

# Estimation of Change-point and Post-change Parameters after Adaptive Sequential CUSUM Test in an Exponential Family

Yanhong Wu<sup>1</sup>

<sup>1</sup> Department of Mathematics, California State University Stanislaus, Turlock, CA, USA

Correspondence: Yanhong Wu, Department of Mathematics, California State University Stanislaus, Turlock, CA 95382, USA. E-mail: ywu1@csustan.edu

Received: July 13, 2016 Accepted: August 1, 2016 Online Published: August 16, 2016

doi:10.5539/ijsp.v5n5p43 URL: <http://dx.doi.org/10.5539/ijsp.v5n5p43>

## Abstract

In this paper, we consider an adaptive sequential CUSUM procedure in an exponential family where the change-point and post-change parameters are estimated adaptively. It is shown that the adaptive CUSUM procedure is efficient at the first order. The conditional biases of the estimation for the change-point and post-change parameter are studied. Comparison with the classical CUSUM procedure in the normal case is made. Nile river flow and average global temperature data sets are used for demonstration.

**Keywords:** adaptive CUSUM procedure, biases, change-point, exponential family, nonlinear renewal theorem.

## 1. Introduction

Let

$$dF_\theta(x) = \exp(\theta x - c(\theta))dF_0(x)$$

be a standard exponential family with  $c(0) = c'(0) = 0$  and  $c''(0) = 1$ . Under  $P_\theta(\cdot)$ ,  $E_\theta X = c'(\theta)$  and  $\text{Var}_\theta(X) = c''(\theta)$  for  $|\theta| \leq K_0$  for a positive constant  $K_0 > 0$ . We shall assume that  $|c''(\theta)| \leq K$  for  $|\theta| \leq K_0$ . To detect a change in the parameter  $\theta$  from  $\theta = 0$  to  $\theta > 0$  at the change point  $\nu$ , the regular CUSUM procedure is to select a reference values  $\delta > 0$  for  $\theta$  and form an one-sided CUSUM process  $\tilde{T}_n$  as

$$\tilde{T}_n = \max(0, \tilde{T}_{n-1} + \delta X_n - c(\delta))$$

with  $\tilde{T}_0 = 0$  and an alarm is raised at

$$\tilde{N} = \inf\{n > 0 : \tilde{T}_n > d\},$$

for some predesigned boundary  $d$  which satisfies certain condition such as  $ARL_0 = E_\infty[\tilde{N}]$ . For detecting two-sided change, another one-sided CUSUM process is constructed by using the conjugate value  $\tilde{\delta} < 0$  such that  $c(\tilde{\delta}) = c(\delta)$  to detect negative change.

However, when  $\theta \neq \delta$ , the procedure is no longer efficient. Three approaches have been used to increase efficiencies. One is to use the GLRT (generalized likelihood ratio test) by treating  $\theta$  as a unknown parameter (Siegmund & Venkatramen, 1991). To overcome the memory problem, Lai (1995) considered the window-limited GLRT. The second is to use integrated likelihood ratio by treating the post-change parameter as a nuisance parameter. However, explicit forms are typically difficult to obtain. The third, considered in this paper, is to use the adaptive CUSUM procedure by estimating the change-point and the post-change parameter adaptively (Draglin, 1990; Wu, 2005, 2015; Lorden & Pollak, 2005, 2008). The advantage for this approach is the change-point and post-change parameters are easily identified after the detection as the basic form of the CUSUM procedure is kept.

Many forms of adaptive CUSUM control charts in on-line quality control have been proposed and discussed. Draglin (1998) suggested to use the sample mean. Yakir et al. (1999) and Krieger et al. (2003) considered the linear post-change model. Capizzi and Mascalotto (2003) proposed an adaptive EWMA procedure. An adaptive Shirayev-Roberts procedure using the adaptive estimators is considered in Lorden and Pollak (2005). Yashchin (1995) and Jiang et al. (2008) used the EWMA as the adaptive post-change mean estimator. Han et al. (2010) proposed to use the last current observation as the estimator for the mean.

