Statistica Sinica **6**(1996), 275-287

CHANGE-POINT DETECTION FOR CORRELATED OBSERVATIONS

Hyune-Ju Kim

Syracuse University

Abstract. This paper considers a likelihood ratio test for a change in mean when observations are not independent. First, the effect of correlation on the performance of the likelihood ratio test derived under the assumption of no correlation is examined. Then, the likelihood ratio statistic for testing for a change in mean is obtained under a general structure of nonzero correlation. For general correlation and some serial correlations such as $AR(p)$, distributional properties of the test statistic are examined and methods to compute approximate p-values are discussed. Finally, the power of the likelihood ratio test is compared with that of the test proposed by Henderson (1986).

Key words and phrases: Approximate p-value, autoregressive process, boundary crossing probability, likelihood ratio test.

1. Introduction

Change-point problems for a sequence of random variables $X_1, \ldots, X_n$, are concerned with inference for an unknown change-point τ, $(1 \leq \tau \leq n)$ such that the first τ observations come from a distribution F and the remaining ones come from another distribution F^*. Since Page (1954) developed a cumulative sum test to detect a location change, considerable attention has been given to this problem in a variety of settings. Most authors have assumed that the observations are independent and studied the case where two distributions differ only in location. Some important papers on this topic include those of Hinkley (1970), Sen and Srisvastava (1975), Siegmund (1986), who used likelihood ratio approaches, and Chernoff and Zacks (1964), Smith (1975), who used Bayesian approaches. Some nonparametric methods have been discussed by Carlstein (1988).

In many applications, however, the observations are correlated in various ways. With a special correlation structure such as serial correlation, the problem is to detect a level shift in time series data. Sometimes we are interested in studying an abrupt or a gradual change in repeatedly measured data, which show natural dependency among the observations. The change-point problem when the data are correlated first appeared in Box and Tiao (1965), who discussed tests for the change in level assuming a known change-point. Although several authors have studied this problem since then, most of them used a cumulative sum statistic or a Bayesian approach because of the intractability of the exact distribution of the likelihood ratio statistic. Bagshow and Johnson (1975) studied the effect of serial correlation on the performance of the cusum test. Yashchin (1993) also considered a cusum test with serially correlated observations and suggested a method to compute an approximate average run length in a sequential detection. Henderson (1986) developed a Bayesian testing procedure for correlated observations, made some comments on serial correlations and applied his approach to material accountancy data. Nagaraj (1990) formulated the likelihood ratio statistics under several different assumptions and observed analytic difficulties associated with the likelihood ratio statistics. Zhao (1993) indicated the applicability of this problem in software reliability, but left the analytic treatment of the likelihood ratio statistic open. Our main goal in this paper is to study analytic properties of the likelihood ratio statistics in the case of a general correlation structure and to suggest approximations for p-values of the likelihood ratio test.

This paper is organized as follows. In Section 2, we examine the effect of correlation on the performance of the likelihood ratio statistic derived under the assumption of independence. Then, we obtain the likelihood ratio statistic for a change in mean when $\mathrm{Cov}(X_1,\ldots,X_n) = \sigma^2\Lambda$, where Λ is not an identity matrix. Section 3 discusses how to evaluate the p-value of the likelihood ratio test. Using large deviation theory, we develop approximations for the p-value of the likelihood ratio test and provide numerical examples to assess the accuracy of the approximations. The results are specialized to settings where Λ has a structure such as a serial correlation of $AR(p)$. In Section 4, the power of the likelihood ratio test is compared to that of the test suggested by Henderson (1986) and material accountancy data is discussed as an example.

2. Model and Likelihood Ratio Statistics

Let X_i be a sequence of observations from a normal distribution with $E[X_i] = \mu_i$, for $i = 1,\ldots,n$. The hypotheses under consideration are

$$H_0 : \mu_i = \mu \quad \text{for all} \quad i = 1,\ldots,n$$
$$H_1 : \mu = \mu_1 = \cdots = \mu_\tau \neq \mu_{\tau+1} \cdots = \mu = \mu^* \quad \text{for some} \quad 1 \le \tau < n.$$

Many authors have worked on this problem when the X's are independent. Hinkley (1970) studied the asymptotic behavior of the likelihood ratio statistic (LRS) and suggested a large sample approximation for the p-value of the test. Siegmund

(1986) proposed a generalized likelihood ratio test (LRT) to avoid the case where the likelihood ratio statistic does not have a valid limiting distribution. Using a method developed to solve boundary crossing problems in sequential analysis, Siegmund also derived an approximation for the p-value of the test, which is accurate even for small samples. James, James, and Siegmund (1987) extended the results of Siegmund (1986) to the unknown variance case and also compared the performance of the LRT with other tests. Worsley (1983, 1986) proposed a numerical method for computing the p-value of the LRT to detect a change in binomial probability and in location of an exponential family distribution.

