ACTUARIAL RESEARCH CLEARING HOUSE
1994 VOL. 1

# Diagnostics and Tests for Abrupt Change with an Application to a Two-state Markov Chain

J. L. VAN WYK AND F. LOMBARD
DEPARTMENT OF STATISTICS
RAND AFRIKAANS UNIVERSITY
P. O. BOX 524
AUCKLAND PARK, 2006, SOUTH AFRICA

## 1. Introduction

Sometimes a shift in parameter of a process can be related to an apparent causal event at a time which can be precisely specified. However, in many cases potential causal events cannot be identified with reasonable confidence, and timing of the changepoint is uncertain. We present diagnostics and tests for the retrospective identification of abrupt change in the mean level of a process where the mean and variance of the underlying random variables cannot be functionally separated. For example, in the case of binary data $X_1, \ldots, X_n$ with $\Pr(X_i = 1) = 1 - \Pr(X_i = 0) = p$, we have $\mathrm{E}(X_i) = p$ and $\mathrm{var}(X_i) = p(1-p)$. A simple time-sequence plot of such observations is usually uninformative regarding possible changes in the parameter. Consequently we propose to smooth the data by taking a moving average and then applying a variance stabilizing transformation. This will ensure that the observations have constant variance (at least asymptotically). Various test statistics can be constructed from the transformed process. The two proposed diagnostics involve cumulative sums and principal components, respectively.

## 2. Development of diagnostics

Suppose $X_1, \ldots, X_n$ is a sequence of independent random variables having means $\mu_i$ and variances $\sigma^2(\mu_i)$, $i = 1, \ldots, n$. Consider the null hypothesis of constancy of the mean

$$H_0 : \ \mu_i = \mu \ ; \ i = 1, \ldots, n$$

versus the changepoint alternative

$$H_\tau : \ \mu_i = \mu + \Delta Ind(\tau < i \leq n),$$

where the initial level, $\mu$, the extent of the change, $\Delta$ and the changepoint $\tau$ is unknown. $Ind(\cdot)$ denotes the indicator function.

Define the $l$-dependent sequence of moving averages

$$Y_{r,l_n} = l_n^{-1} \sum_{j=0}^{l_n-1} X_{r+j} \quad ; \qquad r = 1, 2, \ldots, m, \tag{2.1}$$

where $m = n - l_n + 1$. To simplify the notation we choose to write $l$ for $l_n$. Let $f(\cdot)$ denote a variance stabilizing transformation, chosen to satisfy the differential equation $f'(\mu)\sigma(\mu) = 1$. Assume that $l \to \infty$, $m \to \infty$ in such a way that $l/m \to 0$, $l^3/m^2 \to 0$ and $l^2/m \to \infty$. It follows from these assumptions that $l = n^{1-\delta}$, where $\frac{1}{3} \le \delta \le \frac{1}{2}$. Also, assume that $\mu_6 \equiv \mathrm{E}(X_i - \mu)^6 < \infty$ and that $|f''(x)|$ is a convex function of $x$.

Let $\overline{f(Y_l)} = m^{-1} \sum_{i=1}^m f(Y_{i,l})$, and define

$$B_{m,l}(k) = m^{-1/2} \sum_{r=1}^{k} \left\{ f(Y_{r,l}) - \overline{f(Y_l)} \right\}, \quad k = 1, \ldots, m.$$

Constancy of the mean can then be checked via a test statistic of the form

$$S_{m,l} = \sum_{k=1}^{m} a_m(k) B_{m,l}^2(k) \,, \tag{2.2}$$

where

$$a_m(k) = \int_{(2k-1)/2m}^{(2k+1)/2m} \varphi(u) du$$

with $\varphi(u)$, $0 < u < 1$, a non-negative weight function such that $\int_0^1 u(1-u)\varphi(u)du < \infty$. It follows from Theorem 2 of MacNeill (1974) that the limit distribution of this statistic is that of

$$\int_0^1 \varphi(t) B^2(t) dt,$$

with $(B(t)\,;\ 0 \le t \le 1)$ denoting a standard Brownian bridge process. Percentage points for a wide range of weight functions are known. Below we only consider the weight function $\varphi(u) \equiv 1$, $0 \le u \le 1$, which gives $a_m(k) = m^{-1}$.

