MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TESTS FOR A CHANGE-POINT

by

Barry James
and
Kang Ling James
IMPA, Rio de Janeiro

David Siegmund
Stanford University

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

AD-A159 381

DTIC FILE COPY

DTIC
ELECTE
SEP 2 4 1985
S D

85 09 23 175

TESTS FOR A CHANGE-POINT

by

Barry James
and
Kang Ling James
IMPA, Rio de Janeiro

David Siegmund
Stanford University

$\sqrt{\dagger}$

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DTIC
COPY
INSPECTED

TESTS FOR A CHANGE-POINT

Barry James, Kang Ling James, and David Siegmund

The problem considered is that of testing a sequence of independent normal variables with constant, known or unknown, variance for no change in mean versus alternatives with a single change-point. Various tests, such as those based on the likelihood ratio and recursive residuals, are studied. Power approximations are developed by integrating approximations for conditional boundary crossing probabilities. A comparison of several tests is made, and the power approximations obtained are compared with Monte Carlo values.

Some key words: Change-point; Likelihood ratio test; Recursive residuals; Power approximations; Boundary crossing probabilities.

TESTS FOR A CHANGE-POINT

Barry James, Kang Ling James, and David Siegmund

1. Introduction and Summary.

Let $z_1, z_1, \cdots, z_m$ be independent random variables. The purpose of this paper is to discuss tests of the hypothesis that the z's are identically distributed against the alternative that for some value j, $1 \leq j < m$, $z_1, \cdots, z_j$ are identically distributed and $z_{j+1}, \cdots, z_m$ are also identically distributed, but with a distribution different from that of z_1. For the most part we consider only the special case where the z_n are normally distributed with mean μ_n and variance 1. Then the hypotheses can be described more formally as

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m,$$

$$H_1 : \text{For some } j, 1 \leq j < m, \ \mu_1 = \cdots = \mu_j \neq \mu_{j+1} = \cdots = \mu_m. \tag{1}$$

One important goal in studying this very simple problem in detail is to gain insight into more complicated models and related problems. Thus, although we do not consider regression models explicitly, we attempt to keep this generalization in mind and comment on it when appropriate. Similarly, we do not study the related problems of estimating j and/or the magnitude of the change, but when other things are equal we prefer test statistics which seem to be useful for estimation as well. For discussions of confidence sets for the change-point j see Cobb (1978) and Siegmund (1986).

The sampling distributions of most of the statistics described below are quite complicated, and other authors have often studied these problems by numerical or Monte Carlo methods (e.g. Sen and Srivastava, 1975; Hawkins, 1977; Worsley, 1983). Using methods developed to solve boundary crossing problems in sequential analysis, we give analytic approximations to the sampling distributions of various test statistics. This facilitates our

comparisons of different tests and allows one to make informal use of the procedures without any programming effort. The process of obtaining the approximations also yields some qualitative insights into the tests themselves.

The paper is organized as follows. Section 2 introduces several test statistics and describes their behavior qualitatively. The most important are (1) the likelihood ratio statistic, (2) an *ad hoc* statistic proposed by Pettitt (1980), which can be interpreted as a kind of score statistic, (3) the recursive residuals statistic of Brown, Durbin, and Evans (1975), and (4) the quasi Bayes statistic of Chernoff and Zacks (1964). Approximations for the significance levels of these tests are given in Section 3 and approximations for the power in Section 4. Section 5 contains numerical examples. Section 6 contains miscellaneous remarks and speculations concerning extensions of our results to non-normal data, survival analysis and regression problems. An appendix gives some mathematical results. A more detailed presentation of the underlying mathematics will be given in a future paper.

The reader more interested in our conclusions than in the theoretical development may wish to turn directly from Section 2 to Sections 5 and 6, which can be read independently of Sections 3 and 4.

2. Test Statistics and Qualitative Behavior.

Let $x_1, x_2, \cdots, x_m$ be independent random variables, and assume that x_n is normally distributed with mean μ_n and variance 1 $(n = 1, 2, \cdots, m)$. Let $S_n = x_1 + \cdots + x_n$. The square root of the log likelihood ratio statistic for testing the hypotheses (1) is easily calculated to be

$$\max_{1 \leq k < m} \{|kS_m/m - S_k|/[k(1 - k/m)]^{1/2}\}. \tag{2}$$

Without the max, (2) is the normalized difference between the mean of the first k observations and the overall mean, i.e. it is the standard two sample test statistic for testing

2

that the means of the first k observations and the last $m - k$ are equal. The max searches for the most plausible place to separate the sample into two subsamples having different means. A derivation of this statistic which suggets some alternative possibilities goes as follows. The testing problem (1) is invariant under common shifts in location of all the observations, so one might restrict consideration to invariant procedure, i.e., those which depend not directly on the x's but only on the differences $y_n = x_n - x_1$, $n = 2, \cdots, m$ (cf. Lehmann, 1959, p. 216). For given values of j and $\delta = \mu_m - \mu_1$, the likelihood ratio of the y's under H_1 to the y's under H_0 is easily calculated to be

$$\exp\{\delta[jS_m/m - S_j] - j(1 - j/m)\delta^2/2\}. \tag{3}$$

Maximizing (3) first over δ and then over j yields (2).

In what follows it wll be more convenient to consider one sided alternatives, for which we assume that the sign of δ is known, say $\delta > 0$. Then the likelihood ratio statistic is (2) without the absolute values. For reasons given below we consider the generalization

$$\max_{m_0 \leq k \leq m_1} [\{kS_m/m - S_k\}/[k(1 - k/m)]^{1/2}], \tag{4}$$

where $1 \leq m_0 < m_1 < m$.

By differentiating the logarithm of (3), setting $\delta = 0$ and then maximizing over j, we obtain a score-like statistic suggested by Pettitt (1980),

$$\max_{1 \leq k \leq m} (kS_m/m - S_k). \tag{5}$$

Still another possibility is to take the log of (3) for some arbitrary value δ_0, which might be interpreted as a typical change or the minimal change one is interested in detecting, and maximize over j. This yields the statistic

$$\max_k \{kS_m/m - S_k - k(1 - k/m)\delta_0/2\}. \tag{6}$$

3

An interesting *ad hoc* test statistic is the so-called recursive residual statistic proposed by Brown, Durbin, and Evans (1975). We consider the standardized residual of x_{n+1} from the mean of $x_1, \cdots, x_n$, to wit

$$z_n = [n/(n+1)]^{1/2}(x_{n+1} - \bar{x}_n), \quad n = 1, 2, \cdots, m-1. \tag{7}$$

We form the cumulative sum $\tilde{S}_n = z_1 + \cdots + z_n$, and we use as a test statistic

$$\max_{m_0 \leq n < m} (\tilde{S}_n/n^{1/2}). \tag{8}$$

Actually there is considerable arbitrariness in this definition. One might equally well cumulate sums "from the right" to obtain

$$\max_{m_0 < k \leq m} (\tilde{S}_{m-1} - \tilde{S}_{m-k})/(k-1)^{1/2}, \tag{9}$$

and one might consider either (7) or (8) with z_n defined as the residual of x_n from the mean of $x_{n+1}, \cdots, x_m$. We shall argue below that (9) is typically preferable to (8), but in general there seems to be no preferred way to define z_n.

It is easy to see that under H_0 the recursive residuals $z_1, \cdots, z_{m-1}$ are independent standard normal random variables. The appeal of the recursive residual concept is that this property persists under general regression models (Brown, Durbin, and Evans, 1975). By way of contrast, in a regression context the null hypothesis distribution of the likelihood ratio statistic depends on the spacings between the independent variables.