In this paper, we consider a generalized version of the LRT as in Siegmund (1986) and discuss a detection procedure when the X's are correlated. We first examine the effect of nonzero correlation on the performance of (3.6) in Siegmund (1986), the likelihood ratio statistic derived under the assumption of independence (we call it LRS_0), and then derive a likelihood ratio statistic under nonzero correlation.

Table 1. Tail Distribution of LRS_0 for $x_i \sim AR(1)$ with autocorrelation parameter β, $n = 40$, $n_0 = 4$, $n_1 = 36$

	$\beta = -0.5$		$\beta = 0.2$		$\beta = 0.7$	
true probability	b	SA	b	SA	b	SA
0.25	1.41	0.828	2.16	0.259	3.77	3×10^{-3}
0.10	1.67	0.596	2.57	0.104	4.67	6×10^{-5}
0.05	1.83	0.468	2.82	0.055	5.24	3×10^{-6}
0.01	2.17	0.254	3.37	0.010	6.42	3×10^{-9}

Note: b is the tail percentile estimated by a Monte Carlo experiment with 10,000 repititions and SA is the approximation evaluated by (3.12) of Siegmund (1986).

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and suppose that $\mathrm{Cov}[\boldsymbol{X}] = \sigma^2\Lambda(\neq \sigma^2 I)$. Table 1 shows the sensitivity of the LRS_0 to the nonzero correlation among the observations. As an example, we consider observations from the $AR(1)$ process $X_i = \beta X_{i-1} + \epsilon_i$, where ϵ's are iid normal random variables with mean zero and variance one. For different values of β, Table 1 includes the tail percentiles of the LRS_0 and their corresponding p-values. The percentiles, b, of the LRS_0 are obtained by Monte Carlo experiments with 10,000 repetitions and their p-values are approximated by (3.12) of Siegmund (1986) and are labeled as SA. It is observed that we significantly underestimate/overestimate the true p-values when we ignore positive/negative autocorrelation. This motivates us to develop the LRT for a change-point in mean taking the nonzero correlation structure into consideration.

We first assume that σ and Λ are known for mathematical simplicity and will discuss later how to handle a situation where these values are unknown. For some quality control data sets and specifically for the material accountancy data in Henderson (1986), it is not unnatural to assume σ and Λ to be known. Although Λ is assumed to be known and a transformation to independent observations is possible, the mean vector after the transformation will not present a structural change described under H_1. Hence it is required to obtain the LRS with the existing covariance matrix. Let $\Lambda^{-1/2}$ be the symmetric positive-definite square root of Λ^{-1}, $\boldsymbol{j}_k$ be an $n \times 1$ vector with 0 in the first k components and 1 in the last $n-k$ components and let $Z_k = (\boldsymbol{j}_0; \boldsymbol{j}_k)$. Straightforward computation shows that the LRT rejects the null hypothesis of no change for large values of

$$M = \sigma^{-2} \max_{n_0 \le k \le n_1} \boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y}, \tag{2.1}$$

where $1 \le n_0 < n_1 < n$,

$$\begin{aligned} \boldsymbol{Y} &= \Lambda^{-1/2}\boldsymbol{X}, \\ P_k &= \Lambda^{-1/2} Z_k (Z_k'\Lambda^{-1} Z_k)^{-1} Z_k' \Lambda^{-1/2'} \\ \text{and} \quad P_n &= \Lambda^{-1/2}\boldsymbol{j}_0(\boldsymbol{j}_0'\Lambda^{-1}\boldsymbol{j}_0)^{-1}\boldsymbol{j}_0'\Lambda^{-1/2'}. \end{aligned}$$

It is easy to check that under the null hypothesis of no change, $Q_k = \sigma^{-2}\boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y}$ has a chi-squared distribution with 1 degree of freedom for each k. However, complicated covariance structures between the Q_k's or the $Q_k^{1/2}$'s, which depend on Λ, make it difficult to study analytic properties of the LRS.

If we assume that σ is unknown and Λ is known, then it can be observed that the test statistic is $F_1 = \max_{n_0 \le k \le n_1} \boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y}/\boldsymbol{Y}'(I - P_k)\boldsymbol{Y}$ or equivalently $F_2 = \max_{n_0 \le k \le n_1} n\boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y}/\boldsymbol{Y}'(I - P_n)\boldsymbol{Y}$. Under the null hypothesis, it is straightforward to show that $\boldsymbol{Y}'(I - P_n)\boldsymbol{Y}/n$ converges to σ^2 in probability, and thus large sample properties of (2.1) will be approximately the same as those of F_2. In a general case of unknown σ and Λ, it may be possible to develop the LRS when Λ is assumed to have a special structure. Even in the case of $AR(1)$, however, the maximum likelihood estimates of parameters cannot be obtained in closed forms, and thus the derivation of the LRS would require some iterations or approximate computation. If the observations form an $AR(1)$ process and we further assume that σ is known, it would be possible to derive the LRT and study its asymptotic properties.