The cusum diagnostic consists of plotting $B_{m,l}(k)$ against $k$. A sustained change of direction in the plot is indicative of a changepoint. When the change is judged significant after formal testing, an estimate of the point of change is given by the maximizer over $k$ of $|B_{m,l}(k)|$.

The second diagnostic will only be described briefly since formal test statistics are still under investigation. Let $\check{Y}_{r,l} = Y_{r,l} - \overline{Y_l}$, and let $\check{Y}_l$ denote the random vector $(\check{Y}_{1,l}, \ldots, \check{Y}_{m,l})^{\mathsf{T}}$. Under the null hypothesis the elements of the covariance matrix of $\check{Y}_l$ are given by $\sigma^2(\mu)K^*(r,s)$. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_{n-l} > 0$ denote the non-zero eigenvalues of $K^*$, with corresponding normalized eigenvectors $v_1, \ldots, v_{n-l}$. (There also exists a non-zero eigenvector $v_{n-l+1}$, corresponding to the eigenvalue $\lambda_{n-l+1} = 0$.) The principal components of $\check{Y}_l$ are given by

$$\phi_j(\check{Y}_l) \equiv \phi_j = v_j^{\mathsf{T}} \check{Y}_l, \qquad j = 1, \ldots, n - l.$$

Since the eigenvectors are orthonormal it follows that $\mathrm{cov}(\phi_j, \phi_k) = \sigma^2(\mu)\lambda_j \,\mathrm{Ind}(j = k)$, i.e. the standardized principal components, $\phi_j/\lambda_j^{1/2}$, are uncorrelated with common variance $\sigma^2(\mu)$.

Now, consider the case where $X_1, \ldots, X_n$ are independent, but with non-constant mean. Put $\nu_r = \mathrm{E}(Y_{r,l})$, $\tilde{\nu}_r = \mathrm{E}(\tilde{Y}_{r,l})$, $\nu = (\nu_1, \ldots, \nu_m)^\top$ and $\tilde{\nu} = (\tilde{\nu}_1, \ldots, \tilde{\nu}_m)^\top$. Then

$$\mathrm{E}\left(\phi_j\left(\tilde{Y}_l\right)\right) = \mathrm{E}\left(v_j^\top \tilde{Y}_l\right) = v_j^\top \tilde{\nu} = \phi_j\left(\tilde{\nu}\right).$$

It follows easily that

$$\begin{aligned} &\nu_1 = \cdots = \nu_m \text{ if and only if} \\ &\phi_j\left(\tilde{\nu}\right) = v_j^\top \tilde{\nu} = 0 \text{ for } j = 1, \ldots, n - l. \end{aligned} \tag{2.3}$$

Hence, the stationarity or non-stationarity of $\mathrm{E}(Y_{r,l})$ can be determined by analyzing the principal components of $\tilde{Y}_l$.

For the variance-stabilized process, define $\tilde{f}(Y_l) = \left(\tilde{f}(Y_{1,l}), \ldots, \tilde{f}(Y_{m,l})\right)^\top$, where $\tilde{f}(Y_{r,l}) = f(Y_{r,l}) - \overline{f(Y_l)}$. The covariance matrix of $\tilde{f}(Y_l)$ is then approximately $K^*$. Letting

$$\psi_j \equiv \psi_j(\tilde{f}(Y_l)) = v_j^\top \tilde{f}(Y_l) \approx \left(v_j^\top \tilde{Y}_l\right) f'(\mu) = \phi_j f'(\mu), \tag{2.4}$$

we see that

$$\mathrm{cov}\left(\psi_j \lambda_j^{-1/2};\ j = 1, \ldots, m\right) \approx I,$$

the identity matrix.

