Journal of Statistical Computation & Simulation
Vol. 00, No. 00, January 2008, 1–26

REVIEW ARTICLE

The Computation of Average Run Length and Average Time to Signal: An Overview

Zhonghua Li[a], Changliang Zou[a], Zhen Gong[a] & Zhaojun Wang [a]*

[a]*LPMC and School of Mathematical Sciences, Nankai University, Tianjin 300071, China*

(*v4.2 released May 2008*)

Control charts are widely used in industries to monitor a process for quality improvement. Evaluation of the average run length (ARL) or average time to signal (ATS) plays an important role in the design of control charts and performance comparison. In this paper, we review several basic and popular procedures, including the Markov chain and integral equation methods for computing ARL, ATS and associated run length distributions for cumulative sum (CUSUM) charts, exponentially weighted moving average (EWMA) charts, and combined control charts, respectively. Some important references and key formulations are provided for practitioners.

Keywords: Average Run Length; Average Time to Signal; Markov Chain; Integral Equation; Statistical Process Control

AMS Subject Classification: 62P30; 65C20; 65C60; 68U20

1. Introduction

Control charts are very important tools in statistical process control (SPC), whose main objective is to improve and guarantee the quality of processes so as to satisfy customer requirements. Several charting schemes have been proposed to try to simplify the process of SPC, in particular the Shewhart $\bar{X}$ control chart ([1]), the CUSUM control chart ([2]) and the EWMA control chart ([3]). Interested readers are referred to [4–8] for the voluminous research, comprehensive discussions and broad applications of control charts.

The basic idea of SPC chart is similar to the following sequential hypothesis test:

$$H_0 : X_i \sim N(\mu_0, \sigma_0^2), i \geq 1 \longleftrightarrow H_1 : X_i \sim \begin{cases} N(\mu_0, \sigma_0^2), \ 1 \leq i \leq \tau, \\ N(\mu_1, \sigma_1^2), \quad i > \tau, \end{cases}$$

where $\mu_1 \neq \mu_0$ and/or $\sigma_1 \neq \sigma_0$, τ is an unknown change point. If there is no enough confidence to reject the null hypothesis H_0, we will say that the production line is in "statistical control" (in control, IC) and stable with "common causes", which can not be removed easily from the process without fundamental changes in the process itself. On the other hand, if it shows enough support to reject H_0, we will say that the control chart issues a signal and the production line is out of "statistical control" (out-of-control, OC) and undergoes an unusual variation due to "assignable causes". One purpose of a control chart is to detect unusual

*Corresponding author. Email: zjwang@nankai.edu.cn

ISSN: 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/0094965YYxxxxxxxx
http://www.informaworld.com

variation as soon as possible, keeping at the same time the probability of erroneous signal below a reasonable level.

Extensive research in recent years has developed variable sample rate (VSR) control charts that vary the sampling rate as a function of current and prior sample results, which include variable sampling intervals (VSI) chart, variable sample size (VSS) chart and variable sample sizes and sampling intervals (VSSI) chart. The advantage of using a VSR chart instead of a fixed sampling rate (FSR) chart is that a VSR chart provides much faster detection of small and moderate process changes, for a given IC ARL or ATS and a given IC average sampling rate. There have been lots of research on control chart using VSR features in the literature, for the $\overline{X}$ control chart, see [9–12]; for the CUSUM control chart, see [13–15]; for the EWMA control chart, see [16, 17]; for the CUSUM of Q chart, see [18]; for linear profile, see [19]; for the cumulative count of conforming chart, see [20] and for the Hotelling's T^2 control chart, see [21].

Traditionally, the ARL, which is defined as the average number of samples before the chart signals, has been generally employed as a performance indicator to evaluate the effectiveness of various control schemes, provided that the sampling interval remains constant. However, when the sampling interval is variable, the time to signal is not a constant multiple of the ARL, and thus ARL is not appropriate for evaluating the effectiveness of VSI control charts. The widely used performance indicators for control charts with VSI are the ATS, which is defined as the expected time from the start of the process to the time when the charts indicate an OC signal, and the adjusted average time to signal (AATS), which is defined as the expected time from the occurrence of an assignable cause to the time when the charts indicate an OC signal. The AATS is also called the steady-state ATS (SSATS). When the sampling sizes are variable, a necessary indicator is the average number of samples to signal (ANSS), which is the expected number of samples taken from the start of the process to the time the charts signal.

When the process is IC, a chart with a larger IC ARL or ATS indicates a lower false alarm rate than other charts. When the process is OC, a chart with a smaller OC ARL or ATS indicates a better detection ability of process shifts than other charts. Therefore, in comparison of various candidate control charts, ARL or ATS is very important and also popular used criterion. Given the voluminous research in various areas of control charts, the purpose of this paper is to give detailed review of the computation of ARL and ATS, encourage research of control charting and SPC, and provide handy references for practitioners. The main methods for computing ARL in the literature are Markov chain approach ([22]), integral equation approach ([23]) and Monte Carlo simulation. Monte Carlo simulation is not reviewed in this paper, because in some cases, the computation burden is considerable. For example, in the low defect rate process, even the IC ARL can be quite large ([24, 25]). Some approximations to the run length distribution are made in [2, 26–29], among others. The comparison between Markov chain approach and integral equation approach is made in [30].

This paper is the result of an extensive literature review of the most recent developments in the area of computation of ARL, ATS and related indexes, and the rest of this paper is organized as follows. In Sections 2-4, we give detailed review of the Markov chain approach and integral equation approach for CUSUM control charts, EWMA control charts and combined control charts, respectively. In Sections 5 and 6, we turn our focus on adaptive control charts and show some differences between steady state ARL (SSARL) and ARL, SSATS and ATS, and also some basic methods for correlated data. Finally, we summarize our conclusions in Section 7. Two numerical methods for integral equations are deferred to the Appendix.

