A Martingale Approach to the Changepoint Problem

Göran BROSTRÖM

The changepoint problem for a binary sequence is considered. A test statistic based on recursive residuals is compared to the test statistic suggested by Pettitt. The new test statistic has more interesting properties for use in sequential testing. However, neither of the two test statistics dominates the other. Sequential versions of the martingale-based test, forward and reverse, are given and compared to other tests by means of a simulation study. The reverse martingale tests detect a shift earlier, if it is detected. The price to be paid is a slightly higher probability of not detecting a shift.

KEY WORDS: Binary data; Bootstrap; Brownian motion; Conditional inference; Recursive residuals; Reliability.

1. INTRODUCTION

The problem of detecting a shift in a constant level of probability of success has been considered by many researchers in the 1970s and 1980s (see, e.g., Carlstein 1988; Hinkley 1970; Pettitt 1979, 1980; Smith 1975; Worsley 1983). Generally, the problem can be formulated as follows: $X_1, \ldots, X_n$ are independent Bernoulli random variables with probabilities $p_1, \ldots, p_n$ of success, and it is known that for some $\kappa, 0 < \kappa \leq n$, the p's have the following relation:

$$p_1 = p_2 = \cdots = p_\kappa < p_{\kappa+1} = \cdots = p_n.$$

The null hypothesis to be tested is that $\kappa = n$; that is, all of the p's are equal. Only one-sided alternatives are considered. The applications in mind are mainly from reliability theory, where concern is about detecting deteoriation in a production, so that the probability of failure of produced items is supposed to be constant at a low lewel but at some time point, produced items suddenly have a higher failure probability.

Pettitt's test and a weighted version thereof are reviewed in Section 2, and the likelihood-ratio based test is described in Section 3. The martingale approach to the changepoint problem is introduced in Section 4. The idea is very close to a solution to the problem of testing constancy of regression relationships over time, as given by Brown, Durbin, and Evans (1975). The difference is that the analysis is done conditionally on the total sum of successes. The involved filtration thus is nonstandard.

In Section 5, seven fixed sample size tests are compared by simulation; two versions of Pettitt's test, four versions of the martingale-based test, and the likelihood ratio test. The results indicate that perhaps the martingale-based tests detect change faster. This is the theme of Section 6, where a simulation study is performed. The comparisons include a version of the classical cumulative sum (cusum) approach.

2. PETTITT'S TEST

The general setup is as follows: $X_1, \ldots, X_n$ are iid Bernoulli variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i, i = 1, \ldots, n$. The null hypothesis is that $p_1 = \cdots = p_n$, and the alternative is that for some $\kappa, 0 < \kappa < n, p_1 = \cdots = p_\kappa < p_{\kappa+1} = \cdots = p_n$. This is a reasonable alternative in a reliability context, where interest is focused on detecting an increased probability of failure.

Suppose for a moment that κ is known; then there exists a uniformly most powerful unbiased test (Lehmann 1986, pp. 154–155). The test statistic is $U = \sum_{i=1}^{\kappa} X_i$, and the test is performed conditional on $T = \sum_{i=1}^{n} X_i$. The result is Fisher's exact test for a 2×2 table.

Let $S_i = \sum_1^i X_j, i = 1, \ldots, n$. Then $U = S_\kappa$ and $T = S_n$, and, under the null hypothesis,

$$P_0(S_\kappa = u | S_n = t) = \frac{\binom{\kappa}{u}\binom{n-\kappa}{t-u}}{\binom{n}{t}};$$

that is, the hypergeometric distribution.

It follows that under the null hypothesis, the first two conditional moments for $k = 0, \ldots, n$, are

$$E(S_k | S_n) = k\hat{p}$$

and

$$V(S_k | S_n) = k(n-k)\hat{p}(1-\hat{p})/(n-1),$$

where $\hat{p} = S_n/n$.

The test statistic is

$$R(\kappa, n) = \frac{\sqrt{(n-1)}(\kappa\hat{p} - S_\kappa)}{\sqrt{\kappa(n-\kappa)\hat{p}(1-\hat{p})}},$$

and its limiting conditional distribution under the null, as $\kappa, n \to \infty$ in such a manner that n, κ tends to a limit bounded from 0 and 1, is standard normal. Holst (1979) gave a strict derivation of this seemingly trivial statement (in a more general context). The test rejects the null hypothesis for large values of the test statistic.

What if the time point for the shift is unknown? Then we have the changepoint problem, and the natural test statistic is

$$R_n^{(2)} = \max_{1 \leq k \leq n} R(k, n),$$

Göran Broström is Associate Professor of Statistics, Department of Statistics, Umeå University, S-90187 Umeå, Sweden. Research was supported by The Bank of Sweden Tercentenary Foundation. The author thanks a referee, the associate editor, and the editor for helpful suggestions.