3. Computation of p-value

When the test statistic is too difficult to handle analytically, one way to compute the p-value of the test is to use simulation or a resampling technique.

Since the null distribution of (2.1) does not depend on the location parameter, μ, we may generate n observations with mean 0 and a given covariance matrix $\sigma^2\Lambda$ and use those for $\boldsymbol{X}$ in (2.1).

Table 2. Accuracy of Approximations for $x_i \sim AR(1)$ with autocorrelation parameter β, $n = 40$, $n_0 = 4$, $n_1 = 36$

true	$\beta=-0.7$			$\beta=-0.4$			$\beta=0$	
probability	b	(3.3)	SA	b	(3.3)	SA	b	(3.2)
0.10	2.51	0.105	0.120	2.53	0.108	0.116	2.57	0.104
0.05	2.76	0.055	0.065	2.77	0.056	0.063	2.82	0.055
0.01	3.33	0.010	0.012	3.34	0.010	0.012	3.37	0.010

true	$\beta=0.2$			$\beta=0.8$		
probability	b	(3.3)	SA	b	(3.3)	SA
0.10	2.62	0.091	0.092	2.92	0.112	0.041
0.05	2.87	0.046	0.048	3.16	0.050	0.020
0.01	3.45	0.009	0.009	3.61	0.010	0.005

Note: b is the tail percentiles estimated by a Monte Carlo experiment with 10,000 repititions and SA is the approximation evaluated by (3.12) of Siegmund (1986).

Compared to simulation, analytic approximations are more efficient and more importantly, help to understand properties of the test statistic. Since (2.1) with an identity matrix for Λ is equivalent to the LRS_0, it is expected that the approximation (3.12) of Siegmund (1986), derived under the assumption of independence, works well when the correlation between the observations is weak. Table 2 includes the tail percentiles of (2.1) and approximate p-values. As in Table 1, selected percentiles of the tail of the distribution of (2.1) are estimated by the Monte Carlo experiments with 10,000 repetitions; and the naive approximation (3.12) of Siegmund, labeled SA in Table 2, are evaluated through numerical integration. The Monte Carlo results indicate that the percentile of (2.1) increases as the autocorrelation parameter, β, gets larger. Table 2 also shows that the approximation of Siegmund works reasonably well if the correlation is weak, but loses its accuracy as the correlation becomes strong.

The main purpose of this section is to derive analytic approximations for the p-value of the LRT taking a nonzero correlation into consideration. To evaluate the p-value of the test, $\alpha = \Pr\{\sigma^{-2}\max_{n_0\le n\le n_1}\boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y} \ge b^2\}$, we may assume that σ is equal to one without loss of generality. We first note that for any $n_0 \le k \le n_1$, the null distribution of $T_k = \boldsymbol{j}_k' \Lambda^{-1/2} \boldsymbol{Y}$ does not depend on the location parameter and that $T_0 = \boldsymbol{j}_0' \Lambda^{-1/2} \boldsymbol{Y}$ is a complete sufficient statistic for the location parameter. Thus Basu's theorem (Lehmann (1986, Theorem 5.2)) tells us that under the null hypothesis, (2.1) is independent of T_0. Therefore,

$$\begin{aligned}\alpha &= \Pr\{\max_{n_0 \le k \le n_1} \boldsymbol{Y}'(P_k - P_n)\boldsymbol{Y} \ge b^2 | \boldsymbol{j}_0' \Lambda^{-1/2} \boldsymbol{Y} = 0\} \\ &= \Pr\{\max_{n_0 \le k \le n_1} (D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{-1/2} |T_k| \ge b | T_0 = 0\},\end{aligned}$$

where $D_k = (\boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)(\boldsymbol{j}_k' \Lambda^{-1} \boldsymbol{j}_k) - (\boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_k)^2$. Since $\boldsymbol{Y}$ is a Gaussian random vector with identity covariance matrix, straightforward computation of the conditional density of T_k given $T_0 = 0$, shows that

$$\alpha = \Pr\{|U_k| \ge b(D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2} \quad \text{for some} \quad n_0 \le k \le n_1\},$$

where the U_k's are normally distributed with mean zero, variance $D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0$, and

$$\mathrm{Cov}(U_k, U_l) = \boldsymbol{j}_k' \Lambda^{-1} \boldsymbol{j}_l - (\boldsymbol{j}_k' \Lambda^{-1} \boldsymbol{j}_0)(\boldsymbol{j}_l' \Lambda^{-1} \boldsymbol{j}_0) / (\boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0) \quad \text{for} \quad k < l.$$