2. Methods for CUSUM Control Charts

Markov chain method was originally proposed by [22] for CUSUM control chart, and from then on the Markov Chain idea, due to its ease to implement, was widely applied in various control schemes. Integral equation method ([23]) is another basic and popular method. Although in most cases the integral equations cannot be solved directly, they can be converted into a set of linear equations with the help of Gauss quadrature, and then the ARL can be obtained indirectly. In this section, we will give detailed review of the Markov chain approach and integral equation approach for one-sided and two-sided CUSUM and adaptive CUSUM ([31]).

2.1. *Markov Chain Method*

Noting that the operation of a CUSUM scheme forms a Markov process with a continuous state space, [22] found that good approximations to the various characteristics of the run length distribution can be obtained by discretizing the probability distribution of the monitoring statistics so that the CUSUM is restricted to a finite set of values. Then the transition probability matrix can be easily constructed from the given probability distribution. Hence, the exact probability distribution of run length and its moments can be determined.

For different control charts, the differences in Markov chain method mainly lie in the transition probability matrix. We will give elaborate introductions on the Markov chain method proposed by [22] first, and then show its applications in other practical control charts.

2.1.1. Markov Chain Method ([22])

For one-sided CUSUM control chart, [22] used the V-mask type of scheme at time n: $S_n = \sum_{i=1}^{n}(X_i - k)$, where X_i is the observed process characteristic variable and k is reference value. The control chart issues a signal if $S_n > h$, where h is control limit. They first considered the discrete case, that is, X_i, k, h are all positive integers, and so that S_n can only take one of the integral values $0, 1, 2, \ldots, h$. If $S_n = i$, the scheme is said to be in state E_i, where E_h is an absorbing state. The initial state is assumed to be E_0.

The transition probabilities from state E_i to state E_j are determined only by the probability distribution of X as follows:

$$\begin{aligned}
p_{ij} &= P\{S_{n+1} \in E_j | S_n \in E_i\} = P\{S_n + X_{n+1} - k = j | S_n = i\} \\
&= P\{X_{n+1} - k = j - i\}, \quad i \neq h, j \neq h, j \neq 0, \\
p_{i0} &= P\{X \leq k - i\}, \\
p_{ih} &= P\{X \geq k + h - i\}, \\
p_{hj} &= 0, \quad j = 0, 1, ..., h-1, \\
p_{hh} &= 1.
\end{aligned}$$

Let $p_r = P\{X - k = r\}$ and $F_r = P\{X - k \leq r\}$ and assume the values of k, h and the probability distribution of X are given. Then the transition probability matrix $\mathbf{P}$ has the following form

$$\mathbf{P} = \begin{bmatrix} F_0 & p_1 & \cdots & p_{h-1} & 1-F_{h-1} \\ F_{-1} & p_0 & \cdots & p_{h-2} & 1-F_{h-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ F_{1-h} & p_{2-h} & \cdots & p_0 & 1-F_0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Many of the results we require can be obtained by working with the matrix $\mathbf{R}$ obtained from $\mathbf{P}$ by deleting the final row and column, so $\mathbf{P}$ is rewritten in the form of partitioned matrix

$$\mathbf{P} = \begin{pmatrix} \mathbf{R} & (\mathbf{I}-\mathbf{R})\mathbf{1} \\ \mathbf{0}^{\mathbf{T}} & 1 \end{pmatrix},$$

where $\mathbf{I}$ is the $h \times h$ identity matrix and the vector $\mathbf{1}$ has each of its h elements equal to unity. Let T_i be the number of steps taken starting from E_i to reach the absorbing state E_h for the first time and $\boldsymbol{T} = (T_0, T_1, \ldots, T_{h-1})'$. Then ARL with initial state E_i is just the ith component of $E(\boldsymbol{T})$. For $r = 1, 2, \ldots$, define

$$\begin{aligned} \boldsymbol{F}_r &= (P\{T_0 \le r\}, P\{T_1 \le r\}, \ldots, P\{T_{h-1} \le r\})^T, \\ \boldsymbol{L}_r &= (P\{T_0 = r\}, P\{T_1 = r\}, \ldots, P\{T_{h-1} = r\})^T. \end{aligned}$$

Again by the properties of Markov chain, for $r = 1, 2, \ldots$,

$$\begin{aligned} \boldsymbol{F}_r &= (\mathbf{I} - \mathbf{R}^r)\mathbf{1}, \\ \boldsymbol{L}_r &= \mathbf{R}\boldsymbol{L}_{r-1} = \mathbf{R}^{r-1}(\mathbf{I}-\mathbf{R})\mathbf{1}. \end{aligned}$$

By the definition of expectation,

$$ET_i = \sum_{m=1}^{\infty} mP\{T_i = m\} = \sum_{m=1}^{\infty} P\{T_i \ge m\}.$$

So the ARL can be obtained as

$$ARL = E(\boldsymbol{T}) = (\mathbf{I} - \mathbf{R})^{-\mathbf{1}}\mathbf{1}. \tag{1}$$

From Eq. (1), the distributional function of run length has a form very similar to the univariate geometric distribution.

As for the commonly used form of upper-sided CUSUM control chart, i.e., decision interval (DI) form, which is equivalent to the V-mask form above ([7]) and has the following definition:

$$S_n = \max\{0, S_{n-1} + X_n - k\}.$$

When $S_n \ge h$, the control chart issues a signal. For convenience, we denote this chart by $C^+(S_0, k, h)$. In a similar way, denote lower-sided CUSUM control chart by $C^-(s_0, k, h)$, which is defined as

$$s_n = \min\{0, s_{n-1} + X_n + k\}.$$

When the observed process characteristic variables X are theoretically continuous, [22] represented the continuous scheme by a Markov chain having $t+1$ states by dividing the state space of S_n into $t+1$ subintervals:

$$[0,\omega/2) \cup [\omega-\omega/2, \omega+\omega/2) \cup \cdots \cup [(t-1)\omega-\omega/2, (t-1)\omega+\omega/2) \cup [h,\infty),$$

where $\omega = 2h/(2t-1)$. When the statistics S_n fall in the ith ($i = 0,1,2,\ldots,t$) subinterval I_i, it is said that S_n is in state E_i. Obviously, E_t is the absorbing state.