Journal of the American Statistical Association
September 1997, Vol. 92, No. 439, Theory and Methods

which is a weighted version of the test statistic given by Pettitt (1980):

$$R_n^{(1)} = \max_{1\le k\le n} \frac{k\hat{p} - S_k}{\sqrt{n\hat{p}(1-\hat{p})}}.$$

It is well known that the exact null distribution of $R_n^{(1)}$ is the same as the null distribution of (a multiple of) the Kolmogorov–Smirnov two-sample test statistic (Steck 1969). Thus numerous small sample tables and asymptotic results are available. The asymptotic distribution is the one of the maximum of a Brownian bridge on (0, 1).

However, at sample sizes where tables stop (usually at sample sizes around 100), asymptotic results are still unreliable—mainly on the conservative side, causing low power. Therefore, relying on the bootstrap as soon as tables are not available is recommended. The bootstrap is very simple to implement in the present situation, and it is used throughout.

Pettit (1980) claimed that $R_n^{(2)}$ is "generally inferior to" $R_n^{(1)}$, a claim needing closer evaluation.

3. THE LIKELIHOOD RATIO TEST

Regarding the likelihood function in the shift model,

$$L(p,p',\kappa) = p^{S_\kappa}(1-p)^{k-S_\kappa}p'^{S_n-S_\kappa}(1-p')^{n-\kappa-(S_n-S_\kappa)},$$

as a function of three parameters, it is possible to test the hypothesis

$$H_0\colon \kappa = n$$

(no shift) with the likelihood ratio test. A difficulty is that ordinary asymptotic distributional results do not apply, and the exact distribution is awkward. The bootstrap approach poses no problems, however. The test statistic is

$$R_n^{(5)} = 2(\max_{1\le k\le n} \sup_{0<p\le p'<1} \ln L(p,p',k) - \sup_{0<p<1} \ln L(p,p,n))$$

when the alternative is one sided.

4. THE MARTINGALE APPROACH

Consider the iid Bernoulli sequence $X_1, X_2, \ldots, X_n$ and its corresponding cusum process $S_1, S_2, \ldots, S_n$, the natural filtration is

$$\mathcal{F}_j = \sigma\{X_1, \ldots, X_j\}, \qquad j = 1, \ldots, n,$$

and it is obvious that $\{S_j, \mathcal{F}_j\}_0^n$ is a submartingale. But we want to perform the analysis conditional on S_n. Therefore, the filtration

$$\mathcal{F}_j^{(n)} = \mathcal{F}_j \vee \sigma\{S_n\} = \sigma\{X_1, \ldots, X_j, S_n\},$$

$$j = 1, 2, \ldots, n$$

is appropriate. For completeness, let $S_0 = 0, \mathcal{F}_0^{(n)} = \sigma\{S_n\}$, and $\mathcal{F}_{-1}^{(n)} = \{\Omega, \emptyset\}$.

The sequence $\{S_j, \mathcal{F}_j^{(n)}\}_{j=0}^n$ is a submartingale because (a) $\mathcal{F}_j^{(n)} \subset \mathcal{F}_{j+1}^{(n)}$ and $S_j \in \mathcal{F}_j^{(n)}, j = 0, \ldots, n-1$; (b) $0 < E(S_j) < \infty, j = 0, \ldots, n$; and (c) $E(S_{j+1}|\mathcal{F}_j^{(n)}) \ge S_j$ a.e.

Doob's decomposition theorem states that $\{S_j\}_{j=0}^n$ can be decomposed into a martingale $\{Z_j^{(n)}, \mathcal{F}_j^{(n)}\}_{j=0}^n$ and an increasing process $\{A_j^{(n)}\}_{j=0}^n$, where the latter is predictable—meaning that for each j, $A_j^{(n)} \in \mathcal{F}_{j-1}^{(n)}$.

Theorem 1. The compensator $A^{(n)}$ of the submartingale $\{S_j, \mathcal{F}_j^{(n)}\}$ is given by the partition

$$S_j = Z_j^{(n)} + A_j^{(n)}, \qquad j = 0, \ldots, n,$$

with

$$A_0^{(n)} = 0, \qquad A_j^{(n)} = A_{j-1}^{(n)} + \frac{S_n - S_{j-1}}{n-j+1},$$

$$j = 1, \ldots, n,$$

and

$$Z_j^{(n)} = S_j - A_j^{(n)}, \qquad j = 0, \ldots, n.$$

Proof. The proof is straightforward, by construction.

The goal now is to show that $\{Z_j^{(n)}\}_0^n$, suitably normalized and time transformed, converges weakly to the standard Wiener process on (0, 1) as n tends to infinity. Thus the variance function is needed.