Noting that α is the probability that a discrete Gaussian process U crosses either one of upper or lower boundaries between n_0 and n_1, we now discuss how to approximate the p-value of the test. Since

$$\begin{aligned}\alpha/2 &= (1/2) \Pr\Big\{ \max_{n_0 \le k \le n_1} \frac{|U_k|}{(D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2}} \ge b \Big\} \\ &\sim \Pr\Big\{ \max_{n_0 \le k \le n_1} \frac{U_k}{(D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2}} \ge b \Big\} \\ &= \Pr\Big\{ U_{n_1} \ge b(D_{n_1} / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2} \Big\} + \sum_{k=n_0}^{n_1 - 1} \int_{\xi > b(D_k / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2}} \\ &\quad \Pr\Big\{ \max_{k \le k+i \le n_1} \frac{U_{k+i}}{(D_{k+i} / \boldsymbol{j}_0' \Lambda^{-1} \boldsymbol{j}_0)^{1/2}} < b | U_k = \xi \Big\} \Pr(U_k \in d\xi), \qquad (3.1)\end{aligned}$$

we need to develop approximations for the conditional probability in the above integral.

Let $a_{kl} = \boldsymbol{j}_k' \Lambda^{-1} \boldsymbol{j}_l$, $A_{l_0, m_0}(l_1, m_1) = \sum_{m=m_0+1}^{m_1} \sum_{l=l_0+1}^{l_1} \Lambda^{-1}(l, m)$ and $A(i,j) = A_{k,k}(k+i, k+j)$. Then simple algebra shows that

$$E[U_{k+i} | U_k = \xi] = \xi[1 - \{a_{0,0} A_{k,k}(k+i, n) - a_{k,0} A_{k,0}(k+i, n)\} / D_k]$$

and

$$\begin{aligned}\mathrm{Cov}[U_{k+i}, U_{k+j} | U_k = \xi] = {} & A(i,j) - A_{k,0}(k+i, n) A_{k,0}(k+j, n) / a_{0,0} \\ & - [a_{0,0} A_{k,k}(k+i, n) - a_{k,0} A_{k,0}(k+i, n)] \\ & \cdot [a_{0,0} A_{k,k}(k+j, n) - a_{k,0} A_{k,0}(k+j, n)] / (D_k a_{0,0}).\end{aligned}$$

Let $\Lambda^{-1}(i,j) = f(i/n, j/n)$ for some integrable function f. Then as $k, l \to \infty$ in such a way that $k/n \to t$ and $l/n \to s$, define

$$G(t,s) = \lim_{n\to\infty} \boldsymbol{j}_k' \Lambda^{-1} \boldsymbol{j}_l / n^\rho,$$
$$D_t = G(0,0)G(t,t) - G^2(0,t),$$
$$\text{and} \qquad B_t = \{G(0,0)/D_t\}^{1/2},$$

where, for fixed i and j and some functions C_1, C_2 and C_3,

$$\rho = \begin{cases} 1, & \text{if } A(i,j) = C_1(t)\min(i,j) + C_2(t) + o(1), \\ 2, & \text{if } A(i,j) = C_3(t)ij + o(1). \end{cases}$$

Since the conditional distribution of (U_{k+i}, U_{k+j}) given U_k is determined by the matrix Λ^{-1} or $A(i,j)$, we consider the three different cases of $A(i,j)$ in Theorem 1.

Theorem 1. *Suppose that n, n_0, n_1 and $b \to \infty$ in such a way that $n_i/n \to t_i$ $(i = 0,1)$ and $b^2/n \to c^2$. Let ν be a function defined to be*

$$\nu(x) = 2x^{-2}\exp\Big\{-2\sum_{n=1}^{\infty} n^{-1}\Phi(-xn^{1/2}/2)\Big\},$$

where Φ and ϕ are the standard normal distribution and density functions, respectively. Then for some functions $C_1, C_{21}, C_{22}(>0)$ and C_3, as $k/m \to t$,
Case 1. *If $A(i,j) = C_1(t)\min(i,j) + o(1)$,*

$$\alpha \sim 2(1-\Phi(b)) + b\phi(b)\int_{t_0}^{t_1} \nu(c\{\gamma_t/D_t\}^{1/2})\gamma_t/D_t\,dt, \tag{3.2}$$