Our discussion is mainly focused on the change-point and post-change parameter estimation after detection under the exponential family model which extends the results of Wu (2005) and Lorden and Pollak (2008). By treating the CUSUM procedure as a sequence of sequential tests, we can use the adaptive sequential tests (Robbins & Siegmund, 1974, 1975) by estimating the post-change parameters adaptively for each test. More specifically, we use the notations from stochastic

approximation. For observations  $X_{k+1}, \dots, X_n$ , define recursively the mean estimator

$$\mu_{k+1,n}(\delta, t) = \mu_{k+1,n-1} + \gamma_{k,n}(t)(X_n - \mu_{k+1,n-1}(\delta, t)),$$

where  $\mu_{k+1,k} = \mu_0$  is the initial mean value and  $\gamma_{k,n}(t)$  is the gain defining the rate of convergence. Two most used ones are (i)(recursive moment estimation)  $\gamma_{k,n} = \frac{1}{t+n-k}$ ; and (ii) (exponentially weighted moving average)  $\gamma_{k,n} = \gamma$ , a constant. Correspondingly, the post-change parameter  $\theta$  is estimated through the equation

$$c'(\theta_{k+1,n}(\delta, t)) = \mu_{k+1,n}(\delta, t).$$

The adaptive CUSUM procedure is defined as follows:

- (i) Set  $v = 0$ ,  $\mu = \delta$ , and  $T_0 = 0$ ;
- (ii) For  $\theta = c'^{-1}(\mu)$ , define

$$T_n \leftarrow \max\{0, T_{n-1} + \theta x_n - c(\theta)\}$$

- (iii) If  $T_n = 0$ , reset  $v = n$  and  $\mu = \delta$ ;

If  $T_n > 0$ , update  $\theta$  by letting

$$c'(\theta) \leftarrow c'(\theta) + \gamma_{k,n}(t)(x_n - c'(\theta)).$$

- (iv) The procedure stops at

$$N = \inf\{n > 0 : T_n > d\}.$$

And the change-point and post-change parameter are estimated as

$$\hat{v} = v_N; \hat{\theta} = \theta_{v_N+1,N}.$$

The discussion is organized as follows. In Section 2, we first present a nonlinear renewal theorem for the adaptive random walk which gives asymptotic results for the  $ARL_0$  under the changed adaptive measure and also provides an adaptive importance sampling technique for simulating  $ARL_0$ . The first order result for  $ARL_1$  is given by using a martingale structure related to the adaptive random walk which shows the adaptive CUSUM procedure is asymptotically efficient. The biases for the change-point and post-change parameter estimation are studied theoretically in Section 3 by using the renewal property of the adaptive CUSUM process. Simulation comparison with the classical CUSUM procedure in the normal case in terms of average delay detection time and bias of change-point estimation is conducted in Section 4. Nile river flow and average global temperature data sets are used for illustration in Section 5.

## 2. Operating Characteristics

### 2.1 A Nonlinear Renewal Theorem

We first present a nonlinear renewal theorem under the changed adaptive measure for an adaptive random walk, which helps to derive the first order result for  $ARL_0$  and also provides an adaptive importance sampling technique for simulation. Denote by  $H_k = \sigma(\mu_0, X_1, \dots, X_k)$  the history up to time  $k$ , where  $\{X_n\}$  follows the exponential family distribution  $F_{\theta_{k+1,n}}(x)$  with adaptively estimated parameter  $\theta_{k+1,n-1} = \theta_{k+1,n-1}(\theta_0, t)$ . Define

$$S_{k+1,n} = S_k + \sum_{i=k+1}^n (\theta_{k+1,i-1} X_i - c(\theta_{k+1,i-1})),$$

as the adaptive random walk. We are interested in evaluating the probability  $P_0(\tau_d < \infty | H_k)$ , where

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

Suppose

(A1) under  $P_\theta(\cdot)$ ,  $\theta_{k+1,n} \rightarrow \theta$  with probability one such that  $E_\theta(\theta_{k+1,n} - \theta)^2 = O(1/n)$ .

We first write

$$S_{k+1,n} = S_k + \sum_{i=k+1}^n (\theta_{k+1,i-1} X_i) - \sum_{i=k+1}^n c(\theta_{k+1,i-1}).$$

Under  $P_0(\cdot | H_k)$ , Since  $E_0[\sum_{i=k+1}^n (\theta_{k+1,i-1} X_i)] = 0$  and

$$Var_0(\sum_{i=k+1}^n (\theta_{k+1,i-1} X_i)) = E_0[\sum_{i=k+1}^n \theta_{k+1,i-1}^2] = O(\ln(n)),$$

under Assumption (A1), the normal law for martingale states that  $\frac{1}{\sqrt{\ln(n)}} \sum_{i=k+1}^n (\theta_{k+1,i-1} X_i)$  converges weakly to a normal variance. On the other hand,

$$E_0 \left[ \sum_{i=k+1}^n c(\theta_{k+1,i-1}) \right] = O(\ln(n)),$$

under Assumption (A1). Thus,

$$S_{k+1,n}/\sqrt{\ln(n)} = \frac{1}{\sqrt{\ln(n)}} \sum_{i=k+1}^n (\theta_{k+1,i-1} X_i) - \frac{1}{2} O_p(\sqrt{\ln n}),$$

which converges almost surely to  $-\infty$ . Thus

$$P_0(\tau_d < \infty | H_k) \rightarrow 0,$$

as  $d \rightarrow \infty$ . To obtain a more accurate approximation, we introduce the adaptive changed measure

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

Assume

(A2) under  $P_0^*(\cdot | H_k)$ ,  $\theta_{k+1,n}$  converges with probability 1 to a random variable, say  $\theta_{k+1,\infty}$ , as  $d \rightarrow \infty$ .