Chernoff and Zacks (1964) assume that j has a uniform prior distribution over $\{1, 2, \cdots, m\}$ and that δ is close to zero. An expansion for small δ gives the quasi Bayesian test statistic

$$C = \sum_{n=1}^{m-1} [n(n+1)]^{1/2} z_n, \tag{10}$$

where z_n is the recursive residual defined in (7). Gardner (1969) follows the Chernoff-Zacks prescription to obtain a test statistic for two sided alternatives, which turns out to be quite

4

different from (10) in appearance. A disadvantage of Gardner's statistic is that in general it gives no idea whether δ is positive or negative. For this reason one might prefer to use $|C|$ for a two sided test. For simplicity we consider only the statistic (10) for a one sided alternative. Unlike the other statistics suggested above, the sampling distribution of (10) is normal under both the null and alternative hypotheses. Other things being equal, this would be a point in favor of (10). We shall see that other things are not equal.

Under the alternative hypothesis, we have the following relations:

$$E(kS_m/m - S_k) = \begin{cases} k(1 - j/m)\delta & \text{for } k \le j \\ (j/m)(m - k)\delta & \text{for } k \ge j; \end{cases} \tag{11}$$

$$E(z_k) = \begin{cases} 0 & \text{for } k < j \\ j\delta/\{k(k+1)\}^{1/2} & \text{for } k \ge j, \end{cases} \tag{12}$$

so

$$E(\tilde{S}_k) = \begin{cases} 0 & \text{for } k < j \\ j\delta \sum_{n=j}^{k} 1/\{n(n+1)\}^{1/2} & \text{for } k \ge j, \end{cases} \tag{13}$$

and

$$E(C) = j(m - j)\delta. \tag{14}$$

Some qualitative insights follow from (11)-(13). If we represent the rejection region of (2) (without the absolute value signs) as indicated in Figure 1, it seems intuitively clear that the primary contribution to the power of that test comes from the probability that the process $kS_m/m - S_k$ exceeds $b\{k(1 - k/m)\}^{1/2}$ for some k in a neighborhood of $k = j$. If we superimpose the rejection region of (5), i.e. $\{\max_k(kS_m/m - S_k) > b_1\}$, on the same picture, in order that the two tests have the same significance level, b_1 must be less than $b\{k(1 - k/m)\}^{1/2}$ in a neighborhood of $k = m/2$. Hence we expect that (5) has greater power than (2) when j is about $m/2$ and the line of drift is more likely to carry the process above the constant b_1 than above $b\{k(1 - k/m)\}^{1/2}$. The converse is true when j is near 0 or m, and the curve $b\{k(1 - k/m)\}^{1/2}$ near the change-point lies below b_1. Introduction

5

of m_0 and m_1 in (4) gives the statistician the flexibility to trade some decrease of power to detect changes occurring near $j = 0$ and $j = m$ for an increase in power to detect changes occurring near $j = m/2$.

Similar reasoning suggests that (6), like (4), is a compromise between (2) and (5); and preliminary calculations suggest this is indeed the case. Since (6) seems less easily adapted to multiparameter problems, we shall not discuss it in this paper.

One can make a similar crude comparison of (8) and (9). Now, in effect the rejection regions are the same, but the processes are different (cf. Figure 2). It is easy to see that if $j = mt^*$ for some fixed $0 < t^* < 1$ and m is large, then

$$\mathrm{pr}[\tilde{S}_{m-1} - \tilde{S}_{mt^*} \geq b_2\{m(1 - t^*)\}^{1/2}] = \max_n \mathrm{pr}(\tilde{S}_{m-1} - \tilde{S}_{m-1-n} \geq b_2 n^{1/2})$$

$$> \max_n \mathrm{pr}(\tilde{S}_n \geq b_2 n^{1/2}),$$

provided $\{t^*/(1 - t^*)\}^{1/2}\log(1/t^*) > 2e^{-1}$. To a considerable extent the power of (8) and (9) is determined by the maximum marginal probabilities, and hence it appears that (9) is usually preferable to (8). Although some additional investigation of (8) may be warranted, we do no pursue it here.

Remark. Although Brown, Durbin, and Evans (1975) and Sen in several papers (e.g. Sen, 1982) consider recursive residuals as in (8), often normalized by $a + cn$ rather than $n^{1/2}$, Cox in the discussion to Brown, Durbin, and Evans (1975) implicitly proposes (9).

It is worth noting that (11) and (4) or (5) suggest simple estimates of j and δ. For example, (4) suggests estimating j by the value $\hat{j}$ which yields the maximum, and then (11) suggests estimating δ by

$$(\hat{j}S_m/m - S_{\hat{j}})/[\hat{j}(1 - \hat{j}/m)].$$

Similarly, a comparison of $\tilde{S}_{m-1} - \tilde{S}_{m-k}$, $k = 1, 2, \cdots, m$, with the corresponding expected values (cf. Figure 2) gives some idea of the values of j and δ, albeit less well defined than
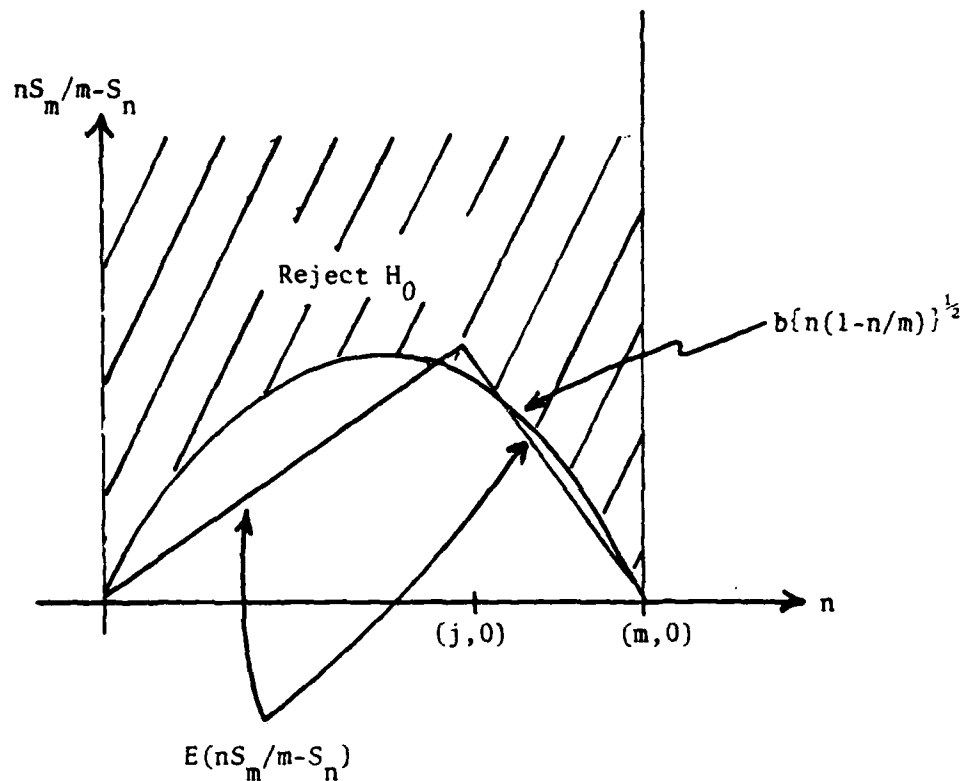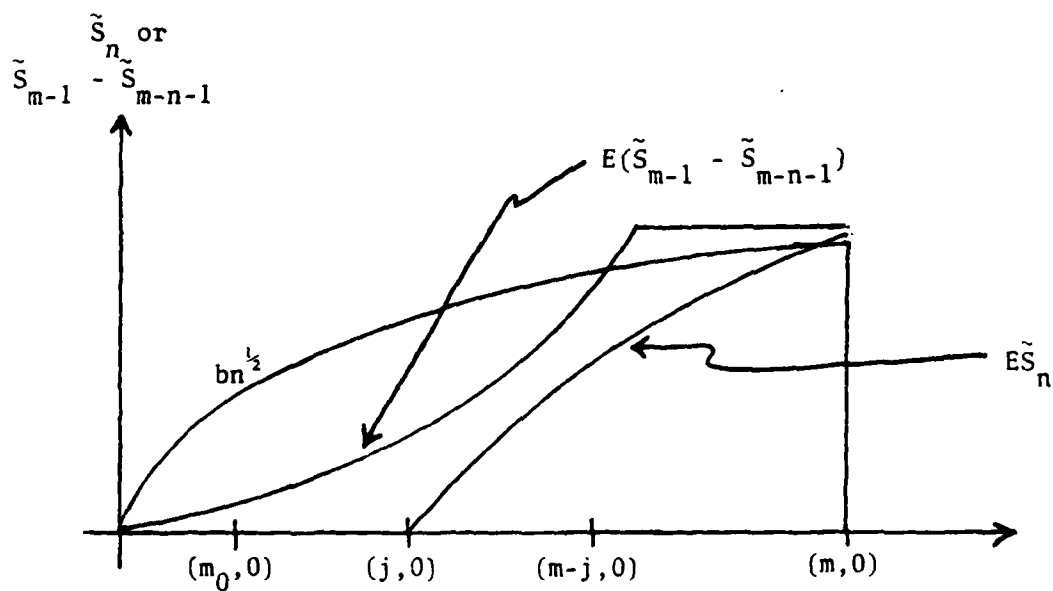
6

