Estimation of Change-point and Post-change Means by Adaptive CUSUM Procedures

Yanhong Wu
Department of Mathematics, California State University Stanislaus,
Turlock, CA 95382 USA

ABSTRACT

For a sequence of normal random variables, adaptive CUSUM procedures are used to detect mean change by using a sequence of adaptive sequential tests. An alarm is made when the adaptive CUSUM process crosses the boundary. We first give the asymptotic results for the average in-control and out-of-control run lengths. Then the biases of the change-point and post-change means are studied. Particular attentions are paid to the cases when the post-change mean is a sudden shift or linearly increasing. Nile river flow data and global temperature data are used for illustration.

Keywords: Adaptive CUSUM procedure; Biases; Nonlinear renewal theorem; Strong renewal theorem.

1 Introduction

For a sequence of independent normal random variables $\{X_i\}$ for $i = 1, 2, ...$ of unit variance with pre-change mean 0 for $i \le \nu$ and post-change μ for $i > \nu$, the CUSUM procedure is defined to make an alarm at the point

$$\tilde{N} = \inf\{n > 0 : T_n = \max(0, T_{n-1} + \delta(X_n - \delta/2)) > d\},$$

for $T_0 = 0$ and δ is selected as a target value for μ. Note that the classical procedure can be interpreted as a sequence of sequential probability ratio tests with two-sided boundary $[0, d)$. Whenever a test ends with crossing the lower boundary 0, a new test starts again until a test ends with crossing the boundary d. The change-point is conveniently estimated as the starting point of the last test, and the post-change mean is estimated as the sample mean of the last test.

When the post-change mean μ is different from δ, the efficiency becomes low in terms of the average delay detection time. Also, it is only restricted to the constant post-change mean with a single jump. To overcome these limitation, the adaptive CUSUM procedure can be used by using a sequence of adaptive sequential tests (Robbins and Siegmund (1973, 1974)) in which the post-change parameters are adaptively estimated for each test.

More specifically, we consider a slightly general model by assuming that the post change mean $\mu_i(\theta)$ for $i > \nu$ which depends on some unknown parameter θ and $i - \nu$.

Based on the observations $\{X_{k+1}, ..., X_n\}$ we define $\theta_{k+1,n} = \theta(X_{k+1}, ..., X_n)$ as an estimator of θ and $\theta_{k+1,k} = \theta_0 = \delta$ for $k \geq 1$. Also for notational convenience, we denote by $\theta_n = \theta_{1,n}$. The following algorithm defines the adaptive CUSUM procedure:

Adaptive CUSUM Procedure:

(i) Set $T_0 = 0$ and $\nu_0 = 0$. Define recursively for $n > 0$,

$$T_n = \max\{T_{n-1} + \mu_n(\theta_{\nu_{n-1}+1,n-1})(X_n - \frac{1}{2}\mu_n(\theta_{\nu_{n-1}+1,n-1}))\}.$$

(ii) If $T_n > 0$, reset $\nu_n = \nu_{n-1}$ and $\theta_{\nu_n+1,n} = \theta(X_{\nu_n+1}, ..., X_n)$;

(iii) If $T_n = 0$, update $\nu_n = n$, and $\theta_{n+1,n} = \delta$.

(iv) An alarm is made at the time $N = \inf\{n > 0 : T_n > d\}$. The change-point and post-change parameter are estimated as

$$\hat{\nu} = \nu_N; \quad and \quad \hat{\theta} = \theta_{\nu_N+1,N}.$$

A variety of adaptive CUSUM procedures have been discussed. Dragalin (1997) considered to use the sample mean. Yakir, Krieger, and Pollak (1999) and Krieger, Pollak, and Yakir (2003) considered the linear post-change model. Capizzi and Mascrotto (2003) considered an adaptive EWMA procedure. An adaptive Shiryayev-Roberts procedure using the adaptive estimators is considered in Lorden and Pollak (2005). Yashchin (1995) and Jiang, Shu, and Apley (2008) used the EWMA as the adaptive post-change mean estimator. Han, Tsung, and Wang (2010) proposed to use the last current observation as the estimator for the mean. Window-limited likelihood-based detection procedures can be seen in Lai (1995) by generalizing Siegmund and Venkatramen (1993).

In this paper, we consider the inference problem after the alarm by extending the results of Wu (2004, 2005, 2014). In Section 2, we first present some preliminary results of renewal theory related to the adaptive sequential test. In particular, a renewal theorem for a random walk with random drift is stated. The asymptotic results for average in-control and out-of-control run lengths are given in Section 3. In Section 4, we study the bias of the change point estimation and post-change parameter estimation. In Section 5, the results are generalized to more general post-change mean case. In particular, the case when the post-change mean is gradually linear increasing is considered. Simulation results are presented in Section 6 and applications to detecting changes in Nile river flow data and global warming are used for illustration.

2 Adaptive sequential tests and a nonlinear renewal theorem

In this section, we briefly review some basic results on the adaptive sequential tests which are presented in Robbins and Siegmund (1973, 1974). Lorden and Pollak (2005) also presented some basic results.

For given S_0 and μ_0, we define for $n \geq 1$,

$$S_n = S_0 + \sum_{i=1}^{n} \mu_{i-1}(X_i - \mu_{i-1}/2),$$

where $\mu_i = \mu_i(\mu_0, X_1, ..., X_i)$ is estimated adaptively from the history $H_i = \sigma(\mu_0, X_1, ..., X_i)$. To present the results clearly, in this section, we denote by $P_0(.)$ the probability measure when the mean of X_i's is $\mu = 0$, and by $P_\mu(.)$ when the mean is equal to μ.