Intuitively, under  $P_0^*(\cdot | H_k)$  the adaptive random walk behaves asymptotically like a conditional random walk with a random drift. Thus, no matter what the sign of the drift is,  $P_0^*(\tau_d < \infty | H_k) = 1$  since for  $\theta \neq 0$ ,

$$E_\theta[\theta X - c(\theta)] = \theta c'(\theta) - c(\theta) > 0.$$

Define

$$S_n^* = \sum_{i=1}^n (\theta_{k+1,\infty} X_i - c(\theta_{k+1,\infty})),$$

as the random walk with random drift  $\theta_{k+1,\infty}$  and  $\tau_d^* = \inf\{n > 0 : S_n^* > d\}$ . The following theorem is deducted from Woodroffe (1990).

**Theorem 1.** As  $d \rightarrow \infty$ , given  $H_k$ , in distribution under  $P_0^*(\cdot | H_k)$ ,

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

where  $R_\infty^*(H_k) = \lim(S_{\tau_d^*} - d)$  denotes the overshoot.

By using Wald's likelihood ratio identity, we have

$$\begin{aligned} P_0(\tau_d < \infty | H_k) &= E_0^*[\exp(-(S_{\tau_d} - S_k)) | H_k] \\ &\approx e^{-(d-S_k)} E_0^*[e^{-R_\infty^*} | H_k]. \end{aligned}$$

**Example 1.** (Recursive moment estimation) First, we notice that the assumption (A1) is obviously satisfied. Second, since

$$\mu_{k+1,n} = \mu_0 + \sum_{i=k+1}^n \frac{X_i - \mu_{k+1,i-1}}{n-i+t},$$

under  $P_0^*(\cdot | H+k)$ ,  $X_i \sim \exp(\theta_{k+1,i-1} x - c(\theta_{k+1,i-1}))$ . So by using the martingale property of the recursive structure,

$$E_0^*[\mu_{k+1,n}] = \mu_0,$$

and

$$\begin{aligned} Var_0^*(\mu_{k+1,n}) &= Var_0^*(\mu_{k+1,n-1}) + \frac{E[c''(\theta_{k+1,n-1})]}{(n-k+t)^2} \\ &= \sum_{i=k+1}^n \frac{E[c''(\theta_{k+1,i-1})]}{(i-k+t)^2} \\ &\leq K \sum_{i=k+1}^n \frac{1}{(i-k+t)^2}, \end{aligned}$$

under the assumption that  $|c''(\theta)| \leq K$ . The almost sure convergence of  $\mu_{k+1,n}$  implies the almost convergence of  $\theta_{k+1,n}$ .

## 2.2 ARL<sub>0</sub>

Define

$$N_x = \inf\{n > 0 : S_n \leq 0; \text{ or } S_n > d\}, \text{ for } S_0 = x.$$

Before the change occurs, the time epoches at which  $T_n = 0$  consists a sequence of renewal points for  $T_n$ . The renewal argument shows that

$$ARL_0 = E_0[N_0] + ARL_0 P_0(S_{N_0} \leq 0),$$

which implies

$$ARL_0 = \frac{E_0[N_0]}{P_0(S_{N_0} > d)}. \quad (1)$$

As  $d \rightarrow \infty$ , by using the renewal theorem and Wald's likelihood ratio repeatedly, we have

$$\begin{aligned} P_0(S_{N_0} > d) &= P_0(\tau_d < \infty) - P_0(\tau_d < \infty; S_{N_0} \leq 0) \\ &= E_0^*[e^{-S_{\tau_d}}] - E_0[P_0(\tau_d < \infty | S_{N_0}); S_{N_0} \leq 0] \\ &\approx e^{-d} E_0^*[e^{-R_\infty^*(H_0)}] - E_0[e^{-d+S_{N_0}} E_0^*[e^{-R_\infty^*(H_{N_0})}]; S_{N_0} \leq 0] \\ &= e^{-d}(E_0^*[e^{-R_\infty^*(H_0)}] - E_0^*[e^{-R_\infty^*(H_{N_0})}; S_{N_0} \leq 0]) \\ &\approx e^{-d}(E_0^*[e^{-R_\infty^*(H_0)}] - E_0^*[e^{-R_\infty^*(H_{\tau_-})}; \tau_- < \infty]). \end{aligned}$$