The transition probabilities $p_{ij} = P\{S_{n+1} \in I_j | S_n \in I_i), i = 0,1,\ldots,t-1$ for the Markov chain are then as follows:

$$\begin{aligned}
p_{i0} &= P\{S_n \in I_0 | S_{n-1} = i\omega\} = P\{X_n \le k - iw + w/2\}, \\
p_{ij} &= P\{S_n \in I_j | S_{n-1} = i\omega\} \\
&= P\{(j-i)w - w/2 < X_n - k \le (j-i)w + w/2\}, 1 \le j \le t-1, \\
p_{it} &= P\{S_n \in I_t | S_{n-1} = i\omega\} = P\{X_n - k > (t-i)w - w/2\}.
\end{aligned}$$

Then ARL can be computed by Eq. (1). [32] proposed a more accurate numerical method for the transition probabilities $p_{ij} = P\{S_{n+1} \in I_j | S_n \in I_i)$.

2.1.2. *Computation of ARL for Adaptive CUSUM*

Assume observations X_i are independently and identically distributed (i.i.d.) as normal with mean μ and variance σ^2, and a step shift δ (in units of the standard deviation) is issued at some unknown point during the process. Because the CUSUM statistics are obtained based on likelihood ratio test ([7]), it is well known that if the IC parameters, say μ and σ are assumed to be known, the CUSUM with reference value $k = \delta/2$ is optimal for detecting this shift ([33–36]). However, the reference value is designed for a given shift and is generally difficult to be determined especially in the start-up stages, where even the IC parameters are not accurately identified. In order to overcome this limitation, [37] proposed the idea of adaptive CUSUM (ACUSUM). [37] first used EWMA to estimate the size of shift, and then chose the reference value adaptively according to the estimated size of shift. For $C^+(S_0,k,h)$, the monitoring statistics are $S_t^+ = \max\{0, S_{t-1}^+ + (X_t - Q_t^+/2)/h(Q_t^+/2)\}$, where $Q_t^+ = \max\{\delta_{\min}^+, (1-\lambda)Q_{t-1}^+ + \lambda X_t\}$, $0 < \lambda \le 1$ is the smoothing constant, $h(k)$ is a known operating function of k with the aim to make the control limit c close to 1, $\delta_{\min}^+ > 0$ is a minimum location shift of high importance for early detection, Q_t^+ is an estimation of shift for the unknown mean, and in general $Q_0^+ = \delta_{\min}^+$.

Compared with CUSUM charts, the ACUSUM chart proposed by [37] can be efficient in signaling a broader range of mean shifts, by dynamically adjusting its reference values according to current process information. However, the ARL values are obtained through Monte Carlo simulation in [37]. [31] developed a two-dimensional Markov chain model to analyze run length performance of the ACUSUM chart, which improves on the theoretical understanding of the ACUSUM schemes and also allows the analysis without running extensive Monte Carlo simulations.

[31] modeled $(S_t^+, Q_t^+)^T$ as a two-dimensional Markov chain. They chose

a large control limit L for the process of Q_t^+ so that the IC region can be partitioned within a two-dimensional rectangle $[0,c] \times [\delta_{\min}^+, L]$ to obtain a discretized Markov chain. Assume that the number of states along the axis S_t^+ over the range $[0,c]$ is m_1, then the width of each segment is $\omega = 2c/(2m_1 - 1)$, except that the width of the first segment is $\omega/2$. Similarly, the axis Q_t^+ over the interval $[\delta_{\min}^+, L]$ is segmented into m_2 states such that the width of each segment is $\Delta = 2(L - \delta_{\min}^+)/(2m_2 - 1)$, except that the width of the first segment is $\Delta/2$. The states along the axis S_t^+ and the axis Q_t^+ are labelled by $i = 0,1,2,...,m_1 - 1$ and $j = 0,1,2,...,m_2 - 1$, respectively.

Therefore, the IC region is divided into a number of $N = m_1 \times m_2$ two-dimensional rectangles.

Let $p_{(i,j)(k,l)}$ be the transition probability of $(S_t^+, Q_t^+)^T$ from state (i,j) to state (k,l). Let S_t^+ be $i\omega$, which is the midpoint of the ith state along axis S_t^+, when S_t^+ is in the ith state and similarly let Q_t^+ be $\delta_{\min}^+ + j\Delta$, which is the midpoint of the jth state along axis Q_t^+, when Q_t^+ is in the jth state. When neither of k, l is 0, the transition probability $p_{(i,j)(k,l)}$ can be evaluated by

$$\begin{aligned} p_{(i,j)(k,l)} &= P\{S_{t+1}^+ \in k, Q_{t+1}^+ \in l | S_t^+ \in i, Q_{t+1}^+ \in j\} \\ &= P\{(k-0.5)\omega < S_{t-1}^+ + \frac{[X_t - (\delta_{\min}^+ + l\Delta)/2]}{h([\delta_{\min}^+ + l\Delta]/2)} < (k+0.5)\omega, \\ &\qquad \delta_{\min}^+ + (l-0.5)\Delta < (1-\lambda)Q_{t-1}^+ + \lambda X_t < \delta_{\min}^+ + (l+0.5)\Delta \mid \\ &\qquad\qquad S_{t-1}^+ = i\omega, Q_{t-1}^+ = \delta_{\min}^+ + j\Delta\}. \end{aligned}$$

When $k \neq 0, l = 0$ or $k = 0, l \neq 0$ or $k = 0, l = 0$, the transition probability $p_{(i,j)(k,l)}$ can be obtained in the same way but with a simpler representation. Therefore, if the distribution of observations X_t is given, it is immediate to compute the above transition probabilities. Finally, the ARL values can be evaluated by formula (1).