where $\gamma_t = C_1(t)G(0,0)$,
Case 2. *If $A(i,j) = C_{21}(t)\min(i,j) + C_{22}(t) + o(1)$,*

$$\begin{aligned}\alpha &\sim 2n\int_{t_0}^{t_1}\int_{-\infty}^{0}\{\Phi(b - v/\sigma_t) - \Phi(b)\}\phi((v-\mu_{v,t})/\sigma_{v,t})/\sigma_{v,t}\,dvdt \\ &\sim \frac{2n\phi(b)}{b}\int_{t_0}^{t_1}\{\Phi(-\mu_{v,t}/\sigma_{v,t}) - \Phi(\mu_{v,t}/\sigma_{v,t})\}dt, \end{aligned}\tag{3.3}$$

where $\sigma_t = n^{1/2}B_t$, $\mu_{v,t} = -cC_{22}(t)/(2B_t)$, and $\sigma^2_{v,t} = C_{22}(t)$,
Case 3. *If $A(i,j) = C_3(t)ij + o(1)$,*

$$\alpha \sim 2(1-\Phi(b)) + 2\phi(b)\int_{t_0}^{t_1}\{h(t)\}^{1/2}dy/\pi^{1/2}, \tag{3.4}$$

where

$$h(t) = B_t^2 \left\{ f(t,t) - \left(\int_0^1 f(t,w)dw \right)^2 / G(0,0) \right\} /2 + B_t^4 \left\{ \int_t^1 f(t,w)dw + G(0,t) \int_0^1 f(t,w)dw/G(0,0) \right\}^2 .$$

If the X's are iid observations, then $A(i,j) = \min(i,j) + o(1)$, $\gamma_t = 1$ and $D_t = t(1-t)$. It is easy to check that (3.2) with these γ and D is equivalent to (3.12) of Siegmund (1986), whose accuracy has been proved in a series of papers on change-point problems such as Siegmund (1986) and James, James and Siegmund (1987). If the X's are from an $AR(1)$ process, then it can be easily checked that Λ^{-1} satisfies the condition of Case 2 in Theorem 1 with $C_{21} = (1-\beta)/(1+\beta)$, $C_{22} = ((1+\beta^2)/(1-\beta^2))I(\beta < 0) + (2\beta/(1-\beta^2))I(\beta > 0)$ and $B_t = \{C_{21}t(1-t)\}^{1/2}$, where $I(\cdot)$ is an indicator function. Table 2 indicates the accuracy of the approximations (3.2) and (3.3) for some specific correlation functions. We observe that the approximation (3.3) works reasonably well even for a strong autocorrelation. Approximate p-values for (2.1) with $AR(p)$ observations can be obtained in a similar way as in the $AR(1)$ case. Implementing (3.4) requires a little more work, and we will add discussions on it in the following section.

Remark 1. As is to be shown in the Appendix, an approximate p-value in Case 2 may include more terms. Since the other terms are usually negligible, Theorem 1 reports only the term with a significant contribution.

Remark 2. Although the Gaussian property of U has been obtained via the normality of original observations, the zero mean Gaussian process U might also serve as a large sample approximation to the conditional distribution of T_k given $T_n = 0$ even for a nonnormal underlying distribution. Even if the x's are not normally distributed, T_k and T_0 will be asymptotically normal unless the departure from normality is severe. However, it is anticipated that the convergence rate would depend on Λ and we plan to pursue this problem in a future paper.

Table 3. Accuracy of the Approximation (3.3) with unknown σ and $x_i \sim AR(1)$ with autocorrelation parameter β, $n = 40$, $n_0 = 4$, $n_1 = 36$

	$\beta = -0.7$		$\beta = -0.4$		$\beta = 0.2$		$\beta = 0.8$	
probability	b	(3.3)	b	(3.3)	b	(3.3)	b	(3.3)
0.10	2.46	0.125	2.47	0.130	2.55	0.109	2.90	0.119
0.05	2.67	0.073	2.68	0.075	2.74	0.066	3.07	0.068
0.01	3.12	0.020	3.13	0.020	3.09	0.024	3.40	0.022

Note: b is the tail percentiles of (2.1) and is estimated by a Monte Carlo experiment with 10,000 repititions.