Theorem 2. The conditional variance function, given S_n, is, for $j = 1, \ldots, n$, given by

$$\begin{aligned} V(Z_j^{(n)}|\mathcal{F}_0^{(n)}) &= \hat{p}(1-\hat{p})\frac{n}{n-1}\sum_{i=1}^{j}\frac{n-i}{n-i+1} \\ &= j\hat{p}(1-\hat{p}) + \mathrm{O}(n). \end{aligned}$$

Proof. The proof is provided in the Appendix.

Suppose now that the changepoint occurs at time κ; that is, $X_1, \ldots, X_\kappa$ are iid Bernoulli variables with success probability p, and $X_{\kappa+1}, \ldots, X_n$ are iid Bernoulli variables with success probability p'. Let $\Delta Z_j^{(n)} = Z_j^{(n)} - Z_{j-1}^{(n)}, j = 1, \ldots, n$. Then

$$E(\Delta Z_j^{(n)}) = \begin{cases} \frac{n-\kappa}{n-j+1}(p-p'), & j = 1, \ldots, \kappa \\ 0, & j = \kappa+1, \ldots, n. \end{cases}$$

In other words, $\{Z_j^{(n)}\}_{j=1}^{\kappa}$ has negative drift, because $(p - p') < 0$, whereas $\{Z_j^{(n)} - Z_k^{(n)}\}_{j=\kappa+1}^n$ has zero drift independent of the value of p_1 and p_2. Therefore, a reasonable test statistic is

$$R_n^{(4)} = \max_{1\le k<n} \frac{-Z_k^{(n)}}{\sqrt{k\hat{p}(1-\hat{p})}}. \tag{1}$$

However, we also consider its unweighted analog,

$$R_n^{(3)} = \max_{1\le k<n} \frac{-Z_k^{(n)}}{\sqrt{n\hat{p}(1-\hat{p})}}. \tag{2}$$

In the light of Theorem 2, the convergence of $R_n^{(3)}$ under the null hypothesis to the maximum of a Wiener process on

Table 1. Comparison of Power for Seven Tests, n = 50, p_1 = .2, p_2 = .4, Size 5%, One-Sided

Change-point	Pettitt Unweighted	Pettitt Weighted	Martingale Forward Unweighted	Martingale Forward Weighted	Martingale Reverse Unweighted	Martingale Reverse Weighted	Likelihood ratio
25	**.393**	.311	.345	.364	.381	.293	.309
30	**.386**	.319	.367	.352	.352	.314	.306
35	.333	.306	**.351**	.312	.292	.310	.282
40	.236	.269	**.289**	.241	.207	.282	.239
45	.123	.199	.180	.144	.118	**.217**	.169
48	.074	.126	.097	.080	.074	**.153**	.098
No	.052	.051	.050	.049	.050	.055	.053

Table 2. Comparison of Power for Seven Tests, n = 100, p_1 = .2, p_2 = .4, Size 5%, One-Sided

Change-point	Pettitt Unweighted	Pettitt Weighted	Martingale Forward Unweighted	Martingale Forward Weighted	Martingale Reverse Unweighted	Martingale Reverse Weighted	Likelihood ratio
50	**.628**	.505	.558	.577	.597	.472	.492
60	**.609**	.505	.583	.554	.549	.491	.479
70	.528	.472	**.552**	.490	.449	.474	.433
80	.373	.404	**.455**	.375	.303	.413	.354
90	.172	.283	.275	.208	.153	**.299**	.234
95	.097	.193	.156	.118	.093	**.207**	.156
No	.052	.051	.051	.051	.050	.052	.052

(0, 1) is straightforward. (See general theorems on martingale convergence.)

The test statistics $R_n^{(3)}$ and $R_n^{(4)}$ are the *forward* versions of the martingale based test. The *reversed* versions are obtained by looking at failures in reverse order. Let $Y_i = 1 - X_{n-i+1}, i = 1, \ldots, n$, and apply the theory of this section to this new binary sequence. The resulting tests are not the same as the original tests. It turns out that the reversed versions often are preferred; see the numerical comparisons.

5. NUMERICAL COMPARISONS

Seven tests were compared in a simulation study: four variations of the martingale-based test (unweighted/unweighted and forward/backward), two versions of Pettitt's test (unweighted and weighted), and the likelihood ratio test.

Bootstrap tables of critical values were constructed as follows. For fixed n and S_n, 100,000 permutations of S_n 1s and $(n - S_n)$ 0s were simulated, and for each combination, the value of the test statistic was computed. In this manner, generated values were regarded as an iid sample from the test statistic under the null hypothesis. From this sample, the appropriate percentile was estimated and written to a table. The procedure was performed for $n = 50, 100$, and 200 and, in principle, for each $1 \leq S_n < n$. A lazy attitude was taken, though, in that only requested table values were computed.

