Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 1258–1280

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On the biases of change point and change magnitude estimation after CUSUM test[☆]

Keyue Ding[a,*], Yanhong Wu[b]

[a]*National Cancer Institute of Canada Clinical Trials Group, Queen's University, Kingston, Ontario, Canada K7L 3N6*

[b]*Department of Mathematics, University of the Pacific Stockton, CA 95211, USA*

Received 3 September 2002; accepted 13 September 2004
Available online 27 October 2004

Abstract

The present paper investigates the biases of estimates of change point and change magnitude after CUSUM test. By assuming that the change point is far from the beginning and the in-control average run length of samples is large, second order approximations for the biases of both estimates are obtained by conditioning on detection, and biases of both estimates are very significant. Simulation studies show the approximations to be quite accurate in the case of detecting an increase in mean or variance when sampling from a normal distribution. The results demonstrate the fundamental differences between fixed sample size test and sequential test.

MSC: primary 62L12; secondary 62L10; 62N10

Keywords: Bias; CUSUM process; Change-point; Standard one parameter exponential family; Random walk; Renewal theorem

1. Introduction

A sequence $\{X_k\}$ of random quantities—e.g., batch sample means or variances—are observed sequentially. The objectives are to detect the occurrence change point $\nu = k$ in

☆ This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

* Corresponding author.
E-mail address: kding@ctg.queensu.ca (K. Ding).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.09.005

the distribution of X_k, and to estimate the change point v and the change magnitude of the process mean. We suppose that $\{X_k\}$ are independent random variables with distribution function F_{θ_0} for $k \leqslant v$ and F_{θ_1} for $k > v$, where $\theta_0 < 0 < \theta_1$, v is called the change point, and F_{θ_0} and F_{θ_1} belong to a standard one parameter exponential family. Such a family is defined by $\mathrm{d}F_\theta(x) = \exp(\theta x - \psi(\theta))\mathrm{d}F_0(x)$, $\theta \in \Theta$, which contains an interval with 0 in it, and the function ψ is normalized so that $\psi(0) = \psi'(0) = 0$, $\psi''(0) = 1$, and $\psi'(\theta) <, =,$ or > 0 according to $\theta <, =,$ or > 0; and we also assume that the distribution F_0 is strongly non-lattice, i.e. $\lim_{|\lambda|\to\infty} \sup |E_0(\exp(\mathrm{i}\lambda X_1))| < 1$.

For quick detection of the change point, Page (1954) proposed the CUSUM procedure which makes an alarm at $N = \min\{n > 0 : T_n > d\}$, where T_n is the CUSUM process, defined as $T_n = \max(0, T_{n-1} + X_n)$, with $T_0 = 0$, and d is the control limit, a prescribed constant determining the in-control average run length of samples.

Optimality of the CUSUM procedure has been studied by Lorden (1971) and Moustakides (1986), while comparison with other procedures can be seen in Pollak and Siegmund (1985), Roberts (1966) and Srivastava and Wu (1993). The biases of estimates of change-point and change magnitude was studied by Srivastava and Wu (1999) for detecting a change in the drift of a Brownian motion process. Ding (2003) constructed a lower confidence bound for the change point after CUSUM test.

In this paper, we investigate the biases of estimates of change point and change magnitude after CUSUM test.

It is well known that when $\psi(\theta_0) = \psi(\theta_1)$ the maximum likelihood estimator for v is

$$\hat{v} = \max\{n < N : T_n = 0\},$$

i.e. the first zero point of T_n counting backward from the detection time N (Hinkley, 1970, 1971). However, as θ_1 is usually unknown, we would be interested in estimating θ_1 and v simultaneously. A natural way is still to use $\hat{v}$ as the estimator of v and to estimate θ_1 by first estimating $\mu_1 = \psi'(\theta_1)$ with

$$\hat{\mu}_1 = T_N/(N - \hat{v}),$$

and then solving the equation $\hat{\mu}_1 = \psi'(\theta_1)$ for θ_1 to obtain the estimate of the interested parameter, θ_1. In this aspect, Hinkley (1971) considered the biases of $\hat{\theta}$ and $\hat{v}$ in the cases of fixed and sequential sampling from a normal distribution. For the sequential case, a simulation study was taken without conditioning on $N > v$. The results show that negative bias for $\hat{v}$ appears mainly due to the false alarm possibility; and positive biases for $\hat{\theta}$. He also pointed out the necessity for further investigation.

It is well known that the maximum likelihood estimator following sequential test overestimates the underlining parameter (Cox, 1952; Siegmund, 1978; Whitehead, 1986; Woodroofe, 1992). There is no exception here. Conditioning on the change has happened, we find that the biases are quite substantial for both estimates.

The rest of the paper is organized as follows. In Section 2, we give some notations and assumptions which will be used throughout the paper. In Section 3, we present the main results of this paper, while give detailed proofs in Section 5. Applications of these results are given in Section 4. Finally, some concluding remarks are given in Section 6.

2. Notations and assumptions

In this section, we introduce some notations that will be used repeatedly, and the basic assumptions under which the main results of this paper are developed.

Denote

$$S_n = S_0 + \sum_{i=1}^{n} X_i \quad \text{and} \quad M = \sup_{0 \leqslant k < \infty} S_n \quad \text{with } S_0 = 0.$$

Let

$$\tau_x = \begin{cases} \inf\{n : S_n > x \mid S_0 = 0\} & \text{for } x > 0, \\ \inf\{n : S_n \leqslant x \mid S_0 = 0\} & \text{for } x < 0 \end{cases}$$

be the boundary crossing time, $R_x = S_{\tau_x} - x$ be the overshoot at the boundary x, and

$$\tau_+ = \inf\left\{n : S_n > 0 \mid S_0 = 0\right\} \quad \text{and} \quad \tau_- = \inf\left\{n > 0 : S_n \leqslant 0 \mid S_0 = 0\right\}$$

be the ascending and descending ladder epochs. Define the two-sided boundary crossing time by

$$N_x = \min\left\{n > 0 : S_n \leqslant 0 \text{ or } > d \mid S_0 = x\right\} \quad \text{for } 0 \leqslant x < d.$$

Let $P^{\nu}(.)$ denote the probability measure when the change point is ν and $P_\theta(.)$ the probability when the change occurs at zero and X_i's have distribution F_θ. It is well known that, for $\theta_0 < 0 < \theta_1$, there exist $\tilde{\theta}_1 < 0 < \tilde{\theta}_0$, s.t. $\psi(\theta_1) = \psi(\tilde{\theta}_1)$ and $\psi(\theta_0) = \psi(\tilde{\theta}_0)$. Denote $\Delta_i = \theta_i - \tilde{\theta}_i$, $\mu_i = \psi'(\theta_i)$ and $\tilde{\mu}_i = \psi'(\tilde{\theta}_i)$ for $i = 0, 1$.

Under any distribution of P_θ for $\theta \geqslant 0$, R_x approaches a limiting distribution as $x \to \infty$. Let R_∞ denote a random variable whose distribution is this limiting distribution. The following strong renewal theorem about the overshoot R_x will be used repeatedly (see Siegmund, 1979; Chang, 1992): *For* $x \geqslant 0$, *and* $y \geqslant 0$, *there exist* $\beta > 0$, *and* $\theta^* > 0$ *such that*, *uniformly for* $\theta \in [0, \theta^*]$,

$$\begin{aligned} P_\theta(R_x < y) &= P_\theta(R_\infty < y) + \mathrm{O}(\mathrm{e}^{-\beta(x+y)}) \\ &= \frac{1}{E_\theta S_{\tau_+}} \int_0^y P_\theta(S_{\tau_+} > z)\,\mathrm{d}z + \mathrm{O}(\mathrm{e}^{-\beta(x+y)}). \end{aligned} \tag{1}$$

Using the above result and Wald's likelihood ratio identity (Siegmund, 1985, p. 13), we obtain that: *For* $x \geqslant 0$, *and* $y \geqslant 0$, *there exist a* $\theta^* > 0$ *and positive constants* $C > 0$ *and* $\beta > 0$ *such that*, *uniformly for* $\theta \in [0, \theta^*]$,

$$|P_\theta(\tau_{-x} < \infty)/\mathrm{e}^{-\Delta x} E_{\tilde{\theta}} \mathrm{e}^{\Delta R_{-\infty}} - 1| \leqslant C \Delta \mathrm{e}^{-\beta x}, \tag{2}$$

where $\tilde{\theta} \leqslant 0$ *satisfies* $\psi(\theta) = \psi(\tilde{\theta})$ *and* $\Delta = \theta - \tilde{\theta}$.

