Statistica Sinica Preprint No: SS-2020-0333	
Title	Optimal Sequential Tests for Monitoring Changes in the Distribution of Finite Observation Sequences
Manuscript ID	SS-2020-0333
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0333
Complete List of Authors	Dong Han, Fugee Tsung, Jinguo Xian and Miaomiao Yu
Corresponding Author	Fugee Tsung
E-mail	season@ust.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Optimal Sequential Tests for Monitoring Changes in the Distribution of Finite Observation Sequences

Dong Han, Fugee Tsung, Jinguo Xian, Miaomiao Yu

Shanghai Jiao Tong University

Hong Kong University of Science and Technology

East China Normal University

Abstract: This study proposes a method for constructing an optimal sequential test that monitors changes in the distribution of finite observation sequences with a general dependence structure. This method allows us to prove that different optimal sequential tests can be constructed for different performance measures of detection delay times. A formula is presented to calculate the value of the generalized out-of-control average run length for every optimal sequential test. Moreover, we show that there is an equivalent optimal control limit that does not depend on the test statistic directly when the post-change conditional densities (probabilities) of the observation sequences do not depend on the change time. The detection performance of six sequential tests, including two optimal sequential tests, are illustrated using numerical simulations and a real-data example.

Key words and phrases: Optimal sequential test, Change-point detection , Dependent observation sequence.

1. Introduction

One of the basic problems in statistical process control (SPC) is designing an effective

doi:10.5705/ss.202020.0333

sequential test (or a control chart), as proposed by Shewhart (1931), to detect possible changes at some instant (change-point) in the behavior of a series of sequential observations. The objective is to raise an alarm as soon as a change occurs, while keeping the rate of false alarms to an acceptable level. Detecting abrupt changes in a stochastic system quickly without exceeding a specified false alarm rate is an important issue, not only in industrial quality and process control applications, but also in nonindustrial processes (Bersimis et al. (2018)), biology (Siegmund (2013)), clinical trials and public health (Woodall (2006); Chen and Baron (2014); Rigdon and Fricker (2015)) , econometrics and financial surveillance (Frisén (2009)), and graph and network data (Akoglu et al. (2015); Woodall et al. (2017); Hosseini and Noorossana (2018)), among others.

A great variety of sequential tests have been proposed, developed, and applied to detect changes in the distribution of sequential observations quickly in various fields; see, for example, Siegmund (1985), Basseville and Nikiforov (1993), Lai (1995), Stoumbos et al. (2000), Chakraborti et al. (2001), Lai (2001), Bersimis et al. (2007), Montgomery (2009), Poor and Hadjiliadis (2009), Woodall and Montgomery (2014), Qiu (2014), and Tartakovsky et al. (2015). This raises two questions: What is the optimal sequential test? How do we design or construct an optimal sequential test?

First, we recall the main results of the known optimal sequential tests. A sequential test T^* is optimal for detecting changes in the distribution if the average value of some detection delay time $(T-k+1)^+$ of T^*, for all possible change times $k \geq 1$, is the smallest of all sequential tests T with a given probability of a false alarm that is no greater than

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

a preset level (or with a given false alarm rate that is no less than a given value), where $x^+ = \max\{0, x\}$. In the literature, there are four main kinds of optimal sequential tests: the Shiryaev test, $T_S(c_1)$ (Shiryaev (1963); Shiryaev (1978, P. 193–200)); two sum of the log likelihood ratio (SLR) tests, $T_{SLR_1}(c_2)$ (Chow et al. (1971, P. 108)) and $T_{SLR_2}(c_3)$ (Frisén (2003)); the CUSUM test, $T_C(c_4)$ (Page (1954); Moustakides (1986)); and the Shiryaev–Roberts test, $T^r_{SR}(c_5)$ (Polunchenko and Tartakovsky (2010)). Here the five positive numbers $c_i > 0$, for $1 \leq i \leq 5$, denote the five constant control limits or the threshold limits. Thus, to prove the optimality of these tests, we need to assume there is an infinite independent or Markov observation sequence (Han et al. (2017)).

In fact, it is not realistic for us to have an infinite observation sequence; that is, people can only obtain finite observation sequences in reality. For example, consider a production line that produces one product per minute. If the production line works eight hours a day, then the number of products or observations per day is $N = 480$. Our task is to design or construct an effect test that can detect whether the 480 observations (usually not independent) are abnormal in real time. However, when we have only N finite independent observation sequences $\{X_n, 1 \leq n \leq N\}$ $(N \geq 2)$, none of the five optimal sequential tests mentioned above are optimal.

Based on the work of Chow, Robbins, and Siegmund (Chow et al. (1971, Chap. 3)), we develop a method for constructing various optimal sequential tests under different performance measures of detection delay times in order to detect changes in the probability distribution of finite observation sequences. Moreover, we determine a formula to calculate

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

the value of the generalized out-of-control average run length for each optimal test, and obtain an equivalent optimal control limit that may not depend on the test statistic directly.

The rest of this paper is organized as follows. Section 2.1 presents a generalized Shiryaev measure that evaluates how well a sequential test detects changes in the distribution of finite observation sequences. Section 2.2 constructs the optimal sequential test and gives the formula for calculating the generalized out-of-control average run length. The equivalent optimal control limit is presented and proved in Section 3. The detection performance of two optimal tests is illustrated by the comparison and analysis of numerical simulations for 60 observations in Section 4. Section 5 provides a real example. The four performance measures and the proofs of the three theorems are given in Appendix and the online Supplementary Material, respectively.

2. Optimal sequential tests for finite observations

In this section, we first present the performance measure and optimization criterion, and then construct the optimal sequential tests.

Consider finite observations, $X_0, X_1, X_2, ..., X_N$. Without loss of generality, we assume $N \geq 2$. Let $\tau = k$ $(1 \leq k \leq N)$ be the change-point. Let $p_0(x_0, x_1, ..., x_N)$ and $p_k(x_0, x_1, ..., x_N)$ be the pre-change and post-change joint probability densities, respectively. Denote the post-change joint probability distribution and the expectation by $\mathbf{P}_k$ and $\mathbf{E}_k$, respectively, for $1 \leq k \leq N$. When $\tau > N$, that is, a change never occurs

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

in the N observations $X_1, X_2, ..., X_N$, the probability distribution and the expectation are denoted by $\mathbf{P}_0$ and $\mathbf{E}_0$, respectively, for all observations $X_0, X_1, X_2, ..., X_N$ with the pre-change joint probability density $p_0(x_0, x_1, ..., x_N)$. Moreover, when the observations X_n, for $0 \leq n \leq N$, are discrete random variables, the above joint probability densities and conditional probability densities are considered to be the joint probabilities and conditional probabilities, respectively.

In order to construct the optimal sequential tests in Section 2.2, we assume that the likelihood ratio of the post-change conditional probability density to the pre-change conditional probability density, $\Lambda_j^{(k)}$, satisfies

$$\Lambda_j^{(k)} = \frac{p_{\mathbf{1}j}^{(k)}(X_j|X_{j-1}, ..., X_0)}{p_{\mathbf{0}j}(X_j|X_{j-1}, ..., X_0)} < \infty \ \ (a.s.\mathbf{P}_0) \tag{2.1}$$

and has no atoms with respect to $\mathbf{P}_0$, for $1 \leq k \leq N$ and $k \leq j \leq N$, where $p_{\mathbf{0}j}(x_j|x_{j-1}, ..., x_0)$ for $1 \leq j \leq N$ and $p_{\mathbf{1}j}^{(k)}(x_j|x_{j-1}, ..., x_0)$ for $1 \leq k \leq N$, $k \leq j \leq N$ denote the pre-change and post-change conditional probability densities, respectively, and the notation (k) in $p_{\mathbf{1}j}^{(k)}$ denotes that the post-change conditional probability densities $p_{\mathbf{1}j}^{(k)}$ rely on the change-point k, for $k \leq j \leq N$. If $\Lambda_j^{(k)} = \Lambda_j$, for $1 \leq k \leq j \leq N$, then the post-change conditional densities (probabilities) of the observation sequence do not depend on the change-point.

2.1 Performance measures of sequential tests

Let $T \in \mathfrak{T}_N$ be a sequential test, where $\mathfrak{T}_N$ is a set of all sequential tests satisfying $1 \leq T \leq N+1$ and $\{T \leq n\} \in \mathfrak{F}_n = \sigma\{X_j, 0 \leq j \leq n\}$, for $1 \leq n \leq N$, where $\{T = N+1\}$ denotes the random event $\{T > N\}$, which means that the change point

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

occurs after N observation samples. Thus $\{T = N+1\} \in \mathfrak{F}_N$.

Let $W = \{w_j, 1 \leq j \leq N+1\}$ and $V = \{v_j, 1 \leq j \leq N+1\}$ be two series of nonnegative random variables satisfying $w_k, v_k \in \mathfrak{F}_{k-1}$, for $1 \leq k \leq N+1$. Denote the indicator function by $I(.)$. We may regard the two nonnegative random variables w_k and v_k as random weights of the detection delay $(T-k)^+$ and the event $I(T \geq k)$, respectively, such that the time of a false alarm is greater than or equal to the change-point k. Here, $w_k, v_k \in \mathfrak{F}_{k-1}$ means that both weights w_k and v_k can be determined using the observation information before the time k, for $1 \leq k \leq N$. Using the concept of the randomization probability of the change-point and the definition describing the average detection delay proposed by Moustakides (2008), we can define a performance measure $\mathcal{J}_{M,N}(.)$ for every given weighted pair $M = (W, V)$ to evaluate the detection performance of each sequential test $T \in \mathfrak{T}_N$, as follows:

$$\mathcal{J}_{M,N}(T) = \frac{\sum_{k=1}^{N+1} \mathbf{E}_k(w_k(T-k)^+)}{\sum_{j=1}^{N+1} \mathbf{E}_0(v_j I(T \geq j))} = \frac{\sum_{k=1}^{N} \mathbf{E}_k(w_k(T-k)^+)}{\mathbf{E}_0(\sum_{j=1}^{T} v_j)}. \tag{2.2}$$

Here, the denominator comes from $T \leq N+1$ and $\sum_{j=1}^{N+1} \mathbf{E}_0(v_j I(T \geq j)) = \mathbf{E}_0(\sum_{j=1}^{T} v_j)$. Because we only consider the detection delay after the change-point $\tau = k \geq 1$, the commonly used detection delay $(T-k+1)^+$ is replaced with $(T-k)^+$ hereafter. Note that W and V may not be the randomization probability of the change-point.