We are interested in evaluating the boundary crossing probability $P_0(\tau_d < \infty|H_k)$, where

$$\tau_d = \inf\{n > 0 : S_n > d\}.$$

Define a changed measure $P_0^*(.|H_k)$ by

$$dP_0^*(.|H_k) = exp(S_n - S_k)dP_0(.|H_k).$$

We are interested in evaluating the probability $P_0(\tau_d < \infty|H_k)$ as $d \to \infty$. Suppose under $P_0^*(.|H_k)$, $\mu_{k+1,n} \to \mu_{k+1,\infty}$ almost surely as $n \to \infty$. The following theorem is due Woodroofe (1990) which states that the renewal theorem still holds by conditioning on the value of $\mu_{k+1,\infty}$.

Theorem 1: *As $d \to \infty$, assume $\mu_{k+1,n}$ converges to $\mu_{k+1,\infty}$ almost surely given H_k. Given $\mu_{k+1,\infty} = \mu$, define $S_n^* = \sum_{i=1}^n \mu(X_i^* - \mu/2)$ where $X_1^*, ..., X_n^*, ...$ are conditional i.i.d. $N(\nu, 1)$ random variables, and*

$$\tau_d^* = \inf\{n > 0 : \quad S_n^* > d\}.$$

Then as $d \to \infty$ in distribution,

$$(\mu_{k+1,\tau_d}, S_{\tau_d} - d) \to (\mu_{k+1,\infty}, R_\infty^*),$$

where $R_\infty^* = \lim R_d^* = \lim(S_{\tau_d^*} - d)$.

By using the renewal theorem, we can find

$$\begin{aligned} \lim P_0(\tau_d < \infty|H_k) &= \lim E_0^*[exp(-(S_{\tau_d} - S_k)|H_k] \\ &= e^{-d+S_k} E_0^*[\nu(\mu_{k+1,\infty})] \\ &= e^{-d-r_1(H_k)+S_k}, \end{aligned}$$

where $r_1(H_k) = -\ln E_0^*[\nu(\mu_{k+1,\infty})]$.

Example 1: (Recursive Mean Estimation) Let $\{X_1, ..., X_n, ...\}$ be i.i.d. normal random variables with mean μ and variance σ^2. For given tuning value $t > 0$ and initial estimator $\mu_k = \delta$ given H_k, we define

$$\mu_{k+1,n}(t, \delta) = \frac{t}{t + n - k}\delta + \frac{n - k}{t + n - k}\bar{X}_{k+1,n},$$

where $\bar{X}_{k+1,n} = (X_{k+1} + ... + X_n)/(n - k)$. In special, when $k = 0$, we denote by $\mu_n = \mu_n(t, \delta) = \mu_{1,n}(t, \delta)$. Recursively, one can show that

$$\begin{aligned} \mu_{k+1,n}(t, \delta) &= \mu_k + \sum_{j=k+1}^{n} \frac{1}{t + j - k}(X_j - \mu_{j-1}) \\ &= \mu_{k+1,n-1} + \frac{1}{t + n - k}(X_n - \mu_{k+1,n-1}). \end{aligned}$$

As $n \to \infty$, it can be shown that $\mu_{k+1,n}(t, \delta)$ almost surely converges to a normal variable with mean μ_k and variance $\sigma_k^2 = \sum_{j=1}^\infty 1/(t + k + j)^2$. Thus,

$$e^{-r_1(H_k)} = \int \nu(y)\frac{1}{\sigma_k}\phi\left(\frac{y - \mu_k}{\sigma_k}\right) dy = \int \nu(\mu_k + \sigma_k y)\phi(y)dy.$$

3 Average run lengths

Define for $S_0 = x$,

$$N_x = \inf\{n > 0 : \quad S_n \leq 0, \quad \text{or} \quad S_n > d\},$$

where $S_n = S_0 + \sum_{i=1}^n \mu_{i-1}(X_i - \mu_{i-1}/2)$ and $\mu_i = \mu_i(\mu_0, X_1, ..., X_{i-1})$ with $\mu_0 = \delta$.

When there is no change, each time the CUSUM process T_n comes back to zero consists a renewal point. Similar to the proof for the regular CUSUM procedure (Siegmund(1985, Ch.2)), the average in-control run length is equal to

$$ARL_0 = E_0[N] = E_0[N_0]/P_0(S_{N_0} > d),$$

as N is in distribution equal to $N_0^{(1)} + ..., N_0^{(K)}$, where $K = \inf\{i > 0 : \quad S_{N_0^{(i)}} > d\}$ and $N_0^{(i)}$ are i.i.d. copies of N_0. Since

$$\begin{aligned} P_0(S_{N_0} > d) &= P_0(\tau_d < \infty) - P_0(\tau_d < \infty; S_{N_0} \leq 0) \\ &= E_0^*[e^{-S_{\tau_d}}] - E_0[P_0(\tau < \infty | S_{N_0}); S_{N_0} \leq 0] \\ &= e^{-d-r_1} - E_0[e^{-d-r_1(H_{N_0})+S_{N_0}}; S_{N_0} \leq 0] \\ &= e^{-d}(e^{-r_1} - E_0^*[e^{-r_1(H_{N_0})}; S_{N_0} \leq 0]) \\ &\approx e^{-d}(e^{-r_1} - E_0^*[e^{-r_1(H_{\tau_-})}; \tau_- < \infty]), \end{aligned}$$

as $d \to \infty$. Similarly, we can see

$$\begin{aligned} E_0[N_0] &= E_0^*[N_0 e^{-S_{N_0}}] \\ &\approx E_0^*[\tau_- e^{-S_{\tau_-}}; \tau_- < \infty]. \end{aligned}$$

Thus,

$$ARL_0 \approx e^d \frac{E_0^*[\tau_- e^{-S_{\tau_-}}; \tau_- < \infty]}{e^{-r_1} - E_0^*[e^{-r_1(H_{N_0})}; S_{N_0} \leq 0]}.$$