FIGURE 1

FIGURE 2

for (4) or (5). Although one can compute numerically a formal Bayes estimate, there do not appear to be natural estimators associated with (10).

For future reference we also record the log likelihood ratio statistic for testing (1) when the variance of the x's is an unknown, but unchanging, constant σ^2. The statistic is

$$\max_{1 \le k < m} \left\{ -\frac{1}{2} m \log \left[1 - \frac{(S_k - kS_m/m)^2}{k(1 - k/m) \sum_1^m (x_i - \bar{x}_m)^2} \right] \right\}, \qquad (15)$$

where $\bar{x}_m = S_m/m$.

Approximations to the significance level of (4), (5), (9), and (15) are given in the following section.

3. Approximations to Significance Levels.

In this section we give approximations to the right hand tail of the distributions under H_0 of (4), (5), and (9), or equivalently, (8). These approximations have been developed in the context of sequential analysis. For example, the null hypothesis distribution of (9) yields the significance level of a so-called repeated significance test, first studied by Armitage, McPherson, and Rowe (1969) by numerical methods. For derivation of the approximations and documentation of their accuracy for very small samples, e.g. $m = 5$, see Siegmund (1985, 1986).

We also give approximations for the significance level of (15) and appropriate modifications of (5) and (9) for the case of an unknown variance. The derivation of these approximations requires some new techniques, which are described in the simplest context in an appendix and will be given in more detail in a future publication.

Several authors have noted that the likelihood ratio statistic in various change-point problems, not restricted to the normal case considered here, has a large sample approximation under the null hypothesis of no change, which corresponds to (4) with $\{S_k, \ k =$

8

$0, 1, \cdots, m\}$ replaced by standard Brownian motion $\{W(t),\ 0 \le t \le m\}$. See, for example, Kendall and Kendall (1980) and Matthews, Farewell, and Pyke (1985). Although this approximation is often fairly crude, its generality makes it useful. A simple and accurate approximation to the Brownian motion probability is given below. With a view towards more general problems, it is given for the d-dimensional case $(d \ge 1)$.

In order to state approximations to the significance levels of (4), (5), and (9) it is helpful to introduce the function

$$\nu(x) = 2x^{-2} \exp\left\{ -2\sum_1^\infty n^{-1}\Phi\left(-\frac{1}{2}xn^{1/2}\right)\right\} \qquad (x > 0), \qquad (16)$$

where Φ denotes the standard normal distribution function. The function ν is easily evaluated numerically; or alternatively in the range $0 \le x \le 2$ one can use the local expansion

$$\nu(x) = \exp(-\rho x) + o(x^2) \qquad (x \to 0), \qquad (17)$$

where ρ is a numerical constant which approximately equals .583. See Siegmund (1985, Chapter X).

Let $y_1, y_2, \cdots, y_m$ be independent standard normal random variables and $S_n = y_1 + \cdots + y_n$ $(n = 1, 2, \cdots, m)$. Then for $1 \le m_0 < m_1 < m$ and $b > 0$

$$\text{pr}\left(\max_{m_0 \le n \le m_1} [(nS_m/m - S_n)/\{n(1 - n/m)\}^{1/2}] \ge b\right)$$
$$\cong 1 - \Phi(b) + b\varphi(b) \int_{b(m_1^{-1}-m^{-1})^{1/2}}^{b(m_0^{-1}-m^{-1})^{1/2}} x^{-1}\nu(x + b^2/mx)dx, \qquad (18)$$

and

$$\text{pr}(\max_{m_0 \le n \le m} n^{-1/2}S_n \ge b) \cong 1 - \Phi(b) + b\varphi(b) \int_{bm^{-1/2}}^{bm_0^{-1/2}} x^{-1}\nu(x)dx, \qquad (19)$$

where Φ is the standard normal distribution, $\varphi = \Phi'$, and ν is given by (16) or approximately by (17). Also

$$\text{pr}\{\max_{1 \le n \le m} (nS_m/m - S_n) \ge b\} \cong \exp[-2m^{-1}(b + \rho)^2]. \qquad (20)$$

where $\rho \cong .583$, as above.

The approximations (18)–(20) are respectively (11.33), (4.40), and (10.43) of Siegmund (1985). They can be shown numerically to provide excellent approximations by comparing them to exact numerical computations of Pocock (1977) and Worsley (1983). Some comparisons are given by Siegmund (1985, 1986). Siegmund (1985, p. 83) provides a table for evaluating the integral in (19).

In the case of an unknown and constant variance, if the statistics (4), (5), and (9) are Studentized in the "obvious" way, we obtain the following approximations as analogues of (18)–(20) respectively. Let $\gamma = b/m^{1/2}$, and assume that $0 < \gamma < 1$. Then

$$
\mathrm{pr}\left(\max_{m_0 \leq n \leq m_1} \left[(nS_m/m - S_n)/\{n(1 - n/m)\}^{1/2} \left\{ m^{-1} \sum_1^m (y_n - \bar{y}_m)^2 \right\}^{1/2} \right] \geq b \right)
$$
$$
\cong (m/2\pi)^{1/2} \int_\gamma^1 (1 - x^2)^{(m-4)/2} dx
$$
$$
+ (2\pi)^{-1/2} b(1 - b^2/m)^{(m-4)/2} \int_{b\{(m_1^{-1} - m^{-1})/(1-\gamma^2)\}^{1/2}}^{b\{(m_0^{-1} - m^{-1})/(1-\gamma^2)\}^{1/2}} x^{-1} \nu[x + b^2/\{m(1 - \gamma^2)x\}] dx
$$
$$
\tag{21}
$$

and

$$
\mathrm{pr}\left\{ \max_{m_0 \leq n \leq m} S_n / \left(nm^{-1} \sum_1^m y_i^2 \right)^{1/2} \geq b \right\} \cong (m/2\pi)^{1/2} \int_\gamma^1 (1 - x^2)^{(m-3)/2} dx
$$
$$
+ (2\pi)^{-1/2} b(1 - b^2/m)^{(m-3)/2} \int_{b/\{m(1-\gamma^2)\}^{1/2}}^{b/\{m_0(1-\gamma^2)\}^{1/2}} x^{-1} \nu(x) dx.
$$
$$
\tag{22}
$$