Also, note that

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

Thus,

$$ARL_0 \approx e^d \frac{E_0^*[\tau_- e^{-S_{\tau_-}}; \tau_- < \infty]}{E_0^*[e^{-R_\infty^*(H_0)}] - E_0^*[e^{-R_\infty^*(H_{\tau_-})}; \tau_- < \infty]}.$$

## 2.3 ARL<sub>1</sub>

The most common measure for the operating characteristics of a detecting procedure is the average out-of-control run length  $ARL_1$ . By using the same renewal argument as in Equation (1), we can show

$$ARL_1 = \frac{E_\theta(N_0)}{P_\theta(S_{N_0} > d)}.$$

To evaluate  $E_\theta(N_0)$ , we note that (Robbins & Siegmund, 1975)  $\{\sum_{k=1}^n \theta_{n-1}(X_n - c'(\theta))\}$  is a martingale with mean 0 under  $P_\theta(\cdot)$ . We rewrite it as

$$\begin{aligned} \sum_{k=1}^n \theta_{n-1} X_n - \sum_{k=1}^n c'(\theta) \theta_{n-1} &= \sum_{k=1}^n (\theta_{n-1} X_n - c(\theta_{n-1})) \\ &+ \sum_{k=1}^n (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta)) + n(\theta c'(\theta) - c(\theta)). \end{aligned}$$

By using the martingale property, we get

$$E_\theta(N_0) = \frac{E_\theta(S_{N_0}) + E[\sum_{k=1}^{N_0} (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta))]}{\theta c'(\theta) - c(\theta)}.$$

The second term appears because of the adaptive estimation. The first order result can be obtained as follows. First,

$$\begin{aligned} E_\theta[S_{N_0}] &= E_\theta[S_{N_0}; S_{N_0} > d] + E_\theta[S_{N_0}; S_{N_0} \leq 0] \\ &\approx E_\theta[S_{N_0}; S_{N_0} > d] + E_\theta[S_{\tau_-}; \tau_- < \infty]. \end{aligned}$$

Second, using the same technique, we have

$$\begin{aligned} E[\sum_{k=1}^{N_0} (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta))] &= E[\sum_{k=1}^{N_0} (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta)); S_{N_0} > d] \\ &\quad + E[\sum_{k=1}^{\tau_-} (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta)); \tau_- < \infty] \end{aligned}$$

Note that

$$E_\theta[S_{N_0}|S_{N_0} > d] \approx d + E_\theta[R_\infty].$$

On the other hand, given  $S_{N_0} > d$ ,  $N_0 = O_p(d/(\theta c'(\theta) - c(\theta)))$  (see Wu (2004)). Under the boundness of  $c''(\theta)$

$$|c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta)| \leq K'(\theta_{n-1} - \theta)^2,$$

and the assumption that  $E_\theta(\theta_{n-1} - \theta)^2 = O(1/n)$ , we have

$$E \left[ \sum_{k=1}^{N_0} (c(\theta_{n-1}) - c(\theta) - (\theta_{n-1} - \theta)c'(\theta)) | S_{N_0} > d \right] = O(\ln(d)).$$

It follows that at the first order,

$$ARL_1 = \frac{d}{\theta c'(\theta) - c(\theta)} (1 + o(1)).$$

**Remark.** For the recursive moment estimation,  $E_\theta(\theta_{n-1} - \theta)^2 = O(1/n)$  is obviously satisfied.

### 3. Bias of $\hat{\nu}$ and $\hat{\theta}$

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).

#### 3.1 Bias of $\hat{\nu}$

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

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

and given  $L = k$ ,  $M$  follows the same distribution as  $S_k$  given  $S_1 > 0, \dots, S_k > 0$  and  $\mu_0 = \delta$ . In particular, if  $L = 0$ ,  $M = 0$ . By splitting on whether  $\hat{\nu} > \nu$  or  $\hat{\nu} \leq \nu$ , we can write

$$\begin{aligned} E^\nu[\hat{\nu} - \nu | N > \nu] &= E^\nu[\hat{\nu} - \nu; \hat{\nu} > \nu | N > \nu] \\ &\quad - E^\nu[\nu - \hat{\nu}; \hat{\nu} < \nu | N > \nu]. \end{aligned}$$