To evaluate ARL_1, we can write

$$ARL_1 = E_\mu[N_0]/P_\mu(S_{N_0} > d).$$

From Robbins and Siegmund (1974), $\{\sum_{n=1}^m \mu_{n-1}(X_n - \mu)\}$ is a martingale. Thus

$$E_\mu\left[\sum_{n=1}^{N_0} \mu_{n-1} X_n\right] = \mu E_\mu\left[\sum_{n=1}^{N_0} \mu_{n-1}\right].$$

Rewriting it, we have

$$\begin{aligned} E_0[N_0] &= \frac{2}{\mu^2}\left[E_\mu\left[\sum_{n=1}^{N_0} \mu_{n-1}(X_n - \mu_{n-1}/2)\right] + \frac{1}{2} E_\mu\left[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2\right]\right] \\ &= \frac{2}{\mu^2}\left[E_\mu[S_{N_0}] + \frac{1}{2}\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2\right] \\ &= \frac{2}{\mu^2}\left[E_\mu[S_{N_0} | S_{N_0} > d] P_\mu(S_{N_0} > d) + E_\mu[S_{N_0}; S_{N_0} \leq 0]\right. \\ &\quad \left. + \frac{1}{2} E_\mu\left[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2\right]\right]. \end{aligned}$$

As $d \to \infty$, we have the following approximation for ARL_1,

$$\begin{aligned} ARL_1 &= \frac{2}{\mu^2} E_\mu[S_{N_0} | S_{N_0} > d] \\ &\quad + \frac{2}{\mu^2} \frac{E_\mu[S_{N_0}; S_{N_0} \le 0] + \frac{1}{2} E_\mu[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2]}{P_\mu(S_{N_0} > d)} \\ &\approx \frac{2}{\mu}(d + \rho(\mu)) + \frac{2}{\mu^2} \frac{E_\mu[S_{\tau_-}; \tau_- < \infty]}{P_\mu(\tau_- = \infty)} + \frac{1}{\mu^2} \frac{E_\mu[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2]}{P_\mu(S_{N_0} > d)}, \end{aligned}$$

where

$$\rho(\mu) = \lim_{d \to \infty} E_\mu[S_{N_0} - d | S_{N_0} > d].$$

The last term is the one caused the difference between the regular CUSUM procedure when μ is known and the adaptive CUSUM procedure for the same control limit d and needs more careful treatment, we first can write

$$\frac{E_\mu[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2]}{P_\mu(S_{N_0} > d)} \approx E_\mu \left[\sum_{n=1}^{N_0} (\mu_{n-1} - \mu)^2 | S_{N_0} > d \right] + \frac{E_\mu[\sum_{n=1}^{\tau_-} (\mu_{n-1} - \mu)^2; \tau_- < \infty]}{P_\mu(\tau_- = \infty)}.$$

Given $S_{N_0} > d$, we see that $N_0/d \approx 1/(\mu^2/2)$, Thus, since $E_\theta(\mu_{n-1} - \mu)^2 = O(1/n)$, the first term will be at the order of $\sum_{n=1}^{d/(\mu^2/2)} (1/n) = O(\ln(d/(\mu^2/2)))$. Thus, we see that the adaptive CUSUM procedure is efficient at the first order, but not on the second order. Unfortunately, it seems difficult to develop second order approximations for the related quantities, so we only show simulated results for comparison.

4 Biases of estimation

In this section, we study the biases of the estimation for the change point and post-change mean in the recursive mean estimation case. The main ideas follow the lines of Srivastava and Wu (1999) and Wu(2004).

4.1 Bias of $\hat{\nu}$

From the renewal theorem, as $\nu \to \infty$, $(\nu - \nu_n, T_n)$ converges in distribution to (L, M) where L follows distribution

$$P_0(L = k) = P_0(\tau_- \ge k)/E_0\tau_-, \quad \text{for} \quad k \ge 0,$$

and given $L = k$, M follows the same distribution as S_k given $S_1 > 0, ..., S_k > 0$ and $\mu_0 = \delta$. In particular, if $L = 0$, $M = 0$. Similar to Wu (2004), we can write

$$E^\nu[\hat{\nu} - \nu | N > \nu] = E^\nu[\hat{\nu} - \nu; \hat{\nu} > \nu | N > \nu] - E^\nu[\nu - \hat{\nu}; \hat{\nu} < \nu | N > \nu].$$

The event $\{\hat{\nu} > \nu\}$ is asymptotically equivalent to $\tau_M < \infty$ with initial state (L, M). Given $\hat{\nu} > \nu$, $\hat{\nu} - \nu$ is equivalent to τ_M plus the total length of cycles of T_n coming back to zero afterwards with total expected length

$$\frac{E_\mu[\tau_-; \tau_- < \infty]}{P_\mu(\tau_- = \infty)}.$$

On the other hand, given $\hat{\nu} < \nu$, $\nu - \hat{\nu}$ is asymptotically equal to L. Thus, we have the following result:

Theorem 3: *As* $\nu, d \to \infty$,

$$E^{\nu}[\hat{\nu} - \nu | N > \nu] \to E_0[E_{\mu}[\tau_{-M}; \tau_{-M} < \infty | L, M]] + E_0[P_{\mu}(\tau_{-M} < \infty | L, M)] \frac{E_{\mu}[\tau_{-}; \tau_{-} < \infty]}{P_{\mu}(\tau_{-} = \infty)} - E_0[L P_{\mu}(\tau_{-M} = \infty | L, M)].$$

4.2 Bias of $\hat{\mu}$

To evaluate the bias of $\hat{\mu}$, we again write

$$E^{\nu}[\hat{\mu} - \mu | N > \nu] = E^{\nu}[\hat{\mu} - \mu; \hat{\nu} > \nu | N > \nu] + E^{\nu}[\hat{\mu} - \mu; \hat{\nu} < \nu | N > \nu].$$