According to the definition of $\mathcal{J}_{M,N}(T)$, the smaller $\mathcal{J}_{M,N}(T)$, the better is the detection performance of the test T satisfying $\sum_{j=1}^{N+1} \mathbf{E}_0(v_j I(T \geq j)) \geq \gamma$, for some given positive constant γ.

Remark 1. The numerator and denominator of $\mathcal{J}_{M,N}(T)$ can be regarded as a gen-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

eralized out-of-control average run length (ARL_1) and a generalized in-control ARL_0, respectively. Moreover, the measure $\mathcal{J}_{M,N}(.)$ can be considered a generalization of the following Shiryaev measure:

$$\mathcal{J}_S(T) = \frac{\sum_{k=1}^{\infty} \rho_k \mathbf{E}_k((T-k)^+)}{\sum_{j=1}^{\infty} \rho_j \mathbf{E}_0(I(T \geq j))} = \mathbf{E}(T-\tau | T \geq \tau),$$

for $T \leq N+1$, where $\rho_k = \mathbf{P}(\tau = k)$, for $k \geq 1$.

It is clear that taking various weighted pairs $M = (W, V)$ yields different measures $\mathcal{J}_{M,N}(.)$. Next, we list four known measures by taking the appropriate weighted pairs, $M_i = (W_i, V_i)$, for $1 \leq i \leq 4$:

$$\begin{aligned}
\mathcal{J}_{M_1,N}(T) &= \frac{\sum_{k=1}^{N+1} \rho_k \mathbf{E}_k(T-k)^+}{\sum_{j=1}^{N+1} \rho_j \mathbf{P}_0(T \geq j)}, \quad \mathcal{J}_{M_2,N}(T) = \frac{\mathbf{E}_1(T-1)}{\mathbf{P}_0(T \geq N+1)}, \\
\mathcal{J}_{M_3,N}(T) &= \frac{\sum_{k=1}^{N} \mathbf{E}_k((1-Z_{k-1})^+(T-k)^+)}{\mathbf{E}_0(T)}, \\
\mathcal{J}_{M_4,N}(T) &= \frac{r\mathbf{E}_1(T-1) + \sum_{k=1}^{N} \mathbf{E}_k((T-k)^+)}{r + \mathbf{E}_0(T)},
\end{aligned}$$

where $W_1 = V_1 = \{\rho_k, 1 \leq k \leq N+1\}$, $\rho_{N+1} := 1 - \sum_{k=1}^{N} \rho_k$, $W_2 = \{w_1 = 1, w_k = 0,\ 2 \leq k \leq N+1\}$, $V_2 = \{v_j = 0,\ 1 \leq j \leq N,\ v_{N+1} = 1\}$, $W_3 = \{w_j = v_j = (1-Z_{j-1})^+,\ 1 \leq j \leq N+1\}$, $V_3 = \{v_k = 1, 1 \leq k \leq N+1\}$, $W_4 = V_4 = \{w_1 = v_1 = r,\ w_k = v_k = 1, 2 \leq k \leq N+1\}$, and $Z_k = \max\{1, Z_{k-1}\}\Lambda_k$, for $1 \leq k \leq N$, are the statistics of the CUSUM test with $Z_0 = 0$ (see Moustakides (1986)). A further four performance measures $\mathcal{J}_{M_j,N}(.)$, for $5 \leq j \leq 8$, are listed in Appendix, where $\mathcal{J}_{M_7,N}(.)$ and $\mathcal{J}_{M_8,N}(.)$ are new measures. Because the in-control ARL_0, $\mathbf{E}_0(T)$, is easier to calculate than the generalized in-control

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

ARL$_0$, $\mathbf{E}_0(\sum_{j=1}^{T}(1-Z_{j-1})^+)$, in $\mathcal{J}_{M_6,N}(T)$, we often use $\mathcal{J}_{M_3,N}(T)$ instead of the measure $\mathcal{J}_{M_6,N}(T)$.

Note that for an infinite independent observation sequence, the five measures $\mathcal{J}_{M_i,\infty}$, for $i = 1, 2, 5, 6, 4$ and $N = \infty$, have been used by Shiryaev (1978, P. 193–200), Chow et al. (1971, P. 108), Frisén (2003), Moustakides (1986), and Polunchenko and Tartakovsky (2010) to prove the optimality of the sequential tests, T_S, T_{SLR_1}, T_{SLR_2}, T_C, and T_{SR}^r, respectively.

2.2 Optimal sequential tests

For a given weighted pair $M = (W, V)$, we first define the optimization criterion of the sequential tests for N observations.

Definition 1. *A sequential test $T^* \in \mathfrak{T}_N$, with $\boldsymbol{E}_0(\sum_{k=1}^{T^*} v_k) \geq \gamma$, is optimal under the measure $\mathcal{J}_{M,N}(T)$ if*

$$\inf_{T\in\mathfrak{T}_N,\ \boldsymbol{E}_0(\sum_{j=1}^{T} v_j)\geq\gamma} \mathcal{J}_{M,N}(T) = \mathcal{J}_{M,N}(T^*), \tag{2.3}$$

where γ satisfies $\boldsymbol{E}_0(v_1) < \gamma < \boldsymbol{E}_0(\sum_{j=1}^{N+1} v_j)$.

To construct the optimal sequential test under the measure $\mathcal{J}_{M,N}(T)$ in (2.2) with a given weighted pair $M = (W, V)$, we need to present a series of nonnegative test statistics, Y_n, for $0 \leq n \leq N + 1$, as follows:

$$Y_n = \sum_{k=1}^{n} w_k \prod_{j=k}^{n} \Lambda_j^{(k)} \tag{2.4}$$

for $0 \leq n \leq N + 1$, where $Y_0 = 0$, $Y_{N+1} := Y_N$, $W = \{w_k, 1 \leq k \leq N + 1\}$, and $\Lambda_j^{(k)}$ satisfies (2.1). It can be seen that the statistics Y_n, for $1 \leq n \leq N$, depend not only on the likelihood ratio $\{\Lambda_j^{(k)}\}$, but also on the weight of the detection delay $\{w_k\}$. In particular, if $\Lambda_j^{(k)} = \Lambda_j$, for $1 \le k \le j \le N$, that is, the post-change conditional densities (probabilities) of the observation sequences do not depend on the change-point, then

$$Y_n = \sum_{k=1}^{n} w_k \prod_{j=k}^{n} \Lambda_j = (Y_{n-1} + w_n)\Lambda_n, \tag{2.5}$$

for $1 \le n \le N$.

Remark 2. Even if (2.5) holds, the test statistic sequence $\{Y_n, 0 \le n \le N\}$ is not necessarily a Markov chain. For example, let both the pre-change observation sequence $X_1, ..., X_{k-1}$ and the post-change observation sequence $X_k, ...X_N$, be independent and identically distributed (i.i.d.). Therefore, (2.5) holds, and it is clear that the statistic $\{Y_n, 0 \le n \le N\}$ is not a Markov chain when we take $w_1 = 1, w_n = \frac{1}{n-1}\sum_{j=1}^{n-1} e^{X_j}$, for $2 \le n \le N$, in (2.5).

Note that $(T-k)^+ = \sum_{m=k+1}^{N+1} I(T \ge m)$, for $T \in \mathfrak{T}_N$ and $I(T \ge m) \in \mathfrak{F}_{m-1}$, and the post-change joint probability density $p_k(x_0, x_1, ..., x_n)$ for the change-point k $(1 \le k \le N)$ can be written as

$$p_k(x_0, x_1, ..., x_n) = p(x_0) \prod_{j=1}^{(k-1)\wedge n} p_{\mathbf{0}j}(x_j|x_{j-1}, ..., x_0) \prod_{j=k}^{n} p_{\mathbf{1}j}^{(k)}(x_j|x_{j-1}, ..., x_0),$$

for $1 \le n \le N$, where $p(x_0)$ is the probability density (or probability) of X_0 at the initial time $k = 0$, $(k-1) \wedge n$ denotes $\min\{k-1, n\}$, $\prod_{j=1}^{(k-1)\wedge n} = 1$ for $k = 1$, and $\prod_{j=k}^{n} = 1$ for

$n < k$. For the nonnegative test statistic Y_n in (2.4), we have that

$$\begin{aligned} & \sum_{k=1}^{N} \mathbf{E}_k(w_k(T-k)^+) \\ = & \mathbf{E}_0\Big(\sum_{k=1}^{N}\sum_{m=k+1}^{N+1} w_k I(T \geq m) \prod_{j=k}^{m-1} \Lambda_j^{(k)}\Big) \\ = & \mathbf{E}_0\Big(\sum_{m=1}^{N} Y_m I(T \geq m+1)\Big) = \mathbf{E}_0\Big(\sum_{m=1}^{T} Y_{m-1}\Big), \end{aligned}$$

for all $T \in \mathfrak{T}_N$. This equality means that the generalized out-of-control ARL_1 (the numerator of the measure $\mathcal{J}_{M,N}(T)$) is equal to the generalized in-control ARL_0, in which the weight $\{v_m\}$ is replaced by the statistic $\{Y_{m-1}\}$. Thus, finding an optimal sequential test T^* under the measure $\mathcal{J}_{M,N}(T)$ in (2.2) is equivalent to constructing an optimal sequential test T^* that satisfies the following equation:

$$\inf_{T\in\mathfrak{T}_N,\ \mathbf{E}_0(\sum_{j=1}^{T} v_j)\geq\gamma} \{\mathbf{E}_0\Big(\sum_{m=1}^{T} Y_{m-1}\Big)\} = \mathbf{E}_0\Big(\sum_{m=1}^{T^*} Y_{m-1}\Big), \tag{2.6}$$

for $\mathbf{E}_0(\sum_{j=1}^{T^*} v_j) = \gamma$, where γ satisfies $\mathbf{E}_0(v_1) < \gamma < \mathbf{E}_0(\sum_{j=1}^{N+1} v_j)$.

