt^{1/2}-y\right)/t^{1/2}\right)\phi\left(y/m_0^{1/2}\right)dy\,dt/m_0^{1/2}$$

$$\le 2b\int_{m_2}^{m}t^{-1/2}(m_0+t)^{-1/2}\phi\left(bt^{1/2}/(m_0+t)^{1/2}\right)dt$$

$$\le 2b\phi(b)\exp\left(\left(b^2m_0/2m_2\right)\right)\log(m/m_2).$$ □

LEMMA 5. *The order of magnitude of the probability of the second set on the right-hand side of equation* (A.3) *is bounded as follows*:

$$P\left\{\max_{0<i\le m_0,\, m_1<j\le m_2}\frac{(S_i+S'_2)}{(i+j)^{1/2}}>b,\ \max_{0<i\le m_0,\, m_2<j\le m}\frac{(S_i+S'_j)}{(i+j)^{1/2}}\le b\right\}$$

$$=O\left(m_0^{1/2}b^2\phi(b)\right)+O\left(b\phi(b)m_0^{1/2}\log m_2\right),$$

*where $\phi$ is the standard normal density.*

PROOF. Observe that $(i+j+k+l)^{1/2} - (i+j)^{1/2} \le (k+l)/2(i+j)^{1/2}$. Let $Z_{ij} = S_i + S'_j$. Let $A_1$, $A_2$ and $A_3$ be as in Lemma 3. Also let

$$A_4 = \{(i',j'): i < i' \le m_0 \text{ and } j' = j, \text{or } 0 < i' \le m_0 \text{ and } j < j' \le m\}$$

and

$$A_5 = \{(k,l): k + l \ge 1, 0 \le k \le m_0 - i, 0 \le l \le m - j\}.$$

It will also be convenient to let $W_n$, $n = 1, 2, \dots,$ denote a random walk with $N(\mu, 1)$ increments and write $P_\mu$ when we are concerned with this auxiliary random walk, to denote the dependence of probabilities on $\mu$. For $-\infty < x < \infty$, let $\tau(x) = \inf\{n\colon n \ge 1, W_n > x\}$. This notation will also be in force in Lemmas 6–8, which follow.

We have

$$\begin{aligned}
&P\left\{\max_{i\in A_1, j\in A_2} \frac{Z_{ij}}{(i+j)^{1/2}} > b, \max_{i\in A_1, j\in A_3} \frac{Z_{ij}}{(i+j)^{1/2}} \le b\right\} \\
&\quad= \sum_{i\in A_1, j\in A_2} \int_0^\infty \frac{\phi\big(b + x/(i+j)^{1/2}\big)}{(i+j)^{1/2}} \\
&\qquad\times P\left\{\max_{(i',j')\in A_4} \frac{Z_{i'j'}}{(i'+j')^{1/2}} \le b \,\middle|\, Z_{ij} = b(i+j)^{1/2} + x\right\} dx \\
&\quad\le \sum_{i\in A_1, j\in A_2} \int_0^\infty \frac{\phi\big(b + x/(i+j)^{1/2}\big)}{(i+j)^{1/2}} \\
&\qquad\times P\left\{\max_{(k,l)\in A_5} \frac{Z_{i+k,j+l}}{(i+j+k+l)^{1/2}} \le b \,\middle|\, Z_{ij} = b(i+j)^{1/2} + x\right\} dx \\
&\quad= \sum_{i\in A_1, j\in A_2} \int_0^\infty \frac{\phi\big(b + x/(i+j)^{1/2}\big)}{(i+j)^{1/2}} \\
&\qquad\times P\Big\{Z_{i+k,j+l} - Z_{ij} \le b\big((i+j+k+l)^{1/2} - (i+j)^{1/2}\big) \\
&\qquad\qquad -x, \forall (k,l) \in A_5\Big\}\, dx \\
&\quad\le \sum_{i\in A_1, j\in A_2} \int_0^\infty \frac{\phi\big(b + x/(i+j)^{1/2}\big)}{(i+j)^{1/2}} \\
&\qquad\times P_\mu\Big\{\max_{1\le k\le m_0 - i} W_k \le -x\Big\} P_\mu\Big\{\max_{1\le l\le m-j} W_l \le -x\Big\}\, dx \\
&\quad\le \phi(b) \sum_{i\in A_1, j\in A_2} \frac{P_\mu\{\max_{1\le k\le m_0-i} W_k \le 0\}}{(i+j)^{1/2}} \\
&\qquad\times \int_0^\infty \exp\left(-x\left(\frac{b}{(i+j)^{1/2}}\right)\right) P_\mu\Big\{\max_{1\le l\le m-j} W_l < -x\Big\}\, dx,
\end{aligned}$$

where $\mu = -b/2(i+j)^{1/2}$.