For the recursive mean estimation, given $\hat{\nu} > \nu$, $\hat{\mu} = \mu_{N_0}(t)$ conditioning on $S_{N_0} > d$ with $\mu_0 = \delta$ and $S_0 = 0$. On the other hand, given $\hat{\nu} < \nu$, $\hat{\mu}$ is equivalent to $\mu_{N_0}(t+L)$ with initial mean

$$\mu'_L = \frac{\delta t + L \bar{X}'_L}{t + L},$$

with $S_0 = M$, where L, M, and μ'_L are defined from another independent copy $\{X'_1, ..., X'_n, ...,\}$ of $\{X_1, ..., X_n, ...\}$. Therefore, as $\nu, d \to \infty$,

$$E^{\nu}[\hat{\mu} - \mu | N > \nu] \to E_{\mu}[\mu_{N_0} - \mu | S_{N_0} > d] E_0[P_{\mu}(\tau_{-M} < \infty | L, M)] + E_0[E_{\mu}[\mu_{N_0}(t + L) - \mu | S_{N_0} > d; S_0 = M; \mu_0 = \mu'_L] P_{\mu}(\tau_{-M} < \infty | L, M)].$$

It seems difficult to derive the second order approximation for the bias, and we only give the first order result:

Theorem 4: *As* $\nu, d \to \infty$, *uniformly for* μ *in a compact positive interval,*

$$E^{\nu}[\hat{\mu} - \mu | N > \nu] = \frac{1}{d}\left[\frac{\mu^2}{2} t(\delta - \mu) + \mu + \frac{\mu^2}{2} E_0[L(\bar{X}'_L - \mu) P_{\mu}(\tau_{-M} = \infty)] + \frac{\mu}{2} \frac{P_{0\mu}(\tau_{-M} < \infty)}{P_{\mu}(\tau_{-} = \infty)} \frac{\partial}{\partial \mu} P_{\mu}(\tau_{-} = \infty) + E_0[\frac{\partial}{\partial \mu} P_{0\mu}(\tau_{-M} = \infty)]\right](1 + o(1)).$$

Proof. Note that conditioning on $S_{N_0} > d$, uniformly for μ in a compact positive interval, as $d \to \infty$,

$$N_0 = (d/\mu^2/2)(1 + o_p(1)),$$

and

$$\begin{aligned} \mu_{N_0} &= \frac{t}{t + N_0}\delta + \frac{N_0}{t + N_0}\bar{X}_{N_0} \\ &= \left(\frac{t(\delta - \mu)}{N_0} + \bar{X}_{N_0}\right)(1 + o_p(1)). \end{aligned}$$

Thus, we can write

$$E_{\mu}[\mu_{N_0} - \mu; S_{N_0} > d] = \frac{t(\delta - \mu)}{2} \frac{\mu^2}{2} P_{\mu}(S_{N_0} > d)$$

$$+\frac{1}{d}\frac{\partial}{\partial\mu}\left(\frac{\mu^2}{2}P_\mu(S_{N_0}>d)\right)(1+o(1)).$$

Thus, we have

$$\begin{aligned} E_\mu[\mu_{N_0}-\mu|S_{N_0}>d] &= \frac{1}{d}\left[\frac{\mu^2}{2}t(\delta-\mu)+\mu+\frac{\mu^2(\partial/\partial\mu)P_\mu(S_{N_0}>d)}{2P_\mu(S_{N_0}>d)}\right](1+o(1)) \\ &= \frac{1}{d}\left[\frac{\mu^2}{2}t(\delta-\mu)+\mu+\frac{\mu^2(\partial/\partial\mu)P_\mu(\tau_-=\infty)}{2P_\mu(\tau_-=\infty)}\right]\times(1+o(1)). \end{aligned}$$

Similarly,

$$\begin{aligned} & E_0[E_\mu[\mu_{N_0}(t+L)-\mu;S_{N_0}>d]] \\ =\; & E_0\left[E_\mu\left[\frac{t\delta+L\bar{X}'_L+N_M\bar{X}_{N_M}}{N_M+t+L}-\mu;S_{N_M}>d\right]\right] \\ =\; & \frac{1}{d}\left[\frac{t\mu^2}{2}(\delta-\mu)P(\tau_{-M}=\infty)+\frac{\mu^2}{2}E[L(\bar{X}'_L-\mu);\tau_{-M}=\infty]\right. \\ & \left.+\frac{\partial}{\partial\mu}(\frac{\mu^2}{2}P(\tau_{-M}=\infty))\right](1+o(1)). \end{aligned}$$

5 Generalizations

The main advantage for the adaptive CUSUM procedure is that it can not only be flexibly adapted to more general post-change mean models, but also be used for unknown initial mean case.

5.1 Unknown initial mean

Let μ_0 and μ be the pre-change and post-change means which are unknown and $\mu-\mu_0>0$ be the change magnitude. We can update the estimate for μ_0 after each sequential test when it goes below zero and trace the change magnitude recursively when a new sequential test is formed. More specifically, with a little abuse of notations, let $\mu_0^{(0)}=\mu_0$ and $\delta_0^{(0)}$ be the assigned starting value for the pre-change mean and change magnitude. Define

$$N^{(i)}=\inf\{n>0:\ S_n^{(i)}=\sum_{j=1}^{n}\delta_{j-1}^{(i-1)}(X_j^{(i)}-\mu_0^{(i-1)}-\delta_{j-1}^{(i-1)}/2)\le 0;\quad \text{or}\quad >d\},$$

where

$$\delta_{j-1}^{(i-1)}=\mu(\delta_0,X_1^{(i)}-\mu_0^{(i-1)},X_2^{(i)}-\mu_0^{(i-1)},...,X_{j-1}^{(i)}-\mu_0^{(i-1)}),$$

and if $S_{N_{(i)}}^{(i)}\le 0$, we update $\mu_0^{(i-1)}$ to

$$\mu_0^{(i)}=\frac{(N^{(1)}+...+N^{(i-1)})\mu_0^{(i-1)}+X_1^{(i)}+...+X_{N^{(i)}}^{(i)}}{N^{(1)}+...+N^{(i)}}.$$

An alarm will be made at $N^{(1)}+...+N^{(K)}$ where

$$K=\inf\{i\ge 1:\quad S_{N^{(i)}}^{(i)}>d\}.$$

The change-point ν and post-change mean will be estimated as

$$\hat{\nu} = N^{(1)} + ... + N^{(K-1)}, \quad \text{and} \quad \hat{\mu} = \mu_0^{(K-1)} + \delta^{(K)}_{N^{(K)}},$$

with $\mu_0^{(K-1)}$ being the pre-change mean estimation.