[38] proposed a new self-starting approach that integrates the CUSUM of the Q chart with the feature of adaptively varying the reference value, to better detect a range of shifts with unknown process parameters. Since [39] had shown that the Q statistics are i.i.d. normal random variables under IC conditions, that is, the IC mean and variance of the process are both known, the two-dimensional Markov chain model developed by [31] can be used to evaluate the IC ARL of the proposed adaptive CUSUM of the Q chart.

Note that to evaluate the overall ARL of a two-sided ACUSUM scheme, a four-dimensional Markov chain based on the vector $(S_t^+, S_t^-, Q_t^+, Q_t^-)^T$ need to be developed. However, this would be computationally burdensome. Instead, [31] suggested that the ARL of a two-sided ACUSUM scheme be derived from those of two one-sided schemes, i.e.,

$$\frac{1}{ARL} = \frac{1}{ARL^+} + \frac{1}{ARL^-} \tag{2}$$

where ARL^+ and ARL^- are the respective ARLs for the upper- and lower-sided ACUSUM schemes. We will defer the conditions on which this formula can hold to next subsection.

2.1.3. Methods for Two-sided CUSUM

In the previous subsections, the Markov chain method is illustrated with upper-sided CUSUM ($C^+(S_0^+, k^+, h^+)$), and it is straightforward to apply the Markov chain method to lower-sided CUSUM ($C^-(s_0^-, k^-, h^-)$). If practitioners want to detect both the increase and decrease in shifts simultaneously, a pair of one-sided schemes (two-sided schemes) is needed. Although the overall ARL of a two-sided CUSUM can also be evaluated by a two-dimensional Markov chain based on vector $(S_n^+, s_n^-)^T$, the computation would be burdensome and even impossible due to computer limitation when the number of states is large. Fortunately, under the condition of non-interaction, the ARL of two-sided schemes can be obtained from the ARL of the upper-sided and lower-sided schemes by Eq. (2). Thus, the crucial question in this situation is whether the upper-sided and lower-sided schemes interact. Note that the upper-sided and lower-sided schemes do not interact implies whenever the monitoring statistics of one of the upper-sided and lower-sided schemes issue a signal, the monitoring statistics of the other one will reset to zero.

[40] proved the relationship in Eq. (2) under the condition of non-interaction. [41] derived the following formula for computing the ARL of CUSUM control chart with fast initial response (FIR)

$$ARL(s) = \frac{L_H(s)L_L(0) + L_H(0)L_L(s) - L_H(0)L_L(0)}{L_H(0) + L_L(0)}, \tag{3}$$

where L_H and L_L denote ARLs for upper- and lower-sided CUSUM, respectively and the "s" or "0" in the parentheses represents the initial value of the monitoring statistics. Note that the formula in Eq. (2) is just a special case of the formula in Eq. (3), if the initial value s is set to 0. [42] derived necessary and sufficient conditions for non-interaction of upper- and lower-sided schemes and also obtained two-sided ARL expression similar to Eq. (2) through Laplace transformation.

When $k^+ = k^-, h^+ = h^-$, [41] extended the Markov chain method of [22] directly to two-sided CUSUM control chart. The vector (S_n^+, s_n^-) is modeled as a two-dimensional Markov Chain. Suppose that the state is denoted by $E_{ij}, i, j = 0, 1, \ldots, t-1$. Note that E_{00} represents both of the statistics (S_n^+, s_n^-) are in initial states zero, and E_{ij} represents the statistic S_n^+ is in state i, while the other statistic s_n^- is in state j. All the absorbing states (i.e., $i = t$ or $j = t$) are labelled as one state. Then, the transition probability matrix has dimension $(t^2 + 1) \times (t^2 + 1)$.

From the discussions above, we can know that the difficulty in computing ARL for two-sided CUSUM lies in the large number of states, which will lead to tedious computation of the inverse matrix of the transition matrix. And what is worse, when the dimension of the transition probability matrix is so high that it is ill-conditioned, the inverse matrix does not exist. [43] studied the state space of two-sided CUSUM control chart $C^+(S_0^+, k^+, h^+) \cup C^-(s_0^-, k^-, h^-)$, and minimized the number of states included in the Markov chain in order to make the Markov chain methods as efficient as possible.

2.2. ***Integral Equation Method***

The idea of computing ARL through integral equation method was firstly proposed in EWMA control chart, in which the monitoring statistics are cumulated with exponential weight and naturally effective for upper- and lower-sided shifts simultaneously. From then on, the integral equation method also finds its application in CUSUM control charts.

2.2.1. Integral Equation Method: General Idea

For one-sided CUSUM control chart $C^+(S_0, k, h)$, let $f(x)$ and $F(x)$ be the probability density function (PDF) and cumulative distribution function (CDF) of the process variable X, $L(\mu)$ denote ARL when the CUSUM starts with initial value μ, and suppose the initial value of the monitoring statistics $S_0 = \mu$.

Note that, for the first observation X_1, one and only one of the following three events happens

$$\{X_1 - k + \mu \geq h\}, \quad \{X_1 - k + \mu \leq 0\}, \quad \{0 < X_1 - k + \mu < h\}.$$

Then

$$L(\mu) = 1 \cdot P\{X_1 \geq h+k-\mu\} + (1+L(0))P\{X_1 \leq k-\mu\} + \int_0^h (1+L(y))f(y+k-\mu)\mathrm{d}y,$$

i.e.,

$$L(\mu) = 1 + L(0)F(k-\mu) + \int_0^h L(y)f(y+k-\mu)\mathrm{d}y, \tag{4}$$

which is the integral equation for computing the ARL of one-sided CUSUM $C^+(S_0, k, h)$.

Although ARL is an important and widely employed criterion for evaluating control charts, it is known that the distribution of run length is skewed and has long right tail, so it may be necessary to have further information on the distribution of run length. Let $p(n, \mu)$ denote the probability of run length T equaling n with initial value μ.