Remark 3. The approximations in Theorem 1 are obtained under the assumption of known σ and Λ. Hence they cannot be applied directly to the case of unknown σ and Λ. A natural modification would be to evaluate the approximations with suitably estimated values, say $\hat{\sigma}$ and $\hat{\Lambda}$. We simulated the tail percentiles, $(\tilde{b})$, of (2.1) with $\hat{\sigma}$ defined in Section 2 and $\hat{\beta} = \sum_{i=2}^{n}(x_i - \bar{x})(x_{i-1} - \bar{x})/\sum_{i=2}^{n-1}(x_i - \bar{x})^2$, and evaluated (3.3) with these $\tilde{b}$ and $\hat{\beta}$. Since $\hat{\sigma}$ and $\hat{\beta}$ are consistant estimators of σ and β at least asymptotically, (3.3) is expected to serve as a large sample approximation for the p-value of the test. However, it is observed in Table 3 that for the sample of size 40, (3.3) may be used as an approximate p-value for moderate sizes of the tail probability, usually greater than 0.1, but it overestimates the small tail probabilities about 120-250%. Kim (1988) has observed similar overestimation in the case of independent observations. Simulation studies of Kim (1988) have shown that $\hat{\sigma}$ tends to significantly overestimate σ when the observations yield the value of the test statistic at the very end of the right tail of its distribution. It has been, however, observed that the bias gets smaller as sample size increases. With correlated observations, our simulation study also indicates similar behavior of $\hat{\sigma}$ and it is expected that if the sample size is over 100, the bias should be within the practically tolerable range. As discussed in Henderson (1986), the estimators which are robust to the mean change can improve the accuracy of the approximations, but this problem needs further work.

Remark 4. If the alternative hypothesis specifies a special correlation structure such as $AR(p)$, the LRS would have a different form. For example, if we wish to test a level shift in an $AR(p)$ process with unknown original mean level and unknown autocorrelation parameter, it can be shown that the LRS converges to the maximum of the pinned-down Brownian Bridge process. In such cases, the statistic (2.1) may lose some power, but it has the advantage of generality.

Table 4. Powers of the LRT and Henderson's Test $x_i \sim AR(1)$ with autocorrelation parameter β, $n = 40$, $n_0 = 4$, $n_1 = 36$ $\delta = 1$, significance level = .05

	$\beta = -0.8$		$\beta = -0.4$		$\beta = -0.2$		$\beta = 0$	
τ	LRT	Bayes	LRT	Bayes	LRT	Bayes	LRT	Bayes
10,30	1.0000	0.9999	0.9683	0.9249	0.8637	0.8012	0.7116	0.6591
20	1.0000	0.9999	0.9937	0.9939	0.9441	0.9550	0.8359	0.8632

	$\beta = 0.2$		$\beta = 0.4$		$\beta = 0.8$	
τ	LRT	Bayes	LRT	Bayes	LRT	Bayes
10,30	0.5677	0.5235	0.4277	0.4035	0.2594	0.2174
20	0.6861	0.7283	0.5264	0.5727	0.2805	0.2720

4. Power Comparisons and Discussion

Henderson (1986) used a Bayesian approach to handle the problem discussed in Section 2. Table 4 compares the power of the likelihood ratio test discussed in Section 2 with that of the Bayes test proposed by Henderson. Since the alternative distribution of (2.1) depends only on the size of the shift, $\delta = \mu^* - \mu$, we do not need to know the value of the initial mean. In Table 4, only the case of sample size 40, $\delta = 1$ and two-sided significance level 0.05 is considered. The power of the LRT with $n_0 = 4$ and $n_1 = 36$ is estimated by Monte Carlo experiments with 10,000 repetitions and the power of the Henderson's test evaluated by the formula in Section 2 of Henderson with u replaced by w in (2.2) of Henderson. It is observed that the statistics behave similarly, but the LRT performs a little better when the change occurs near the end points.

Henderson also applies his method to material accountancy data assuming that both σ and Λ are known. The data discussed in Henderson such that correlation between X_i and X_j is ρ if $|i-j| = 1$ and is zero otherwise, satisfies Case 3 in Theorem 1 with very weak dependency between the observations. Instead of putting extra effort on computing the approximations suggested in Section 3, we estimate the p-value by simulation and obtain the p-value of .015 which is highly significant. Examining the values of the likelihood ratio statistic at each i of the material accountancy data, we can obtain the point estimate of the change-point. Since the maximum likelihood estimate of τ is the value of k at which Q_k is maximized, the point-estimate is 11, which is equivalent to Henderson's point estimate.

The purpose of this paper has not been to present a polished solution to particular problems, but rather to indicate the applicability of the likelihood ratio test to detect a mean change in correlated observations. It has been shown that the p-value of the test for a level shift in an $AR(p)$ process can be approximated by (3.3). In case of more general correlation, implementation of (3.4) requires further computation for each specific case of interest. For example, if we are interested in detecting a mean change in repeatedly measured data, are able to parameterize Λ using parameters $\lambda_1, \ldots, \lambda_q$, and are able to assume that such parameters are known, then it would be possible to compute the right hand side of (3.4). If it is preferred to assume the λ's are unknown, we need to derive the LRT, at least in an approximate form, and to study each case separately. The implementation of (3.4) will be studied in a future paper with some other interesting correlation structures.