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  $\tau_M$  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 2** As  $\nu, d \rightarrow \infty$ ,

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

#### 3.2 Bias of $\hat{\theta}$

To evaluate the bias of  $\hat{\theta}$ , we first consider the bias of  $\hat{\mu}$  by writing

$$\begin{aligned} E^\nu[\hat{\mu} - c'(\theta) | N > \nu] &= E^\nu[\hat{\mu} - c'(\theta); \hat{\nu} > \nu | N > \nu] \\ &\quad + E^\nu[\hat{\mu} - c'(\theta); \hat{\nu} < \nu | N > \nu]. \end{aligned}$$

To develop explicit formula, we only consider the case of recursive mean estimation. Given  $\hat{\nu} > \nu$ ,  $\hat{\mu} = \mu_{N_0}(\delta, t)$  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  $\mu_{L+1, N_0}$  with initial value

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

where  $L, M$ , and  $\mu'_L$  are defined from another independent copy  $\{X'_1, \dots, X'_n, \dots\}$  of  $\{X_1, \dots, X_n, \dots\}$  and  $S_0 = M$ . Therefore, as  $v, d \rightarrow \infty$ ,

$$\begin{aligned} E^v[\hat{\mu} - c'(\theta)|N > v] &\rightarrow E_\theta[\mu_{N_0} - c'(\theta)|S_{N_0} > d]E_0[P_\theta(\tau_{-M} < \infty|L, M)] \\ &\quad + E_0[E_\theta[\mu_{L+1, N_0} - c'(\theta)|S_{N_0} > d; S_0 = M; \mu_0 = \mu'_L] \\ &\quad \times P_\theta(\tau_{-M} < \infty|L, M)]. \end{aligned}$$

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

**Theorem 3** As  $v, d \rightarrow \infty$ , uniformly for  $\theta$  in a compact positive interval,

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

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

$$N_0 = \frac{d}{\theta c'(\theta) - c(\theta)} (1 + o_p(1)),$$

and

$$\begin{aligned} \mu_{N_0} &= \frac{t}{t + N_0} \delta + \frac{N_0}{t + N_0} \bar{X}_{N_0} \\ &= \left( \frac{t(\delta - c'(\theta))}{N_0} + \bar{X}_{N_0} \right) (1 + o_p(1)). \end{aligned}$$

Thus, we can write

$$\begin{aligned} E_\theta[\mu_{N_0} - c'(\theta); S_{N_0} > d] &= \frac{t(\delta - c'(\theta))}{2} (\theta c'(\theta) - c(\theta))P_\theta(S_{N_0} > d) \\ &\quad + \frac{1}{d} \frac{\partial}{\partial \theta} [(\theta c'(\theta) - c(\theta))P_\theta(S_{N_0} > d)] (1 + o(1)). \end{aligned}$$

Thus, we have

$$\begin{aligned} &E_\theta[\mu_{N_0} - c'(\theta)|S_{N_0} > d] \\ &= \frac{1}{d} \left[ (\theta c'(\theta) - c(\theta))t(\delta - c'(\theta)) + \frac{1}{P_\theta(S_{N_0} > d)} \frac{\partial}{\partial \theta} [(\theta c'(\theta) - c(\theta))P_\theta(S_{N_0} > d)] \right] (1 + o(1)) \\ &= \frac{1}{d} \left[ (\theta c'(\theta) - c(\theta))t(\delta - c'(\theta)) + \frac{1}{P_\theta(\tau_- = \infty)} \frac{\partial}{\partial \theta} [(\theta c'(\theta) - c(\theta))P_\theta(\tau_- = \infty)] \right] (1 + o(1)). \end{aligned}$$

Similarly,

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

The bias of  $\hat{\theta}$  can be obtained by using delta-method as

$$E^v[\hat{\theta} - \theta|N > v] \approx \frac{1}{c''(\theta)} E^v[\hat{\mu} - \mu|N > v].$$

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

$(\delta, t)$	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

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

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

#### 4. Simulation Comparison in the Normal Case

##### 4.1 Normal Mean Shift

We compare the classical CUSUM procedure with the adaptive CUSUM procedure for detecting the mean shift in a normal model with unit variance.

For the classical CUSUM procedure, the design of  $d$  can use the following simple accurate approximation (Siegmund, 1985, Equation (2.56))

$$ARL_0 = E_0[N] = (e^{d+1.166\delta} - 1 - (d + 1.166\delta)) / (\delta^2/2),$$

where

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

So we shall first simulate  $ARL_0$ 's for the adaptive CUSUM procedure and then find the value of  $d$  for the classical CUSUM procedure by matching the corresponding  $ARL_0$ 's.