The moments of R_∞ are given by

$$\rho^{(k)}(\theta) = E_\theta(R_\infty^k) = \frac{E_\theta(S_{\tau_+}^{k+1})}{(k+1) E_\theta(S_{\tau_+})}$$

for $k>0$. For convenience, let $\rho_+(\theta)=\rho^{(1)}(\theta)$, $\rho_+=\rho_+(0)$. Analogously, we define $\rho_-(\theta)$ and ρ_- upon replacing τ_+ by τ_-.

A relation that will be used in this paper is $(E_0 S_{\tau_-})(E_0 S_{\tau_+})=-\frac{1}{2}$, which holds for any standard one parameter exponential family (Chang, 1992, p. 716).

The asymptotic biases of both estimates are developed under the following assumption: (A): Both θ_0 and θ_1 approach zero at the same order, and, for some $\gamma>0$, $|\theta_0|^{1+\gamma}d\to\infty$.

3. Main results

In this section, we investigate the asymptotic biases of the estimate of change point $\hat{v}$ and the estimate of the process mean $\hat{\mu}_1$ under assumption (A) given in previous section.

3.1. Bias of $\hat{v}$

For the bias of the estimate of change point, we have the following result.

Theorem 1. *Under Condition* (A),

$$E[\hat{v}-v\,|\,N>v]=\frac{\Delta_0}{\tilde{\mu}_1(\Delta_1-\Delta_0)^2}\mathrm{e}^{\Delta_0(\rho_++\rho_-)}-\tilde{p}_0\frac{1}{\tilde{\mu}_1\Delta_1}+\frac{1}{\tilde{\mu}_0\Delta_0}$$
$$-\frac{\Delta_0}{\tilde{\mu}_0(\Delta_1-\Delta_0)^2}\mathrm{e}^{\Delta_1(\rho_++\rho_-)}+\mathrm{O}(1).$$

When $\psi(\theta_1)=\psi(\theta_0)$, *then*

$$E[\hat{v}-v\,|\,N>v]=-2\frac{\rho_++\rho_-}{\Delta_0}+\mathrm{O}(1).$$

Remark. From Theorem 1, we see that if the distribution of X_i is not symmetric under P_0, even in the case $\psi(\theta_1)=\psi(\theta_0)$, the bias of $\hat{v}$ is not negligible for small θ, and it is positive or negative according to $\rho_++\rho_->$ or <0. This gives us another typical example to show the effect of the sequential sampling rule. However, this bias becomes less significant for larger θ's. Comparisons with simulation results will be given in Section 4.

The proof of the theorem is presented in a series of Lemmas. At first, we write

$$E[\hat{v}-v\,|\,N>v]=E[\hat{v}-v;\,\hat{v}>v\,|\,N>v]-E[v-\hat{v};\,\hat{v}<v\,|\,N>v]. \tag{3}$$

To evaluate the first term on the right-hand side of (3), from the renewal property of T_n at the zero point, we have

$$\begin{aligned}E[\hat{v}-v;\,\hat{v}>v|N>v]&=E[\hat{v}-v;\,S_{N_{T_v}}\leqslant 0;\,\hat{v}>v|N>v]\\&=E_{\theta_1}[N_{T_v};\,S_{N_{T_v}}\leqslant 0]+P_{\theta_1}(S_{N_{T_v}}\leqslant 0)E_\theta(v_2),\end{aligned} \tag{4}$$

where v_2 is the length from the first zero point to the last zero point of S_n after v.

The following two lemmas give the second order approximations for the three terms on the right-hand side of (4) and their proofs can be found in Lemmas 2, 3 and 4 of Ding (2001).

Lemma 1. *Under Condition* (A),

$$P_{\theta_1}(S_{N_{T_v}} \leqslant 0) = P_{\theta_1}(\tau_{-M} < \infty) + \mathrm{O}(\mathrm{e}^{-\Delta_1 d}) = \tilde{p}_0 + \mathrm{O}(\Delta_1^3), \tag{5}$$

where M is the maximum value of another independent copy of S_n with drift μ_0 and

$$\tilde{p}_0 = C_1 \mathrm{e}^{(C_2/C_1)\Delta_0\Delta_1}, \tag{6}$$

with

$$C_0 = \int_0^\infty E_0(R_{-x} - \rho_-)\,\mathrm{d}E_0(R_x - \rho_+), \quad C_1 = \frac{-\Delta_0}{\Delta_1 - \Delta_0}\mathrm{e}^{\Delta_1(\rho_+ + \rho_-)},$$

$$C_2 = \tfrac{1}{2}\rho_+^2 + \tfrac{1}{2} + \rho_- E_0 S_{\tau_+} - \tfrac{1}{2}(\rho_+^2 - r_1) + \tfrac{1}{2}(\rho_-^2 - r_0) - C_0,$$

and

$$r_1 = \frac{E_0(S_{\tau_+}^3)}{3E_0(S_{\tau_+})} - \rho_+^2, \quad r_0 = \frac{E_0(S_{\tau_-}^3)}{3E_0(S_{\tau_-})} - \rho_-^2.$$

Lemma 2. *Under Condition* (A),

$$E_{\theta_1}[N_{T_v}; S_{N_{T_v}} \leqslant 0] = E_{\theta_1}[\tau_{-M}; \tau_{-M} < \infty] + \mathrm{o}(1)$$
$$= \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2}\mathrm{e}^{\Delta_0(\rho_+ + \rho_-)} + \mathrm{O}(1). \tag{7}$$

$$E_\theta(v_2) = -\frac{1}{\tilde{\mu}_1 \Delta_1} + \mathrm{O}(1).$$

Summarizing Lemmas 1–2, we have

Lemma 3. *Under Condition* (A),

$$E[\hat{v} - v; \hat{v} > v \mid N > v] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2}\mathrm{e}^{\Delta_0(\rho_+ + \rho_-)} - \tilde{p}_0 \frac{1}{\tilde{\mu}_1 \Delta_1} + \mathrm{O}(1).$$

To evaluate the second term on the right-hand side of (3), we first note that under Condition (A)

$$P(T_v < x \mid N > v) \to P(M < x);$$

see Lemma 1 of Ding (2003), and

$$P_{\theta_1}(S_{N_M} > d) = P_{\theta_1}(\tau_{-M} = \infty) + \mathrm{O}(\mathrm{e}^{-\Delta_1 d}).$$

On the other hand, by looking at $\{S_k\}$ backward in time starting from v, we see that $\hat{v}-v$ is actually the maximum point of $S'_n = S_v - S_{v-n}$ for $0 \leqslant n \leqslant v$ with drift θ_0. In fact, $T_{\hat{v}} = 0$ and $S_{\hat{v}} = \min_{0 \leqslant k \leqslant v} S_k$. Suppose $T_v = S_v - \min_{0 \leqslant k \leqslant v} S_k = x$. Then, We have

$$\max_{0 \leqslant n \leqslant v} S'_n = \max_{0 \leqslant k \leqslant v} (S_v - S_k) = x = S_v - S_{\hat{v}}.$$

Let σ_x denote the maximum point of S'_n with a maximum value of x. Then under Condition (A), by Wald's likelihood ratio identity, the second term of on the right-hand side (3) is equal to

$$E_{\theta_0}[\sigma_M P_{\theta_1}(\tau_{-M} = \infty)] + \mathrm{o}(1).$$

From Lemma 3 of Ding (2003), we have

Lemma 4. *Under Condition* (A),

$$E[v - \hat{v};\, \hat{v} \leqslant v \,|\, N > v] = -\frac{1}{\Delta_0 \tilde{\mu}_0} + \frac{\Delta_0}{\tilde{\mu}_0 (\Delta_1 - \Delta_0)^2} \mathrm{e}^{\Delta_1(\rho_+ + \rho_-)} + \mathrm{O}(1).$$

Combining the results of Lemmas 3 and 4, we complete the proof of Theorem 1.

3.2. Bias of $\hat{\mu}_1$

In this subsection, the second order approximation for the bias of $\hat{\mu}_1$ is developed, and the following results is obtained.