When $T = 1$,

$$p(1, \mu) = 1 - F(h + k - \mu); \tag{5}$$

When $T = n = 2, 3, \ldots$

$$p(n, \mu) = p(n-1, 0)F(k-\mu) + \int_0^h p(n-1, x)f(x - \mu + k)\mathrm{d}x. \tag{6}$$

Moreover, for theoretically continuous variables, [8] showed that the following formula

$$p(n, 0) \simeq \frac{1}{L(0)} \exp\left\{-\frac{n-1}{L(0)}\right\}$$

should be a good approximation to the distribution of run length.

When the process variable X_i follows exponential distribution, the analytic solutions for the above integral equations are derived in [44].

2.2.2. *Integral Equation Methods for CUSUM with Estimated Parameters ([45])*

The performance of the CUSUM is generally evaluated with the assumption that the process parameters are known. In practice, the parameters are rarely known and are frequently replaced with estimates from an IC reference sample. In this case, it is with high probability that the monitoring statistics would be correlated, because the monitoring statistics are constructed based on the same historical reference data. [45] discussed the run length distribution of the CUSUM with estimated parameters and provided a method for approximating the distribution and moments of run length.

Assume the IC process parameters μ_0 and σ_0^2 are unknown and a sample of m subgroups of size n, $(X_{i1}, \ldots, X_{in})$, $i = 1, \ldots, m$, from an IC process is collected. Commonly used estimators for μ_0 and σ_0 are

$$\hat{\mu}_0 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}, \qquad \hat{\sigma}_0 = \frac{S_p}{c_{4,m}},$$

where $S_p = \sqrt{\frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{i\cdot})^2}$ and $c_{4,m} = \frac{\sqrt{2}\Gamma(\frac{m(n-1)+1}{2})}{\sqrt{m(n-1)}\Gamma(\frac{m(n-1)}{2})}$ to make $\hat{\sigma}_0$ an unbiased estimator of σ. Taking $C^+(S_0, k, h)$ for example, the monitoring statistics in [45] are

$$S_t = \max(0, S_{t-1} + y_t - k),$$

where $y_t = \frac{\bar{X}_t - \hat{\mu}_0}{\hat{\sigma}_0/\sqrt{n}}$ and $\bar{X}_t$ is the mean of the t^{th} subgroup of n process observations $(n \geq 1)$.

In order to derive the run length distribution of the CUSUM chart when parameters are estimated, it is helpful to rewrite y_t as

$$y_t = \frac{1}{W}(\gamma Z_t + \delta - \frac{Z_0}{\sqrt{m}}),$$

where $W = \frac{\hat{\sigma}_0}{\sigma_0}$ is a random variable, the distribution of which is related to the square root of a χ^2 random variable, representing the ratio of the estimator of the IC standard deviation to the IC standard deviation, $Z_0 = \sqrt{m}\frac{\hat{\mu}_0-\mu_0}{\sigma_0/\sqrt{n}}$ is a standard normal random variable, representing the standardized distance from the estimated IC mean to the true IC process mean, $Z_t = \frac{\sqrt{n}(\bar{X}_t-\mu)}{\sigma}$ is also a standard normal random variable, representing the standardized subgroup average at time t, where μ and σ are the respective mean and standard deviation of the process at time t, the constant $\gamma = \frac{\sigma}{\sigma_0}$ is the ratio of the standard deviation at time t to the true IC standard deviation and the constant $\delta = \frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}$ represents a standardized shift in the mean. Note that if the process is IC, then $\gamma = 1$ and $\delta = 0$.

With T denoting run length and by Eq. (5) and Eq. (6), it is shown that the conditional PDF of T is

$$P\{T=1|w,z_0,\gamma,\delta,\mu\} = 1-\Phi\left(\frac{w}{\gamma}[h-\mu+k]+\frac{\delta}{\gamma}-\frac{z_0}{\gamma\sqrt{m}}\right),$$

$$\begin{aligned}P\{T=t|w,z_0,\gamma,\delta,\mu\} &= P\{T=t-1|w,z_0,\gamma,\delta,0\}\Phi\left(\frac{w}{\gamma}[k-\mu]-\frac{\delta}{\gamma}+\frac{z_0}{\gamma\sqrt{m}}\right)\\ &+\frac{w}{\gamma}\int_0^h P\{T=t-1|w,z_0,\gamma,\delta,s\}\phi\left(\frac{w}{\gamma}[s-\mu+k]-\frac{\delta}{\gamma}+\frac{z_0}{\gamma\sqrt{m}}\right)ds,\end{aligned}$$

where Φ and ϕ are the CDF and PDF of standard normal random variables. Note that W and Z_0 are independent due to the independence of $\hat{\mu}_0$ and $\hat{\sigma}_0$. The marginal distribution of run length T can be given by

$$P\{T=t|\gamma,\delta,\mu\} = \int_{-\infty}^{\infty}\int_0^{\infty} P\{T=t|w,z_0,\gamma,\delta,\mu\}f_w(w)\phi(z_0)dwdz_0,$$

where $f_w(w)$ is the PDF of W.

Then ARL (the first moment of T) can be computed as follows

$$
\begin{aligned}
E(T|\gamma,\delta,\mu) &= \sum_{t=1}^{\infty} tP\{T=t|\gamma,\delta,\mu\} \\
&= \int_{-\infty}^{\infty}\int_{0}^{\infty} M(w,z_0,\gamma,\delta,\mu) f_w(w)\phi(z_0)dwdz_0, \qquad (7)
\end{aligned}
$$

where $M(w,z_0,\gamma,\delta,\mu)$ is the ARL conditioned on particular observations of W and Z_0. Similarly, the second moment of T is given by

$$
E(T^2|\gamma,\delta,\mu) = \int_{-\infty}^{+\infty}\int_{0}^{+\infty} M_2(\omega,z_0,\gamma,\delta,\mu) f_\omega(\omega)\phi(z_0)d\omega dz_0, \qquad (8)
$$

where $M_2(w,z_0,\gamma,\delta,\mu)$ is the second moment of T conditioned on particular values of W and Z_0.