The simulations were made on a Sun SPARCstation 5 running Solaris 2.4 and on a Dell Pentium 90 MHz running MS-DOS 6.2. In both cases the programming language was FORTRAN 90, the NAG version. The traditional 16807 congruential generator was used for random number generation. The errors in the estimated probabilities are at most three units in the third decimal, with 95% confidence.

5.1 Testing for a Changepoint

Focus was on small p's and late changes, and on increases in the probability of success. 100,000 simulated samples were drawn for each combination of $n = 50, 100$, and 200, and changes from $p = .2$ to .4, .6, and .8, representing small, medium and large shifts. Then the powers were estimated. The results are shown in Tables 1, 2, and 3, where figures in boldface represent the tests with highest power. Pettitt's

Table 3. Comparison of Power for Seven Tests, n = 200, p_1 = .2, p_2 = .4, Size 5%, One-Sided

Change-point	Pettitt Unweighted	Pettitt Weighted	Martingale Forward Unweighted	Martingale Forward Weighted	Martingale Reverse Unweighted	Martingale Reverse Weighted	Likelihood ratio
100	**.883**	.784	.825	.836	.850	.753	.770
120	**.867**	.774	.839	.811	.804	.760	.754
140	.796	.728	**.806**	.741	.691	.729	.697
160	.613	.629	**.697**	.595	.475	.641	.581
180	.267	.440	.439	.324	.211	**.454**	.368
190	.120	.290	.235	.161	.113	**.299**	.225
No	.050	.054	.051	.050	.049	.055	.051

Table 4. Bias × 100 of Seven Estimators of the Changepoint, n = 50, p_1 = .2, p_2 = .6, in the Case of Rejection of the Null Hypothesis, Size 5%, One-Sided

Changepoint	Pettitt Unweighted	Pettitt Weighted	Martingale Forward Unweighted	Martingale Forward Weighted	Martingale Reverse Unweighted	Martingale Reverse Weighted	Likelihood ratio
25	56	120	1,300	270	−960	**21**	37
30	−44	73	940	250	−1,300	6	**−1**
35	−170	34	630	190	−1,700	**−13**	−44
40	−400	**−12**	330	88	−2,200	−50	−96
45	−1,100	−93	**26**	−130	−2,500	−130	−210
48	−1,500	−390	**−170**	−540	−2,700	−480	−880

Table 5. Mean Squared Error × 10 of Seven Estimators of the Changepoint, n = 50, p_1 = .2, p_2 = .6, in the Case of Rejection of the Null Hypothesis, Size 5% , One-Sided

Changepoint	Pettitt Unweighted	Pettitt Weighted	Martingale Forward Unweighted	Martingale Forward Weighted	Martingale Reverse Unweighted	Martingale Reverse Weighted	Likelihood ratio
25	**31**	47	150	63	120	51	54
30	**30**	42	110	57	150	46	50
35	40	**37**	77	48	190	41	50
40	63	**33**	44	37	230	36	50
45	120	30	**16**	41	270	34	56
48	160	56	**27**	83	280	64	120

test is symmetric, so only the forward version is given in the tables.

The martingale versions, especially the weighted and reverse version, are obviously most powerful in detecting late shifts.

5.2 Estimation of the Changepoint

If the test for a changepoint rejects the null hypothesis, then estimating the location of the changepoint is of interest. A simulation study conducted along the same lines as in the case of testing gave the results displayed in Tables 4 and 5.

Table 6. Achieved Type I Error Probabilities With Bonferroni Limits, M = 50, p = .2

i	Test	α_i
1.	Pettitt's unweighted	.027
2.	Pettitt's weighted	.056
3.	Forward martingale, unweighted	.045
4.	Forward martingale, weighted	.035
5.	Reverse martingale, unweighted	.032
6.	Reverse martingale, weighted	.067
7.	Likelihood ratio	.036
8.	Cusum	.019

Table 7. Achieved Type I Error Probabilities With Limits Given by $\alpha^{(0)}$, M = 50, p = .2

i	Test	α_i
1.	Pettitt's unweighted	.049
2.	Pettitt's weighted	.048
3.	Forward martingale, unweighted	.052
4.	Forward martingale, weighted	.049
5.	Reverse martingale, unweighted	.050
6.	Reverse martingale, weighted	.048
7.	Likelihood ratio	.049
8.	Cusum	.051