From (2.4) and (2.3) we can check for stationarity of the mean by plotting $v_j^\top \tilde{f}(Y_l)$ against $j$.

The procedures above are applied to the Dow Jones Industrial average for the period July 1, 1971 through August 2, 1974, a data set also considered by Hsu (1979). Our results agree with those obtained by Hsu who analyzed the data by different methods.

## 3. Two-state Markov chains

Consider a first order Markov chain $X_0, X_1, \ldots, X_n$ with state space $\{0, 1\}$. Denote the transition probability matrix at time $t$ by $P(t)$, where

$$p_{ij}(t) = \Pr\left(X_t = j \mid X_{t-1} = i\right),\ i, j = 0, 1.$$

Consider the null hypothesis of stationarity,

$$\mathbf{H}_0 : P(t) = P, \qquad t = 1, \ldots, n$$

where $p_{ij} > 0$, and the alternative hypothesis

$$\mathbf{H}_\tau : P(t) = \begin{cases} P & t = 1, \ldots, \tau \\ P^* & t = \tau + 1, \ldots, n \end{cases}$$

where

$$P^* = P + \begin{bmatrix} \delta_0 & -\delta_0 \\ -\delta_1 & \delta_1 \end{bmatrix}, \tag{3.1}$$

with $P$, $\delta_0$, $\delta_1$ and $\tau \in \{1, \ldots, n-1\}$ unknown.
Assume the null hypothesis of stationary transition probabilities, in other words,

$$P(t) = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix},$$

independent of $t$. The waiting times between transitions are

$$\begin{aligned} T_0 &\equiv 0 \\ T_1 &\equiv \min(t > T_0 \ : \ X_t \neq X_{T_0}) \\ T_2 &\equiv \min(t > T_1 \ : \ X_t \neq X_{T_1}), \ etc. \end{aligned}$$

Let, for $i = 1, 2, \ldots$

$$W_i = \begin{cases} T_{2i-1} - T_{2i-2} & \text{if } X_{T_0} = 0 \\ T_{2i} - T_{2i-1} & \text{if } X_{T_0} = 1 \end{cases}$$

and

$$V_i = \begin{cases} T_{2i} - T_{2i-1} & \text{if } X_{T_0} = 0 \\ T_{2i-1} - T_{2i-2} & \text{if } X_{T_0} = 1 \end{cases}.$$

Then, conditional on the value of $X_{T_0}$, $(W_1, W_2, \ldots)$ and $(V_1, V_2, \ldots)$ are two independent sequences. $W_1, W_2, \ldots$ are independent copies of a geometric random variable $W$ with mass function

$$\Pr(W = w) = p_{00}^{w-1} p_{01} \ ; \ w \geq 1$$

while $V_1, V_2, \ldots$ are independent copies of a geometric random variable $V$ with mass function

$$\Pr(V = v) = p_{11}^{v-1} p_{10} \ ; \ v \geq 1.$$

Note that each of the times $W_1, W_2, \ldots$ is computed from the same initial state $X_{T_0}$, and thus the Markov property guarantees that $W_1, W_2, \ldots$ are independent and identically distributed.
To test for abrupt change in the transition probability from state 0 to state 1 we have

$$\mathrm{E}(W) = p_{01}^{-1} \equiv \mu$$

and

$$\operatorname{var}(W) = p_{00}p_{01}^{-2} = \mu(\mu-1) \equiv \sigma^2(\mu).$$

Note that E($W$) increases when $p_{01}$ decreases, with a subsequent clustering of zeros. From this it follows that for small off-diagonal elements of the transition probability matrix, $n$ must be quite large in order to obtain a reasonable number of $W$'s and $V$'s.
The transformation $f$ must therefore be chosen so that

$$f'(\mu) = (\mu(\mu-1))^{-1/2}.$$

The solution to this equation is given by $f(\mu) = 2\cosh^{-1}\left(\sqrt{\mu}\right)$.[1] Letting