5.2 Restricted adaptive estimations

For practical application, a shortcoming of the recursive post-mean estimation is that it may become negative. Robbins and Siegmund (1974) proposed to use

$$\max(\delta, \mu_{k+1,n}(\delta, t))$$

as the adaptive estimation where δ is treated as the minimum shift amount to detect.

Sparks (2000) and Jiang,Shu, and Apley (2008) proposed to use restricted exponentially weighted moving average as the adaptive estimation. More specifically, instead of using the sample mean we define

$$\mu_{k+1,n}(\delta, \beta) = (1 - \beta)\mu_{k+1,n-1}(\delta, \beta) + \beta x_n,$$

as the exponentially weighted moving average and use $\max(\delta, \mu_{k+1,n}(\delta, \beta))$ as the adaptive estimation. The EWMA as a control charting tool has been extensively studied in the literature and an adaptive EWMA procedure can be seen in Capizzi and Mascrotto (2003). An advantage of EMMA estimation is that it gives the current mean estimation for more flexible post-change mean structures.

5.3 Detecting slope change

Suppose the means follow the model

$$\mu_k(\nu) = I_{[k\leq\nu]} + \beta(k - \nu)I_{[k>\nu]}.$$

Following the same idea as for the mean shift case, we define the adaptive estimator for β based on $X_{k+1}, ..., X_n$ as

$$\begin{aligned}\beta_{k+1,n}(\beta_0, t) &= \frac{t\beta_0 + \sum_{j=k+1}^{n}(j - k)X_i}{t + \sum_{j=k+1}^{n}(j - k)^2} \\ &= \beta_{k+1,n-1} + \frac{n - k}{t + \sum_{j=k+1}^{n}(j - k)^2}(X_n - (n - k)\beta_{k+1,n-1}),\end{aligned}$$

where $\beta_{k+1,k} = \beta_0$ by default. The CUSUM process can be defined as

$$T_n = \max\{0, T_{n-1} + \beta_{\nu_{n-1}+1,n-1}(n - \nu_{n-1})(X_n - \frac{1}{2}\beta_{\nu_{n-1}+1,n-1}(n - \nu_{n-1}))\},$$

where the adaptive change-point estimation is updated as $\nu_n = \nu_{n-1}$ if $T_n > 0$, and $\nu = 0$ if $T_n = 0$. After an alarm is raised at N, the change-point is estimated as ν_N, and the post-change slope is estimated as

$$\beta_{\nu_N+1,N} = \frac{t\beta_0 + \sum_{j=\nu_N+1}^{N}(j - \nu_N)X_i}{t + \sum_{j=\nu_N+1}^{N}(j - \nu_N)^2}.$$

Table 1: Simulated ARL_0 for recursive mean estimation with $d = 4.8$

(δ,t)	A	B	C	ARL_0
(1.0,0.0)	0.519	0.395	1.885	1117.5
(1.0,0.5)	0.518	0.447	1.890	993.7
(0.5,0.0)	0.367	0.442	2.132	1596.2
(0.5,0.5)	0.343	0.490	2.198	1589.1

6 Simulation study and application

In this section, we run some simulation studies to show the performance of the adaptive CUSUM procedure in the simple recursive mean estimation case and then apply the procedure to the Nile river flow data and global warming data by fitting into mean shift and piecewise linear change models.

6.1 Simulation study

For $\delta = 1.0, 0.5$ and $t = 0.0, 0.5$, we let $d = 4.8$. Table 1 gives the simulated results for ARL_0 where we use the adaptive importance sampling technique by simulating ARL_0 as

$$ARL_0 = \frac{E_0^*[N_0 e^{-S_{N_0}}]}{E_0^*[e^{-S_{N_0}}; S_{N_0} > d]} = e^d \frac{C}{A \times B},$$

where $A = P_0^*(S_{N_0} > d)$, $B = E_0^*[e^{-(S_{N_0}-d)}|S_{N_0} > d]$, and $C = E_0^*[N_0 e^{-S_{N_0}}]$. The simulation is replicated for 10,000 times. The results show that the effect of t is not significant.

Table 2 gives the corresponding ARL_1 for several typical values of μ where $E[R_{N_0}|.] = E[S_{N_0} - d|S_{N_0} > d]$.

Finally, we simulate the biases for the change-point and post-change mean estimators. For the same designs given in Table 2, the simulation is replicated 5000 times and and only those stopping times with $N > \nu$ are counted to calculate the conditional expectations. Reported also includes the average delay detection time

$$ADT = E^\nu[N - \nu|N > \nu],$$

as an alternative to ARL_1. By comparing Table 3 with Table 2, we see that there are very little differences between ALR_1 and ADT. Also, the bias for the change-point estimation gets larger when the post-change mean gets smaller, so is the bias for the post-change mean estimation.