Motivated by the Chow–Robbins–Siegmund method of backward induction (Chow et al. (1971, P. 49)), we present a nonnegative random dynamic control limit $\{l_n(c)$, for $0 \leq n \leq N+1\}$, that is defined by the following recursive equations:

$$\begin{aligned} l_{N+1}(c) &= 0,\ \ l_N(c) = cv_{N+1} \\ l_n(c) &= cv_{n+1} + \mathbf{E}_0\Big([l_{n+1}(c) - Y_{n+1}]^+|\mathfrak{F}_n\Big), \end{aligned} \tag{2.7}$$

for $0 \leq n \leq N-1$, where $c > 0$ is a constant and $V = \{v_j, 1 \leq j \leq N+1\}$. It is clear that $l_n(c) \geq cv_{n+1}$ and $l_n(c) \in \mathfrak{F}_n$, for $0 \leq n \leq N$. The positive number c can be

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

regarded as an adjustment coefficient for the random dynamic control limit, because $l_n(c)$ is increasing on $c \geq 0$ with $l_n(0) = 0$ and $\lim_{c\to\infty} l_n(c) = \infty$ for $v_{n+1} > 0$.

Now, for a given weighted pair $M = (W, V)$, we define a sequential test $T^*_M(c, N)$ using the test statistic Y_n, for $1 \leq n \leq N+1$, and the control limits $l_n(c)$, for $1 \leq n \leq N+1$, as follows:

$$T^*_M(c, N) = \min\{1 \leq n \leq N+1 : Y_n \geq l_n(c)\}. \tag{2.8}$$

It is easy to check that $T^*_M(c, N) \in \mathfrak{T}_N$.

The following theorem shows that for any given performance measure $\mathcal{J}_{M,N}$ in (2.2), the sequential test $T^*_M(c, N)$ constructed above is optimal.

Theorem 1. *Assume that the ratio* $\Lambda^{(k)}_j$ *satisfies* (2.1), *for* $1 \leq k \leq N$ *and* $k \leq j \leq N$. *Let* γ *be a positive number satisfying* $\boldsymbol{E}_0(v_1) < \gamma < \sum_{j=1}^{N+1} \boldsymbol{E}_0(v_j)$. *Then:*

(i) There exists a positive number c_γ *such that* $T^*_M(c_\gamma, N)$ *is optimal in the sense of* (2.2) *(or* (2.6)*) with* $\boldsymbol{E}_0(\sum_{j=1}^{T^*_M(c_\gamma,N)} v_j) = \gamma$*; that is,*

$$\inf_{T\in\mathfrak{T}_N,\, \boldsymbol{E}_0(\sum_{j=1}^{T} v_j)\geq\gamma} \mathcal{J}_{M,N}(T) = \mathcal{J}_{M,N}(T^*_M(c_\gamma, N)). \tag{2.9}$$

(ii) If $T \in \mathfrak{T}_N$ *satisfies* $T \neq T^*_M(c_\gamma, N)$*, that is,* $\boldsymbol{P}_0(T \neq T^*_M(c_\gamma, N)) > 0$ *and* $\boldsymbol{E}_0(\sum_{j=1}^{T} v_j) = \gamma$*, then*

$$\mathcal{J}_{M,N}(T) > \mathcal{J}_{M,N}(T^*_M(c_\gamma, N)). \tag{2.10}$$

(iii) Moreover,

$$\mathcal{J}_{M,N}(T^*_M(c_\gamma, N)) = c_\gamma\Big(1 - \frac{\boldsymbol{E}_0(v_1)}{\gamma}\Big) - \frac{\boldsymbol{E}_0[l_1(c_\gamma) - Y_1]^+}{\gamma}. \tag{2.11}$$

Here, the random dynamic control limit $\{l_n(c), 0 \le n \le N+1\}$ of the optimal test $T^*_M(c, N)$ can be called an optimal dynamic control limit.

It follows from (2.9) and (2.11) that the minimum value of the generalized out-of-control ARL_1 (the numerator of the measure $\mathcal{J}_{M,N}(T)$), for all $T \in \mathfrak{T}_N$, can be calculated using the following formula:

$$\begin{aligned}
& \inf_{T\in\mathfrak{T}_N,\, \mathbf{E}_0(\sum_{j=1}^{T} v_j)\ge\gamma} \sum_{k=1}^{N} \mathbf{E}_k(w_k[T-k]^+) \\
= & \sum_{k=1}^{N} \mathbf{E}_k(w_k[T^*_M(c_\gamma, N)-k]^+) \qquad (2.12)\\
= & \; c_\gamma(\gamma - \mathbf{E}_0(v_1)) - \mathbf{E}_0([l_1(c_\gamma) - Y_1]^+).
\end{aligned}$$

As an application of Theorem 1, we have the following corollary.

Corollary 1. *The eight sequential tests $T^*_{M_i}(c, N)$, for $1 \le i \le 8$, defined in (2.8), which correspond to the eight weighted pairs M_i, for $1 \le i \le 8$, are optimal under the measures $\mathcal{J}_{M_i,N}$, for $1 \le i \le 8$, respectively.*

Note that the optimality of the two tests $T^*_{M_4}(c, N)$ and $T^*_{M_6}(c, N)$ with the optimal dynamic control limits $\{l^{(4)}_n(c), 0 \le n \le N+1\}$ and $\{l^{(6)}_n(c), 0 \le n \le N+1\}$, respectively, is not under Lorden's measure (see Lorden (1971); Moustakides (1986)), but under the corresponding measures $\mathcal{J}_{M_4,N}$ and $\mathcal{J}_{M_6,N}$, respectively.

3. Optimal control limits

It is clear that the optimal control limit $\{l_n(c), 0 \le n \le N+1\}$ of the optimal sequential test $T^*_M(c, N)$ plays a key role in detecting changes in a distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Because $\mathbf{E}_0([l_{n+1}(c)-Y_{n+1}]^+|\mathfrak{F}_n)$ and v_{n+1} are measurable with respect to $\mathfrak{F}_n$, it follows that there are $2N+1$ nonnegative functions $h_n = h_n(c, x_0, x_1, ..., x_n)$, for $0 \le n \le N-1$, and $v_n = v_n(x_0, x_1, ..., x_{n-1})$, for $0 \le n \le N-1$, such that

$$\begin{aligned} h_n &= h_n(c, x_0, x_1, ..., x_n) \\ &= \mathbf{E}_0([l_{n+1}(c) - Y_{n+1}]^+|X_n = x_n, X_{n-1} = x_{n-1}, ..., X_0 = x_0), \end{aligned}$$

for $0 \le n \le N-1$. Therefore, the optimal control limit $l_n(c)$ in (2.7) can be written as

$$l_n(c) = cv_{n+1}(x_0, X_1, ..., X_n) + h_n(c, x_0, X_1, ..., X_n),$$

for $0 \le n \le N$, where $X_0 = x_0$ is a constant. It can be seen that the optimal control limit $\{l_n(c), 0 \le n \le N+1\}$ of the optimal sequential test $T^*_M(c, N)$ is not easy to calculate for a general dependence observation sequence $\{X_n,\, 0 \le n \le N\}$.

To reduce the number of observation variables on which the control limit $\{l_n(c),\, 0 \le n \le N\}$ depends, we let the observation sequence $\{X_n,\, 0 \le n \le N\}$ be at most a q-order Markov process, where $q = \max\{i,\, j\}$, $0 \le q \le N$; that is, the pre-change observations $X_1, ..., X_{k-1}$ and the post-change observations $X_k, ..., X_N$ are i-order and j-order Markov processes, respectively, with transition probability density functions $p_{\mathbf{0}n}(x_n|x_{n-1}, ..., x_{n-i})$ and $p^{(k)}_{\mathbf{1}m}(x_m|x_{m-1}, ..., x_{m-j})$, respectively, that satisfy the following Markov property:

$$\begin{aligned} p_{\mathbf{0}n}(x_n|x_{n-1}, ..., x_{n-i}) &= p_{\mathbf{0}n}(x_n|x_{n-1}, ..., x_{n-i}, ..., x_0) \\ p_{\mathbf{1}m}(x_m|x_{m-1}, ..., x_{m-j}) &= p_{\mathbf{1}m}(x_m|x_{m-1}, ..., x_{m-j}, ..., x_0) \\ &= p^{(k)}_{\mathbf{1}m}(x_m|x_{m-1}, ..., x_{m-j}, ..., x_0), \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

for $n \geq i$, $m \geq j$, and $1 \leq k \leq m \leq N$. The first equality of conditional probability denotes that the current situation x_n depends only on what happened in the previous i periods for pre-change observations; this is an i-order Markov process. Obviously, the second quality means that the post-change sequence of observations is a j-order Markov process. The last equation means that the post-change conditional densities of the observation sequences do not depend on the change-point. Here, a zero-order Markov process means that the pre-change observations $X_1, ..., X_{k-1}$ and the post-change observations $X_k, ..., X_N$ are mutually independent. When $q = N$, we consider that at least one of the pre-change observations $X_1, ..., X_{k-1}$ and the post-change observations $X_k, ..., X_N$ is not a Markov process of any order, because we have only N observations. In this case, the test statistic sequence $\{Y_0, Y_1, ..., Y_N\}$ is not a Markov process of any order.

Theorem 2 shows that the optimal control limit $l_n(c)(0 \leq n \leq N)$ depends on Y_n and q observation variables if the observation sequence $\{X_n, 0 \leq n \leq N\}$ is at most a q-order Markov process.