Theorem 2. *Under Condition* (A),

$$E[\hat{\mu}_1 - \mu_1 | N > v] = \frac{1}{d}\left(2 - \frac{\Delta_1^3}{2\Delta_0^2(\Delta_1 - \Delta_0)}\right) + \mathrm{o}\left(\frac{1}{d}\right).$$

In particular, when $\psi(\theta_1) = \psi(\theta_0)$, we have

$$E[\hat{\mu}_1 - \mu_1 | N > v] = \frac{7}{4d} + \mathrm{o}\left(\frac{1}{d}\right).$$

As we shall see in the practical situation (Section 4), the bias is quite significant and the bias correction for the estimate is definitely necessary.

To prove the theorem, we write

$$E[\hat{\mu}_1 | N > \hat{v}] = E\left[\frac{T_N}{N - \hat{v}};\, v > \hat{v} | N > v\right] + E\left[\frac{T_N}{N - \hat{v}};\, v \leqslant \hat{v} | N > v\right]. \tag{8}$$

The proof of theorem boils down to evaluate the two terms on the right-hand side of (8).

Note that conditioning on $\{v \leqslant \hat{v}\}$, $\{T_n\}$ for $n > \hat{v}$ behaves stochastically equivalent to a random walk $\{S_n\}$ for $n > 0$ conditioning on $\{S_{N_0} > d\}$. Thus, from Lemma 1, we have

$$\begin{aligned} E\left[\frac{T_N}{N-\hat{v}}; v < \hat{v} \,|\, N > v\right] &= E_{\theta_1}\left[\frac{S_{N_0}}{N_0}\middle| S_{N_0} > d\right] P_{\theta_1}(S_{N_{T_v}} \leqslant 0) \\ &= \left(P_{\theta_1}(\tau_{-M} < \infty) + \mathrm{O}(\mathrm{e}^{-\Delta_1 d})\right) \\ &\quad \times E_{\theta_1}\left[\frac{S_{N_0}}{N_0}\middle| S_{N_0} > d\right]. \end{aligned} \tag{9}$$

From a similar analysis as in Section 3.1, we have

$$\begin{aligned} E\left[\frac{T_N}{N-\hat{v}}; v > \hat{v} | N > v\right] &= E\left[\frac{T_N}{N-v+v-\hat{v}}; S_{N_{T_v}} > d\right] \\ &= E\left[\frac{S_{N_M}}{N_M + \sigma_M}\middle| S_{N_M} > d\right] \\ &\quad \times \left(P_{\theta_1}(\tau_{-M} = \infty) + \mathrm{O}(\mathrm{e}^{-\Delta_1 d})\right), \end{aligned} \tag{10}$$

where σ_M is as given in the section before Lemma 4.

To approximate $E_\theta[\frac{S_{N_0}}{N_0}|S_{N_0} > d]$, we use the following Taylor series expansion for $f(x, y) = \frac{x}{y}$, i.e.,

$$\begin{aligned} f(x,y) &= \frac{x_0}{y_0} + \frac{1}{y_0}(x-x_0) - \frac{x_0}{y_0^2}(y-y_0) + \frac{x_0}{y_0^3}(y-y_0)^2 - \frac{1}{y_0^2}(x-x_0)(y-y_0) \\ &\quad - \frac{x^*}{2y^{*4}}(y-y_0)^3 + \frac{1}{y^{*3}}(y-y_0)^2(x-x_0), \end{aligned} \tag{11}$$

where $|x^* - x_0| \leqslant |x - x_0|$ and $|y^* - y_0| \leqslant |y - y_0|$.

By letting $X = \frac{S_{N_0}}{E[N_0|S_{N_0}>d]}$, $Y = \frac{N_0}{E[N_0|S_{N_0}>d]}$, $x_0 = E[X|S_{N_0} > d]$ and $y_0 = E[Y|S_{N_0} > d]$, we get

$$\begin{aligned} E_{\theta_1}\left[\frac{S_{N_0}}{N_0}\middle| S_{N_0} > d\right] &\approx \frac{E[S_{N_0}|S_{N_0} > d]}{E[N_0|S_{N_0} > d]} + \frac{E[S_{N_0}|S_{N_0} > d]}{(E[N_0|S_{N_0} > d])^3}\mathrm{Var}(N_0|S_{N_0} > d) \\ &\quad - \frac{1}{(E[N_0|S_{N_0} > d])^2}\mathrm{Cov}(N_0, S_{N_0}|S_{N_0} > d). \end{aligned} \tag{12}$$

It will be shown in the Appendix that the error of this approximation is at the order of $\mathrm{o}(\frac{1}{d})$.

Approximations for the terms on the right-hand side of (12) will be given in next lemma.

Lemma 6. *Under Condition* (A),

$$E[S_{N_0}|S_{N_0} > d] = d + \rho_+ + \mathrm{o}(1). \tag{13}$$

$$E_{\theta_1}[N_0|S_{N_0} > d] = \frac{d+\rho_+}{\mu_1} - \frac{1}{\mu_1^2} + \mathrm{O}\left(\frac{1}{\Delta_1}\right). \tag{14}$$

$$\mathrm{Var}_{\theta_1}(N_0|S_{N_0} > d) = \frac{d+\rho_+}{\mu_1^3}\psi''(\theta_1) - \frac{2}{\mu_1^4} + \mathrm{O}\left(\frac{1}{\Delta_1^3}\right). \tag{15}$$

Eq. (13) is given by Siegmund (1985, (10.22)). The proofs of (14) and (15) will be given in Section 5.

By Cauchy–Schwarz inequality, we have

$$|\mathrm{Cov}(N_0, S_{N_0}|S_{N_0} > d)| \leqslant (\mathrm{Var}(N_0|S_{N_0} > d))^{1/2}(\mathrm{Var}(S_{N_0}|S_{N_0} > d))^{1/2}.$$

Since $\mathrm{Var}(S_{N_0}|S_{N_0} > d)$ approaches a constant, the last term of (12) is thus bounded by the order of $(\frac{\mu_1}{d+\rho_+})^2(\frac{d+\rho_+}{\mu_1^3})^{1/2} = \mathrm{o}(\frac{\mu_1}{d+\rho_+})$. Combining the results of Lemma 6, the following result is obtained after some simplifications.

Lemma 7. *Under Condition* (A),

$$E\left[\frac{S_{N_0}}{N_0}\middle| S_{N_0} > d\right] = \mu_1 + \frac{2}{d} + \mathrm{o}\left(\frac{1}{d}\right).$$

By similar arguments leading to (12), we have

$$\begin{aligned} E\left[\frac{S_{N_M}}{N_M + \sigma_M}\middle| S_{N_M} > d\right] = &\frac{E[S_{N_M}|S_{N_M} > d]}{E[N_M + \sigma_M|S_{N_M} > d]} \\ &+ \frac{E[S_{N_M}|S_{N_M} > d]}{(E[N_M + \sigma_M|S_{N_M} > d])^3} \\ &\times \mathrm{Var}(N_M + \sigma_M|S_{N_M} > d) + \mathrm{o}\left(\frac{1}{d}\right). \end{aligned} \tag{16}$$

Again, we shall show in the Appendix that the approximation is indeed at the order of $\mathrm{o}(\frac{1}{d})$. The following lemma gives the corresponding results as in Lemma 6 and its proof will be given in Section 5.

Lemma 8. *Under Condition* (A),

$$E[S_{N_M}|S_{N_M} > d] = d + \rho_+ + \mathrm{o}(1), \tag{17}$$

$$E[N_M|S_{N_M} > d] = \frac{d+\rho_+}{\mu_1} + \frac{1}{\mu_1\Delta_0} - \frac{1}{\mu_1\Delta_1} - \frac{\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)} + \mathrm{O}\left(\frac{1}{\Delta_1}\right), \tag{18}$$

$$E[\sigma_M|S_{N_M} > d] = -\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)} + \mathrm{O}\left(\frac{1}{\Delta_1}\right), \tag{19}$$

$$\mathrm{Var}(N_M + \sigma_M|S_{N_M} > d) = \frac{d+\rho_+}{\mu_1^3}\psi''(\theta_1) + \mathrm{O}\left(\frac{1}{\Delta_1^4}\right). \tag{20}$$

Combining the results of Lemma 8, we obtain

Lemma 9. *Under Condition* (A),

$$E\left[\frac{S_{N_M}}{N_M + \sigma_M}\middle| S_{N_M} > d\right] = \mu_1 + \frac{1}{d}\left(2 - \frac{\Delta_1^2}{2\Delta_0^2}\right) + \mathrm{o}\left(\frac{1}{d}\right).$$

Summing up the results of Lemmas 7 and 9, we complete the proof of the Theorem 2.

4. Application

In this section, we demonstrate how to use our results in a practical situation, and conduct simulation studies to check the accuracy of results given in Section 3.