Acknowledgment

This research was supported in part by NSF Grant DMS 9111904 and NIH Grant CA 56713. A part of the results, was derived during the author's stay at the Dana-Farber Cancer Institute and the Department of Biostatistics at Harvard University.

Appendix

Since the marginal distributions of the U_k's are known, it is required to derive approximations for the conditional probabilities in the integrand of (3.1). Depending on the nature of the matrix Λ, we obtain different approximations. We follow the notation used in the main text.

Lemma. *Suppose that n, n_0, n_1 and $b \to \infty$ in such a way that $n_i/n \to t_i$ $(i = 0, 1)$ and $b^2/n \to c^2$. Then for fixed i and j, and some functions $C_1, C_{21}, C_{22} (> 0)$ and C_3 defined in Theorem 1, as $k/n \to t$,*

Case 1.

$$\begin{aligned}&\Pr(U_{k+i} < b\{D_k/\boldsymbol{j}_0'\Lambda^{-1}\boldsymbol{j}_0'\}^{1/2} \quad \textit{for all} \quad 1 \le i \le n_1 - k | U_k = \xi)\\ \sim\ &\Pr{}_{-\eta_t,\lambda_t^2}(\max_{i\ge1} S_i < -x),\end{aligned}$$

where S_k is the partial sum of k-iid variables with mean $\eta_t = cf(t,t)B_t/2$ and variance $\lambda_t^2 = f(t,t)$.

Case 2.

$$\begin{aligned}&\Pr(U_{k+i} < b\{D_k/\boldsymbol{j}_0'\Lambda^{-1}\boldsymbol{j}_0'\}^{1/2} \quad \textit{for all} \quad 1 \le i \le n_1 - k | U_k = \xi)\\ \sim\ &\Pr{}_{-\eta_t,\lambda_t^2,\mu_{v,t},\sigma_{v,t}^2}(\max_{i\ge1} S_i + V < -x),\end{aligned}$$

where V is the random variable whose mean $\mu_{v,t} = -cC_{22}(t)/(2B_t)$, variance $\sigma_{v,t}^2 = C_{22}(t)$ and is independent of the S_k's $k = 1, 2, \ldots$.

Case 3.

$$\mathrm{Cov}[\tilde{U}_k, \tilde{U}_{k+i}] = 1 - (i/n)^2 A_t + o\Big(|\frac{i}{n}|^2\Big),$$

where

$$\tilde{U}_k = U_k/\{D_k/\boldsymbol{j}_0'\Lambda^{-1}\boldsymbol{j}_0'\}^{1/2},$$

$$\textit{and} \quad A_t = \left\{C_3(t) - \Big(\int_0^1 f(t,w)dw\Big)^2/G(0,0)\right\} G(0,0)/(2D_t).$$

Proof of the Lemma is straightforward and is omitted here.

Proof of Theorem 1.

Case 1. Using the results in problem (8.13) of Siegmund (1985).

$$\begin{aligned}\alpha &\sim \sum_{k=n_0}^{n_1} \int_{x>0} \Pr{}_{-\eta_t,\lambda_t^2}(\max_{i\ge1} S_i < -x)\phi(b + x/\sigma_k)/\{\sigma_k(2\pi)^{1/2}\}dx,\\ &\sim \sum_{k=n_0}^{n_1} \frac{\phi(b)}{\sigma_k} \int_{x>0} \Pr{}_{-\zeta,1}(\max_{i\ge1} S_i < -x)\exp(-bx/\sigma_k)dx\\ &\sim \frac{b\phi(b)}{2} \int_{t_0}^{t_1} \nu(x\{\gamma_t/D_t\}^{1/2})\gamma_t/D_t\, dt,\end{aligned}$$

where $\sigma_k^2 = D_k/\boldsymbol{j}_0'\Lambda^{-1}\boldsymbol{j}_0$, $\zeta = \eta_t/\lambda_t$.
Case 2. Using arguments in Chapter 8 of Siegmund (1985),

$$\begin{aligned}
\alpha &\sim 2\sum_{k=n_0}^{n_1}\int_{x>0}\Pr{}_{-\eta_t,\lambda_t^2}(\max_{i\geq 1} S_i + V < -x)\phi(b+x/\sigma_k)/\{\sigma_k(2\pi)^{1/2}\}dx,\\
&\sim 2\sum_{k=n_0}^{n_1}\int_{x>0}\int_{-\infty}^{\infty}\Pr{}_{-\eta_t,\lambda_t^2}\Big(\max_{i\geq 1} S_i < -(x+v)\Big)\\
&\qquad\qquad \phi((v-\mu_{v,t})/\sigma_{v,t})\phi(b+x/\sigma_k)/\{\sigma_{v,t}\sigma_k(2\pi)\}dx,\\
&\sim 2\sum_{k=n_0}^{n_1}\Big[\int_{v>0}\int_{x>0}\cdot\, dxdv + \int_{v<0}\int_{x>-v}\cdot\, dxdv + \int_{v<0}\int_{0<x<-v}\cdot\, dxdv\Big]\\
&\sim 2(\mathrm{I}+\mathrm{II}+\mathrm{III}-\mathrm{IV}),
\end{aligned}$$