For large $b$, by Lemmas 6–8, this last expression is

$$\le \phi(b) \sum_{i,j} (i+j)^{-1/2} P_\mu\Big( \max_{1 \le k \le m_0 - i} W_k < 0 \Big) \frac{b}{(i+j)^{1/2}}$$

$$\le b\phi(b) \bigg[ O\big(bm_0^{1/2}\big) + \sum_{i,j} j^{-1} O\big((m_0 - i)^{-1/2}\big) \bigg]$$

$$= O\big(b^2 m_0^{1/2} \phi(b)\big) + O\big(bm_0^{1/2} \log m_2 \phi(b)\big). \qquad \square$$

LEMMA 6. *$P_0(\tau_0 > n) \le n^{-1/2}$ for large $n$.*

PROOF. We can get this by Theorem 1a of Feller [(1972), Chapter 12, Section 7], which states

$$P_0(\tau_0 > n) \sim (\pi n)^{-1/2}. \qquad \square$$

LEMMA 7. *There exists $k > 0$ such that, for any $\mu > 0$,*

$$P_{-\mu}\{\max_{1 \le i \le n} W_i < 0\} \le 2^{1/2}\mu + kn^{-1/2}$$

*for all $n$.*

PROOF. Observe that

$$P_{-\mu}\Big\{ \max_{1 \le i \le n} W_i < 0 \Big\} = P_{-\mu}(\tau_0 = \infty) + P_{-\mu}(n < \tau_0 < \infty).$$

From Siegmund [(1985), page 175–176] we see that $P_{-\mu}(\tau_0 = \infty) = 2^{1/2}\mu[\nu(2\mu)]^{1/2} \le 2^{1/2}\mu$. For the other probability we will use the likelihood ratio $dP_{-\mu}/dP_0$ to obtain

$$P_{-\mu}(n < \tau_0 < \infty) = \int_{(n < \tau_0 < \infty)} \exp\big(-\mu W_{\tau_0} - \mu^2 \tau_0/2\big)\, dP_0$$

$$\le P_0(n < \tau_0 < \infty).$$

Now Lemma 6 yields the result. $\square$

LEMMA 8. *Let $\mu = b/2(i+j)^{1/2}$. For all large $b$,*

$$\int_0^\infty \exp(-2\mu x) P_{-\mu}\Big\{ \max_{1 \le n \le m - m_2} W_n \le -x \Big\}\, dx \le 2\mu,$$

*where $0 < i \le m_0$, $m_1 \le j \le m_2$ and $m_0$, $m_1$, $m_2$ and $m$ are as defined before.*

PROOF. Observe that

$$P_{-\mu}\Big\{ \max_{1 \le k \le m - m_2} W_k \le -x \Big\} = P_{-\mu}\{\tau_{-x} = \infty\} + P_{-\mu}\{m - m_2 < \tau_{-x} < \infty\}.$$

By Lemma 19 in Siegmund (1992), we know

$$\int_0^\infty \exp(-2\mu x) P_{-\mu}\{\tau_{-x} = \infty\}\, dx = \mu\nu(2\mu) \le \mu.$$

Hence it is enough to prove

$$\mu^{-1} \int_0^\infty \exp(-2\mu x) P_{-\mu}\{m - m_2 < \tau_{-x} < \infty\}\, dx \le 1$$

for all large $b$. We will show this using the likelihood ratio $dP_{-\mu}/dP_0$ as follows:

$$\begin{aligned}
&\mu^{-1} \int_0^\infty \exp(-2\mu x) P_{-\mu}\{m - m_2 < \tau_{-x} < \infty\}\, dx \\
&\quad = \mu^{-1} \int_0^\infty \exp(-2\mu x) \int_{(m - m_2 < \tau_{-x} < \infty)} \exp\left(-\mu W_{\tau_{-x}} - \mu^2 \tau_{-x}/2\right) dP_0\, dx \\
&\quad \le \mu^{-1} \int_0^\infty \exp(-2\mu x) \int_{(m - m_2 < \tau_{-x} < \infty)} \exp\left(\mu x - \mu^2 (m - m_2)/2\right) dP_0\, dx \\
&\quad \le \mu^{-1} \exp\left(-\mu^2 (m - m_2)/2\right) \int_0^\infty \exp(-\mu x) P_0\{m - m_2 < \tau_{-x} < \infty\}\, dx \\
&\quad \le \mu^{-2} \exp\left(-\mu^2 (m - m_2)/2\right) \to 0
\end{aligned}$$

as $b \to \infty$ uniformly in $(i, j)$. □

**Acknowledgment.** The authors would like to thank the referee for his careful reading of the manuscript.

## REFERENCES

ALDOUS, D. (1989). *Applications of the Poisson Clumping Heuristic*. Springer, New York.

BARNARD, G. A. (1959). Control charts and stochastic processes. *J. Roy. Statist. Soc. Ser. B* **21** 239–257.

BASSEVILLE, M. (1988). Detecting changes in signals and systems—a survey. *Automatica* **24** 309–326.

BROOK, D. and EVANS, D. A. (1972). An approach to the probability distribution of CUSUM run lengths. *Biometrika* **59** 539–549.

DARLING, D. A. and ERDÖS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

FELLER, W. (1972). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

ITÔ, K. and MCKEAN, H. P., JR. (1965). *Diffusion Processes and Their Sample Paths*. Springer, New York.

LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7** 60–76.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

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

LUCAS, J. M. (1982). Combined Shewhart–CUSUM quality control schemes. *Journal of Quality Technology* **14** 51–59.

PAGE, E. A. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

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

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. (1986). Approximations to the average run length of CUSUM tests. Technical report, Stanford Univ.

POLLAK, M. and SIEGMUND, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist.* **19** 394–416.

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. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. in Appl. Probab.* **11** 701–719.

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

SIEGMUND, D. (1988). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16** 487–501.

SIEGMUND, D. (1992). Tail approximations for maxima of random fields. In *Proceedings of the Singapore Probability Conference* (L. H. Y. Chen, K. P. Choi, K. Hu and J. Lou, eds.) 147–158. de Gruyter, Berlin.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

BIOSTATISTICS SERVICE
MEMORIAL SLOAN-KETTERING
CANCER CENTER
NEW YORK, NY 10021