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  $\mu$  where  $E[R_{N_0}| \cdot ] = E[S_{N_0} - d | S_{N_0} > d]$ .

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

$(\delta, t)$	$\mu$	$ADT$	$E[\hat{\nu} - \nu]$	$E[\hat{\mu} - \mu]$
(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

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

#### 4.2 Unknown Initial Mean

Let  $\mu_0$  and  $\mu$  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  $\mu_0$  after each sequential test when it goes below zero and track the change magnitude recursively when a new sequential test is formed. More specifically, with a slight different 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

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

where

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

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

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

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

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

The change-point  $\nu$  and the post-change mean will be estimated as

$$\hat{\nu} = N^{(1)} + \dots + N^{(K-1)}, \text{ and } \hat{\mu} = \mu_0^{(K-1)} + \delta_{N^{(K)}}^{(K)},$$

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

#### 4.3 Restricted Adaptive Estimations

For practical application, the recursive post-mean estimation may become negative. Robbins and Siegmund (1974) proposed to use

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

as the adaptive estimation where  $\delta$  is treated as the minimum shift amount to detect.

Sparks (2000) and Jiang et al. (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 Mascalotto (2003). An advantage of EMMA estimation is that it gives the most current mean estimation for more flexible post-change mean structures. However, its convergence in probability under the adaptive change probability measure  $P^*(.)$  can not be established.

#### 4.4 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  $\beta$  based on  $X_{k+1}, \dots, X_n$  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  $\nu_N$  and the post-change slope is estimated as

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

## 5. Examples

### 5.1 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 in Figure 1 shows that there is an obvious decrease after year 1900.

(i) To use the adaptive CUSUM procedure, we use the first 20 data from year 1871 to 1890 as the training sample to estimate the pre-change mean and stdev as 1070 and 143, respectively. We standardize the data by letting

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

for  $i = 1, 2, \dots, 100$ , and a negative sign is added in order to detect a decrease 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 the ones by using the regular CUSUM procedure with known post-change mean (Wu [19]). Also, the post-change mean is estimated as 1.63, which gives post-change mean  $1070 - 143 * 1.62 \approx 837$ .

(ii) To detect whether a second change occurs, we use the data from 29 to 52 to calculate the mean and standard deviation as 837 and 149.5. So we standardize the data as

$$x_i = -(y_i - 837)/149.5,$$

for  $i = 29, \dots, 100$ . For  $t = 0.5$  and  $\delta = 0.5$  and 1.0 with  $d = 30$ , no further change point is detected by the adaptive CUSUM procedure.

With change-point  $\hat{\nu} = 28$ , the global pre-change mean is given as 1097.75 and the post-change mean is 849.97. Note that we implicitly assumed that the post-change variance is the same as the pre-change variance. Figure 1 also shows that the residuals have no significant correlation.

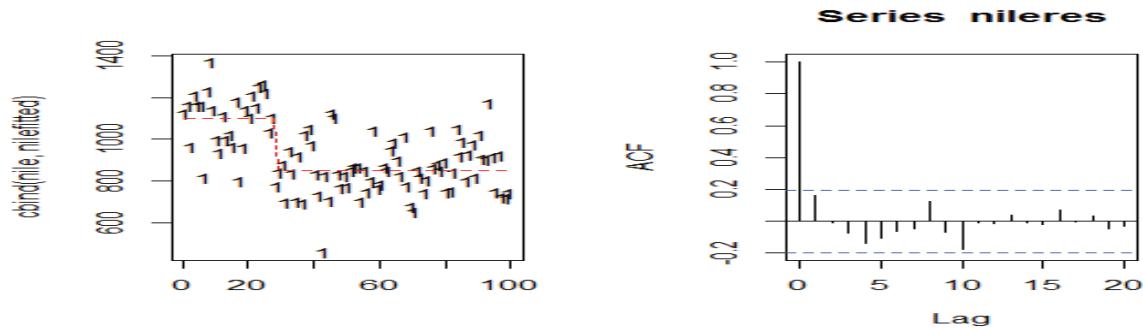


Figure 1. Nile river date

### 5.2 Global Warming Data

In this subsection, we apply the technique to detect the change in global average temperature. Figure 2 plots the global average temperature from 1880 to 2013 which is downloaded from <http://data.giss.nasa.gov/gistemp/>. A piecewise linear model is used in Karl et al. (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 scatter-plot of the data  $\{y_i\}$  for  $i=1, \dots, 134$  (years 1880-2013) shows that there are at least two increment periods. Here we use the sequential adaptive CUSUM procedure to detect the change-point one-by-one. After a change is detected, the post-change model is fitted by using R programming by using the data from the delay detection time.