In the field of quality control, we might be monitoring some characteristic of a manufacturing process, and in many practical situations, the quality characteristic process Z_k is assumed to be normally distributed with mean μ_0 and standard deviation σ_0, where μ_0 is the target value and σ_0 reflects the variation of quality. Any shift from the target value μ_0 or increase in the process variation results in poor quality, and we want to detect the change as soon as possible. To monitor the quality characteristic process, random samples of size, say m, are usually taken at some regular time interval, and CUSUM charts based on sample mean $X_k = \bar{Z}_k$ or sample variance $X_k = s_k^2$ are plotted with the appropriate control limit(s), respectively. We would like to stop the process for inspection and repair at the point that the CUSUM chart falls out side of control limit. Since false alarm are often costly, we therefore assume the in-control average run length (ARL) of samples to be very large.

In the following, we will apply our results to detect change in the mean or the standard deviation, respectively. For simplicity, we only consider detecting an increase.

4.1. Detect increase in mean

Without lose of generality, we assume that the observed process $X_k = \bar{Z}_k$ comes from a normal population with mean 0 and standard deviation 1 when the process is in control, and with mean μ and standard deviation 1 when the process is out of control. Then the CUSUM procedure is defined as making alarm at $N = \inf\{n > 0 : T_n > d\}$, where T_n is the CUSUM process $T_n = \max(0, T_{n-1} + Y_n)$, with $T_0 = 0$, and $Y_n = \bar{X}_n - \frac{\delta}{2}$, and d is the control limit with reference value δ. Usually, the reference value δ is the change magnitude which we are interested in detecting quickly, and serves as a preliminary estimate of μ, the true change magnitude.

In this case,

$$\psi(\theta) = \frac{1}{2}\theta^2, \quad \theta_0 = -\frac{\delta}{2}, \quad \theta_1 = u - \frac{\delta}{2},$$

$$\mu_0 = -\tilde{\mu}_0 = -\frac{\delta}{2}, \quad \mu_1 = -\tilde{\mu}_1 = u - \frac{\delta}{2},$$

$$E_0 S_{\tau_+} = -E_0 S_{\tau_-} = \frac{1}{\sqrt{2}}, \quad \rho_+ = -\rho_- \approx 0.583,$$

$$r_0 = r_1 = \frac{1}{4} \quad \text{and} \quad C_0 = \frac{1}{2}\left(\frac{1}{\sqrt{2}} - \rho_+\right)^2.$$

Table 1
Comparison of approximated and simulated bias(normal)

$\theta_0(d)$	θ	Bias ($\hat{v}$)	Bias ($\hat{\theta}$)
−0.2	0.2	0	0.1757
(9.96)		−0.816(0.28)	0.1755(0.0022)
	0.25	−4.5	0.1628
		−3.99(0.28)	0.1533(0.0023)
	0.30	−6.94	0.1431
		−6.05 (0.24)	0.1252(0.0024)
−0.25	0.25	0	0.1975
(8.86)		−0.131(0.21)	0.1938(0.0026)
	0.30	−2.44	0.1865
		−2.47(0.18)	0.1685(0.0026)
	0.40	−4.875	0.1498
		−3.77(0.16)	0.1463(0.0027)
−0.3	0.3	0	0.2267
(7.72)		−0.28(0.14)	0.2096(0.0028)
	0.35	−1.47	0.2166
		−1.41(0.12)	0.1876(0.0029)
	0.4	−2.43	0.2025
		−2.44(0.11)	0.1616(0.0029)

Insert the related quantities into Theorem 1 and 2, we have

Corollary 1. *Under Condition* (A),

$$E[\hat{v} - v]|N > v] = \frac{1}{2(\mu - \delta/2)^2} - \frac{2}{\delta^2} + \mathrm{O}(1), \tag{21}$$

$$E[\hat{\mu}_1 - \mu_1|N > v] = \frac{1}{d}\left(2 - \frac{2(\mu - \delta/2)^3}{\mu\delta^2}\right) + \mathrm{o}\left(\frac{1}{d}\right), \tag{22}$$

when $\mu = \delta$,

$$E[\hat{v} - v]|N > v] = \mathrm{O}(1), \tag{23}$$

$$E_[\hat{\mu}_1 - \mu_1|N > v] = \frac{7}{4d} + \mathrm{o}\left(\frac{1}{d}\right). \tag{24}$$

Simulation results based on 10,000 replications are presented as follows. We choose the average in-control run length $\mathrm{ARL}_0 = E_{\theta_0}N = 1000$, and $v = 100$. The control limit d is determined by using Siegmund's approximation (Siegmund, 1985 (2.57)). For $\theta_0 = -0.2, -0.25, -0.3$, we obtain that $d = 9.96$, 8.82 and 7.72, respectively. Table 1 presents the comparisons of simulated and approximated biases for $\hat{v}$ and $\hat{\mu}_1$. In each cell, the top number is the approximation value while the bottom number is the simulated value and its standard error is given in the parentheses. We can see that the theoretical values are quite close to simulation values.

4.2. Detect increase in variance

Without lose of generality, we assume that the observed process $X_k = s_k^2$ comes from a population of $\chi^2(p)$ when the process is in control, and from a population of $(1+\varepsilon)^2\chi^2(p)(\varepsilon > 0)$ when the process is out of control. Then the CUSUM procedure is defined as making alarm at $N = \inf\{n > 0 \ : \ T_n > d\}$, where $T_n = \max(0, T_{n-1} + Y_n)$, with $T_0 = 0$, and $Y_n = \frac{s_n^2[(1+\varepsilon_0)^2-1]}{2\sqrt{2p}(1+\varepsilon_0)^2\ln(1+\varepsilon_0)} - \sqrt{\frac{p}{2}}$, and d is the control limit with reference value ε_0. The reference value ε_0 is the relative increase change magnitude we are interested in detecting quickly, and serves as an preliminary estimate of ε, the true relative increase.

In this case, $\psi(\theta) = -\sqrt{\frac{p}{2}}\theta - \frac{p}{2}\ln(1 - \sqrt{\frac{2}{p}}\theta)$, and

$$\theta_0 = \sqrt{\frac{p}{2}}\left(1 - \frac{2(1+\varepsilon_0)^2\ln(1+\varepsilon_0)}{(1+\varepsilon_0)^2 - 1}\right),$$

$$\theta_1 = \sqrt{\frac{p}{2}}\left(1 - \frac{2(1+\varepsilon_0)^2\ln(1+\varepsilon_0)}{(1+\varepsilon)^2[(1+\varepsilon_0)^2 - 1]}\right),$$

$$\mu_0 = \sqrt{\frac{p}{2}}\left(\frac{(1+\varepsilon_0)^2 - 1}{2(1+\varepsilon_0)^2\ln(1+\varepsilon_0)} - 1\right),$$

$$\mu_1 = \sqrt{\frac{p}{2}}\left(\frac{(1+\varepsilon)^2[(1+\varepsilon_0)^2 - 1]}{2(1+\varepsilon_0)^2\ln(1+\varepsilon_0)} - 1\right).$$

For simplicity, we only give results of a special case, i.e. $p=2$. As it becomes an important case of detecting a change in the mean of an exponential distribution family, which plays a critical role in reliability. In this case, S_{τ_+} is an exponential random variable, and S_{τ_-} is uniformly distributed on $(-1, 0)$ under P_0. So, we have that $\rho_+ = 1$, $r_1 = 1$, $\rho_- = -\frac{1}{3}$, $r_0 = \frac{1}{18}$ and $C_0 = 0$.

Substituting the above corresponding values into Theorems 1 and 2, the following results are obtained.