Theorem 2. *Let the observation sequence be at most a q-order Markov chain for $0 \leq q \leq N$. Let $A_{n,q} := \{X_n, ..., X_{n-q+1}\}$ and $B_{n,0} := \{X_n, ..., X_0\}$. Assume that the post-change conditional densities of the observation sequences do not depend on the change-point, and that the weighted pair $M = (W, V)$ satisfies $w_{n+1} = w_{n+1}(Y_n, A_{n,q_1})$ and $v_{n+1} = v_{n+1}(Y_n, A_{n,q_2})$, for $0 \leq n \leq N$, where $0 \leq q_1, q_2 \leq q$, $w_{n+1} = w_{n+1}(Y_n)$ for $q_1 = 0$, and $v_{n+1} = v_{n+1}(Y_n)$ for $q_2 = 0$. Then:*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

(i) For $1 \le q \le N$, the optimal control limit $\{l_n(c), 0 \le n \le N\}$ can be written as

$$\begin{aligned} l_n(c) &= cv_{n+1}(Y_n, A_{n,q_2}) \\ &+ \boldsymbol{E}_0\Big([l_{n+1}(c) - (Y_n + w_{n+1}(Y_n, A_{n,q_1}))\Lambda_{n+1}]^+ | Y_n, B_{n,0}\Big) \end{aligned}$$

for $0 \le n \le q-1$, and as

$$\begin{aligned} l_n(c) &= cv_{n+1}(Y_n, A_{n,q_2}) \\ &+ \boldsymbol{E}_0\Big([l_{n+1}(c) - (Y_n + w_{n+1}(Y_n, A_{n,q_1}))\Lambda_{n+1}]^+ | Y_n, A_{n,q}\Big) \end{aligned}$$

for $q \le n \le N$, where we replace X_{n-q_1+1} or X_{n-q_2+1} with X_0, as long as $n - q_1 + 1 < 0$ or $n - q_2 + 1 < 0$, respectively.

(ii) For $q = 0$, we have

$$l_n(c) = cv_{n+1}(Y_n) + \boldsymbol{E}_0\Big([l_{n+1}(c, Y_{n+1}) - (Y_n + w_{n+1}(Y_n))\Lambda_{n+1}]^+ | Y_n\Big),$$

for $0 \le n \le N$.

Note that the optimal control limit $l_n(c)$ depends not only on $A_{n,q}$, but also on the test statistic Y_n, for $1 \le n \le N$. Can we find a control limit $\widetilde{l_n}(c)$ that has the same property as $l_n(c)$, but that does not directly depend on the test statistic Y_n, for $1 \le n \le N$? To answer this question, we first define an equivalent control limit.

Definition 2. *Let the observation sequence $\{\widetilde{l_n}(c), 1 \le n \le N\}$ be a control limit of a sequential test $\widetilde{T} \in \mathfrak{T}_N$, where $\widetilde{T} = \min\{1 \le n \le N+1 : Y_n \ge \widetilde{l_n}(c)\}$. If $\widetilde{T}$ is equal to the optimal sequential test $T_M^*(c, N)$ (a.s. $\boldsymbol{P}_0$), then we call the control limit $\{\widetilde{l_n}(c)\}$ an equivalent control limit of the optimal sequential test $T_M^*(c, N)$.*

The following theorem answers the above question.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Theorem 3. *Let the observation sequence and the weighted pair $M = (W, V)$ satisfy the conditions of Theorem 2. Let $a_{n,q} := \{x_n, ..., x_{n-q+1}\}$ and $b_{n,0} := \{x_n, ..., x_0\}$. Assuming that $q_1 = q_2 = q$, $y + w_{n+1}(y, a_{n,q})$ and $v_{n+1}(y, a_{n,q})$ are continuous nondecreasing and non-increasing on $y \geq 0$, respectively, for given $a_{n,q}$, $0 \leq n \leq N$. Then,*

(i) For $1 \leq q \leq N$, there is an equivalent control limit $\widetilde{l_n}(c)$ of the optimal sequential test $T^*_M(c, N)$ that does not depend directly on the statistic Y_n, for $1 \leq n \leq N$, such that $\widetilde{l_n}(c) = y_n(c, B_{n,0})$ for $0 \leq n \leq q-1$, and $\widetilde{l_n}(c) = y_n(c, A_{n,q})$ for $q \leq n \leq N$, where the nonnegative functions $y_n = y_n(c, b_{n,0})$ for $0 \leq n \leq p-1$ and $y_n = y_n(c, a_{n,q})$ for $q \leq n \leq N$ satisfy

$$
\begin{aligned}
y_n &= cv_{n+1}(y_n, b_{n,0}) \\
&+ \boldsymbol{E}_0\Big([l_{n+1}(c) - (y_n + w_{n+1}(y_n, b_{n,0}))\Lambda_{n+1}]^+ | Y_n = y_n, B_{n,0} = b_{n,0}\Big),
\end{aligned}
$$

for $0 \leq n \leq q-1$, and

$$
\begin{aligned}
y_n &= cv_{n+1}(y_n, a_{n,q}) \\
&+ \boldsymbol{E}_0\Big([l_{n+1}(c) - (y_n + w_{n+1}(y_n, a_{n,q}))\Lambda_{n+1}]^+ | Y_n = y_n, A_{n,q} = a_{n,q}\Big),
\end{aligned}
$$

for $q \leq n \leq N$.

(ii) Let $q = 0$. There is a series of nonnegative nonrandom numbers, y_n, for $1 \leq n \leq N$, such that the equivalent control limit $\widetilde{l_n}(c) = y_n$ and y_n satisfies

$$
y_n = cv_{n+1}(y_n) + \boldsymbol{E}_0\Big([l_{n+1}(c) - (y_n + w_{n+1}(y_n))\Lambda_{n+1}]^+ | Y_n = y_n\Big), \tag{3.1}
$$

for $1 \leq n \leq N$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Remark 3. Using a similar method to the proof of Theorem 3, we can prove that the results of Theorem 3 are still true for $0 \leq q_1, q_2 \leq q \leq N$.

It is clear that the weighted pairs M_i, for $1 \leq i \leq 6$, satisfy the conditions of Theorem 3. As an application of Theorem 3, we have the following corollary.

Corollary 2. *Let the observation sequence be at most a q-order Markov chain for $0 \leq q \leq N$, and let the post-change conditional densities of the observation sequences not depend on the change-point. Then, the six optimal sequential tests $T^*_{M_i}(c, N)m$ for $1 \leq i \leq 6$, have equivalent control limits. In particular, when $q = 0$, the equivalent control limits consist of a series of dynamic nonrandom numbers.*

None of the equivalent control limits of the optimal sequential tests $T^*_{M_i}(c, N)$, for $1 \leq i \leq 5$, are constants when $q = 0$. This means that $T^*_{M_1}(c, N) \neq T_{S,N}(c_1)$, $T^*_{M_2}(c, N) \neq T_{SLR_1,N}(c_2)$, $T^*_{M_3}(c, N) \neq T_{C,N}(c_4)$, $T^*_{M_4}(c, N) \neq T^r_{SR,N}(c_5)$, and $T^*_{M_5}(c, N) \neq T_{SLR_2,N}(c_3)$, because the control limits c_i, for $1 \leq i \leq 5$, are constants, where $T_{S,N}(c_1) = \min\{T_S(c_1), N+1\}$, $T_{SLR_1,N}(c_2) = \min\{T_{SLR_1}(c_2), N+1\}$, $T_{SLR_2,N}(c_3) = \min\{T_{SLR_2}(c_3), N+1\}$, $T_{C,N}(c_4) = \min\{T_C(c_4), N+1\}$, and $T^r_{SR,N}(c_5) = \min\{T^r_{SR}(c_5), N+1\}$.

Thus, from (ii) of Theorem 1, we obtain the following corollary.

Corollary 3. *The optimal sequential tests $T^*_{M_i}(c, N)$, for $1 \leq i \leq 5$, are strictly superior to the tests $T_{S,N}(c_1)$, $T_{SLR_1,N}(c_2)$, $T_{C,N}(c_4)$, $T^r_{SR,N}(c_5)$, and $T_{SLR_2,N}(c_3)$ under the measures $\mathcal{J}_{M_i,N}$, for $1 \leq i \leq 5$, respectively, when they all have the same (generalized) in-control ARL_0.*

Remark 4. Sections 4.1 and 4.2 illustrate that the CUSUM and Shiryaev–Roberts

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

tests with appropriate dynamic control limits can be superior to the CUSUM test $T_{C,N}$ under Lorden's measure and the Shiryaev–Roberts test $T^r_{SR,N}$ under Pollak's measure (see Pollak (1985)), respectively, for finite independent observations. Thus, the reason why the optimal sequential tests $T_S(c_1)$, $T_{SLR_1}(c_2)$, $T_{SLR_2}(c_3)$, $T_C(c_4)$, and $T^r_{SR}(c_5)$ for a sequence of infinite independent observations are no longer optimal for finite independent observation sequences is that their control limits c_k, for $1 \leq k \leq 5$, are all constants.

Next, in the following example, we illustrate how to find an equivalent control limit by analyzing the optimal control limit of the optimal sequential test $T^*_{M_2}(c, N)$. In particular, for some special pre-change and post-changes probability densities, we can get the closed-form optimal control limit.