(i) To detect the first change, we use the initial value  $\mu_0 = -0.387$  as the mean of the first 30 observations with s.d. = 0.130. Then the data are standardized by letting

$$x_i = (y_i + 0.387)/0.13,$$

which are assumed to be i.i.d.  $N(0,1)$  random observations by ignoring the correlations at first. As it is not clear whether the change is shift or linearly increasing, we use the sudden mean shift model. For  $t = 0.5$ ,  $\delta = 1.0$ , and control limit  $d = 20$ , the alarm is raised at year 1937 ( $N = 58$ ) with the estimated change-point at year 1909 ( $\hat{\nu} = 30$ ).

(ii) To detect the second change-point, the 28 data after the change-point estimation are used to fit a model by using R and it shows that a linear model  $-0.30867 + 0.009267(i - 30)^+$  is the better fit with estimated standard deviation 0.10143. So we standardize the data starting from number 31 by subtracting the mean  $-0.30867 + 0.009267(i - 30)^+$  and being divided by 0.10143. The same values of  $\delta$ ,  $t$ , and  $d$  are used and the alarm is raised at time 85 (year 1964) with change-point estimation at 68 (year 1947).

(iii) Similarly, to detect the third change-point, we use the delay detection data from numbers 69 to 85 to fit the post-change model and it shows a constant mean model is a better fit. The mean of these 17 data is -0.05294 with stdev 0.1063. By standardizing rest of the data starting from number 69, the adapted CUSUM procedure detected the third change-point at number 97 (year 1976) with alarm at 103 (year 1982).

(iv) Since no more change-point is detected, we can use the three change-points 30, 68, 97 to fit a global piece-wise linear model. The lm() function in R is used to find the best fit and the final mean is given as

$$\hat{\mu}(t) = -0.3722 + 0.01194(t - 30)I_{[30 < t \leq 68]} + 0.3280I_{[68 < t \leq 97]} + (0.5118 + 0.01997(t - 97))I_{[97 < t \leq 134]}.$$

(v) The residual analysis shows that the AR(1) model fits the residuals well with autocorrelation 0.2033 and stdev 0.1079. So the final fitted model is

$$x_t = \hat{\mu}(t) + \epsilon_t$$

where

$$\epsilon_t = 0.2033\epsilon_{t-1} + 0.1079z_t,$$

with  $z_t$  being i.i.d. normal random variables. Figure 2 also gives the fitted model and the ACFs and qq-norm plot for the residuals before and after the correlation fitting. The analysis under AR(1) model under classical CUSUM procedure can be seen in Wu (2016).