The first integrals in (21) and (22) can themselves be approximated by virtue of the expansion

$$
m^{1/2} \int_\gamma^1 (1 - x^2)^{(m-i)/2} dx = b^{-1}(1 - b^2/m)^{(m-i)/2}\{1 + 2m^{-1} - b^{-2} + o(m^{-1}\}, \tag{23}
$$

which is valid as $b, m \to \infty$ with $b/m^{1/2} = \gamma$ fixed. Now let $\gamma = b/m$ and assume that

$0 < \gamma < 1/2$. Then

$$\text{pr}\left[\max_{0 \le k \le m}(kS_m/m - S_k)/\left\{m^{-1}\sum_1^m(y_i^2 - \bar{y}_m)^2\right\}^{1/2} \ge b\right] \tag{24}$$
$$\cong \nu\{4\gamma/(1 - 4\gamma^2)^{1/2}\}(1 - 4b^2/m^2)^{(m-3)/2}.$$

The approximations (21), (22), and (24) are written in a way to facilitate comparison with the corresponding results, (18)–(20), for known variance. It appears that there is little difference between the two cases. Some numerical results given in the appendix indicate that this is in fact the case except when the probability or the sample size m is quite small. The appendix also contains an informal proof of (24). The more complicated (21) and (22) will be discussed in a future paper.

Remarks. (i) It is easy to see that the probability that the likelihood ratio statistic (15) exceeds a is given approximately by twice (21) with $b = [m\{1 - \exp(-2a/m)\}]^{1/2}$. (ii) At first glance (22) may appear to be an incorrect Studentization of (9). But note that the y's in (22) play the role of the z's in (9); and if z_n is defined by (7), it is easy to use the Helmert orthogonal transformation to show that $\sum_1^{m-1} z_n^2 = \sum_1^m(x_i - \bar{x}_m)^2$. (iii) The approximation (22) can be used for the general regression model of Brown, Durbin, and Evans (1975), but the distribution of the likelihood ratio statistic is quite model dependent.

An easy application of the theory of weak convergence of stochastic processes, e.g. Billingsley (1968), shows that the probabilities discussed above are given approximately by the corresponding probabilities defined in terms of a Brownian motion process $W(t)$, $0 \le t < \infty$. For example, the left hand side of (18) or (21) is approximately

$$\text{pr}[\max_{t_0 \le t \le t_1} W_0(t)/\{t(1 - t)\}^{1/2} \ge b], \tag{25}$$

where $W_0(t) = W(t) - tW(1)$ is a Brownian bridge process on $[0, 1]$ and $t_i = m_i/m$ $(i = 0, 1)$. The advantage of (25) as an approximation is its generality. It would serve also if the

underlying distribution of the observations were not normal, hence for testing the hypothesis of no change in quite general models. This same generality is its disadvantage, because it means that the approximation is often a crude one.

Several authors, e.g. Mandl (1962), Keilson and Ross (1975), and DeLong (1981) have described numerical methods for evaluating (25) and have published numerical tables. We give here an approximation to (25), which is easily evaluated and which is valid in an arbitrary number of dimensions. A proof can be given along the lines of Siegmund's (1985, Theorem 11.1) argument for the one dimensional case.

Let $W_0(t)$, $0 \leq t \leq 1$, be a d-dimensional Brownian bridge process, and let $\| \cdot \|$ denote the d-dimensional Euclidean norm. Let $0 < t_0 < t_1 < 1$ and set $r = t_1(1 - t_0)/t_0(1 - t_1)$. Then as $b \to \infty$

$$
\begin{aligned}
\mathrm{pr}[\max_{t_0 \leq t \leq t_1} &\|W_0(t)\|/\{t(1 - t)\}^{1/2} \geq b] \\
&= \frac{b^d \exp(-b^2/2)}{2^{(d-2)/2}\Gamma(d/2)} \{\frac{1}{2}(1 - d/b^2) \log r + 2b^{-2} + o(b^{-2})\},
\end{aligned}
\tag{26}
$$

where $\Gamma(\cdot)$ is the gamma function.

Remark. It is well known and easily verified that $W(t) = (1 + t)W_0\{t/(1 + t)\}$, $0 \leq t < \infty$, is a standard d-dimensional Brownian motion process, and hence (26) also gives approximations to the probabilities appearing in (19) and (21).

The accuracy of (26) is easily ascertained by comparing it with extensive tables of DeLong (1981) for $d = 1, 2, 3, 4$. For example, for $d = 4$, $r = 50$, and $b \approx 3.85, 4.10$, and 4.58, DeLong gives for the probability in (16) the respective values .1, .05, .01. The right hand side of (26) yields .104, .051, and .0103. In fact, the approximation (26) is moderately good even when the probability is not close to zero, although there is no apparent mathematical reason why this should be the case.

As an approximation to the probability in (18), (26), or more precisely (26) multiplied

12

by 1/2, is much less satisfactory. The numerical example discussed extensively in Section 5 has $b = 2.82$, $m_0 = 5$, $m_1 = 35$, and $m = 40$, so $r = 49$. The approximation (18) yields .025, whereas (26) gives .041. A 2500 repetition Monte Carlo experiment using importance sampling along the lines indicated in Remark 4.45 of Siegmund (1985) yielded $.0239 \pm .0005$.

Remark. Kiefer (1959) has computed $\text{pr}\{\max_{0 \leq t \leq 1} \|W_0(t)\| \geq b\}$ exactly in terms of an infinite series of Bessel functions, and has given tables for dimensions 2, 3, and 4. An approximation similar to (26) but requiring a somewhat different argument is given in Problem 11.1 of Siegmund (1985).

4. Power.

In this section we adapt the methods of Siegmund (1977, 1978, 1985) to obtain approximations to the power of (4), (5), and (9). The basic ideas, which go back to Anscombe (1952), are much simpler than in the preceding section and quite general. They are sketched below without details. We first consider (4) and (5), and later indicate the changes appropriate to handle (9).

The following result is related to Cramér's approximation for the probability of ruin of a risk process. See Feller (1972, Chapter XII) and Siegmund (1985, Chapter VIII).

Proposition 1. Let $\mu > 0$ and assume that $y_1, y_2, \cdots$ are independent $N(-\mu, 1)$. Then as $x \to \infty$

$$\text{pr}\left(\sum_1^n y_k \geq x \text{ for some } n \geq 1 \right) \sim \nu(2\mu) \exp(-2\mu x),$$

where ν is defined by (16) and given approximately in (17).

For the rest of this section $z_1, \cdots, z_j$ are independent $N(\mu_1, 1)$, $z_{j+1}, \cdots, z_m$ are independent $N(\mu_m, 1)$, $\delta = \mu_m - \mu_1$, $S_n = z_1 + \cdots + z_n$, and $S_n^* = nS_m/m - S_n$, $n = 0, 1, \cdots, m$.