By systematically choosing the particular values of the random variables W and Z_0 to be the appropriately scaled abscissae, the double integrals in the equations above can be approximated by Gaussian quadrature or be solved by the method of [46], which is summarized in the Appendix of this paper.

3. Methods for EWMA Control Charts

Generally speaking, the ARL computing procedures by Markov chain approach and integral equation approach for EWMA control charts are quite similar to those for CUSUM control charts. One point practitioners should keep in mind is that the EWMA is naturally effective for both upper- and lower-sided shifts simultaneously, while CUSUM has to be separately designed for one- and two-sided shifts.

3.1. *Markov Chain Method*

3.1.1. Markov Chain Method for EWMA

With X_t being the observed process variable, the monitoring statistic of EWMA control chart is

$$
S_t = (1-\lambda)S_{t-1} + \lambda X_t, \qquad (9)
$$

where $\lambda \in (0,1]$ is the smoothing parameter ([3]).

From the definition of the monitoring statistic of EWMA control chart in Eq. (9), the EWMA is naturally effective for both upper- and lower-sided shifts simultaneously. [47, 48] divided the interval between the upper and lower control limits into subintervals with equal length ω, and labelled the two absorbing states (i.e., the one beyond upper control limit and the one beyond lower control limit) as one state. Then the transition probability p_{ij} can be obtained as follows,

$$
\begin{aligned}
p_{ij} &= P\{S_{t+1} \in I_j | S_t \in I_i\} \\
&= P\{j\omega - \omega/2 < (1-\lambda)S_t + \lambda X_{t+1} \le j\omega + \omega/2 | S_t = i\omega (\text{the centre point of interval} \quad I_i)\} \\
&= P\{j\omega - \omega/2 < (1-\lambda)i\omega + \lambda X_{t+1} \le j\omega + \omega/2\},
\end{aligned}
$$

from which the ARL can be computed by Eq. (1) if the distribution of X_t is given.

3.1.2. Computation of ARL for Adaptive EWMA

Note, from the definition in Eq. (9), that the Shewhart control chart ([1]) is a special case of EWMA control chart with $\lambda = 1$. It is known that ([47]) the smoothing parameter has great effect on EWMA control chart for different shift sizes, i.e., the EWMA control chart with small λ can quickly detect small to moderate shifts, while the EWMA control chart with large λ can signal quickly for large shifts, and a single EWMA scheme cannot have a "nearly minimum" ARL for both small and large shifts simultaneously. Due to the reasons that it is generally difficult to determine the unknown shift size in practice and it is usually necessary for practitioners to detect a range of shifts, it is badly in need of a control scheme that can achieve good performance for a wider range of shifts. [49] proposed adaptive EWMA (AEWMA) control chart that essentially tries to combine an EWMA and a Shewhart chart in a smooth way. The underlying idea is to adapt the weight of the past observations to detect, in a more balanced way, shifts of different sizes.

The monitoring statistic in [49] is

$$S_t = (1 - \varpi(e_t))S_{t-1} + \varpi(e_t)X_t,\ S_0 = \mu$$

where $e_t = X_t - S_{t-1}$, $\varpi(e_t) = \frac{\phi(e_t)}{e_t}$ as a varying weight and $\phi(e)$ is a score function. It can be seen that Shewhart and EWMA charts are special cases of AEWMA chart if we set $\varpi(e_t) = e_t$ or $\varpi(e_t) = \lambda e_t$. [49] considered the following three score functions:

$$\phi_{hu}(e) = \begin{cases} e + (1-\lambda)k, \text{if} & e < -k, \\ e - (1-\lambda)k, \text{if} & e > k, \\ \lambda e, & \text{otherwise.} \end{cases}$$

$$\phi_{bs}(e) = \begin{cases} e[1 - (1-\lambda)(1 - (e/k)^2)^2], \text{if} & |e| \le k, \\ e, & \text{otherwise.} \end{cases}$$

$$\phi_{cb}(e) = \begin{cases} e, & \text{if} \quad e \le -p_1, \\ -\widetilde{\phi}_{cb}(-e), & \text{if} \quad -p_1 < e < -p_0, \\ \widetilde{\phi}_{cb}(e), & \text{if} \quad p_0 < e < p_1, \\ e, & \text{if} \quad e \ge p_1, \\ \lambda e, & \text{otherwise,} \end{cases}$$

where $0 < \lambda \le 1$, $k \ge 0$, $0 \le p_0 < p_1$ $\widetilde{\phi}_{cb}(e) = \lambda e + (1 - \lambda)(\frac{e-p_0}{p_1-p_0})^2 \left[2p_1 + p_0 - (p_0 + p_1)(\frac{e-p_0}{p_1-p_0})\right]$.

The ARL of an AEWMA scheme can be approximated using the approach described by [47] for an EWMA control chart; that is, by discretizing the infinite-state transition probability matrix of the continuous-state Markov chain with the corresponding transition probabilities

$$\begin{aligned} p_{ij} &= P\{(1 - \varpi(e_t))S_{t-1} + \varpi(e_t)X_t \in I_j | S_{t-1} \in I_i\} \\ &= P\{S_{t-1} + \phi(X_t - i\omega) \in I_j | S_{t-1} \in I_i\} \\ &= P\{j\omega - i\omega - \frac{\omega}{2} < \phi(X_t - i\omega) \le j\omega - i\omega + \frac{\omega}{2}\} \\ &= P\{i\omega + \phi^{-1}(j\omega - i\omega - \frac{\omega}{2}) < X_t \le i\omega + \phi^{-1}(j\omega - i\omega + \frac{\omega}{2})\}, \end{aligned}$$

if $\phi(e)$ is not increasing in e. For the functions $\phi_{hu}(e)$, $\phi_{bs}(e)$ and $\phi_{cb}(e)$ defined above, the inverse functions are given respectively by