Example Let $\{X_0, 1 \leq k \leq 60\}$ be an i.i.d observation sequence with the pre-change probability density $p_{\mathbf{0}}$ and the post-change probability density $p_{\mathbf{1}}$; that is, both the pre-change observations $X_1, ..., X_{k-1}$ and the post-change observations $X_k, ..., X_N$ are i.i.d with the probability densities $p_{\mathbf{0}}$ and $p_{\mathbf{1}}$, respectively. Take $W_2 = \{w_1 = 1,\, w_k = 0,\, 2 \leq k \leq N+1\}$ and $V_2 = \{v_k = 0,\, 1 \leq k \leq N,\, v_{N+1} = 1\}$. We know that $\{Y_n = \prod_{j=1}^n \Lambda_j,\, 1 \leq n \leq N\}$ is a Markov process and $\Lambda_n = p_{\mathbf{1}}(X_n)/p_{\mathbf{0}}(X_n)$ and Y_n are mutually independent with Λ_{n+1}, for $0 \leq n \leq N-1$. Because $q = 0$, it follows from (2.7) and (ii) of Theorem 3 that the optimal control limit $\{l_n(c, y) : 1 \leq n \leq N\}$ of $T^*_{M_2}(c, N)$ can be written as

$$
\begin{aligned}
l_{N+1}(c, y) &= 0,\;\; l_N(c, y) = c > 0 \\
l_n(c, y) &= \mathbf{E}_0\Big([l_{n+1}(c, Y_{n+1}) - Y_{n+1}]^+ | Y_n = y\Big) \\
&= \mathbf{E}_0\Big([l_{n+1}(c, y\Lambda_{n+1}) - y\Lambda_{n+1}]^+\Big), \qquad (3.2)
\end{aligned}
$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

for $1 \leq n \leq N-1$, where $l_N(c, Y_N) = c$. It is clear that the function $l_{N-1}(c, y)$ is strictly monotonically decreasing on $y \geq 0$. Hence, $l_n(c, y)$ is also strictly monotonically decreasing on $y \geq 0$, for $1 \leq n \leq N-2$. This means that for each n ($1 \leq n \leq N-1$), there is a unique positive number y_n such that $y_n = l_n(c, y_n)$, for $c > 0$. Thus, $Y_n \geq y_n$ if and only if $Y_n \geq l_n(c, Y_n)$, for $1 \leq n \leq N-1$. In other words, the equivalent control limits $\{\widetilde{l_n}(c),\, 1 \leq n \leq N\}$ of the optimal sequential test $T^*_{M_2}(c, N)$ are a series of positive numbers $\{y_n,\, 1 \leq n \leq N\}$; that is, $\widetilde{l_n}(c) = y_n$, where $y_N = c > 0$ and y_n satisfies $\widetilde{l_n}(c) = y_n = l_n(c, y_n)$, for $1 \leq n \leq N-1$.

Now, we consider the power law distributions that can occur in a diverse range of phenomena. Let $p_{\mathbf{0}}(x) = \alpha/x^{1+\alpha}$ and $p_{\mathbf{1}}(x) = \beta/x^{1+\beta}$, for $x \geq 1$, be the pre-change and post-change probability densities, respectively. Therefore, the likelihood ratio satisfies $\Lambda_n = 1/ax^r$, where $\beta > \alpha > 0$, $r = \beta - \alpha$, and $a = \alpha/\beta$. Let $a \geq (N-1)/N$. Solving the recursive equations in (3.2) above, we obtain the optimal control limit $\{l_n(c, y) : 1 \leq n \leq N\}$ of $T^*_{M_2}(c, N)$, which has the closed form $l_N(c, y) = c > 0$ and

$$l_n(c, y) = \begin{cases} c - (N-n)y & \text{if } y \leq ac/(N-n) \\ c(1-a)\left(\frac{ac}{(N-n)y}\right)^{\beta/r} & \text{if } y > ac/(N-n), \end{cases}$$

for $1 \leq n \leq N-1$. Thus, the equivalent optimal control limits $\{\widetilde{l_n}(c),\, 1 \leq n \leq N\}$ can be written as $\widetilde{l_n}(c) = c_n = c/(N-n+1)$, for $1 \leq n \leq N$, where $\{c_n\}$ satisfies $c_n = l_n(c, c_n)$ and $c_n \leq ac/(N-n)$, for $1 \leq n \leq N$.

4. Comparison and analysis of simulation results

Consider an observation sequence with $N = 60$. Let the change time τ be unknown. By comparing the simulation results in Sections 4.1 and 4.2, we illustrate that the CUSUM test T_C and the Shiryaev–Roberts test T_{SR}^r with a specially designed deterministic initial point r for an exponential model are no longer optimal under Lorden's and Pollak's measures, respectively, for 60 finite independent observations. The detection performance (the generalized out-of-control ARL_1) of six sequential tests, $T^*_{M_3}(c, 60)$, $T^*_{M_4}(c, 60)$, T_C, T_E, $T_C^{-1/60}$, and $T_C^{1/60}$, for independent and dependent observation sequences, are compared in Sections 4.3 and 4.4, respectively, where T_E denotes the exponentially weighted moving average (EWMA) test introduced by Roberts (1959), which, like the CUSUM test T_C, is very popular in statistical process control (see Han and Tsung (2004); Saleh et al. (2015); Hosseini and Noorossana (2018)). Both $T_C^{-1/60}$ and $T_C^{1/60}$ are defined by replacing the constant control limit of the CUSUM test T_C with two straight lines, $c_k^- = c(1-k/60)$ and $c_k^+ = c(1 + k/60)$, respectively, for $1 \le k \le 60$. All numerical simulation results in this section were obtained using 10^5 repetitions.

4.1 Comparison of simulation values of $\mathcal{J}_L(\min\{T, N + 1\})$

Let $\{X_k, 1 \le k \le 60\}$ be an i.i.d observation sequence with a pre-change normal distribution of $N(0, 1)$ and a post-change normal distribution of $N(0.2, 1)$. That is, the likelihood ratio Λ_k of the pre-change and post-change probability densities $p_0(x)$ and $p_1(x)$, respectively, can be written as $\Lambda_k = e^{0.2(X_k-0.1)}$, for $1 \le k \le 60$. We compare the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

performance of the two CUSUM tests $T_C(c, 60)$ and T_{DC} in detecting the mean shift from $\mu_0 = 0$ to $\mu_1 = 0.2$ under Lorden's measure $J_L(\min\{T, N+1\})$ with ARL$_0$=40, where $T_C(c_4, 60) = \min\{T_C(c_4), 61\}$ and

$$T_{DC} = \min\{1 \le k \le N+1 : \; Z_k \ge l_k\},$$

with the dynamic control limits

$$l_k = \begin{cases} 2.53 & \text{if } 1 \le k \le 40 \\ 2.53 + 0.506 * (k - 40) & \text{if } 40 < k \le 60, \end{cases}$$

and $l_{61} = 0$, where $Z_{61} := Y_{60}$, Z_k, for $0 \le k \le 60$, are the CUSUM test statistics; that is, $Z_0 = 0$ and $Z_k = \max\{1, Z_{k-1}\}\Lambda_k$, for $1 \le k \le 60$. It can be calculated that $\mathbf{E}_0(T_{DC}) = 40.02$.

Taking the constant control limit $c_4 = 2.6601$, we have $\mathbf{E}_0(T_C(c_4, 60)) = 40.01$. Note that

$$essup\{\mathbf{E}_k((T_C(c_4, 60) - k)^+ | \mathfrak{F}_{k-1})\} = \mathbf{E}_k((T_C(c_4, 60) - k)^+ | Z_{k-1} \le 1),$$

for $1 \le k \le 60$. Both the simulation values of the detection delay $\mathbf{E}_k((T_C(c_4, 60) - k)^+ | Z_{k-1} \le 1)$ and $\mathbf{E}_k((T_{DC} - k)^+ | Z_{k-1} \le 1)$ are decreasing for $k = 1, 2, ..., 60$; that is, both can arrive at the maximum values at the change-point $k = 1$. Because both $\mathbf{E}_1(T_{DC} - 1) = 22.951$ and $\mathbf{E}_1(T_C(c, 60) - 1) = 23.425$ are the maximum values, it follows that

$$\begin{aligned} & \mathcal{J}_L(T_{DC}) = \max_{1 \le k \le 60} \{\mathbf{E}_k((T_{DC} - k)^+ | Z_{k-1} \le 1)\} = \mathbf{E}_1(T_{DC} - 1) \\ < \; & \mathcal{J}_L(T_C(c, 60)) = \max_{1 \le k \le 60} \{\mathbf{E}_k((T_C(c, 60) - k)^+ | Z_{k-1} \le 1)\} = \mathbf{E}_1(T_C(c, 60) - 1). \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

This means that the CUSUM chart T_C is not optimal under Lorden's measure $\mathcal{J}_L(\min\{T, N+1\})$ restricted in 60 i.i.d. observations.

4.2 Comparison of simulation values of $\mathcal{J}_P(\min\{T, N+1\})$

Let $\{X_k, 1 \le k \le 60\}$ be an i.i.d observation sequence with a pre-change exponential density of $f_0(x) = e^{-x} I(x \ge 0)$ and a post-change exponential density of $f_1(x) = 2e^{-2x} I(x \ge 0)$. The likelihood ratio is $\Lambda_k = 2e^{-X_k}$, for $1 \le k \le 60$. Polunchenko and Tartakovsky (2010) proved that the control chart $T^r_{SR}(c)$ with a specially designed deterministic initial point r for an exponential model is optimal under Pollak's measure $\mathcal{J}_P(T)$, for $1 < \gamma < 2.2188$. Let $T^r_{SR}(c_5, 60) = \min\{T^r_{SR}(c_5), 61\}$. Taking $c_5 = 1.6645$ and $r = \sqrt{2.6645} - 1$, we have $\mathrm{ARL}_0 = \mathbf{E}_0(T^r_{SR}(c_5, 60)) = 2$. It follows from $\mathcal{J}_P(\min\{T, N+1\}) = \max_{1 \le k \le 60}\{\mathbf{E}_k(T - k)^+/\mathbf{P}_0(T \ge k)\}$ that

$$\mathcal{J}_P(T^r_{SR}(c_5, 60)) = \mathbf{E}_1(T^r_{SR}(c_5, 60) - 1) = 1.3165.$$

However, if we define a sequential test as $T^r_{SR}(\{l_k\}, 60)$ with dynamic control limit l_k,

$$l_k = \begin{cases} 1.238 + 0.1238k & \text{if } 1 \le k \le 10 \\ 0 & \text{if } 10 < k \le 60, \end{cases}$$

we obtain

$$\mathcal{J}_P(T^r_{SR}(\{l_k\}, 60)) = \mathbf{E}_1(T^r_{SR}(\{l_k\}, 60) - 1) = 1.2743,$$

with $\mathrm{ARL}_0 = \mathbf{E}_0(T^r_{SR}(\{l_k\}, 60)) = 2.0012$. Thus,

$$\mathcal{J}_P(T^r_{SR}(\{l_k\}, 60)) < \mathcal{J}_P(T^r_{SR}(c_5, 60)).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

This means that the control chart $T_{SR}^{r}(c_5)$ is not optimal under Pollak's measure $\mathcal{J}_P(\min\{T, N+1\})$ restricted in 60 i.i.d. observations.