The process S_n^*, $n = 0, 1, \cdots, m$ has the mean value (11) and the covariance function

$$\text{cov}(S_k^*, S_n^*) = k(1 - n/m) \qquad (k \leq n), \qquad (27)$$

of a discrete time Brownian bridge, tied down to equal 0 at $n = 0$ and $n = m$. Let $c(t)$, $0 \leq t \leq 1$, be a function, and for $1 \leq m_0 < m_1 < m$ let $T_0 = \inf\{n : n \geq m_0, \ S_n^* \geq mc(n/m)\}$. The power of the tests defined by (4) and (5) is of the form

$$\text{pr}(T_0 \leq m_1)$$

with $c(t) = bm^{-1/2}\{t(1-t)\}^{1/2}$ and $c(t) \equiv bm^{-1}$ respectively. Assume that $m_0 < j < m_1$; the other cases can be handled similarly. We begin with the obvious decomposition

$$\text{pr}(T_0 \leq m_1) = \text{pr}\{S_j^* \geq mc(j/m)\} + \int_{-\infty}^{mc(j/m)} \text{pr}(T_0 \leq m_1 \mid S_j^* = \xi) \, \text{pr}(S_j^* \in d\xi). \quad (28)$$

Since the marginal distribution of S_j^* is known, to approximate (28) it suffices to approximate the conditional probability. Moreover, given $S_j^* = \xi$, the processes S_n^*, $n = 0, 1, \cdots, j$, and S_n^*, $n = j, j+1, \cdots, m$ are conditionally independent and are themselves discrete time Brownian bridges with endpoints tied down at 0 and at ξ. Hence in terms of $T_1 = \sup\{n : n \leq m_1, \ S_n^* \geq mc(n/m)\}$, we can write for $\xi < mc(j/m)$

$$\begin{aligned}
\text{pr}(T_0 \leq m_1 \mid S_j^* = \xi) = \ & \text{pr}(T_0 < j \mid S_j^* = \xi) + \text{pr}(T_1 > j \mid S_j^* = \xi) \\
& - \text{pr}(T_0 < j \mid S_j^* = \xi) \, \text{pr}(T_1 > j \mid S_j^* = \xi).
\end{aligned} \qquad (29)$$

Since both probabilities on the right hand side of (29) are of the same form, it suffices to consider the first one. We assume that m is large and that j and $j - m_0$ are proportional to m. For many boundary curves $c(t)$, including those of interest here, the principal contribution to the integral on the right hand side of (28) comes from values of ξ close to $mc(j/m)$, say

$$\xi = mc(j/m) - z \qquad (30)$$

14

with $x = O(\log m)$ as $m \to \infty$. Given $S_j^* = \xi$ of the form (30), if $S_n^* \geq mc(n/m)$ for some $m_0 \leq n < j$, this event with overwhelming probability occurs for some n close to j. For n close to j, say $n = j - k$, we have

$$mc(n/m) = mc(j/m) - kc'(j/m) + 0(k^2/m).$$

Hence for ξ of the form (30) and $k_m = o(m^{1/2})$

$$\mathrm{pr}(T_0 < j \mid S_j^* = \xi) \cong \mathrm{pr}\{S_{j-k}^* - S_j^* \geq x - kc'(j/m) \text{ for some } k \leq k_m \mid S_j^* = \xi\}.$$

For $k << j$, given $S_j^* = \xi$, the process $S_{j-k}^* - S_j^*$, $k = 1, 2, \cdots$ behaves like a sum of independent normally distributed random variables, each having mean $-\xi/j \cong -mc(j/m)/j$ and variance 1. Consequently $\mathrm{pr}(T_0 < j \mid S_j^* = \xi)$ is approximately a probability of the form considered in Propositon 1 with $\mu = (j/m)^{-1}c(j/m) - c'(j/m)$. Although we have reasoned that the important values of x are not large, if we nevertheless use the large x approximation of Proposition 1 together with (17), we obtain

$$\mathrm{pr}(T_0 < j \mid S_j^* = \xi) \cong \exp[-2\{t^{*-1}c(t^*) - c'(t^*)\}(x + \rho)], \tag{31}$$

where $t^* = j/m$.

Consider the special case $c(t) = bm^{-1/2}\{t(1 - t)\}^{1/2}$, and assume that $bm^{-1/2} = \gamma$ and $jm^{-1} = t^*$ are fixed as $m \to \infty$. If we use (31) and a similar approximation for the other conditional probability on the right hand side of (29), substitute into (28), and evaluate the integral asymptotically as $m \to \infty$, we obtain the approximation

$$\mathrm{pr}[S_n^* \geq b\{n(1 - n/m)\}^{1/2} \text{ for some } m_0 \leq n \leq m_1] \cong 1 - \Phi(\varsigma) + m^{-1/2}\varphi(\varsigma)$$
$$\times \left(\frac{2\exp[-\gamma\rho/\{t^*(1 - t^*)\}^{1/2}]}{\delta\{t^*(1 - t^*)\}^{1/2}} - \frac{\exp[-2\gamma\rho/\{t^*(1 - t^*)\}^{1/2}]}{\gamma + \delta\{t^*(1 - t^*)\}^{1/2}} \right), \tag{32}$$

where $\varsigma = m^{1/2}[\gamma - \delta\{t^*(1 - t^*)\}^{1/2}]$, $\gamma = bm^{-1/2}$, and $t^* = j/m$.

15

The analogous approximation when $c(t) \equiv b/m$ is easily obtained, but in this case it is possible to squeeze out a bit more accuracy for small samples by using a slightly different approximation along the lines suggested by Siegmund (1985, Example 8.77) in a similar context. The final approximation is omitted.

According to (13) the mean of the numerator in (9) is a nonlinear function of k and hence it is convenient to center the process to have mean 0. If we also approximate (13) by $j\delta \log^+(k/j)$, the power of (9) can be expressed as

$$\text{pr}\{\tilde{S}_n \geq mc(n/m) \text{ for some } m_0 \leq n \leq m\}. \tag{33}$$

Here $\tilde{S}_n = z_1 + \cdots + z_n$, where the z's are independent standard normal random variables, and in terms of $t^* = j/m$ and $\gamma = bm^{-1/2}$,

$$c(t) = \gamma t^{1/2} + \delta t^* \max\{\log(1-t), \log t^*\}.$$

By conditioning on $\tilde{S}_{m-j}$ one can argue as above to derive an approximation to (33). The approximate drift of the conditional random walk $\tilde{S}_{m-j-k} - \tilde{S}_{m-j}$ given $\tilde{S}_{m-j} = mc(1-t^*) - x$ is $-c(1-t^*)/(1-t^*)$, and hence the appropriate μ for an application of Proposition 1 is $c(1-t^*)/(1-t*) - c'_-(1-t^*)$, where c'_- denotes the left hand derivative of c. However, since $\tilde{S}_{m-j+k} - \tilde{S}_{m-j}$, $k = 1, 2, \cdots, j$ is not tied down at $k = j$, its drift is 0 and the appropriate value of μ for this part of the path is $c'_+(1-t^*)$. The resulting approximation is

$$1 - \Phi(\varsigma) + m^{-1/2}\varphi(\varsigma)$$
$$\times \left\{ \frac{\exp(-\rho[\gamma/(1-t^*)^{1/2} + 2\delta\{t^* \log t^*/1-t^*) + 1\}])}{\delta\{2(1-t^*) + t^* \log t^*\}} + \frac{\exp\{-\rho\gamma/(1-t^*)\}}{\delta t^* \log(1/t^*)} \right.$$
$$\left. - \frac{\exp(-2\rho[\gamma/(1-t^*)^{1/2} + \delta\{t^* \log t^*/(1-t^*) + 1\}])}{\gamma(1-t^*)^{1/2} + \delta\{2(1-t^*) + t^* \log t^*\}} \right\}, \tag{34}$$

where $\varsigma = m^{1/2}\{\gamma + \delta t^* \log t^*/(1-t^*)^{1/2}\}$, and as always ρ is the constant appearing in (17).

16

Remark. The methods of this section easily yield approximations to the power of (5). However, for (8) they appear to work only when $t^* \geq e^{-2}$. Otherwise the asymptotic normalization of Daniels (1974) may be more useful. See also Barbour (1981).

5. Numerical Comparisons.

Tables 1-3 below compare the power of the statistics (4), (5), (9), and (10). To keep the tables digestible, only the case of a sample size $m = 40$ and one-sided significance level .025 is considered. Two issues are involved: (i) the accuracy of the approximations given in Section 4 and (ii) the comparative power of the various test statistics. To verify that the approximations are sufficiently accurate to give a reasonable picture of the relative merits of the various tests, the outcome of a 9999 repetition Monte Carlo experiment is given in parentheses in most of the cells. Other numerical calculations, not reported here, show that the essential conclusions are unchanged over a range of significance levels and sample sizes, although the magnitude of the differences can be more or less for different sample sizes.