$$\phi_{hu}^{-1}(v) = \begin{cases} v - (1-\lambda)k, & \text{if} \quad v < -\lambda k, \\ v + (1-\lambda)k, & \text{if} \quad v > \lambda k, \\ v/\lambda, & \text{otherwise.} \end{cases}$$

$$\phi_{bs}^{-1}(v) = \begin{cases} \phi_{bs}^{*}(v), & \text{if} \quad |v| \le k, \\ v, & \text{otherwise,} \end{cases}$$

where $\phi_{bs}^{*}(v)$ is the unique real root, with absolute value less than k, of the polynomial $y - (1-\lambda)y(1-(y/k)^2)^2] - v$ and

$$\phi_{cb}^{-1}(v) = \begin{cases} v, & \text{if} \quad v \le -p_1, \\ -\phi_{cb}^{*}(-v), & \text{if} \quad -p_1 < v < -\lambda p_0, \\ \phi_{cb}^{*}(v), & \text{if} \quad \lambda p_0 < v < p_1, \\ v, & \text{if} \quad v \ge p_1, \\ v/\lambda, & \text{otherwise,} \end{cases}$$

where $\phi_{cb}^{*}(v)$ is the unique root of $\widetilde{\phi}_{cb}(y) - v$ in the interval (p_0, p_1). The ARL can be computed by Eq. (1).

From the definition of the monitoring statistic of AEWMA, it can detect, in a more balanced way, shifts of different sizes, i.e., the shift size is not a constant step shift but in an interval, say (μ_1, μ_2). For interval shifts, [50] proposed dual CUSUM control chart. For AEWMA, more design parameters are necessary besides the smoothing parameter λ and the control limit h. For example, if $\phi_{hu}(.)$ or $\phi_{bs}(.)$ is used, then the parameters are $\theta = (\lambda, h, k)$. To avoid the flaw that the ARL of a chart designed for a small shift is quite different from that designed for a large shift, [49] devised a design strategy, in which the minimizing of the penalized function involved can be worked out through simulated annealing algorithm ([51]).

Recently, [52] introduced an integral equation technique for evaluating the ARL of the AEWMA chart when the parameters are unknown. They considered six most commonly used estimators of the process standard deviation, which are, following the notations in Section 2.2.2,

$$\hat{\sigma}_1 = \frac{\bar{S}}{c_4(n)}, \hat{\sigma}_2 = c_4(n)\bar{S}, \hat{\sigma}_3 = \frac{\bar{R}}{d_2(n)}, \hat{\sigma}_4 = c_4(\nu+1)S_p, \hat{\sigma}_5 = S_p, \hat{\sigma}_6 = \frac{S_p}{c_4(\nu+1)},$$

where $\nu = m(n-1)$, $c_4(\cdot)$ and $d_2(\cdot)$ are "control chart constants" tabulated in [4]. Then the Markov Chain approach of [52] can be obtained by replacing X_t of [49] by y_t. Noting that the performance of the AEWMA is conditioned on specific estimates of $\hat{\mu}_0$ and $\hat{\sigma}_0$, [52] considered marginal ARL as in Eq. (7). They, moreover, studied the PDF of W $f_w(w)$ in detail.

3.1.3. Methods for Two-sided EWMA with Reflecting Boundaries

It is known that [41] recommended the use of CUSUM charts with FIR in a situation where a manufacturing process usually begins in an OC state, so that an OC signal will be issued earlier. Similarly, [53] evaluated the run length properties of one-sided EWMA with reflecting boundaries. A reflecting boundary can prevent the monitoring statistics of EWMA from drifting to one side indefinitely.

The monitoring statistics of the upper- and lower-sided EWMA with reflecting boundaries are

$$S_t^+ = \max\{A, (1-\lambda_+)S_{t-1}^+ + \lambda_+ X_t\}, \quad S_t^- = \min\{B, (1-\lambda_-)S_{t-1}^- + \lambda_- X_t\},$$

where A and B are two reflecting boundaries, $S_0^+ = u, A \le u < h_+$ and $S_0^- = v, h_- < v \le B$. An OC signal is issued when $S_t^+ \ge h_+$ or $S_t^- \le h_-$.

[54] showed numerically that under the condition of non-interaction, the ARL of two-sided EWMA schemes with reflecting boundaries can be obtained from the ARL of the upper-sided and lower-sided schemes by Eq. (2). The fact that the upper-sided and lower-sided EWMA schemes with reflecting boundaries do not interact implies whenever the monitoring statistics of one of the upper-sided and lower-sided schemes issue a signal, the monitoring statistics of the other one will reset to the corresponding reflecting boundary. Without loss of generality, assuming $\lambda_+ = \lambda_- = \lambda$, $B \le h_+$ and $A \le h_-$, [55] derived the necessary and sufficient conditions for non-interaction of the upper- and lower-sided EWMA with reflecting boundaries, and showed that similar to computing the ARL of CUSUM control chart with FIR ([41]), the ARL of a pair of one-sided EWMA control charts with reflecting boundaries can be computed by Eq. (3).

3.2. *Integral Equation Method*

3.2.1. Integral Equation Method for EWMA

With the monitoring statistics defined in Eq. (9), [23] presented a numerical procedure using integral equations for the tabulation of moments of run lengths of EWMA control chart. In a similar way to the derivation of the integral equations in Eq. (4) for CUSUM, the ARL when the EWMA starts with initial value μ, denoted as $L(\mu)$, satisfies the following integral equation

$$\begin{aligned} L(\mu) &= 1 \cdot \Pr(|(1-\lambda)\mu + \lambda X_1| > h) + \int_{|(1-\lambda)\mu+\lambda y| \le h} [1 + L((1-\lambda)\mu + \lambda y)] f(y) dy \\ &= 1 + \frac{1}{\lambda}\int_{-h}^{h} L(y) f(\frac{y-(1-\lambda)\mu}{\lambda}) dy, \end{aligned}$$

where $f(.)$ is the PDF of observed variable. This integral equation for $L(.)$ is a Fredholm integral equation of the second kind, and can be approximated numerically.