Corollary 2. *Under Condition* (A), *we have*

$$E[\hat{v} - v|N > v] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2}e^{2/3\Delta_0} - \tilde{p}_0\frac{1}{\tilde{\mu}_1\Delta_1} + \frac{1}{\Delta_0\tilde{\mu}_0} - \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2}e^{2/3\Delta_1} + O(1), \tag{25}$$

Table 2
Comparison of approximated and simulated bias (exponential)

$\varepsilon_0(d)$	ε_1	Bias ($\hat{v}$)	Bias ($\hat{\mu}_1$)
0.2	0.2	3.61	0.2188
(10.083)		3.79(0.304)	0.3372(0.0055)
	0.25	−4.27	0.2108
		−3.08(0.24)	0.3019(0.0063)
	0.3	−7.33	0.1935
		−6.08(0.21)	0.2791(0.0070)
0.25	0.25	2.80	0.2659
(8.794)		2.34(0.23)	0.4042(0.0068)
	0.3	−1.67	0.2683
		−1.17(0.18)	0.3834(0.0074)
	0.35	−3.71	0.2660
		−3.59(0.17)	0.3662(0.0080)
0.3	0.3	2.26	0.3169
(7.823)		2.05(0.17)	0.4866(0.0087)
	0.35	−0.51	0.3281
		−0.42(0.15)	0.4522(0.0087)
	0.4	−1.93	0.3375
		−2.07(0.13)	0.4557(0.0098)

where $\tilde{p}_0 = C_1 \mathrm{e}^{25/36 C_1 \Delta_0 \Delta_1}$, *with* $C_1 = -\frac{\Delta_0}{\Delta_1 - \Delta_0}\mathrm{e}^{2/3\Delta_1}$.

$$E[\hat{\mu}_1 - \mu_1 | N > v] = \frac{1}{d}\left(1 + \frac{(1+\varepsilon)^4[(1+\varepsilon_0)^2 - 1]^2}{(1+\varepsilon_0)^4[\ln(1+\varepsilon_0)^2]^2} - \frac{\Delta_1^3}{2\Delta_0^2(\Delta_1 - \Delta_0)}\right) + \mathrm{o}\left(\frac{1}{d}\right). \tag{26}$$

In particular when $\psi(\theta_1) = \psi(\theta_0)$, *i.e.* $\varepsilon_0 = \varepsilon$,

$$E[\hat{v} - v | N > v] = -\frac{4}{3\Delta_0} + \mathrm{O}(1),$$

$$E[\hat{\mu}_1 - \mu_1 | N > v] = \frac{1}{d}\left(\frac{3}{4} + \frac{[(1+\varepsilon_0)^2 - 1]^2}{[\ln(1+\varepsilon_0)^2]^2}\right) + \mathrm{o}\left(\frac{1}{d}\right).$$

Simulation results based on 10,000 replications are reported in Table 2. We choose the average in-control run length $\mathrm{ARL}_0 = E_{\theta_0} N = 1000$, and $v = 100$. The control limit d is obtained by (10.17) of Siegmund (1985), i.e.

$$\mathrm{ARL}_0 \approx \frac{1}{\Delta_0 \mu_0}[\mathrm{e}^{-\Delta_0(d + \rho_+ - \rho_-)} - 1 + \Delta_0(d + \rho_+ - \rho_-)].$$

From this formula, we obtain that $d = 10.083, 8.794$, and 7.823 corresponding to $\varepsilon_0 = 0.2, 0.25$ and 0.3, respectively. The order of the numbers are arranged the same way as those in Table 1.

From Table 2, we can draw the following conclusions. First, the approximations for the bias of $\hat{v}$ is generally very precise. Second, we see that the bias for $\hat{\mu}_1$ is very large and bias correction is definitely required. This is the most important conclusion of this subsection. Third, approximation for the bias of $\hat{\mu}_1$ is systemly less than the simulated bias. This may be due to the following reasons. (1): The expansion of $E[\hat{\mu}_1 - \mu_1|N > v]$ is very weak. (2): The approximation only uses information contained in the first two moments of the population distribution, and reflects nothing about the skewness and kurtosis of the population distribution. Therefore, when the population distribution is skewed, the approximation cannot be very precise.

5. Proof of lemmas

(i) Proof of Lemma 6. Without notational confusion, we shall omit the subscript θ in $E_\theta[.]$ as well as in the rest of discussions.

To prove (14) of Lemma 6, from Lemma 3 of Ding (2003),

$$E_{\theta_1}[N_0; S_{N_0} > d] = \frac{d+\rho_+}{\mu_1} P_{\theta_1}(S_{N_0} > d) + \frac{1}{\mu_1} E_0 S_{\tau_-} - \frac{1}{\tilde{\mu}_1} E_0 S_{\tau_-} + \mathrm{O}(1).$$

Note that under Condition (A),

$$\begin{aligned} P_{\theta_1}(S_{N_0} > d) &= P_{\theta_1}(\tau_- = \infty) + \mathrm{O}(\mathrm{e}^{\Delta_1 d}) \\ &= \frac{1}{E_{\theta_1}(\tau_+)} + \mathrm{O}(\mathrm{e}^{\Delta_1 d}) = \frac{\mu_1}{E_0 S_{\tau_+}} + \mathrm{O}(\Delta_1^2). \end{aligned}$$

By the fact that $E_0 S_{\tau_+} E_0 S_{\tau_-} = -\frac{1}{2}$, we have

$$E[N_0|S_{N_0} > d] = \frac{d+\rho_+}{\mu_1} - \frac{1}{\mu_1^2} + \mathrm{O}\left(\frac{1}{\Delta_1}\right).$$

To prove (15), we first write

$$\mathrm{Var}(N_0|S_{N_0} > d) = E[N_0^2|S_{N_0} > d] - (E[N_0|S_{N_0} > d])^2.$$

Note that

$$E[N_0^2; S_{N_0} > d] = E[N_0^2] - E[\tau_-^2; \tau_- < \infty] + \mathrm{o}(1). \tag{27}$$

By Lemma 5 of Ding (2001),

$$
\begin{aligned}
E[N_0^2] &= \frac{1}{\mu_1^2}[\psi''(\theta_1)EN_0 + 2\mu_1 E(N_0 S_{N_0}) - ES_{N_0}^2] \\
&= \frac{1}{\mu_1^2}[\psi''(\theta_1)E[N_0|S_{N_0} > d]P_{\theta_1}(S_{N_0} > d) \\
&\quad + \psi''(\theta_1)E_{\theta_1}(\tau_-, \tau_- < \infty) \\
&\quad + 2\mu_1(d+\rho_+)E_{\theta_1}[N_0|S_{N_0} > d]P_{\theta_1}(S_{N_0} > d) \\
&\quad + 2\mu_1 E_{\theta_1}(S_{\tau_-}\tau_-, \tau_- < \infty) \\
&\quad - (d+\rho_+)^2 P_{\theta_1}(S_{N_0} > d) - E_{\theta_1}(S_{\tau_-}^2, \tau_- < \infty)] + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right) \\
&= \left[\frac{(d+\rho_+)^2}{\mu_1^2} + \frac{d+\rho_+}{\mu_1^3}\left(-1 + \frac{\mu_1}{\tilde{\mu}_1} + \psi''(\theta_1)\right) - \frac{1}{2\mu_1^4}\psi''(\theta_1)\right] \\
&\quad \times P_{\theta_1}(S_{N_0} > d) + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right).
\end{aligned} \tag{28}
$$

On the other hand,

$$
E_{\theta_1}[\tau_-^2; \tau_- < \infty] = E_{\tilde{\theta}_1}(\tau_-^2 \mathrm{e}^{\Delta_1 S_{\tau_-}}) = E_{\tilde{\theta}_1}(\tau_-^2) + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right).
$$

Similar to the above proof, we have

$$
\begin{aligned}
E_{\tilde{\theta}_1}(\tau_-^2) &= \frac{1}{\tilde{\mu}_1^2}[2\tilde{\mu}_1 E_{\tilde{\theta}_1}(S_{\tau_-}\tau_-) - E_{\tilde{\theta}_1} S_{\tau_-}^2 + \psi''(\tilde{\theta}_1)E_{\tilde{\theta}_1}(\tau_-)] \\
&= \frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3} E_0(S_{\tau_-}) + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right).
\end{aligned} \tag{29}
$$

From (28) and (29), we have

$$
E[N_0^2|S_{N_0} > d] = \frac{(d+\rho_+)^2}{\mu_1^2} + \frac{d+\rho_+}{\mu_1^3}(\psi''(\theta_1) - 2) - \frac{1}{\mu_1^4} + \mathrm{O}\left(\frac{1}{\Delta_1^3}\right),
$$

where in the last equation, we use the fact that $\mu_1 = -\tilde{\mu}_1 + \mathrm{O}(\Delta_1^2)$ and $\psi''(\tilde{\theta}_1) = 1 + \mathrm{O}(\Delta_1)$. Combining the above result with (14) and (15) is proved. $\square$

(ii) Proof of Lemma 8. By Lemma 2, we have

$$\begin{aligned}E[N_M; S_{N_M} > d] &= E[N_M] - E[\tau_{-M}; \tau_{-M} < \infty] + \mathrm{o}(1)\\ &= \frac{1}{\mu_1}[ES_{N_M} - E(M)] - E[\tau_{-M}; \tau_{-M} < \infty] + \mathrm{o}(1)\\ &= \frac{1}{\mu_1}\left[(d+\rho_+)P_{\theta_1}(S_{N_M} > d) + \frac{1}{\Delta_0}\right] - \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2}\\ &\quad + \mathrm{O}\left(\frac{1}{\Delta_1}\right).\end{aligned}$$

So we have

$$E[N_M|S_{N_M} > d] = \frac{d+\rho_+}{\mu_1} + \frac{1}{\mu_1\Delta_0} - \frac{1}{\mu_1\Delta_1} - \frac{\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)} + \mathrm{O}\left(\frac{1}{\Delta_1}\right),$$

where in the last equation we use results of Lemmas 1 and 2, and the fact that $\mu_1 = -\tilde{\mu}_1 + \mathrm{O}(\Delta_1^2)$. Eq. (18) is proved.