Table 1 involves the likelihood ratio statistic (5) with two different choices of m_0 and m_1. Table 2 studies the recursive residual statistic (9). Since this statistic is not symmetric with respect to the ordering of the time scale, i.e. a change at j is not equivalent to a change at $m - j$, this table is slightly more elaborate than the others. Table 3 contains Pettitt's statistic (5) and the Chernoff-Zacks quasi Bayesian statistic (10).

Roughly speaking, the two likelihood ratio statistics and the recursive residual statistic perform about the same, while the Pettitt and Chernoff-Zacks statistics have somewhat greater power to detect changes occurring near $j = m/2$ and less power to detect changes occurring near $j = 0$ or $j = m$. Some of these differences were predicted from qualitative considerations in Section 2, and what the numerical calculations add is a feeling for the magnitude of the differences. The modified likelihood ratio statistic with $m_0 > 1$ and

17

$m_1 < m - 1$ has power at $j = m/2$ which improves over the unmodified likelihood ratio statistic. It must pay for this improvement by having less power for j close to 0 or m, although in the range of j studied, the cost is not apparent.

One possible conclusion is that one should choose a test statistic on a subjective basis, depending on where one "expects" a change to take place, should there be one. A difficulty with this recommendation is our belief that change-point statistics are often applied to retrospective data, frequently after something resembling a change has been noticed in informal investigations. If so, it would not be appropriate to make such a subjective choice of test statistic.

The numerical results lend support to the argument that the recursive residual statistic is not demonstrably inferior to the others, and since it generalizes immediately to a regression context, it seems a reasonably good general purpose statistic. The arbitrariness in its definition noted in Section 2 and reflected in the lack of symmetry about $j = m/2$ in Table 2, and the difficulty in using it for estimation are weak points.

The Chernoff-Zacks statistic does not seem to have any distinct advantage over the Pettitt statistic, except the simplicity of its sampling distribution, and even that vanishes if the variance is unknown. Since the Pettit statistic gives simple and natural estimates of j and δ, it seems preferable.

If estimation of j and/or δ is an important consideration, the preferred statistics appear to be the modified likelihood ratio statistic and Pettitt's statistic, which are not inferior as test statistics and provide natural estimates. Although neither of these tests dominates the other, the modified likelihood ratio test is perhaps slightly preferred because it performs better when j is near 0 or m, where all tests are weak.

Some speculations about the use of these test statistics in different contexts are given

18

in the following section.

Table 1

Likelihood Ratio Statistics

		Power	
δ	j	(4): $b = 2.95$, $m_0 = 1$, $m_1 = 39$	(4): $b = 2.82$, $m_0 = 5$, $m_1 = 35$
0	"∞"	.0254	.0250
.8	20	.482 (.499)	.540 (.541)
	10	.353 (.377)	.409 (.406)
	5	.186 (.209)	.182 (.218)
1.0	20	.706 (.716)	.753 (.758)
	10	.549 (.568)	.605 (.608)
	5	.301 (.319)	.300 (.337)
1.2	20	.872 (.878)	.900 (.905)
	10	.737 (.751)	.781 (.785)
	5	.444 (.470)	.447 (.475)

Table 2

Recursive Residual Statistic

			Power	
δ	j		(9): $b = 2.65$, $m_0 = 5$	
0	"∞"		.0253	
.8	20		.555 (.573)	
	10	30	.364 (.379)	.462 (.477)
	5	35	.172 (.169)	.237 (.270)
1.0	20		.758 (.770)	
	10	30	.534 (.558)	.654 (.660)
	5	35	.244 (.258)	.368 (.398)
1.2	20		.899 (.904)	
	10	30	.702 (.728)	.815 (.821)
	5	35	.334 (.367)	.520 (.547)

19

Table 3

Pettit and Chernoff-Zacks Statistics

		Power	
δ	j	(5): $b = 8.01$	(10): $b = 240$
0	"∞"	.0250	.0250
.8	20	.633	.591
	10	.401	.376
	5	.137	.158
1.0	20	.826	.782
	10	.593	.538
	5	.205	.223
1.2	20	.939	.908
	10	.769	.693
	5	.297	.301

Remark. It is natural to complement these numerical examples with a discussion of asymptotic efficiency, but we do not know a satisfactory formulation. Although it is a simple matter to compute the Bahadur efficiency of the tests, this measure seems too crude to provide much insight for reasonable sample sizes. For example, it would say tht the likelihood ratio statistic (2) is better than the others for every δ and change-points j which are proportional to m as $m \to \infty$. For a discussion of Bahadur efficiency in a regression context, see Deshayes and Picard (1982).

6. Generalizations and Variations.

There are two generalizations of the simple model of the preceding sections which have received considerable attention in the literaure. One involves relaxation of the normality assumption to allow any distribution, say in an exponential family. Change-point problems with Poisson and/or Bernoulli data are discussed by Kendall and Kendall (1980) and Levin and Kline (1985). A more subtle variation occurs in Matthews and Farewell (1982), who

20

introduce a model for survival after therapy. In their model the effect of therapy is not instantaneous, but deaths continue to occur at a constant hazard rate until an unknown time when the therapy becomes effective and the hazard rate for those still alive decreases to a second constant value.

Kendall and Kendall (1980) discuss a modification of the likelihood ratio statistic and observe that one can approximate its null hypothesis distribution by that of the appropriate functional of a Brownian bridge process. On the basis of what appear to be limited comparisons of simulations of the Poisson process with numerical computations for the Brownian bridge given by Mandl (1962), they conclude that the approximation is reasonable.

The asymptotic formula (26) gives an accurate and easily evaluated approximation to the relevant Brownian bridge probability in an arbitrary number of dimensions, but in general one should expect that it is a rather crude approximation for the probability of interest. In particular, for the one dimensional problem with normal data considered throughout this paper, it often overestimates the correct probability by 40 to 100%. It should be possible to give a precise asymptotic approximation similar to (18) for the Poisson case; but it remains to be seen whether more sophisticated mathematical analysis actually leads to a better approximation.

Note that the recursive residual idea depends very heavily on the assumption of normality to be exactly valid, although Sen (1982) has shown that it is asymptotically valid under quite general conditions. We do not know of any attempt to study the accuracy of Sen's approximations for sample sizes of interest.

Regression models are a second generalization which has been discussed in a number of papers, particularly with reference to econometric data. See, for example, Quandt (1958, 1960), Brown, Durbin, and Evans (1975), and Worsley (1983). The recursive residual con-

cept adapts nicely to this setting (Brown, et al., 1975), and indeed the null hypothesis distribution is basically no different than in the simple normal case. A precise asymptotic analysis of likelihood ratio like statistics is rather complicated and depends on the spacings between the dependent variables, which translate into the variances of the normal observations making up a multidimensional statistic similar to (4). Again the multidimensional Brownian bridge provides a crude approximation, which is probably adequate for practical purposes, although rather unsatisfying theoretically.

Remark. In a regression context, there is some ambiguity in the definition of a change-point problem. Scientifically, it may well be reasonable to assume that the regression function is continuous at the change-point. We are assuming, however, along with most authors, that the regression function can jump. It seems plausible that for testing the hypothesis of no change, not much power is lost by allowing this additional degree of freedom under the alternative, but the situation is less clear with regard to estimation. See Hinkley (1969) or Feder (1975a,b) for a discussion when the regression function is required to be continuous at the change-point.

Appendix

Approximate Significance Levels for Studentized Processes

In this appendix we describe an approach to deriving the approximations (21), (22), and (24). A complete development is quite complicated and will be given elsewhere. Here we restrict attention to the simplest case, to wit (24), or more precisely to a slight generalization of (24) which involves one important ingredient of (21) and (22) as well. The method is adapted from Siegmund (1982, 1985).