6.2 Nile river flow data

The Nile river flow data from 1871 to 1970 are reproduced from Cobb(1978) (also see Wu (2005, pg. 27)). A plot shows that there is obvious decrease after year 1900. To use the adaptive CUSUM procedure, we use the first 20 data from year 1871 to

Table 2: Simulated ARL_1 for recursive mean estimation with $d = 4.8$

μ	$P_\mu(S_{N_0} > d)$	$E_\mu[N_0]$	$E_\mu[R_{N_0}\|.]$	ARL_1
$\delta = 1.0$	$t = 1.0$			
0.50	0.119	4.786	0.542	40.32
0.75	0.277	5.322	0.669	19.19
1.00	0.437	5.11	0.861	11.69
$\delta = 1.0$	$t = 0.5$			
0.50	0.134	5.205	0.500	38.76
0.75	0.302	5.516	0.657	18.30
1.00	0.482	5.348	0.837	11.10
$\delta = 0.5$	$t = 0.0$			
0.50	0.121	5.423	0.521	44.74
0.75	0.265	5.583	0.666	21.07
1.00	0.433	5.436	0.833	12.55
$\delta = 0.5$	$t = 0.5$			
0.50	0.154	6.198	0.494	40.35
0.75	0.337	6.654	0.643	19.74
1.00	0.511	6.096	0.788	11.92

Table 3: Simulated bias at $\nu = 75$ with $d = 4.8$

(δ,t)	μ	ADT	$E[\hat{\nu} - \nu\|.]$	$E[\hat{\mu} - \mu\|.]$
(1.0,0.0)	0.50	38.61	12.96	0.448
	0.75	17.93	1.75	0.348
	1.00	10.98	-0.72	0.266
(1.0,0.5)	0.50	36.51	11.14	0.417
	0.75	17.39	1.47	0.319
	1.00	10.42	-1.05	0.226
(0.5,0.0)	0.50	41.98	14.83	0.394
	0.75	19.47	2.60	0.315
	1.00	11.57	-0.59	0.233
(0.5,0.5)	0.50	38.88	10.290	0.337
	0.75	18.43	0.86	0.246
	1.00	11.33	-1.54	0.147

1890 as the training sample to estimate the pre-change mean and stdev as 1070 and 143, respectively. We standardize the data by letting

$$x_i = -(y_i - 1070)/143,$$

and a negative sign is added in order to detect an increase in mean. For $t = 0.5$ and $\delta = 0.5$ and 1.0 with $d = 30$, the adaptive CUSUM procedure gives $N = 52$ and $\hat{\nu} = 28$, which is the same as using the regular CUSUM procedure with known post-change mean (Wu (2005)). Also, the post-change mean is estimated as 1.62, which gives post-change mean $1070 - 143 * 1.62 \approx 838$. Figure 1 plots the data and the CUSUM processes. Note that we implicitly assumed that the post-change variance is the same as the pre-change variance.

6.3 Global warming data

In this subsection, we apply the technique to global warming data. The data set in Figure 2 gives the global average temperature from 1880 to 2013 which is available at http://data.giss.nasa.gov/gistemp/. A piecewise linear model is used in Karl, Knight, and Baker (2000) for a smaller data set (1880-1997) by fitting an AR(1) time series error model with wavelet analysis. It reveals that there are increment periods.

The scatterplot of the data $\{y_i\}$ for i=1,..., 134 (years 1880-2013) shows that there are two increment periods. To detect the first change, we use the initial value $\mu_0 = -0.264$ as the mean of the first 30 observations with s.d. =0.096. Then the data are standardized by letting

$$x_i = (y_i + 0.39)/0.13,$$

which are assumed to be i.i.d. N(0,1) random observations by ignoring the correlations. As it is not clear whether the change is shift or linearly increasing, we use both procedures. For the mean shift model, we use $t = 0.5$ and $\delta = 0.5$ to estimate the post-change mean recursively. For $d = 10$ as the control limit, the alarm is raised at year 1937 ($N = 58$) with change-point at year 1922 ($\hat{\nu} = 43$) and the post-change mean is estimated as 1.224.

For the linear change model, we use $\beta_0 = 0.25$ and $t = 0.5$ and the adaptive CUSUM procedure with recursive least-square estimation for β gives the alarm time $N = 57$ (year 1936) and $\hat{\nu} = 43$ (which corresponds to the year 1922). And the post-change slope is estimated as 0.162. That means, the fitted model without correction is

$$x_i = 0.162(i - 43)^+ + \epsilon_i;$$

$$y_i = -0.264 + 0.0156(i - 44)^+ + 0.096\epsilon_i,$$

for $i = 1, ..., 57$. Figure 3 gives the plots of the CUSUM processes based on the mean shift and linear change model respectively.

Figure 4 also plots the CUSUM processes based on the Robbins-Siegmund adaptive estimator and the EWMA adaptive estimator by treating δ is the minimum post-change mean. For both processes, $\hat{\nu} = 44$ and $N = 58$. The post-change mean and current mean at detection are estimated as 1.272 and 1.674 respectively.

To detect the second increment, we start from $i = 61$(which is year 1940) and use $\mu_0 = -0.018$ as the mean of the temperatures from year 1940 to 1969 with s.d. =0.09. So the data is normalized as

$$x_i = (y_i + 0.018)/0.09,$$

for $i \geq 41$. As a linear post-change mean pattern is so obvious, we use the initial values $\beta_0 = 0.25$ with $t = 0.5$ and $d = 10$, the adaptive CUSUM procedure with recursive least-square estimation for β gives the detection time $N = 48$ (corresponding to year 1987) and $\hat{\nu} = 39$ (corresponding to year 1978). The post-change slope is estimated as 0.344. That means, the fitted model without correction is

$$x_i = 0.344(i - 98)^+ + \epsilon_i;$$

$$y_i = -0.018 + 0.031(i - 39)^+ + 0.09\epsilon_i,$$

for $i = 39, ..., 108$. In fact, we see that the slope is larger for the second increment period. Under the mean shift model with $\delta = 0.5$, $t = 0.5$ and $d = 10$, the alarm time is $N = 44$ (year 1983) and $hat\nu = 37$ (year 1976). Figure 5 gives the plots of the two CUSUM processes until the alarm time.

Note that the method can also be used to detect the decreasing of intercept or slope. An analysis under AR(1) model by considering the correlation without adaptive estimation is conducted in Wu (2015).