6. SEQUENTIAL TESTS

One test was added to the tests considered in the previous sections: a variation of the cumulative sums (cusum) test statistic. The conditional cusum test statistic for data of length M is defined by

$$R_j^{(8)} = \max(0, R_{j-1}^{(8)} + X_j - C), \qquad j = 1, \ldots, M,$$

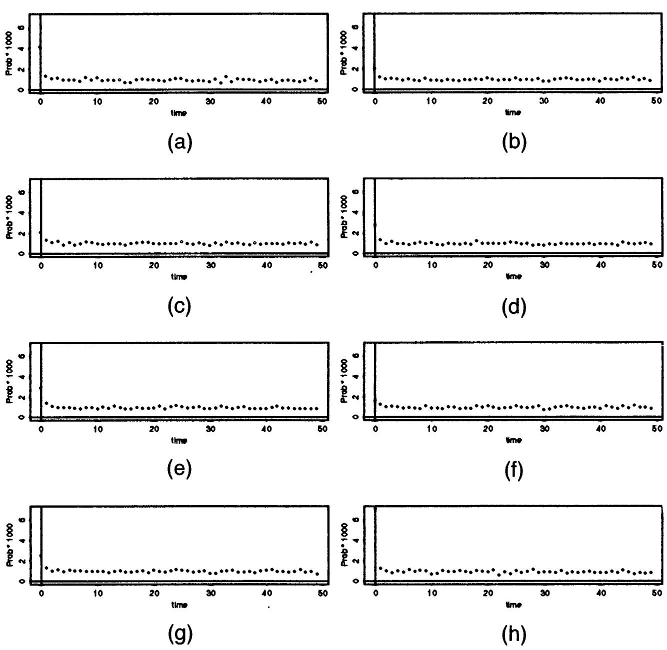

Figure 1. Probability of Alarm Versus Time. M = 50, p = .2. (a) Pettitt's, unweighted; (b) Pettitt's, weighted; (c) forward martingale, unweighted; (d) forward martingale, weighted; (e) reverse martingale, unweighted; (f) reverse martingale, weighted; (g) likelihood ratio; (h) cusum.

Table 8. Probability of Alarm and Conditional Expected Time to Alarm After a Shift From .2 to .4, M = 50

i	*Test*	*Power*	*Time to alarm*
1.	Pettitt's unweighted	.30	19
2.	Pettitt's weighted	.21	**15**
3.	Forward martingale, unweighted	.33	18
4.	Forward martingale, weighted	.31	18
5.	Reverse martingale, unweighted	**.35**	22
6.	Reverse martingale, weighted	.23	16
7.	Likelihood ratio	.27	18
8.	Cusum	.17	22

Table 9. Probability of Alarm and Conditional Expected Time to Alarm After a Shift From .2 to .6, M = 50

i	*Test*	*Power*	*Time to alarm*
1.	Pettitt's unweighted	.80	15
2.	Pettitt's weighted	.66	**13**
3.	Forward martingale, unweighted	.81	14
4.	Forward martingale, weighted	.80	15
5.	Reverse martingale, unweighted	**.84**	19
6.	Reverse martingale, weighted	.69	13
7.	Likelihood ratio	.73	15
8.	Cusum	.59	19

and

$$R^{(8)} = \max_{1 \le j \le M} R_j^{(8)},$$

where C is a suitably chosen constant, preferably close to $E_0(X)$. A possible data-dependent choice is $C = S_n/n$. The null hypothesis is rejected at $j = L$ if

$$\max_{1 \le j < L} R_j^{(8)} < r^{(8)}$$

and

$$R_L^{(8)} \ge r^{(8)},$$

where $r^{(8)}$ is chosen to give the correct size of the test.

Tests of the moving sums type are considered; that is, as time goes by, the latest observation is added and the oldest observation is deleted. If the length of the moving sum is M, then it is assumed that the time horizon is $(2M-1)$, and the tests are compared in a situation where a shift occurs at time $(M+1)$. Each test is applied M times, first on the observations $X_1, \ldots, X_M$, then on the observations $X_2, \ldots, X_{M+1}$,