4.3 Comparison of the generalized out-of-control ARL_1 for independent observations

Let $\{X_k, 1 \le k \le 60\}$ be an i.i.d. observation sequence with a pre-change normal distribution of $N(0,1)$ and a post-change normal distribution of $N(1,1)$. The likelihood ratio is $\Lambda_k = e^{X_k - 1/2}$, for $1 \le k \le 60$. Let $T_3^* = T_{M_3}^*(c, 60)$, and $T_4^* = T_{M_4}^*(c, 60)$, and let the smoothing parameter in the statistics of the EWMA test T_E be 0.1. By Corollary 2, we know that the equivalent control limits of the optimal sequential tests T_3^* and T_4^* consist of a series of nonrandom positive numbers. Fig. 1 shows the constant control limit of T_C (black dots) and the equivalent dynamic control limit of T_3^* (white dots).

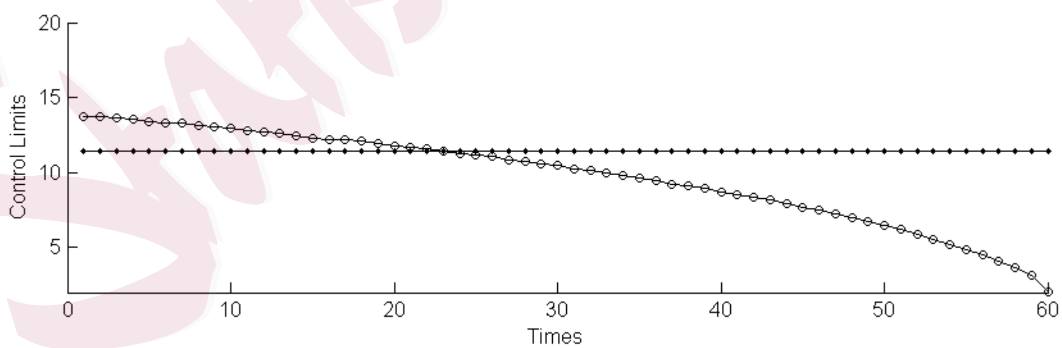

Figure 1: Control limits for T_C and T_3^* with $\mathrm{ARL}_0 \approx 40$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

We use two generalized out-of-control ARL_1s, GARL_3 and GARL_4, to evaluate the detection performance of the sequential tests, where

$$\begin{aligned} GARL_3(T) &= \mathbf{E}_0(T)\mathcal{J}_{M_3,N}(T) = \sum_{k=1}^{N} \mathbf{E}_k((1-Y_{k-1})^+(T-k)^+) \\ GARL_4(T) &= \mathbf{E}_0(T)\mathcal{J}_{M_4,N}(T) = \sum_{k=1}^{N} \mathbf{E}_k((T-k)^+), \end{aligned}$$

where $r = 0$ in $\mathcal{J}_{M_4,N}(T)$. Obviously, for any two sequential tests $T', T \in \mathfrak{T}_N$ with $\mathbf{E}_0(T') = \mathbf{E}_0(T)$, we have $GARL_j(T') \geq GARL_j(T)$ if and only if $\mathcal{J}_{M_j,N}(T') \geq \mathcal{J}_{M_j,N}(T)$, for $j = 3, 4$.

The simulation results of GARL_3 and GARL_4 for the six tests T_3^*, T_4^*, T_C, T_E, $T_C^{-1/60}$, and $T_C^{1/60}$ with the same $\mathrm{ARL}_0 \approx 20, 40, 50$ are listed in Table 1, where the values of ARL_0, the constant control limits of T_C and T_E, and the adjustment coefficients of T_3^*, T_4^*, $T_C^{-1/60}$, and $T_C^{1/60}$ are listed in parentheses. Table 1 shows that both T_3^* and T_4^* have the best detection performance; that is, T_3^* and T_4^* have the smallest GARL_3 and GARL_4 (in bold), respectively, in the six tests with the same $\mathrm{ARL}_0 \approx 20, 40, 50$. This is consistent with the result of Corollary 3: tests T_3^* and T_4^* are optimal under measures $\mathcal{J}_{M_3,N}(T)$ and $\mathcal{J}_{M_4,N}(T)$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Table 1. Simulation values of GARL_3 and GARL_4 with the same ARL_0 for independent observations

$\mathbf{ARL}_0$		**Sequential Tests**					
		T_3^*	T_4^*	T_C	T_E	$T_C^{-1/60}$	$T_C^{1/60}$
20	GARL_3	**17.59**	19.62	18.97	19.98	19.28	19.34
	GARL_4	44.75	**42.10**	45.13	48.02	46.50	47.57
	c	(1.3011)	(0.12216)	(4.4823)	(1.2250)	(6.3900)	(3.629)
	ARL_0	(20.06)	(20.01)	(20.07)	(20.08)	(20.08)	(20.07)
40	GARL_3	**49.26**	55.17	54.44	59.97	54.96	55.99
	GARL_4	145.65	**139.18**	148.07	164.28	148.76	155.80
	c	(2.0251)	(5.5996)	(11.4423)	(1.4064)	(22.1500)	(8.7815)
	ARL_0	(40.06)	(40.02)	(40.06)	(40.04)	(40.01)	(40.02)
50	GARL_3	**80.95**	84.27	83.45	95.52	83.85	85.63
	GARL_4	232.52	**229.26**	240.52	273.29	238.82	248.57
	c	(2.9518)	(0.2656)	(22.8821)	(1.5269)	(52.2500)	(17.2478)
	ARL_0	(50.05)	(50.02)	(50.04)	(50.08)	(50.00)	(50.05)

4.4 Comparison of the generalized out-of-control ARL_1 for a Markov observation sequence

Let $\{X_k, 1 \le k \le 60\}$ be a dependent observation sequence satisfying

$$X_k = \begin{cases} \rho_0 X_{k-1} + \varepsilon_k & \text{if} \quad 1 \le k \le \tau, \\ \rho_1 X_{k-1} + \varepsilon_k & \text{if} \quad k \ge \tau, \end{cases}$$

where $X_0 = 0$, $\{\varepsilon_k,\ 1 \le k \le 60\}$ is i.i.d with a normal distribution, that is, $\varepsilon_k \sim N(0, 1)$ for $1 \le k \le 60$, $\rho_0 = 0.5$, and $\rho_1 = 0.1$. That is, the correlation coefficient changes from

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

0.5 to 0.1. Obviously, $\{X_k, 1 \le k \le 60\}$ is a one-order Markov process. The pre-change and post-change transition probability densities $p_0(x,y)$ and $p_1(x,y)$, respectively, and the likelihood ratio Λ_k can be written as

$$\begin{aligned} p_0(x,y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\rho_0 x)^2}{2}}, \qquad p_1(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\rho_1 x)^2}{2}} \\ \Lambda_k &= \frac{p_1(X_{k-1}, X_k)}{p_0(X_{k-1}, X_k)} = \exp\{[(\rho_1 - \rho_0)X_{k-1}][X_k - (\rho_1 + \rho_0)X_{k-1}/2]\}. \end{aligned}$$

It can be seen that the changes in the variance and covariance of X_k and X_{k-1} occur after the change-point $\tau = k$. Here, the change-point is unknown.

Because $\{X_k, 1 \le k \le 60\}$ is a one-order Markov process, it follows from (i) of Theorem 3 that we need to calculate the equivalent control limits $\widetilde{l}_k = y_k(c, X_k)$, for $1 \le k \le 59$, to get the corresponding optimal tests T_3^* and T_4^*.

We also use the two generalized out-of-control ARL_1s, GARL_3 and GARL_4, to evaluate the detection performance of the six sequential tests T_3^*, T_4^*, T_C, T_E with the smoothing parameter 0.1, $T_C^{-1/60}$, and $T_C^{1/60}$. The simulation results of GARL_3 and GARL_4 for the six tests with the same ARL_0=$20, 40, 50$ are listed in Table 2. The ARL_0 values, the constant control limits of T_C and T_E, and the adjustment coefficients of T_3^*, T_4^*, $T_C^{-1/60}$, and $T_C^{1/60}$ are listed in parentheses. Table 2 shows that tests T_3^* and T_4^* have the best detection performance; that is, T_3^* and T_4^* have the smallest GARL_3 and GARL_4 values (in bold), respectively, of the six tests with the same $\text{ARL}_0 \approx 20, 40, 50$. This is consistent with the result of Corollary 3: sequential tests T_3^* and T_4^* are optimal under measures $\mathcal{J}_{M_3,N}(T)$ and $\mathcal{J}_{M_4,N}(T)$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Note that although the monitoring performance of T_3^* and T_4^* is better than that of $T_C, T_E, T_C^{-1/60}$, and $T_C^{1/60}$ under the measures $\mathcal{J}_{M_3,N}(T)$ and $\mathcal{J}_{M_4,N}(T)$, the constant control limits of T_C, T_E, and $T_C^{-1/60}$ are easier to determine than those of T_3^* and T_4^*.