$$Y_{\tau,l} = l^{-1}\sum_{j=0}^{l-1} W_{\tau+j},$$

we can then test for constancy of the transition probability from 0 to 1 via the statistic (2.2) which, in this special case, we denote by $T_w^2$. In the same way the $V$ - process can be used to test for abrupt change in the transition probability from state 1 to state 0. Denote this statistic by $T_v^2$. The $W$ and $V$ processes are independent. One way of combining them to give a test for constancy of the $0 \to 1$ and $1 \to 0$ transition probabilities is to take

$$T^2 = T_w^2 + T_v^2, \tag{3.2}$$

with asymptotic distribution given by that of

$$\int_0^1 \left(B_1^2(t) + B_2^2(t)\right) dt.$$

Here $B_1$ and $B_2$ are independent Brownian bridge processes. Critical values for this distribution are also known.

## 4. Simulation

The results in the previous section on Markov chains were compared using simulated data. Two methods were considered to generate the Markov chain. Consider a fixed transition probability matrix $P$ and condition on the initial state, $X_0 = i$. Under the null hypothesis of stationarity, for $t = 1; 2; \ldots; n$, if $X_{t-1} = 0$, generate one observation from a Bernoulli distribution with success probability $p_{01}$, that is, $\Pr(X_t = 1 \mid X_{t-1} = 0) = p_{01}$, and if $X_{t-1} = 1$, generate one observation from a Bernoulli distribution with success probability $p_{11}$. Repeat the procedure under the alternative for $t = 1; \ldots; \tau$. For $t = \tau+1; \ldots; n$ we use $P^*$ in (3.1).

[1] Some computer software products do not contain $\cosh^{-1}(x)$ as an implicit function. An easy computational formula is $\cosh^{-1}(x) = \ln\left(x + \sqrt{x^2-1}\right)$; $x \geq 1$.

The second method of simulating the Markov chain was based on the algorithm described by Wang & Scott (1989).
We used $N = 1000$ repetitions, a sample size of 120 and a lag of $l = 11$ to simulate the percentage points. These points were then used to compare powers for different values of $\tau$, $\delta_0$, $\delta_1$. For such a small sample size we focused mainly on the case $p_{00} = p_{11} = 0,5$; the initial state was fixed at $X_0 = 1$.
There did not appear to be any real change in the results produced by the two different methods of simulation - as measured by the percentage points obtained and the relative merits of one test vis-a-vis another. We opted to persist with the first method for simulating the Markov chain.
This study confirmed our intuition regarding the power of our procedures. Based on waiting times, a fairly large amount of data would in general be required to achieve respectable power. As expected, our procedure is most powerful for late changes in the data sequence, and correspondingly less powerful for early changes. This could be partially resolved by employing a weight function which gives more weight to early changes.
If we fix the changepoint and vary the extent of change, our procedure has most power for relatively small diagonals and much less power for larger diagonals. In other words, it is particularly sensitive to frequently varying processes, rather than to processes with greater persistence. The situation improves markedly for larger sample sizes.
Also, our simulation confirmed the known speedy rate of convergence to the asymptotic distribution of the $\sum$-type test criteria. Some simulated percentage points for the $T^2$- test is given below (the asymptotic percentage points are given in brackets).

| $t$ | $\Pr(T^2 \leq t)$ |
|---|---|
| 0.604(0.607) | 0.90 |
| 0.757(0.748) | 0.95 |
| 1.116(1.074) | 0.99 |

## REFERENCES

[1] Hsu, D. A. (1979), Detecting Shifts of Parameter in Gamma Sequences with Applications to Stock Price and Air Traffic Flow Analysis, J.A.S.A., Vol. 74, pp. 31-40.

[2] MacNeill, I. B. (1974), *Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion*, ANN. STATIST., 2, pp. 950-962.

[3] Wang, D. Q. and Scott, D. J. (1989), *Testing a Markov Chain for Independence*, COMMUN. STATIST.-METH., Vol. 18, pp. 4085-4103.