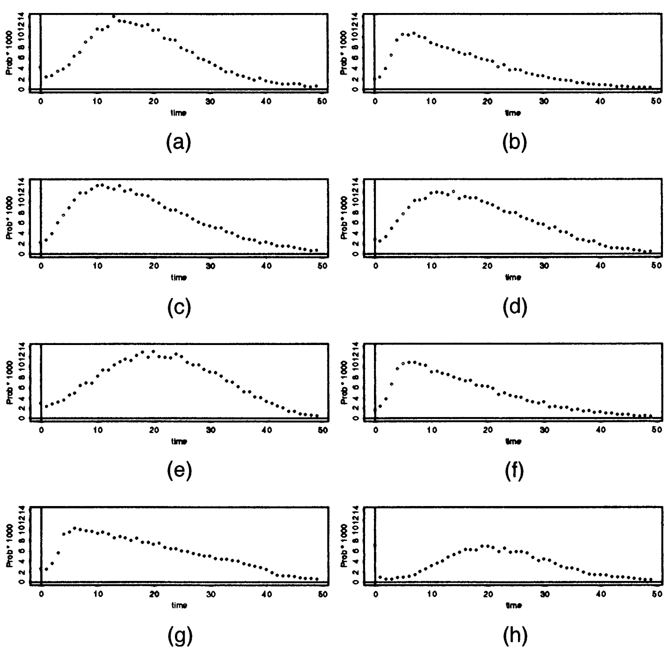

Figure 2. Probability of Alarm Versus Time After a Shift From .2 to .4. M = 50. (a) Pettitt's, unweighted; (b) Pettitt's, weighted; (c) forward martingale, unweighted; (d) forward martingale, weighted; (e) reverse martingale, unweighted; (f) reverse martingale, weighted; (g) likelihood ratio; (h) cusum.

and so on. The last set to consider is $X_M, \ldots, X_{2M-1}$. Thus the first test is actually done under the null hypothesis, and the following $(M-1)$ tests are done when the null hypothesis is false but with different (decreasing) times for the shift. If the $(M+1)$st test is applied, then the null hypothesis would again be true. Remember that the null hypothesis can be formulated as

H_0: There has been no shift during the last M trials.

The type I error is defined as "at least one rejection in the M individual tests, when H_0 is true." The individual tests clearly are not independent, and the degree of dependence varies from test to test. Therefore, it is no easy mater to decide which individual, single-test levels to choose. For a specific situation, $M = 50$ (and $p = .2$) under the null hypothesis, we zoomed in on the individual levels by a sort of trial and error simulation: Aiming at a level of $\alpha = .1$, the Bonferroni level was calculated to $\alpha_i^{(0)} = .1/50 = .002, i = 1, \ldots, 50$, and a simulation gave the total levels for the considered tests (the number of replicates in the simulation was 100,000 for each test); see Table 6.

This was closer to the .05 level than to the target .1, so the subsequent corrections aimed at $\alpha = .05$. It was necessary to choose test-specific levels, and the vector $\boldsymbol{\alpha}^{(0)}$, with

$$10{,}000\boldsymbol{\alpha}^{(0)} = (41, 16, 23, 30, 33, 15, 26, 64),$$

was finally chosen, giving rise to the results presented in Table 7.

Table 10. Probability of Alarm and Conditional Expected Time to Alarm After a Shift From .2 to .8, M = 50

i	*Test*	*Power*	*Time to alarm*
1.	Pettitt's unweighted	.99	10
2.	Pettitt's weighted	.98	9
3.	Forward martingale, unweighted	.99	9
4.	Forward martingale, weighted	.99	10
5.	Reverse martingale, unweighted	**1.0**	13
6.	Reverse martingale, weighted	.98	**9**
7.	Likelihood ratio	.98	9
8.	Cusum	.96	15

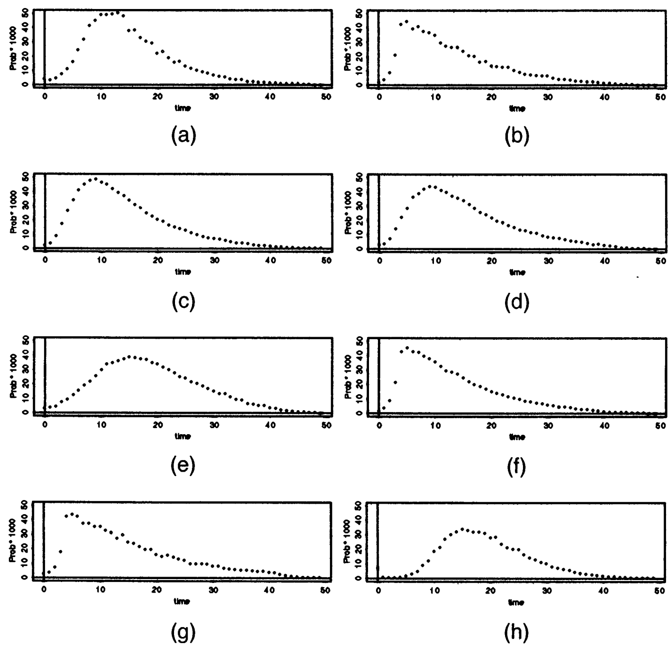

Figure 3. Probability of Alarm Versus Time After a Shift From .2 to .6. M = 50. (a) Pettitt's, unweighted; (b) Pettitt's, weighted; (c) forward martingale, unweighted; (d) forward martingale, weighted; (e) reverse martingale, unweighted; (f) reverse martingale, weighted; (g) likelihood ratio; (h) cusum.

Given (correct) rejection, it is of interest to have the rejection as early as possible after the shift. Under the null hypothesis, the conditional distribution of the time of rejection, given rejection, is almost uniform on $\{1,\dots,(M-1)\}$, as can be seen in Figure 1.