Table 2. Simulation values of GARL$_3$ and GARL$_4$ with the same ARL$_0$ for a one-order Markov observation sequence

		Sequential Tests					
ARL$_0$	**ARL$_1$**	T_3^*	T_4^*	T_C	T_E	$T_C^{-1/60}$	$T_C^{1/60}$
20	GARL$_3$	**21.55**	23.26	22.04	67.46	22.72	23.09
	GARL$_4$	135.25	**115.43**	139.64	551.78	130.92	156.09
	c	(2.075)	(12.016)	(2.3482)	(0.3150)	(3.4500)	(1.8901)
	ARL$_0$	(20.14)	(20.05)	(19.97)	(20.09)	(20.01)	(20.09)
40	GARL$_3$	**57.86**	59.80	59.71	148.09	60.60	60.30
	GARL$_4$	467.17	**409.76**	474.64	1261.20	450.68	490.42
	c	(3.865)	(22.8550)	(4.7828)	(0.5895)	(10.3500)	(3.478)
	ARL$_0$	(40.84)	(40.72)	(40.76)	(40.07)	(40.02)	(40.03)
50	GARL$_3$	**80.42**	84.15	83.32	180.06	87.25	87.57
	GARL$_4$	688.52	**638.15**	705.62	1579.13	722.63	758.57
	c	(5.575)	(32.89)	(7.528)	(0.755)	(23.15)	(5.667)
	ARL$_0$	(49.26)	(49.77)	(49.28)	(49.82)	(49.94)	(50.04)

Remark 5. We now discuss how to choose appropriate performance measures. If the distribution of the change-point τ is known, $\rho_k = P(\tau = k)$, it is better to use the measure $\mathcal{J}_{M_1,N}(T)$. If $\tau = 1$, we use the measure $\mathcal{J}_{M_2,N}(T)$. If the change-point τ is unknown, we

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

should use the measures $\mathcal{J}_{M_3,N}(T)$ or $\mathcal{J}_{M_4,N}(T)$. Because

$$\begin{aligned}GARL_3(T) &= \mathbf{E}_0(T)\mathcal{J}_{M_3,N}(T) = \sum_{k=1}^{N}\mathbf{E}_k((1-Y_{k-1})^+(T-k)^+) \\ &\leq \sum_{k=1}^{N}\mathbf{E}_k((T-k)^+) = \mathbf{E}_0(T)\mathcal{J}_{M_4,N}(T) = GARL_4(T),\end{aligned}$$

where $r = 0$ in $\mathcal{J}_{M_4,N}(T)$, we recommend using the measure $\mathcal{J}_{M_3,N}(T)$ to evaluate the detection performance when the change-point is unknown.

5. A real-data example

Performance monitoring is important for any industry or enterprise to make appropriate evaluations of the past operating cycle and to plan for the next. Sequential tests, or control charts, are commonly used in business to monitor operating indicators, such as customer attrition rates, sales margins, and order numbers.

Consider a real example. The data set is drawn from an actual process of a new e-commerce company providing a retail service. More information can be found in Yu et al. (2018). The parameter being monitored is the daily order quantity in a district in Shanghai. The data period ranges from July 2008 (i.e., when the site first went online) to August 2008, and the data include the order date and user ID. In order to develop customers, new e-commerce companies have offered attractive discounts. However, they cannot sustain these online discounts on a continuous, unlimited, and cost-free basis. They need to observe whether there has been a change in the order volume after a limited period. The aim is to detect any upward shifts in the mean, because these would signal

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

improvements in operating performance.

Because the change-point is unknown, we use the measure $\mathcal{J}_{M_3,N}(T)$ to evaluate the detection performance of each sequential test. The detection performance of the four sequential tests, T_3^*, T_C, $T_C^{1/60}$, and $T_C^{-1/60}$, based on the measure $\mathcal{J}_{M_3,N}(T)$, are illustrated using this real-data example. The data analysis proceeds in several steps:

- Step 1: Exploratory data analysis

  Fig.2 shows the daily order numbers throughout the observation period. The number of orders increases at about the end of July. The goal here is to detect any upward shifts.

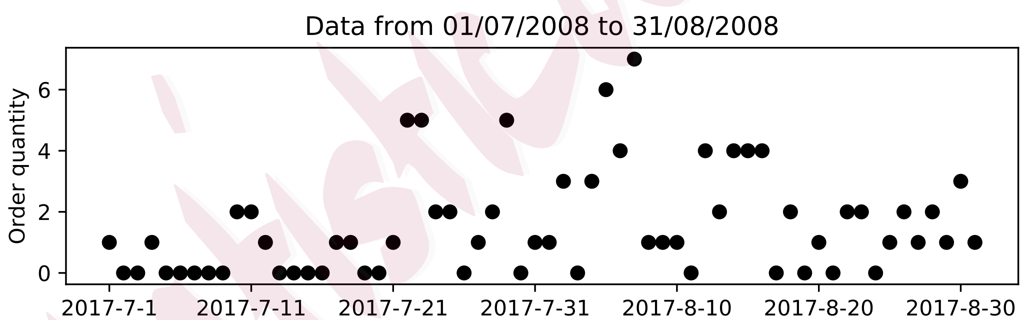

Figure 2: Exploratory data analysis

- Step 2: The test of its Markov property

  Daily order volume data is somewhat correlated. First, we cluster the order quantities ($\{D_n, n = 0, 1, \cdots, 61\}$) into three states, denoted as $X_n \in \{0, 1, 2\}$, as follows:

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

$$X_n = \begin{cases} 0, D_n \leq 1; \\ 1, 2 \leq D_n \leq 3; \\ 2, D_n \geq 4. \end{cases} \tag{5.1}$$

The estimations of the in-control and out-of-control transition probability matrices are

$$P_0 = \begin{pmatrix} 0.8636 & 0.0909 & 0.0455 \\ 0.4 & 0.4 & 0.2 \\ 0.3333 & 0.3333 & 0.3334 \end{pmatrix} \tag{5.2}$$

and

$$P_1 = \begin{pmatrix} 0.4667 & 0.4667 & 0.0666 \\ 0.625 & 0.125 & 0.25 \\ 0.2857 & 0.1429 & 0.5714 \end{pmatrix}, \tag{5.3}$$

respectively, based on data from the previous month and the later month, respectively. The χ^2 statistic is applied to test the Markov property, which the results show is satisfied by both processes.

- Step 3: Detection

  We employ the above four sequential tests with $\mathrm{ARL}_0 = 45$ to detect the observations $X_1, X_2, \cdots, X_{60}$. The parameter c in the control charts of the four tests is shown in Table 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Table 3.Parameter c in the control limit of four sequential tests

Test	c	ARL_0	Change Point
T_6^*	2.64	45.1168	33
T_C	12.90	45.1979	34
$T_C^{1/60}$	32.99	45.1326	34
$T_C^{-1/60}$	10.40	45.1453	34

Figure 3 illustrates the monitoring process in four different tests, and we can find that T_3^* alerts at the 33rd daily record, while the other three tests signal at the 34th record. The reason that the three tests alert at the same day is that there is a relatively bigger change around the 33th day.

The proposed T_3^* scheme performs more sensitively under the measure $\mathcal{J}_{M_3,N}(T)$.

6. Conclusion

By presenting the generalized Shiryaev measures of detection delay $\mathcal{J}_{M,N}(.)$, the statistic Y_n, for $0 \leq n \leq N+1$, the control limit $l_n(c)$, for $0 \leq n \leq N+1$, and the sequential test $T_M^*(c, N)$, for N finite observations, we obtain the following main results. (i) For different measures $\mathcal{J}_{M,N}(.)$ of detection delay, we can construct different optimal sequential tests $T_M^*(c, N)$ under the corresponding measures for a general finite observation sequence. (ii) A formula is presented to calculate the value of the generalized out-of-control ARL_1 for every optimal test $T_M^*(c, N)$ that is the minimum value of the generalized out-of-control ARL_1 of all tests $T \in \mathfrak{T}_N$. (iii) When the post-change conditional densities (probabilities) of the observation sequences do not depend on the change-point, there is an equivalent

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

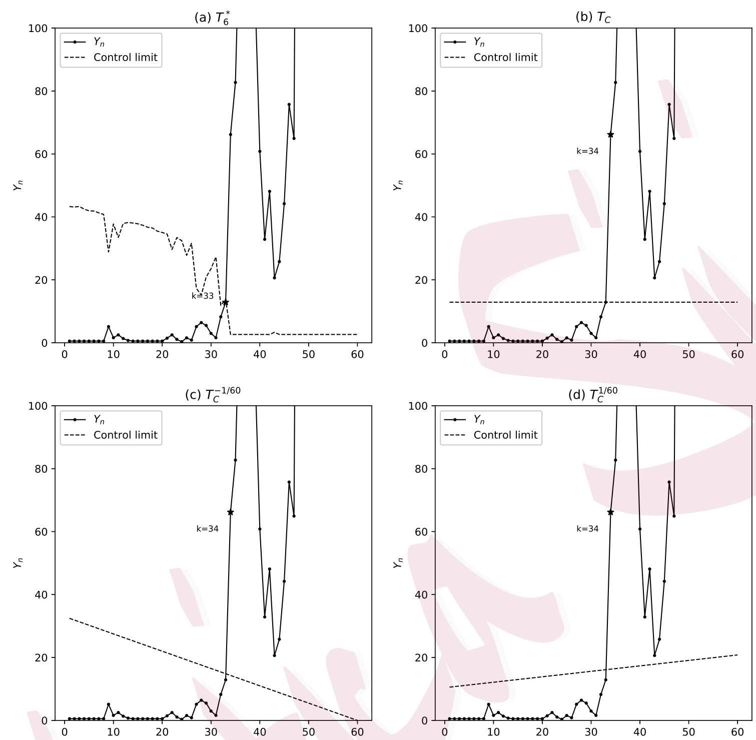

Figure 3: Testing results for Markov observation sequence

control limit that does not depend directly on the statistic of the optimal test $T_M^*(c, N)$ for a q-order Markov process. Specifically, the equivalent control limit can consist of a series of nonnegative nonrandom numbers when the observations are mutually independent.