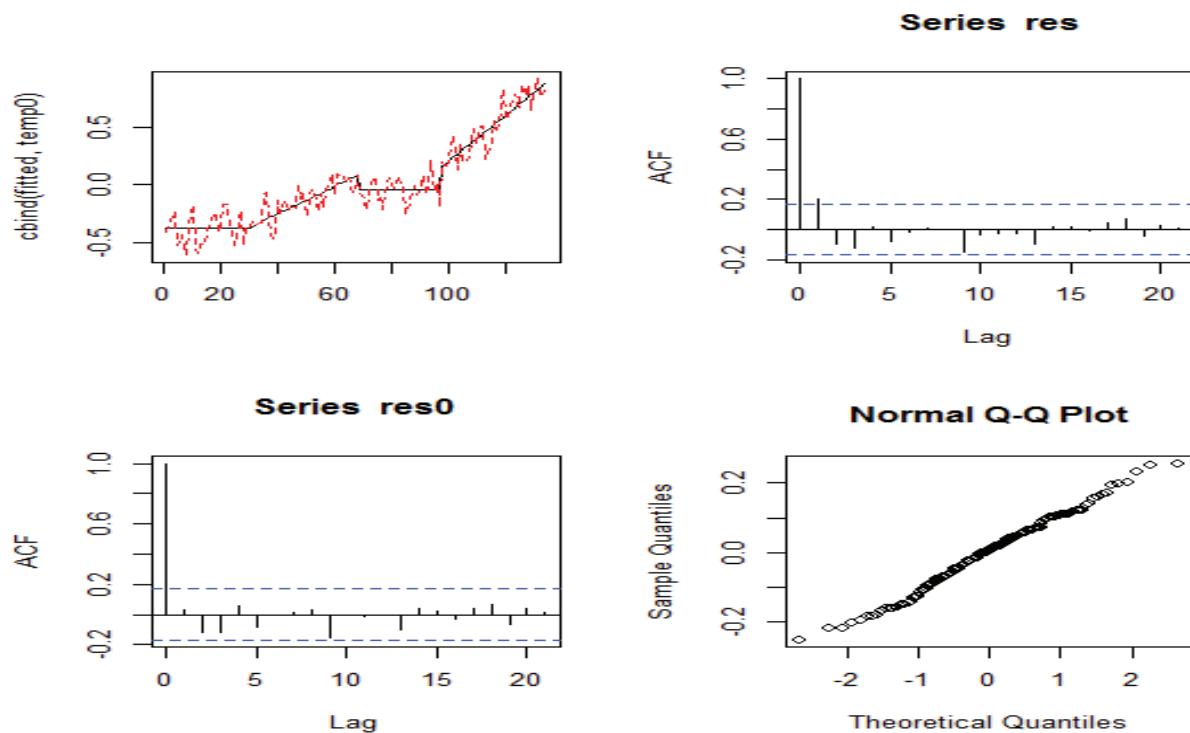


Figure 2. Global average temperature

## 6. Conclusion

In this paper, we discussed an adaptive sequential 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 adaptive CUSUM procedure can easily estimate the change-point estimation and post-change mean comparing with other detecting procedures. Future investigations will consider the generalized multi-parameter exponential family model, e.g. the case when both mean and variance change along with the dependent observation case in order to fit longitudinal data. Also more theoretical comparisons with alternative adaptive CUSUM procedures are under investigation.

## Acknowledgement

The careful reviews from the Editor and a reviewer are gratefully appreciated. This research is partially supported by RSCA of California State University Stanislaus.

## References

- Benveniste, A., Metivier, M., & Priouret, P. (1987). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, Berlin.
- Capizzi, G., & Mascalotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics*, 45, 199-207. <http://dx.doi.org/10.1198/004017003000000023>
- Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a change-point problem. *Biometrika*, 65, 243-251. <http://dx.doi.org/10.1093/biomet/65.2.243>
- Dragalin, V. P. (1998). The sequential change point problem. *Economic Quality Control*, 12, 95-122.
- Han, D., Tsung, F., & 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. <http://dx.doi.org/10.1080/07408170801961412>
- 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.  
<http://dx.doi.org/10.1029/1999GL010877>
- Krieger, A. M., Pollak, M., & Yakir, B. (2003). Surveillance of a simple linear regression. *J. Amer. Statist. Assoc.*, 98, 456-468. <http://dx.doi.org/10.1198/016214503000233>
- 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. <http://dx.doi.org/10.1214/009053605000000183>
- Lorden, G., & Pollak, M. (2008). Sequential Change-Point Detection Procedures that are Nearly Optimal and Computationally Simple. *Seq. Anal.*, 27, 476C512. <http://dx.doi.org/10.1080/07474940802446244>
- Robbins, H., & Siegmund, D. (1974). 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. (1975). The expected sample size of some tests of power one. *Ann. Statist.*, 2, 415-436. <http://dx.doi.org/10.1214/aos/1176342704>
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.  
<http://dx.doi.org/10.1007/978-1-4757-1862-1>
- Siegmund, D., & Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.*, 23, 255-271. <http://dx.doi.org/10.1214/aos/1176324466>
- Sparks, R. S. (2000). CUSUM charts for signalling varying location shift. *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. <http://dx.doi.org/10.1080/07474949908836432>
- Woodroffe, M. (1990). On a nonlinear renewal theorem. *Ann. Probab.*, 18, 1790-1805.  
<http://dx.doi.org/10.1214/aop/1176990649>
- Wu, Y. (2004). Bias of change point estimator detected by a CUSUM procedure. *Ann. Inst. Statist. Math.*, 56, 127-142. <http://dx.doi.org/10.1007/bf02530528>
- 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). Estimation of Change-point and Post-change Means by Adaptive CUSUM Procedures. In *2015 JSM Proceedings*, Alexandria, VA: American Statistical Association, 3675-3689.
- Wu, Y. (2016). Inference for post-change parameters after CUSUM test under AR(1) models. *J. Statist. Plann. Inf.*, 168, 52-67. <http://dx.doi.org/10.1016/j.jspi.2015.06.006>
- Yakir, B., Krieger, A. M., & Pollak, M. (1999). Detecting a change in regression: First order optimality. *Ann. Statist.*, 27, 1896-1913. <http://dx.doi.org/10.1214/aos/1017939243>
- Yashchin, E. (1995). Estimating the current mean of a process subject to abrupt changes. *Technometrics*, 37, 311-323. <http://dx.doi.org/10.1080/00401706.1995.10484337>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).