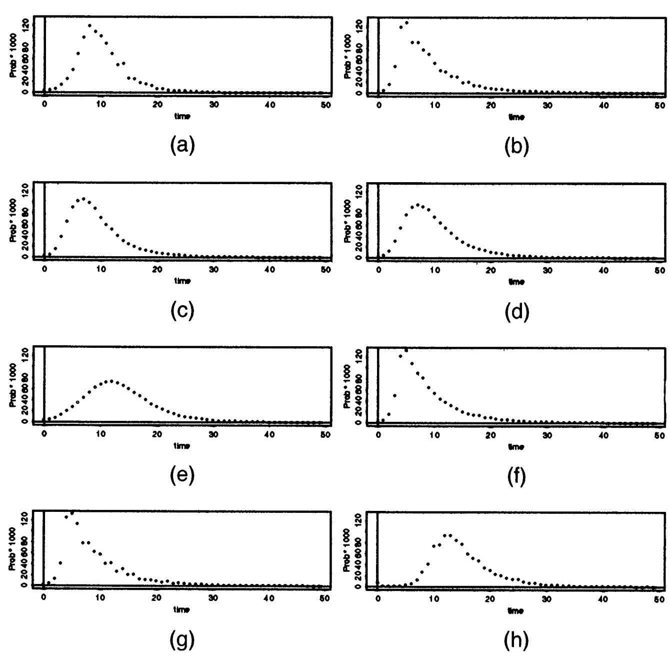

Figure 4. Probability of Alarm Versus Time After a Shift From .2 to .8. M = 50. (a) Pettitt's, unweighted; (b) Pettitt's, weighted; (c) forward martingale, unweighted; (d) forward martingale, weighted; (e) reverse martingale, unweighted; (f) reverse martingale, weighted; (g) likelihood ratio; (h) cusum.

The powers of the tests under the alternative of a shift from .2 to .4 are shown in Table 8. The last column shows the conditional expected time after the shift at which rejection occurs, given rejection after (or at) the shift and no rejection before the shift.

The conditional distribution of the times to alarm under the shift from .2 to .4 alternative is shown in Figure 2. As in Figure 1, the value at $t = 0$ is actually attained under the null hypothesis. It is thus an estimate of the "single-stage" type I error probability.

In Tables 9 and 10, the corresponding figures are shown for a moderate and a huge shift, from $p = .2$ to $p = .6$ and $p = .8$. Figures 3 and 4 display the corresponding conditional distributions of the time to alarm.

7. CONCLUSION

To summarize, the martingale approach is promising, both in a sequential framework and in a fixed sample size setting. If a quick detection is wanted, then the weighted versions are preferred; the unweighted versions have a higher overall power. Pettitt's claim that weighted versions are generally inferior is not supported. The likelihood ratio test does better than expected in the sequential approach. Finally, the cusum test performs surprisingly poorly. However, it is implemented in a nonstandard way here, which may explain its bad performance.

APPENDIX: PROOF OF THEOREM 2

To increase readability, here the superscript (n) of Z is temporarily dropped. For $j = 1,\dots,n$, we have

$$\begin{aligned} V(Z_j|\mathcal{F}_0^{(n)}) &= E(Z_j^2|\mathcal{F}_0^{(n)}) \\ &= E\left(\sum_{i=1}^{j}\{Z_i^2 - Z_{i-1}^2\}|\mathcal{F}_0^{(n)}\right) \\ &= \sum_{i=1}^{j} E(Z_i^2 - Z_{i-1}^2|\mathcal{F}_0^{(n)}) \\ &= \sum_{i=1}^{j} E\left(\frac{S_n - S_{i-1}}{n-i+1}\left\{1 - \frac{S_n - S_{i-1}}{n-i+1}\right\}\Big|\mathcal{F}_0^{(n)}\right). \end{aligned}$$

The last equality follows from the martingale property

$$\begin{aligned} Z_i^2 - Z_{i-1}^2 &= (Z_{i-1} + \Delta Z_i)^2 - Z_{i-1}^2 \\ &= (\Delta Z_i)^2 + 2Z_{i-1}\Delta Z_i. \end{aligned}$$

Taking conditional expectations (with respect to $\mathcal{F}_{i-1}^{(n)}$) throughout gives

$$E(Z_i^2 - Z_{i-1}^2|\mathcal{F}_{i-1}^{(n)}) = E(\{\Delta Z_i\}^2|\mathcal{F}_{i-1}^{(n)}),$$

because

$$E(Z_{i-1}\Delta Z_i|\mathcal{F}_{i-1}^{(n)}) = Z_{i-1}E(\Delta Z_i|\mathcal{F}_{i-1}^{(n)}) = 0(\text{a.s.}).$$