In this study, both the pre-change and post-change joint probability densities are assumed to be known. In fact, we usually do not know the post-change joint probability density before it is detected. However, the potential change domain (including the size and the form of the boundary) and its probability may be determined by engineering

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

knowledge and practical experience. In other words, although the actual post-change joint probability density $p(\theta,k) := p_{\theta,k}(x_0, x_1, ..., x_k, ..., x_N)$ is unknown, that is, the parameter θ is unknown at the change time k, we may assume that there is a known probability distribution $Q_k(.)$ for the known parameter set Θ_k, such that the probability of the post-change joint probability density at change-point k being $p_{\theta,k}$ is $dQ_k(\theta)$, for $1 \le k \le N$, where $p_{\theta,k} \ne p_{\theta',k}$ if and only if $\theta \ne \theta'$. If we have no prior knowledge of the possible parameter θ (corresponding to a possible post-change probability density $p_{\theta,k}$) at the change-point k, it is natural to assume that the probability distribution Q_k may be an equal probability distribution or uniform distribution on Θ_k; that is, $Q_k(\theta = \theta_i) = 1/m$ $(1 \le i \le m < \infty)$ for $\theta_i \in \Theta_k$ or $dQ(\theta)/d\theta = 1/M(\Theta)$, where $dQ/d\theta$ denotes the probability density and $M(\Theta)$ is the measure (length, area, volume, etc.) of the bounded set Θ. Note that the parameter θ may not be the characteristic numbers (the mean, variance, etc.) of the probability distribution. Hence, we can define a new joint probability density

$$p_k := p_k(x_0, x_1, ..., x_k, ..., x_N)$$

as follows:

$$p_k(x_0, x_1, ..., x_N) = \int_{\Theta_k} p_{\theta,k}(x_0, x_1, ..., x_N) dQ_k(\theta),$$

for $1 \le k \le N$. The density function p_k can be considered a known post-change joint probability density at the change-point k, for $1 \le k \le N$.

Supplementary Material

The proofs of Theorems 1, 2, and 3 are shown in the online Supplementary Material.

Acknowledgements

We thank the editor and two referees for their valuable comments and suggestions. Our research work was supported by NSFC 11531001 and 71931006, RGC GRF 16216119 and 16201718, the National Natural Science Foundation of China (No. 11871324), and the National Bureau of Statistics of China (NO. 2020LD03).

References

Akoglu, L., Tong, H. H. and Koutra, D. (2015) Graph based anormaly detection and description: a survey. *Data Min Know Disc.* **29**, 626-688.

Basseville, M. and Nikiforov, I. (1993) *Detection of Abrupt Changes: Theory and Applications.* Prentice-Hall, Englewood Cliffs.

Bersimis, S., Sgora, A. and Psarakis, S. (2018) The application of multivariate statistical process monitoring in non-industrial processes. *Qual. Technol. Quant. Manag.*, **15**, 526-549.

Bersimis,S., Psarakis, S. and Panaretos, J. (2007) Multivariate statistical process control charts: An Overview. *Qual. Reliab. Engng. Int.*, **23**, 517-543.

Chakraborti, S., van der Laan, P. and Bakir, S. T. (2001) Nonparametric control charts: an overview and some results. *J. Qual. Technol.*, **33**, 304-315.

Chen, X. and Baron, M. (2014) Change-point alalysis of survival data with application in clinical trials. *Open J.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

Statistics, **4**, 663-677.

Chow, Y. S., Robbins, H. and Siegmund, D.(1971) *The Theory of Optimal Stopping.* Dover Publications, INC. New York.

Frisén, M. (2003) Statistical Surveillance, Optimality and Methods. *Int. Statist. Rev.*, **71**, 403-434.

Frisén, M. (2009) Optimal Sequential Surveillance for Finance, Public Health, and Other Areas (with Discussion). *Sequent. Analysis*, **28**, 310-337.

Han, D. and Tsung, F. G. (2004) A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes. *Ann. Statist.*, **32**, 316-339.

Han, D., Tsung, F. G. and Xian, J. G. (2017) On the optimality of Bayesian change-point detetion. *Ann. Statist.*, **45**, 1375-1402.

Hosseini, S. H. and Noorossana, R. (2018) Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures. *Qual Reliab Engng Int.*, **34**, 477C500.

Lai, T. L. (1995) Sequential change-point detection in quality control and dynamical systems (with discussion). *J. Roy. Statist. Soc. Ser. B.*, **57**, 613-658.

Lai, T. L. (2001) Sequential analysis: some classical problems and new challenges. *Statist Sinica*, **11**, 303-408.

Lorden, G. (1971) Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, **42**, 1897-1908.

Montgomery, D. C. (2009) *Introduction to Statistical Quality Control.* 6th ed. New York: John Wiley & Sons.

Moustakides, G. V. (1986) Optimal stopping times for detecting changes in distribution. *Ann. Statist.*, **14**, 1379-1387.

Moustakides, G. V. (2008) Sequential change detection revisited. *Ann. Statist.*, **36**, 787-807.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41** 100-114.

Pollak, M. (1985) Optimal detection of a change in distribution. *Ann. Statist.*, **13**, 206-227.

Polunchenko, A. S. and Tartakovsky, A. G. (2010). On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution. *Ann. Statist.*, **38**, 3445-3457.

Poor, H.V. and Hadjiliadis, O. (2009). *Quickest Detection*. Cambridge University Press, Camridge, New York.

Qiu, P. (2014) *Introduction to Statistical Process Control*. Boca Raton, FL: Chapman & Hall/CRC.

Rigdon, S. E. and Fricker, R. D., Jr. (2015) *Innovative Statistical Methods for Public Health Data* (eds by D. G. Chen and J. Wilson). ICSA Book Series in Statistics. Springer International Publishing Switzerland.

Roberts, S.W. (1959) Control chart tests based on geometric moving average. *Technomietrics*, **1**, 239-250.

Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I. and Woodall, W. H. (2015). Another Look at the EWMA Control Chart with Estimated Parameters. *J. Qual. Technol.*, **47**, 363-382.

Shewhart, W. A. (1931) *Economic Control of Quality of Manufactured Product*. New York: Van Nostrand.

Shiryaev, A. N. (1963) On optimum methods in quickest detection problems. *Theory Probab. Appli.*, **13**, 22-46

Shiryaev, A. N. (1978) *Optimal Stopping Rules*. Springer-Verlag, New York.

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

Siegmund, D. (2013) Change-points: from sequential detection to biology and back. *Sequ. Analysis*, **32**, 2-14

Stoumbos, Z. G., Reynolds, M. R., Ryan, T. P., and Woodall, W. H. (2000) The state of statistical process control as we proceed into the 21st century. *J. Amer. Statist. Assoc.*, **95**, 992-998.

Tartakovsky, A. G., Nikiforov, I. and Basseville, M, (2015). *Sequential Analysis: Hypothesis Testing and Change-point Detection*. CRC Press, Taylor & Francis Group.

Woodall, W. H. (2006) The use of control charts in health-care and public-health survellance. *J. Qual. Technol.*, **38**, 89-118.

Woodall, W. H. and Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *J. Qual. Technol.*, **46**, 78-94.

Woodall, W. H., Zhao, M. J., Paynabar, K., Sparks, R. and Wilson, J. D. (2017). An overview and perspective on social network monitoring. *IIE Trans.*, **49**, 354-365.

Yu, M., Wu, C., Wang, Z., and Tsung, F. (2018). A robust CUSUM scheme with a weighted likelihood ratio to monitor an overdispersed counting process. *Comput. Ind. Eng.* **126**, 165-174.

Shanghai Jiao Tong University

E-mail: (donghan@sjtu.edu.cn)

Hong Kong University of Science and Technology

E-mail: (season@ust.hk)

Shanghai Jiao Tong University

E-mail: (jgxian@jgxian@sjtu.edu.cn)

East China Normal University

E-mail: (mmyu@fem.ecnu.edu.cn)

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0333

APPENDIX 1: Four performance measures

It is clear that taking various weighted pairs $M = (W, V)$, we can get various measures $\mathcal{J}_{M,N}(.)$. Take four weighed pairs $M_i = (W_i, V_i), 5 \le i \le 8$, we can get the folowing performance measures

$$\begin{aligned}
\mathcal{J}_{M_5,N}(T) &= \frac{\mathbf{E}_1(T-1)}{\sum_{j=1}^{N+1}\rho_j \mathbf{P}_0(T \ge j)}, \quad \mathcal{J}_{M_6,N}(T) = \frac{\sum_{k=1}^{N}\mathbf{E}_k((1-Z_{k-1})^+(T-k)^+)}{\mathbf{E}_0(\sum_{j=1}^{T}(1-Z_{j-1})^+)}, \\
\mathcal{J}_{M_7,N}(T) &= \frac{\mathbf{E}_1((T-1)) + \sum_{k=2}^{N+1}\mathbf{E}_k(e^{X_{k-1}}(1+e^{X_{k-1}})^{-1}(T-k)^+)}{1+\mathbf{E}_0(\sum_{k=2}^{T}e^{X_{k-1}}(1+e^{X_{k-1}})^{-1})}, \\
\mathcal{J}_{M_8,N}(T) &= \frac{\mathbf{E}_1((T-1)) + \sum_{k=2}^{N}\mathbf{E}_k(\frac{1}{k-1}\sum_{j=1}^{k-1}e^{X_j}(T-k)^+)}{\mathbf{E}_0(T)},
\end{aligned}$$

where $W_5 = \{w_1 = 1, w_k = 0, 2 \le k \le N+1\}$, $V_5 = \{v_k = \rho_k, 1 \le k \le N+1\}$, $\rho_{N+1} := 1 - \sum_{k=1}^{N}$, $W_6 = V_6 = \{w_j = v_j = (1 - Z_{j-1})^+, 1 \le j \le N+1\}$, and $Z_k = \max\{1, Z_{k-1}\}\Lambda_k$ for $1 \le k \le N$, are the statistics of the CUSUM test with $Z_0 = 0$. Here, both $W_7 = V_7 = \{w_k = v_k = e^{X_{k-1}}/(1 + e^{X_{k-1}}), 1 \le k \le N+1\}$ and $W_8 = V_8 = \{w_k = v_k = \frac{1}{k-1}\sum_{j=1}^{k-1} e^{X_j}, 1 \le k \le N+1\}$ in the two new measures $\mathcal{J}_{M_7,N}(T)$ and $\mathcal{J}_{M_8,N}(T)$, can describe some kind of possibility of the changes of the observation values at change-point $k-1$ and the average of the changes of the observation values before the change-point $k \ge 2$, respectively.