Let $y_1, y_2, \cdots$ be independent $N(\mu, \sigma^2)$ variables, and put $S_n = y_1 + \cdots + y_n$, $U_n = y_1^2 + \cdots + y_n^2$. It is convenient to introduce the notation

$$P_{\xi,\lambda}^{(m)}(A) = \mathrm{pr}(A \mid S_m = \xi, \; U_m = \lambda) \qquad (m^{-1}\xi^2 < \lambda),$$

which by sufficiency does not depend on the parameter (μ, σ^2). Let $\lambda_0 = \lambda/m$ and consider

$$P_{0,\lambda}^{(m)}\big(\max_{0 \le k \le m} S_k \ge b\lambda_0^{1/2}\big). \tag{A.1}$$

Let $S^2 = m^{-1} \sum_1^m (y_i - \bar{y}_m)^2$. Since the distribution of the process $\{(S_n - nS_m/m)/S, \; n = 0, 1, \cdots, m\}$ does not depend on (μ, σ^2), the process is independent of the complete sufficient statistic (S_m, U_m), by Basu's theorem (Lehmann, 1959, Theorem 5.2). Therefore the left hand side of (24) is equal to (A.1) for all $\lambda > 0$.

Let η be a real number and define

$$\tau = \inf\{n : S_n \ge b + \eta n\}. \tag{A.2}$$

Then in the special case $\eta = 0$, (A.1) equals $P_{0,m}^{(m)}\{\tau < m\}$. We study this probability more generally in the following theorem.

Theorem. Assume $b = m\varsigma$, $\xi = m\xi_0$, and $\lambda = m\lambda_0$ for some $\xi_0 < \varsigma + \eta$ and $\lambda_0 > \xi_0^2$. Then

as $m \to \infty$

$$P_{\xi,\lambda}^{(m)}\{\tau < m\} \sim \nu \left[\frac{2(2\varsigma + \eta - \xi_0)}{\{\lambda_0 - 4\eta\varsigma - (2\varsigma - \xi_0)^2\}^{1/2}} \right] \left\{ \frac{\lambda_0 - 4\eta\varsigma - (2\varsigma - \xi_0)^2}{\lambda_0 - \xi_0^2} \right\}^{(m-3)/2}, \quad (A.3)$$

where ν is defined in (16) and given approximately in (17).

Proof. Let $\ell_n^{(m)} = \ell^{(m)}(n, S_n, U_n; \xi, \xi_1, \lambda, \lambda_1)$ denote the likelihood ratio of $y_1, \cdots, y_n$ under $P_{\xi,\lambda}^{(m)}$ relative to $P_{\xi_1,\lambda_1}^{(m)}$. A straightforward calculation shows that for $n \leq m - 2$,

$$\ell_n^{(m)} = \left\{ \frac{\lambda - U_n - (\xi - S_n)^2/(m-n)}{\lambda_1 - U_n - (\xi_1 - S_n)^2/(m-n)} \right\}^{(m-n-3)/2} \left(\frac{\lambda_1 - \xi_1^2/m}{\lambda - \xi^2/m} \right)^{(m-3)/2}. \quad (A.4)$$

For $n = m - 1$ the two probabilities are not absolutely continuous, but we do not need to consider this case. By Wald's likelihood ratio identity (e.g. Siegmund, 1985, p. 13), for all $m' < m - 1$

$$P_{\xi,\lambda}^{(m)}\{\tau \leq m'\} = E_{\xi_1,\lambda_1}^{(m)}\{\ell_\tau^{(m)}; \tau \leq m'\}. \quad (A.5)$$

It is easy to see that for any fixed $i = 1, 2, \cdots$ $P_{\xi,\lambda}^{(m)}\{S_{m-i} \geq b + \eta(m-i)\}$ is of smaller order of magnitude than (A.3), so it suffices to show that for suitable ξ_1, λ_1, and $m' = m - i$ the right hand side of (A.5) is asymptotic to (A.3).

Let $\xi_1 = m\{2(\varsigma + \eta) - \xi_0\}$, $\Delta = 4\eta(\varsigma + \eta - \xi_0)$, and $\lambda_1 = m(\lambda_0 + \Delta)$. The choice of ξ_1 has a natural interpretation in terms of the reflection principle (cf. Siegmund, 1985, p. 39 ff.), but our justification for the choice of λ_1 is only that it "works." Note that $\Delta = 0$ if $\eta = 0$, and this case would be adequate to prove (24). But dealing with arbitrary η is a useful warm-up for the proof of (21) and (22).

Some algebraic manipulation shows that

$$\{\lambda_1 - U_r - (\xi_1 - S_r)^2/(m-r)\}/\{\lambda - U_r - (\xi - S_r)^2/(m-r)\} \quad (A.6)$$

$$= 1 + 4(\varsigma + \eta - \xi_0)(S_r - m\varsigma - \eta r)/[(m-r)\{\lambda_0 - U_r/m - (\xi - S_r)^2/m(m-r)\}].$$

24

Law of large numbers arguments indicate that under $P_{\xi_1,\lambda_1}^{(m)}$

$$m^{-1}r \xrightarrow{P} \varsigma/(2\varsigma + \eta - \xi_0), \quad m^{-1}U_r \xrightarrow{P} \varsigma\{\lambda_0 + 4\eta(\varsigma + \eta - \xi_0)\}/(2\varsigma + \eta - \xi_0),$$

and $\hspace{10cm}$ (A.7)

$$(\xi - S_r)^2/m(m-r) \xrightarrow{P} (\varsigma + \eta - \xi_0)(2\varsigma - \xi_0)/(2\varsigma + \eta - \xi_0).$$

The results in (A.7) show that the right hand side of (A.6) is $1 + O_p(m^{-1})$. Hence by writing $(\cdot) = \exp\{\log(\cdot)\}$, using (A.6) and a Taylor series expansion, we see that the first factor on the right hand side of (A.4) has the same limit as

$$\exp[-2(\varsigma + \eta - \xi_0)(S_r - m\varsigma - \eta r)/\{\lambda_0 - U_r/m - (\xi - S_r)^2/m(m-r)\}],$$

which by (A.7) has the same limit as

$$\exp[-2(2\varsigma + \eta - \xi_0)(S_r - m\varsigma - \eta r)/\{\lambda_0 - 4\eta\varsigma - (2\varsigma - \xi_0)^2\}]. \hspace{1cm} (A.8)$$

Assuming that we can take these limits inside the expectation, we see from (A.4), (A.5), and (A.8) that

$$\lim_m P_{\xi,\lambda}^{(m)}\{r \le m-3\}[(\lambda_0 - \xi_0^2)/\{\lambda_0 - 4\eta\varsigma - (2\varsigma - \xi_0^2)\}]^{(m-3)/2}$$

$$= \lim_m E_{\xi_1,\lambda_1}^{(m)}\{\exp[-2(2\varsigma + \eta - \xi_0)R_r/\{\lambda_0 - 4\eta\varsigma - (2\varsigma - \xi_0)^2\}]; \ r \le m-3\},$$

where $R_r = S_r - m\varsigma - \eta r$ is the excess over the stopping boundary. If we were dealing with an unconditional probability making the y's independent and identically distributed with the same mean and variance as under $E_{\xi_1,\lambda_1}^{(m)}$, the renewal theorem would allow us to finish the proof along established lines (e.g. Siegmund, 1985, Chapter VIII). Over the relatively short interval in which $m^{-1}r$ falls with $P_{\xi_1,\lambda_1}^{(m)}$-probability close to one (cf. (A.7)), the conditional and unconditional processes behave essentially the same, leading one to expect the same limiting result for the $P_{\xi_1,\lambda_1}^{(m)}$-distribution of R_r. This anticipated result can be proved by ad hoc methods or by appealing to a general theorem in Inchi Hu's unpublished Stanford

thesis, thus completing our informal proof of the theorem.