Conditional on $\mathcal{F}_{i-1}^{(n)}, \Delta Z_i$ is a 0–1 random variable minus its expected value. Therefore,

$$E(\{\Delta Z_i\}^2 | \mathcal{F}_{i-1}^{(n)}) = \frac{S_n - S_{i-1}}{n-i+1}\left(1 - \frac{S_n - S_{i-1}}{n-i+1}\right).$$

Here the conditioning is on the wrong sigma field. However, this is easily resolved, utilizing standard results on switching sigma fields (see, e.g., Chung 1974, thm. 9.1.5, p. 304). We have, with $Y_i = Z_i^2 - Z_{i-1}^2$,

$$\begin{aligned} E(Y_i | \mathcal{F}^{(n)}) &= E[E(Y_i|\mathcal{F}_0^{(n)})|\mathcal{F}_{i-1}^{(n)}] \\ &= E[E(Y_i|\mathcal{F}_{i-1}^{(n)})|\mathcal{F}_0^{(n)}] \\ &= E\left[\frac{S_n - S_{i-1}}{n-i+1}\left\{1 - \frac{S_n - S_{i-1}}{n-i+1}\right\} |\mathcal{F}_o^{(n)}\right]. \end{aligned}$$

The first equality follows because $E(Y_i|\mathcal{F}_0^{(n)}) \in \mathcal{F}_{i-1}^{(n)}$, when $i \geq 1$; the second is the switching sigma fields theorem; and the third follows from the fact that

$$E(Y_i|\mathcal{F}_{i-1}^{(n)}) = E(\{\Delta Z_i\}^2|\mathcal{F}_{i-1}^{(n)}).$$

The remaining calculations constitute a simple exercise with the hypergeometric distribution. Given a population of size n, consisting of S_n 1s and $(n - S_n)$ 0s, $(n - i + 1)$ items are randomly drawn without replacement. Find $E(X/(n - i + 1)\{1 - X/(n - i + 1)\})$, where X is the number of drawn items of type I. It follows that

$$E\left(\frac{S_n - S_{i-1}}{n-i+1}\,\Big|\mathcal{F}_0^{(n)}\right) = \hat{p}\left(= \frac{S_n}{n}\right) \tag{A.1}$$

and

$$E\left(\left\{\frac{S_n - S_{i-1}}{n-i+1}\right\}^2 \Big|\mathcal{F}_0^{(n)}\right) = \frac{n-(n-i+1)}{(n-i+1)(n-1)}\,\hat{p}(1-\hat{p}) + \hat{p}^2. \tag{A.2}$$

Taking the difference between the right sides of (A.1) and (A.2) gives

$$\begin{aligned} &\hat{p} - \frac{i-1}{(n-i+1)(n-1)}\,\hat{p}(1-\hat{p}) - \hat{p}^2 \\ &= \hat{p}(1-\hat{p})\left\{1 - \frac{i-1}{(n-i+1)(n-1)}\right\} \\ &= \hat{p}(1-\hat{p})\frac{n(n-i)}{(n-1)(n-i+1)}. \end{aligned}$$

Finally, for $j = 1, \ldots, n$, we get

$$\begin{aligned} V(Z_j^{(n)}|\mathcal{F}_0^{(n)}) &= \hat{p}(1-\hat{p})\frac{n}{n-1}\sum_{i=1}^{j}\frac{n-i}{n-i+1} \\ &= j\hat{p}(1-\hat{p}) + \mathrm{O}(n), \end{aligned}$$

which completes the proof.

[Received November 1995. Revised July 1996.]

REFERENCES

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 37, 149–192.

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188–197.

Chung, K. L. (1974), *A Course in Probability Theory*, New York: Academic Press.

Hinkley, D. (1970), "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika*, 57, 1–17.

Holst, L. (1979), "Two Conditional Limit Theorems With Applications," *The Annals of Statistics*, 7, 551–557.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, New York: Wiley.

Pettitt, A. N. (1979), "A Nonparametric Approach to the Change-Point Problem," *Applied Statistics*, 28, 455–464.

——— (1980), "A Simple Cumulative Sum–Type Statistic for the Change-Point Problem With Zero–One Observations," *Biometrika*, 67, 79–84.

Smith, A. F. M. (1975), "A Bayesian Approach to Inference About a Change-Point in a Sequence of Random Variables," *Biometrika*, 62, 407–416.

Steck, G. P. (1969), "The Smirnov Two-Sample Tests as Rank Tests," *Annals of Mathematical Statistics*, 40, 1449–1466.

Worsley, K. J. (1983), "The Power of Likelihood Ratio and Cumulative Sum Tests For a Change in a Binomial Probability," *Biometrika*, 70, 455–464.