7 Conclusion

In this paper, we discussed an adaptive CUSUM procedure in order to deal with more flexible post-change mean structures. Sudden mean shift and linear increasing post-change means are used for illustration. The convenience for the adaptive CUSUM procedure is that it is easy to estimate the change-point estimation and post-change mean comparing with other detecting procedures. Future investigations are underway to consider generalized exponential family model, the case when both mean and variance change, and also the dependent observation case in order to fit longitudinal data, see Wu(2015) for the discussions under AR(1) model. Also more theoretical comparisons between alternative adaptive CUSUM procedures are under investigation.

Acknowledgement

This research is supported by a RSCA grant from California State University Stanislaus.

Reference

Capizzi, G., Mascrotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics* 45, 199-207.

Cobb, G.W.(1978). The problem of the Nile: Conditional solution to a change-point problem. *Biometrika* 65, 243-251.

Dragalin, V. P. (1997). The sequential change point problem. *Economic Quality Control* 12, 95 - 122.

Han, D., Tsung, F. , and Wang, K. (2010). A nonlinear filter control chart for detecting dynamic changes. *Statistica Sinica* 20, 1077-1096.

Jiang, W., Shu, L., Apley, D.W. (2008). Adaptive CUSUM procedures with EWMA-based shift estimation. *IIE Trans.* 40, 992-1003.

Karl, T.R., Knight, R.W., Baker, B. (2000). The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophys. Res. Letters* 27,719-722.

Krieger, A. M., Pollak, M., Yakir, B. (2003). Surveillance of a simple linear regression. *J. Amer. Statist. Assoc.* 98, 456-468.

Lai, T. L. (1995). Sequential change point detection in quality control and dynamic systems (with discussions). *J. Royal Statist. Assoc.* (B) 57, 613-658.

Lorden, G., Pollak, M. (2005). Nonanticipating estimation applied to sequential analysis and change point detection. *Ann. Statist.* 33, 1422-1454.

Robbins, H., Siegmund, D. (1973). A class of stopping rules for testing parametric hypotheses. *Proceedings of 6th Berkeley Symposium of Mathematical Statistics and Probability*, 4, 37-41

Robbins, H., Siegmund, D. (1974). The expected sample size of some tests of power one. *Ann. Statist.* 2, 415-436.

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

Siegmund, D., Venkatraman, E.S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* 23, 255-271.

Sparks, R.S. (2000). CUSUM charts for signalling varying location shifts. *J. Quality Tech.* 32,57-171

Srivastava, M. S., Wu, Y. (1999). Quasi-stationary biases of change point and change magnitude estimation after sequential CUSUM test. *Seq. Anal.* 18, 203-216.

Woodroofe, M. (1990). On a nonlinear renewal theorem. *Ann. Probab.* 18, 1790-1805.

Wu, Y. (2004). Bias of change point estimator detected by a CUSUM procedure. *Ann. Inst. Statist. Math.* 56, 127−142.

Wu, Y. (2005). *Inference for Change-point and Post-change Means After a CUSUM Test.* Lecture Notes in Statistics 180, Springer, New York.

Wu, Y. (2015). Inference after CUSUM test under AR(1) model. To appear on *J. Statist. Plann. Inf.*

Yakir, B., Krieger, A. M., Pollak, M. (1999). Detecting a change in regression: First order optimality. *Ann. Statist.* 27, 1896-1913.

Yashchin , E. (1995). Estimating the current mean of a process subject to abrupt changes. *Technometrics* 37, 311-323.