The fundamental identity (A.5) can be used to provide an effective variance reducing device if one wants to perform a Monte Carlo experiment to check the accuracy of the approximation (A.3). However, there is an interesting pitfall, which is not present in the use of this technique as suggested by Siegmund (1975, 1985, Remark 4.45). In particular, it is apparent from (A.4) and (A.6) that when $r = m - 2$, $\ell_r^{(m)}$ can be extremely large. Although this happens with small probability, it can create an overflow; and even if the likelihood ratio is truncated it can completely distort the Monte Carlo estimator. To avoid the problem, it evidently suffices to use (A.5) with $m' = m - 3$ rather than $m - 2$, even though this slightly increases the bias of the resulting estimator.

Table 4 gives some numerical results. The first approximation in each row is (24), and for comparison the approximation (20) for the case of known variance is also given. The Monte Carlo estimates are based on the identity (A.5) with $m' = m - 3$ in the fifth column, upper entry, and $m' = m - 2$ in the fifth column, lower entry. They result from a 2500 repetition experiment, and are given plus or minus an estimated standard error. The likelihood ratio was truncated to prevent overflow. In some cases the lower entry is somewhat larger and its estimated standard error much larger than the upper entry. This phenomenon is undoubtedly a result of the instability of the Monte Carlo procedure, as described above. In these cases a Monte Carlo estimate based on a direct frequency count in a 9999 repetition experiment is given in the sixth column. In all cases it suggests that the use of (A.5) with $m' = m - 3$ is preferred over $m - 2$. When one takes the substantially smaller standard error into account, use of (A.5) is roughly twenty-five to one hundred times as efficient as direct Monte Carlo.

The analytic approximation is reasonably good in all cases, although it deteriorates somewhat at the smaller sample sizes. It is interesting that the known sigma approximation,

26

(20), is sometimes larger and sometimes smaller than (24). It is also reasonably accurate, but since neither approximation is onerous to compute, one may as well use the theoretically appropriate one.

Table 4

Approximations to $P_{0,m}^{(m)}\{r \leq m\}$

		Analytic		Monte Carlo	
m	b	(24)	(20)	(A.5)	Direct
40	8.01	.0237	.0250	.0236 ± .0002	
				.0234 ± .0002	
20	6.0	.0094	.0131	.0097 ± .0002	
				.0098 ± .0002	
20	5.0	.0442	.0043	.0449 ± .0006	.0444
				.0493 ± .0018	
20	4.0	.1366	.1224	.1414 ± .0014	.1368
				.1810 ± .0091	
15	5.0	.0104	.0157	.0114 ± .0002	.0116
				.0121 ± .0004	
15	4.5	.0287	.0319	.0329 ± .0005	.0288
				.0482 ± .0035	

References

Anscombe, F. J. (1953). Sequential estimation, *J. Roy. Statist. Soc.*, *Ser. B* **15**, 1–21.

Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated significance tests on accumulating data, *J. Roy. Statist. Soc.*, *Ser. A* **132**, 235–244.

Barbour, A. D. (1981). Brownian motion and a sharply curved boundary, *Adv. Appl. Probab.* **13**, 736–750.

Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley & Sons, New York.

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

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time, *Ann. Math. Statist.* **35**, 999–1028.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change point problem, *Biometrika* **62**, 243–251.

Daniels, H. (1974). The maximum size of a closed epidemic, *Adv. Appl. Prob.* **6**, 607–621.

DeLong, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process, *Commun. Statist. - Theor. Meth.*, **A10**, 2197–2213.

Deschayes, J. and Picard, D. (1982). Tests of disorder of regression: asymptotic comparison, *Theory Probab. Appl.* **27**, 100–115.

Feder, P. I. (1975a). On asymptotic distribution theory in segmented regression problems – identified case, *Ann. Statist.* **3**, 49–83.

Feder, P. I. (1975b). The log likelihood ratio in segmented regression, *Ann. Statist.* **3**, 94–97.

Feller, W. (1972). *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, New York.

Gardner, L. A., Jr. (1969). On detecting changes in the mean of normal variates, *Ann. Math. Statist.* **40**, 116–126.

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location, *J. Amer. Statist. Assoc.* **72**, 180–186.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression, *Biometrika* **56**, 494–504.

Keilson, J. and Ross, H. F. (1975). Passage time distributions for Gaussian Markov (Ornstein-Uhlenbeck) processes. In *Selected Tables in Mathematical Statistics* **3**, 233–328 (Inst. Math. Statist., ed.) Amer. Math. Soc., Providence.

Kendall, D. G. and Kendall, W. S. (1980). Alignments in two-dimensional random sets of points, *Adv. Appl. Probab.* **12**, 380–424.

Kiefer, J. (1959). *K*-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests, *Ann. Math. Statist.* **30**, 420–447.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, John Wiley & Sons, New York.

Levin, B. and Kline, J. (1985). The cusum test of homogeneity, with an application to spontaneous abortion epidemiology, to appear in *Statistics in Medicine*.

Mandl, P. (1962). On the distribution of the time which the Uhlenbeck process requires to exceed a boundary (in Czechoslovakian with Russian and German summaries). *Apl. Mat.* **7**, 141-148.

Matthews, D. E., Farewell, V. J., and Pyke, R. (1985). *Asymptotic score statistic processes and tests for constant hazard against a change-point alternative*, *Ann. Statist.* **13**, 583-591.

Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a changepoint alternative, *Biometrics* **38**, 463-468.

Pettit, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations, *Biometrika* **67**, 79-84.

Pocock, S. (1977). Group sequential methods in the design of an analysis of clinical trials, *Biometrika*, **64**, 191-200.

Quandt, R. E. (1958). *The estimation of the parameter of a linear regression system obeying two separate regimes*, *J. Amer. Statist. Assoc.* **53**, 873-880.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes, *J. Amer. Statist. Assoc.* **55**, 324-330.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting changes in means, *Ann. Statist.* **3**, 98-108.

Sen, P. K. (1982). Invariance principles for recursive residuals, *Ann. Statist.* **10**, 307-312.

Siegmund, D. (1975). Importance sampling in the Monte Carlo study of sequential tests, *Ann. Statist.*, **4**, 673-684.

Siegmund, D. (1977). Repeated significance tests for a normal mean, *Biometrika*, **64**, 177-189.

Siegmund, D. (1978). Estimation following sequential tests, *Biometrika*, **65**, 341-349.

Siegmund,D. (1982). Large deviations for boundary crossing probabilities, *Ann. Probab.*, **10**, 581-588.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York-Heidelberg-Berlin.

Siegmund, D. (1986). Boundary crossing probabilities and statistical applications, *Ann. Statist.* **14**.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability, *Biometrika* **70**, 455-464.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 35	2. GOVT ACCESSION NO. *AD-A159 381*	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Tests for a Change-Point		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Barry James, Kang Ling James, David Siegmund		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0306
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-373
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Code (411 (SP)) Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE August 1985
		13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS(*If different from Controlling Office*)		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, If different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Change-point; likelihood ratio test; recursive residuals; power approximations; boundary crossing probabilities

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

See reverse side

The problem considered is that of testing a sequence of independent noraml random variables with constant, known or unknown, variance for no change in mean versus alternatives with a single change-point. Various tests, such as those based on the likelihood ratio and recursive residuals, are studied. Power approximations are developed by integrating approximations for conditional boundary crossing probabilities. A comparison of several tests is made, and the power approximations obtained are compared with Monte Carlo values.

END

FILMED

10-85

DTIC