where

$$\mathrm{I} = n^{1/2}\phi(b)\int_{t_0}^{t_1}B_t\lambda_t\int_{v>0}g(v/\lambda_t)\exp(2v\zeta/\lambda_t)\phi((v-\mu_{v,t})/\sigma_{v,t})/\sigma_{v,t}\,dv,$$

with $g(y) = \{n^{1/2}/[2\phi(b)]\}$ ((9.92) of Siegmund with μ^* and x replaced by 2μ and y, respectively)

$$\begin{aligned}
\mathrm{II} &= n^{1/2}\phi(b)\int_{t_0}^{t_1}B_t\eta_t\nu(2\zeta)\Phi(\mu_{v,t}/\sigma_{v,t})dt,\\
\mathrm{III} &= n\int_{t_0}^{t_1}\int_{v<0}[\Phi(b-v/\sigma_t)-\Phi(b)]\phi((v-\mu_{v,t})/\sigma_{v,t})/\sigma_{v,t}\,dvdt\\
&\sim \frac{n\phi(b)}{b}\int_{t_0}^{t_1}[\Phi(-\mu_{v,t}/\sigma_{v,t})-\Phi(\mu_{v,t}/\sigma_{v,t})]dt,\\
\mathrm{IV} &= n^{1/2}\phi(b)\int_{t_0}^{t_1}B_t\nu(2\zeta)\sigma_{v,t}[-\mu_{v,t}\Phi(\mu_{v,t}/\sigma_{v,t})/\sigma_{v,t}+\phi(-\mu_{v,t}/\sigma_{v,t})]dt.
\end{aligned}$$

Since I+II-IV are negligible compared to III, we may use III for a simple approximation.
Case 3. Since $\mathrm{Cov}[\tilde{U}_k, \tilde{U}_{k+i}] = 1-(i/n)^2A_t + o(|\frac{i}{n}|^2)$, arguments in chapter 12 of Leadbetter, Lindgren, and Rootzen (1983) lead to (3.4). (See (12.1.1) in Leadbetter et al.)

References

Bagshaw, M. and Johnson, R. A. (1975). The effect of serial correlation on the performance of CUSUM tests II. Technometrics **17**, 73-80.

Box, G. E. P. and Tiao, G. C. (1965). A change in level of non-stationary time series. Biometrika **52**, 181-192.

Carlstein, E. (1988). Nonparametric change-point estimation. Ann. Statist. **16**, 188-197.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Mathe. Statist. **35**, 999-1018.

Henderson, R. (1986). Change-point problem with correlated observations, with an application in material accountancy. Technometrics **28**, 381-389.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. Biometrika **57**, 1-17.

James, B., James, K. L. and Siegmund, D. (1987). Tests for a change-point. Biometrika **74**, 71-83.

Kim, H.-J. (1988). Change-point problems in regression. Technical Report, Stanford University.

Leadbetter, M. R., Lindgren, G. and Rootzen, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York.

Lehmann, E. L. (1986). Testing Statistical Hypotheses, 2nd edition. John Wiley, New York.

Nagaraj, N. K. (1990). Two-sided tests for change in level for correlated data. Statistical Papers **31**, 181-194.

Page, E. S. (1954). Continuous inspection schemes. Biometrika **41**, 100-115.

Sen, A. and Srivastava, M. S. (1975). Some one-sided tests for change in level. Technometrics **17**, 61-64.

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

Siegmund, D. (1986). Boundary Crossing probabilities and statistical applications. Ann. Statist. **14**, 361-404.

Smith, A. F. M. (1975). A Bayesian approach to inference about change-point in a sequence of random variables. Biometrika **62**, 407-416.

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

Worsley, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika **73**, 91-104.

Yashchin, E. (1993). Performance of CUSUM control schemes for serially correlated observations. Technometrics **35**, 37-52.

Zhao, M. (1993). Change-point problems in software and hardware reliability. Comm. Statist. Theory Methods **22**, 757-768.

Department of Mathematics, Syracuse University, Syracuse, NY 13244, U.S.A.

(Received August 1993; accepted January 1995)