To prove (19), we write

$$\begin{aligned}E[\sigma_M, S_{N_M} > d] &= E[\sigma_M] - E[\sigma_M, S_{N_M} < 0]\\ &= E[\sigma_M] - E[\sigma_M, \tau_{-M} < \infty] + \mathrm{O}(1).\end{aligned}$$

Combing Lemma 1 and Eqs. (15) and (16) in Lemma 2 of Ding (2003), we complete the proof of (19).

Since the idea in proving (20) is similar to those used in the proof of (15), some details are thus omitted.

First, we write

$$\mathrm{Var}(N_M+\sigma_M|S_{N_M}>d) = E[(N_M+\sigma_M)^2|S_{N_M} > d] - (E[N_M+\sigma_M|S_{N_M} > d])^2,$$

and

$$\begin{aligned}E[(N_M+\sigma_M)^2|S_{N_M} > d] &= E[N_M^2|S_{N_M} > d] + 2E[N_M\sigma_M|S_{N_M} > d]\\ &\quad + E[\sigma_M^2|S_{N_M} > d].\end{aligned}$$

It is easy to verify that

$$E[\sigma_M^2|S_{N_M} > d] = \mathrm{O}\left(\frac{1}{\Delta_0^4}\right),$$

and

$$\begin{aligned}2E[N_M\sigma_M|S_{N_M} > d] &= 2E[N_M|S_{N_M} > d]E[\sigma_M|S_{N_M} > d] + \mathrm{O}\left(\frac{1}{\Delta_1^4}\right)\\ &= 2\left(\frac{d+\rho_+}{\mu_1}\right)\left(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)}\right)\\ &\quad + \mathrm{O}\left(\frac{1}{\Delta_1^4}\right)\end{aligned}$$

from (18) and (19).

Next, we note that

$$E[N_M^2; S_{N_M} > d] = E[N_M^2] - E[N_M^2; S_{N_M} \leqslant 0],$$

and

$$\begin{aligned} E[N_M^2] &= \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_M] + 2\mu_1 E[N_M(S_{N_M} - M)] - E(S_{N_M} - M)^2) \\ &= \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_M] + 2\mu_1 E[N_M S_{N_M}] - 2\mu_1 E[N_M M] - E[S_{N_M}^2] \\ &\quad + 2E[S_{N_M} M] - E[M^2]). \end{aligned}$$

Since

$$\begin{aligned} E[N_M S_{N_M}] &= (d+\rho_+)E[N_M | S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) + E[N_M S_{N_M}; S_{N_M} \leqslant 0] \\ &= (d+\rho_+)E[N_M | S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right), \end{aligned}$$

and

$$\begin{aligned} E[N_M M] &= E[M E(N_M|M)] = \frac{1}{\mu_1}E[M(S_{N_M} - M)] = \frac{1}{\mu_1}E[M S_{N_M}] - EM^2 \\ &= \frac{d+\rho_+}{\mu_1}E[M; S_{N_M} > d] - \frac{1}{\mu_1}EM^2 + \mathrm{O}\left(\frac{1}{\Delta_1^2}\right). \end{aligned}$$

While

$$\begin{aligned} E[M; S_{N_M} > d] &= E(M) - E[M; S_{N_M} < 0] \\ &= -\frac{1}{\Delta_0} - E[M; \tau_{-M} < \infty] + \mathrm{O}(1); \end{aligned}$$

and

$$E[M; \tau_{-M} < \infty] = E[M P_{\theta_1}(\tau_{-M} < \infty)] = E_{\theta_0}[M\mathrm{e}^{-\Delta_1 M}](1 + \mathrm{O}(\Delta_1)).$$

By the similar techniques used in proving Lemma 3 of Ding (2003), we have

$$\begin{aligned} E_{\theta_0}[M\mathrm{e}^{-\Delta_1 M}] &= \sum_{k=2}^{\infty} E_{\theta_0}\left[\sum_{i=1}^{k-1} S_{\tau_i^+}\mathrm{e}^{-\Delta_1 \sum_{i=1}^{k-1} S_{\tau_i^+}}\right](1-p)^{k-1}p \\ &= \sum_{k=2}^{\infty}(k-1)E_{\theta_0}[S_{\tau_+}\mathrm{e}^{-\Delta_1 S_{\tau_+}}; \tau_+ < \infty] \\ &\quad \times (E_{\theta_0}[\mathrm{e}^{-\Delta_1 S_{\tau_+}}; \tau_+ < \infty])^{k-2}p \\ &= p\frac{E_{\theta_0}[S_{\tau_+}\mathrm{e}^{-\Delta_1 S_{\tau_+}}; \tau_+ < \infty]}{(1 - E_{\theta_0}[\mathrm{e}^{-\Delta_1 S_{\tau_+}}; \tau_+ < \infty])^2} \\ &= -\frac{\Delta_0}{(\Delta_1 - \Delta_0)^2} + \mathrm{O}(1), \end{aligned}$$

where $p = P_{\theta_0}(\tau_+ = \infty)$ and $S^{(i)}_{\tau_+}$, for $i > 0$, are iid r.v.s which have the same distribution function as that of $S_{\tau_+}|\tau_+ < \infty$ under P_{θ_0}. Therefore,

$$\begin{aligned}
E[M; S_{N_M} > d] &= -\frac{1}{\Delta_0} + \frac{\Delta_0}{(\Delta_1 - \Delta_0)^2} + \mathrm{O}(1),\\
E[S^2_{N_M}] &= (d+\rho_+)^2 P_{\theta_1}(S_{N_M} > d) + \mathrm{O}(d),\\
E[S_{N_M} M] &= (d+\rho_+)E[M; S_{N_M} > d] + \mathrm{O}(1)\\
&= (d+\rho_+)\left(-\frac{1}{\Delta_0} + \frac{\Delta_0}{(\Delta_1 - \Delta_0)^2}\right) + \mathrm{O}(1).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
E[N_M^2|S_{N_M} > d] = {} & \frac{(d+\rho_+)^2}{\mu_1^2} + \frac{d+\rho_+}{\mu_1^3}(\psi''(\theta_1) + \frac{2(\Delta_1 - \Delta_0)\mu_1}{\Delta_1\Delta_0}\\
& - \frac{2\mu_1^2\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)}) + \mathrm{O}\left(\frac{1}{\Delta_1^4}\right).
\end{aligned}$$

Combining above results, we have

$$\begin{aligned}
\mathrm{Var}(N_M + \sigma_M|S_{N_M} > d) &= E[(N_M + \sigma_M)^2|S_{N_M} > d]\\
&\quad - (E[(N_M + \sigma_M)|S_{N_M} > d])^2\\
&= \frac{d+\rho_+}{\mu_1^3}\psi''(\theta_1) + \mathrm{O}\left(\frac{1}{\Delta_1^4}\right). \qquad \square
\end{aligned}$$

6. Concluding remarks

It should be pointed out that there are still some problems to be solved before these methods are fully practical. The critical issue is that the bias of the usual estimate is quite substantial, and the bias correction is definitely necessary. To this end, we propose the following bias correction method:

First substitute the initial estimation of $\hat{\theta}$ into the approximations for the biases of $\hat{v}$ and $\hat{\theta}$ given in Theorems 1 *and* 2, *respectively*, *and obtain the estimated biases. Then let*

$$\tilde{\theta} = \hat{\theta}/(1 + (\mathrm{est.bias}(\hat{\theta}))/\hat{\theta});$$

and

$$\tilde{v} = \hat{v} - (est.bias(\hat{v})).$$

This method slightly over-corrects the bias of $\hat{v}$ to the left and under-corrects the bias of $\hat{\theta}$, but the formula is very simple and easy to be implemented for practical purposes. Details will be communicated in a future presentation.