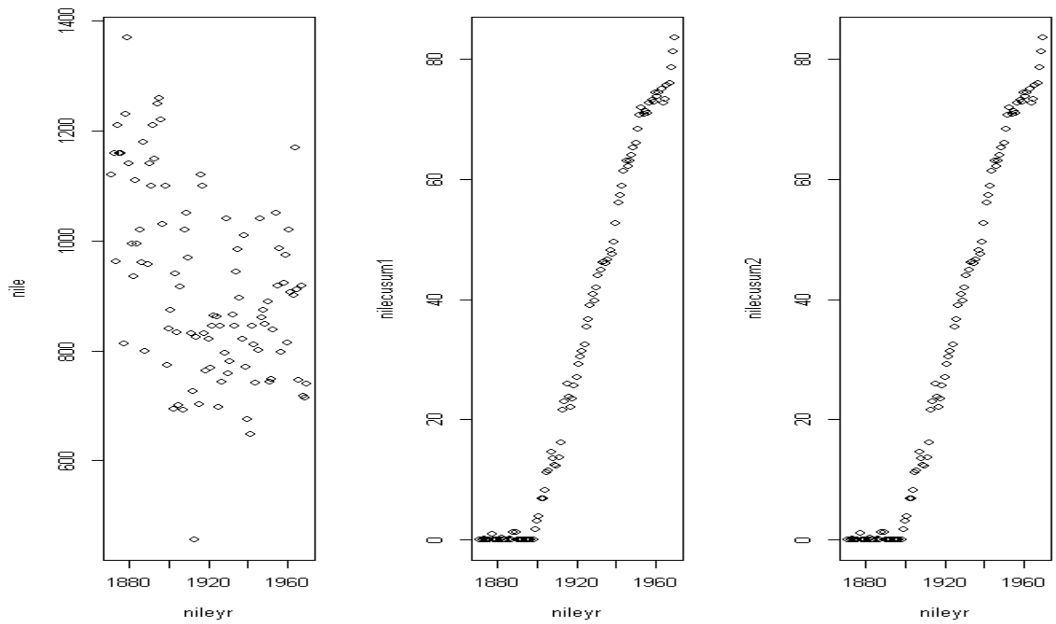

Figure 1. Nile river flow data and adaptive CUSUM process

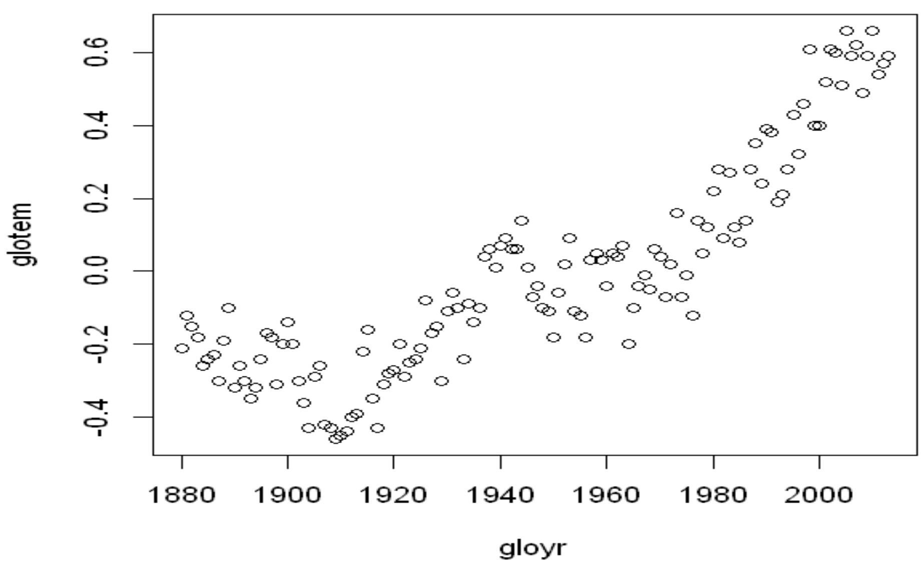

Figure 2. Global temperature from 1880 to 2013

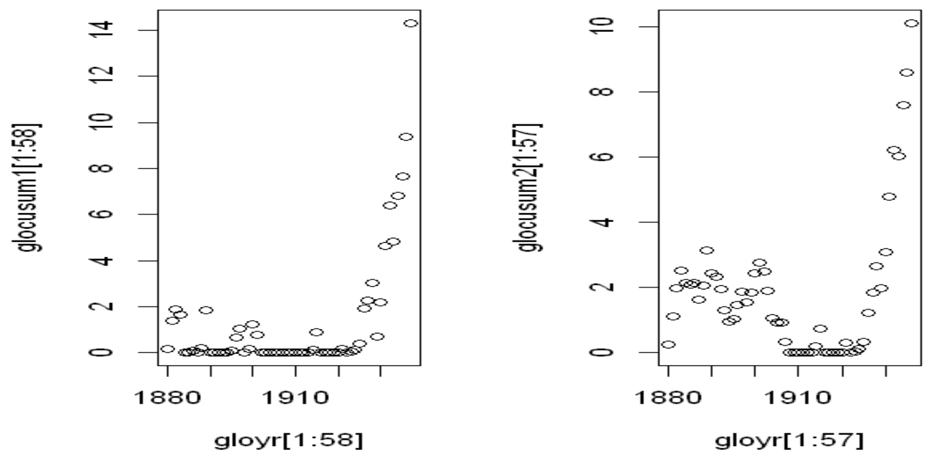

Figure 3 CUSUM process based on mean shift and linear post-change mean

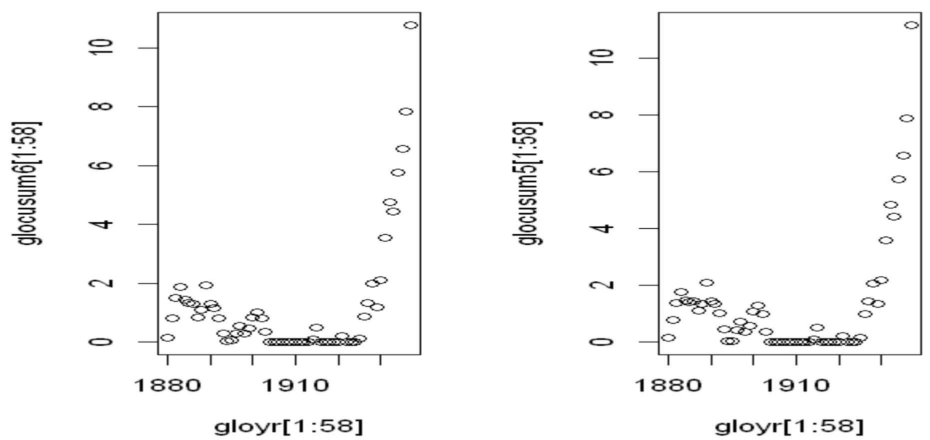

Figure 4. CUSUM process based on Robbins-Siegmund and EWMA estimators

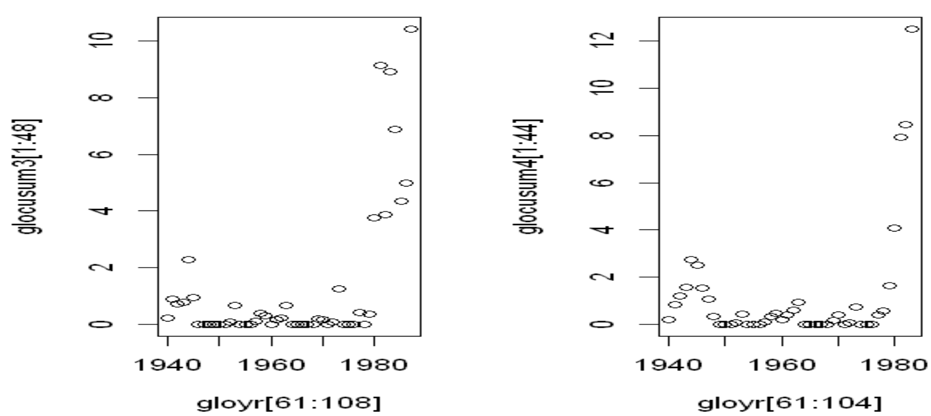

Figure 5. CUSUM process based on linear post-change and mean shift