Acknowledgements

We are very grateful to two referees for their various constructive criticisms on the manuscript which led to substantial improvement in presentation and clarity. Our thanks are also due to Professor Subir Ghosh for many helpful suggestions.

Appendix A. Error checking for (12) and (16)

To show that the remain terms of (12) are at the order of $\mathrm{o}(\frac{1}{d+\rho})$, we note the fact that conditioning on the event $\{S_{N_0} > d\}$, $S_{N_0}/E[N_0|S_{N_0} > d] - E[S_{N_0}|S_{N_0} > d]/E[N_0|S_{N_0} > d]$ and $N_0/E[N_0|S_{N_0} > d] - 1$ converge to 0 in probability. Thus, for the third order terms in the Taylor expansion, the coefficients of $E[(\frac{S_{N_0}}{E[N_0|S_{N_0}>d]} - \frac{E[S_{N_0}|S_{N_0}>d]}{E[N_0|S_{N_0}>d]})(\frac{N_0}{E[N_0|S_{N_0}>d]} - 1)^2|S_{N_0} > d]$ converges to 1 in probability; while the coefficient of $E[(\frac{N_0}{E[N_0|S_{N_0}>d]} - 1)^3|S_{N_0} > d]$ is at the order of $\frac{E[S_{N_0}|S_{N_0}>d]}{E[N_0|S_{N_0}>d]} = \mathrm{O}(\mu_1)$ in probability. Thus, we only need to show that

$$E\left[\left(\frac{S_{N_0}}{E[N_0|S_{N_0} > d]} - \frac{E[S_{N_0}|S_{N_0} > d]}{E[N_0|S_{N_0} > d]}\right)\left(\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right)^2 \middle| S_{N_0} > d\right] = \mathrm{o}\left(\frac{1}{d+\rho}\right), \tag{A.1}$$

and

$$E\left[\left(\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right)^3 \middle| S_{N_0} > d\right] = \mathrm{o}\left(\frac{1}{\mu_1(d+\rho)}\right). \tag{A.2}$$

To show (A.2), we first write

$$\begin{aligned}&E[(N_0 - E[N_0|S_{N_0} > d])^3|S_{N_0} > d]\\&\quad = E[N_0^3|S_{N_0} > d] - 3E[N_0|S_{N_0} > d]E[N_0^2|S_{N_0} > d] + 2(E[N_0|S_{N_0} > d])^3,\end{aligned}$$

and note that

$$\begin{aligned}E[N_0^3; S_{N_0} > d] &= E[N_0^3] - E[N_0^3; S_{N_0} \leqslant 0]\\&= EN_0^3 - E[\tau_-^3; \tau_- < \infty] + \mathrm{o}(1).\end{aligned}$$

It is easy to check that $\{(S_n - \theta n)^3 - 3n\psi''(\theta_1)(S_n - \theta n) - n\psi^{(3)}(\theta_1), \mathfrak{I}_n\}$ is a martingale, where $\mathfrak{I}_n = \sigma\{X_1, \ldots, X_n\}$. Thus, we have

$$\begin{aligned}E_{\theta_1}[N_0^3] = \frac{1}{\mu_1^3}[&ES_{N_0}^3 - 3\mu_1 E(N_0 S_{N_0}^2) + 3\mu_1^2\psi''(\theta_1)E(N_0^2 S_{N_0})\\&- 3\psi''(\theta_1)E(N_0 S_{N_0}) + 3\mu_1\psi''(\theta_1)EN_0^2 - \psi^{(3)}(\theta_1)EN_0]\end{aligned}$$

$$= \frac{1}{\mu_1^3}[(d+\rho_+)^3 P(S_{N_0} > d) - 3\mu_1 (d+\rho_+)^2 E(N_0|S_{N_0} > d) P(S_{N_0} > d) \\ + 3\mu_1 \psi''(\theta_1) E N_0^2] + \mathrm{O}\left(\frac{(d+\rho_+)^2}{\mu_1^3}\right) + \mathrm{O}\left(\frac{1}{\mu_1^6}\right).$$

Combining with the results in Lemma 7, we get

$$E[(N_0 - E[N_0|S_{N_0} > d])^3 | S_{N_0} > d] = 3\psi''(\theta_1)\frac{d+\rho_+}{\mu_1^5} + \mathrm{O}\left(\frac{(d+\rho_+)^2}{\mu_1^3}\right) \\ + \mathrm{O}\left(\frac{1}{\mu_1^6}\right).$$

So the left-hand side of (A.2) is at the order of

$$3\left(\frac{d+\rho_+}{\mu_1}\right)^{-3} \psi''(\theta_1)\frac{d+\rho_+}{\mu_1^5} = \frac{1}{(d+\rho_+)^2\mu_1^2}\psi''(\theta_1) = \mathrm{o}\left(\frac{1}{\mu_1(d+\rho)}\right),$$

(A.2) is proved.

To show (A.1), by Hölder inequality, we have

$$\left| E\left[\left(\frac{S_{N_0}}{E[N_0|S_{N_0} > d]} - \frac{E[S_{N_0}|S_{N_0} > d]}{E[N_0|S_{N_0} > d]}\right)\left(\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right)^2 \middle| S_{N_0} > d\right]\right| \\ \leqslant \left(E\left[\left|\frac{S_{N_0}}{E[N_0|S_{N_0} > d]} - \frac{E[S_{N_0}|S_{N_0} > d]}{E[N_0|S_{N_0} > d]}\right|^3 \middle| S_{N_0} > d\right]\right)^{1/3} \\ \times \left(E\left[\left|\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right|^3 \middle| S_{N_0} > d\right]\right)^{2/3} \\ = \frac{1}{E[N_0|S_{N_0} > d]}(E|R_d - ER_d|^3)^{1/3} \\ \times \left(E\left[\left|\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right|^3 \middle| S_{N_0} > d\right]\right)^{2/3}.$$

By the previous argument, we know that

$$E\left[\left|\frac{N_0}{E[N_0|S_{N_0} > d]} - 1\right|^3 | S_{N_0} > d\right] = \mathrm{O}(1).$$

Thus, the left-hand side of (A.1) is at the order of $\mathrm{O}(\frac{\mu_1}{d+\rho})$, which proves (A.1).

The proof for (16) is technically similar to the proof of (12), but is much more complicated in calculation. By the similar argument in proving (12), we know that, to show (16) is equivalent to show the following two equations.

$$
\begin{aligned}
E\Bigg[\Bigg(&\frac{S_{N_M}}{E[N_M+\sigma_M|S_{N_M}>d]}-\frac{E[S_{N_M}|S_{N_M}>d]}{E[N_M+\sigma_M|S_{N_M}>d]}\Bigg)\\
&\times\left(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1\right)^2\Bigg|S_{N_M}>d\Bigg]=\mathrm{o}\left(\frac{1}{d+\rho}\right),
\end{aligned}
\tag{A.3}
$$

and

$$
E\left[\left(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1\right)^3\Bigg|S_{N_M}>d\right]=\mathrm{o}\left(\frac{1}{\mu_1(d+\rho)}\right). \tag{A.4}
$$

To prove (A.3), we first write

$$
\begin{aligned}
&E[(N_M-E[N_M|S_{N_M}>d]+\sigma_M-E[\sigma_M|S_{N_M}>d])^3|S_{N_M}>d]\\
&\quad=E[(N_M-E[N_M|S_{N_M}>d])^3|S_{N_M}>d]\\
&\qquad+E[(\sigma_M-E[\sigma_M|S_{N_M}>d])^3|S_{N_M}>d]\\
&\qquad+3E[(N_M-E[N_M|S_{N_M}>d])^2(\sigma_M-E[\sigma_M|S_{N_M}>d])|S_{N_M}>d]\\
&\qquad+3E[(N_M-E[N_M|S_{N_M}>d])(\sigma_M-E[\sigma_M|S_{N_M}>d])^2|S_{N_M}>d].
\end{aligned}
\tag{A.5}
$$

Similar to the proof for N_0, we can show that the first term on the right-hand side of (A.5) is equal to

$$
\frac{3(d+\rho_+)}{\mu_1^5}\psi''(\theta_1)+\mathrm{O}\left(\frac{1}{\Delta_0^6}\right).
$$

For the third term on the right-hand side of (A.5), we know that

$$
\begin{aligned}
&E[(N_M-E[N_M|S_{N_M}>d])^2(\sigma_M-E[\sigma_M|S_{N_M}>d])|S_{N_M}>d]\\
&\quad=E[N_M^2\sigma_M|S_{N_M}>d]-E[N_M^2|S_{N_M}>d]E[\sigma_M|S_{N_M}>d]\\
&\qquad-2E[N_M\sigma_M|S_{N_M}>d]E[N_M|S_{N_M}>d]\\
&\qquad+2(E[N_M|S_{N_M}>d])^2E[\sigma_M|S_{N_M}>d].
\end{aligned}
\tag{A.6}
$$

In the following, we shall prove that (A.6) is at the order of $\mathrm{O}(\frac{1}{\Delta_0^6})$.

To evaluate (A.6), with previous obtained results available, we only need to approximate $E[N_M^2\sigma_M|S_{N_M}>d]$. In fact,

$$
\begin{aligned}
E[N_M^2\sigma_M, S_{N_M}>d]&=E[N_M^2\sigma_M]-E[N_M^2\sigma_M, S_{N_M}\leqslant 0]\\
&=E[N_M^2\sigma_M]+\mathrm{O}\left(\frac{1}{\Delta_0^6}\right),
\end{aligned}
\tag{A.7}
$$

and

$$
\begin{aligned}
E[N_M^2\sigma_M] &= E_{\theta_0}[\sigma_M E_{\theta_1}(N_M^2|M)] \\
&= \frac{1}{\mu_1^2}[\psi''(\theta_1)E(N_M\sigma_M) + 2\mu_1 E[N_M\sigma_M(S_{N_M} - M)] \\
&\quad - E[\sigma_M(S_{N_M} - M)^2]].
\end{aligned} \tag{A.8}
$$

From Lemma 8, we have

$$
\begin{aligned}
E(N_M\sigma_M) = \frac{d+\rho_+}{\mu_1}&\left(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)}\right) \\
&\times P_{\theta_1}(S_{N_M} > d) + \mathrm{O}\left(\frac{1}{\Delta_0^4}\right).
\end{aligned}
$$

Since

$$
\begin{aligned}
E[N_M\sigma_M(S_{N_M} - M)] &= E[N_M\sigma_M S_{N_M}] - E[N_M\sigma_M M] \\
&= (d+\rho_+)E[N_M\sigma_M, S_{N_M} > d] \\
&\quad - \frac{1}{\mu_1}[E(\sigma_M M S_{N_M} - E(\sigma_M M^2)] + \mathrm{O}\left(\frac{1}{\Delta_0^4}\right) \\
&= (d+\rho_+)E[N_M\sigma_M|S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) \\
&\quad - \frac{d+\rho_+}{\mu_1}E(\sigma_M M|S_{N_M} > d)P_{\theta_1}(S_{N_M} > d) \\
&\quad + \mathrm{O}\left(\frac{1}{\Delta_0^4}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&E[\sigma_M(S_{N_M} - M)^2] = E[\sigma_M S_{N_M}^2] - 2E[\sigma_M S_{N_M} M] + E[\sigma_M M^2] \\
&\quad = (d+\rho_+)^2 E[\sigma_M|S_{N_M} > d)P_{\theta_1}(S_{N_M} > d) - 2(d+\rho_+)E[\sigma_M M|S_{N_M} > d) \\
&\qquad \times P_{\theta_1}(S_{N_M} > d) + \mathrm{O}\left(\frac{1}{\Delta_0^4}\right).
\end{aligned}
$$

Combining above results, we have

$$
\begin{aligned}
&E[N_M^2\sigma_m|S_{N_M} > d] - E[N_M^2|S_{N_M} > d]E[\sigma_M|S_{N_M} > d] \\
&\quad = \frac{d+\rho_+}{\mu_1^2}\left(-\frac{4}{\tilde{\mu}_0\mu_1\Delta_0^2} + \frac{4}{\tilde{\mu}_0\mu_1\Delta_0\Delta_1} + \frac{4\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)}\right) \\
&\qquad - \frac{d+\rho_+}{\mu_1^2}\left(\frac{2}{\Delta_0} - \frac{2}{\Delta_1} - \frac{2\Delta_0\mu_1}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)}\right) \\
&\qquad \times \left(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)}\right) + \mathrm{O}\left(\frac{1}{\Delta_0^4}\right).
\end{aligned} \tag{A.9}
$$

On the other hand, from Lemma 8, we have

$$\begin{aligned}
&E[N_M\sigma_M|S_{N_M}>d]E[N_M|S_{N_M}>d]-(E[N_M|S_{N_M}>d])^2E[\sigma_M|S_{N_M}>d]\\
&\quad=\frac{d+\rho_+}{\mu_1^2}\left(-\frac{2}{\tilde{\mu}_0\mu_1\Delta_0^2}+\frac{2}{\tilde{\mu}_0\mu_1\Delta_0\Delta_1}+\frac{2\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1\Delta_1(\Delta_1-\Delta_0)}\right)\\
&\qquad-\frac{d+\rho_+}{\mu_1^2}\left(\frac{1}{\Delta_0}-\frac{1}{\Delta_1}-\frac{\Delta_0\mu_1}{\tilde{\mu}_1\Delta_1(\Delta_1-\Delta_0)}\right)\\
&\qquad\times\left(-\frac{1}{\tilde{\mu}_0\Delta_0}+\frac{1}{\tilde{\mu}_0\Delta_1}+\frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1-\Delta_0)}\right)+\mathrm{O}\left(\frac{1}{\Delta_0^4}\right).
\end{aligned}\tag{A.10}$$

Combining Lemma 8, (A.9) and (A.10), we have

$$E[(N_M-E[N_M|S_{N_M}>d])^2(\sigma_M-E[\sigma_M|S_{N_M}>d])|S_{N_M}>d]=\mathrm{O}\left(\frac{1}{\Delta_0^6}\right).$$

Similarly, we can show that

$$E[(N_M-E[N_M|S_{N_M}>d])(\sigma_M-E[\sigma_M|S_{N_M}>d])^2|S_{N_M}>d]=\mathrm{O}\left(\frac{1}{\Delta_0^6}\right).$$

Thus, the right-hand side of (A.5) is at the order of $\mathrm{O}(\frac{d+\rho_+}{\mu_1^5})$.

Similar to the proof of (A.2), we can prove (A.4), and complete the proof of (16).

References

Chang, J.T., 1992. Moments of the first ladder height of random walks with small drift. Ann. Appl. Probab. 2, 714–738.

Cox, D.R., 1952. A note on the sequential estimation of mean. Proc. Camb. Philos. Soc. 48, 447–450.

Ding, K., 2001. Properties of the minimum point of an unbalanced two-sided random walk. Comm. Statist. Theory Methods 30 (11), 2393–2414.

Ding, K., 2003. A lower confidence bound for the change point after a sequential CUSUM test. J. Statist. Plann. Inference 115, 311–326.

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

Hinkley, D.V., 1971. Inference about the change-point from cumulative sum tests. Biometrika 58, 509–523.

Lorden, G., 1971. Procedures for reacting to a change in distribution. Ann. Statist. 6, 1897–1908.

Moustakides, G.V., 1986. Optimal stopping times for detecting changes in distribution. Ann. Statist. 14, 1379–1387.

Page, E.S., 1954. Continuous inspection schemes. Biometrika 41, 100–114.

Pollak, M., Siegmund, D., 1985. A diffusion process and its applications to detecting a change in the drift of Brownian motion. Biometrika 72, 267–280.

Roberts, S.M., 1966. A comparison of some control chart procedures. Technometrics 8, 411–430.

Siegmund, D., 1978. Estimation following sequential test. Biometrika 65, 341–349.

Siegmund, D., 1979. Corrected diffusion approximations in certain random walk problems. Adv. Appl. Prob. 11, 701–719.

Siegmund, D., 1985. Sequential Analysis: Tests and Confidence Intervals, Springer, New York.
Srivastava, M.S., Wu, Y., 1993. Comparison of EWMA, CUSUM and Shiryayev–Roberts procedures for detecting a shift in the mean. Ann. Statist. 21, 645–670.
Srivastava, M.S., Wu, Y., 1999. Quasi-stationary biases of change point and change magnitude estimation after sequential CUSUM test. Sequential Anal. 18, 203–216.
Whitehead, J., 1986. On the bias of maximum likelihood estimation following a sequential test. Biometrika 73, 573–581.
Woodroofe, M., 1992. Estimation after sequential testing: a simple approach for a truncated sequential probability ratio test. Biometrika 79, 347–353.