Furthermore, the probability $p(n,\mu)$, denoting the probability of run length equals n given that the EWMA starts at initial value μ, satisfies the following integral equation

$$\begin{aligned} p(n,\mu) &= \int_{\{|(1-\lambda)\mu+\lambda y| \le h\}} p(n-1, (1-\lambda)\mu + \lambda y) \cdot f(y) dy \\ &= \frac{1}{\lambda}\int_{-h}^{h} p(n-1, y) f(\frac{y-(1-\lambda)\mu}{\lambda}) dy. \end{aligned}$$

3.2.2. Integral Equation Methods for EWMA with Estimated Parameters

[56] derived run length distribution of the EWMA chart with estimated parameters and [57] relaxed the assumption of known parameters and developed accordingly design procedures for the EWMA chart. The general idea and procedure for

EWMA control chart with estimated parameters are nearly the same as those for CUSUM control charts introduced in Section 2.2.2, only bearing in mind that the differences in the form of monitoring statistics and conditional transition probabilities.

4. Methods for Combined Control Charts

Although the computation of ARL of Shewhart control chart is comparatively easy due to the independence of the monitoring statistics, it is known that Shewhart chart is only effective for detecting large process shifts. In order to improve the performance of Shewhart chart for small to moderate shifts, Shewhart control chart with supplementary runs rules using historical data was proposed in the literature. In this section, we will review the Markov chain approach of [58] to model supplementary runs rules used with Shewhart control charts, and some methods for combined Shewhart and CUSUM or EWMA control charts.

4.1. *Markov Chain Method for Shewhart Control Charts with Supplementary Runs Rules*

The reason why Shewhart control chart is not as sensitive in detecting small to moderate shifts is that it only makes use of the information of the current sample and ignores all the information that the past samples can provide. [58] proposed Shewhart control chart with supplementary runs rules $T(k,m,a,b)$, which signals if k of the last m standardized sample means fall in the interval $(a,b), a<b$. Thus the usual Shewhart chart is denoted by $\{(T(1,1,-\infty,-3),T(1,1,3,\infty)\}$, and the Shewhart chart with supplementary runs rules is given by a larger collection of rules of the form $T(k,m,a,b)$. Various combinations of the following rules can be considered and be combined to form most of the control charts suggested in the literature.

The difficulty in computing ARL with Markov chain is to find a procedure that yields the minimal number of states required to represent the charts with Markov chain. Then the nonzero elements of the transition probability matrix can be determined and used to construct recursive equations for obtaining the run length probabilities. [58] gave a simple and efficient method, using Markov chains, to obtain the exact run length properties of Shewhart control charts with supplementary runs rules.

For the rule $T(k_i,m_i,a_i,b_i), m_i>1$, define the vectors

$$\boldsymbol{W}_i' = (W_{i,1},\ldots,W_{i,m_i-1}),$$

where

$$W_{i,j} = \begin{cases} 1, \text{ if the } j\text{th previous observation was in } (a_i,b_i), \\ 0, \text{ otherwise.} \end{cases}$$

It seems that $\boldsymbol{W}_i$ can be used to define the state space. Note, however, that if $\sum_{j=1}^{m_i-2} W_{i,j} < k_i - 1$, and even if the next sample falls into (a_i,b_i), the Shewhart control chart with rule $T(k_i,m_i,a_i,b_i)$ will not issue a signal. To overcome this limitation, define

$$\boldsymbol{X}_i' = (X_{i,1},\ldots,X_{i,m_i-1}),$$

where

$$X_{i,j} = \begin{cases} W_{i,j}, \text{ if } \sum_{l=1}^{j}(1 - W_{i,l}) < m_i - k_i + 1, \\ 0, \quad \text{otherwise.} \end{cases}$$

[58] showed that

> The vector $\boldsymbol{X}'_i$ indicates by 1s only those observations falling in (a_i, b_i) that may contribute to an OC signal. Thus a transient state of a chart using t rules can be represented by $\boldsymbol{S}' = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_t)$, where the subvector $\boldsymbol{X}'_i$ is defined as previously for the rule $T(k_i, m_i, a_i, b_i), i = 1, 2, \ldots, t$. Moreover, to reduce the length of the vector $\boldsymbol{S}'$, it is often convenient to replace the subvector associated with any rule $T(m, m, a, b), m > 1$, with a counter l, where l can take the values $0, 1, ..., m-1$. The number l is the number of consecutive points in the interval (a, b) since the last point not in (a, b).

It follows to compute the transition probability after the state space is determined, and then the ARL can be computed using Markov chain method by Eq. (1).

4.2. *Methods for Combined Shewhart-CUSUM or Shewhart-EWMA*

[59] proposed combined Shewhart-CUSUM control chart by adding Shewhart control limits to a CUSUM control chart and showed that the combined scheme can give an improved ARL curve, because it can be designed to detect more quickly large shifts of the mean with only small changes in the speed of detecting small to moderate shifts or in the IC ARL. Note that the only change in the computation procedure for a combined Shewhart-CUSUM control chart from a standard CUSUM control chart is the insertion of Shewhart control limits. For observations greater than the Shewhart control limits, the transition probability goes to the absorbing state, while for a standard CUSUM control chart some of this probability would be distributed over transition states with higher CUSUM values.

[47] proposed a combined Shewhart-EWMA control chart that provides protection against both large and small shifts in a process and showed that properties of the combined Shewhart-EWMA can be obtained by modifying the transition probability matrix for an EWMA control chart. The modified one-step transition probabilities are given by

$$p_{ij} = P\left\{\min[SCL_U, \max(SCL_L, X_L)] < X_n \leq \max[SCL_L, \min(SCL_U, X_U)]\right\},$$

where

$$X_L = \frac{1}{\lambda}[(j\omega - \omega/2) - (1-\lambda)i\omega], \quad X_U = \frac{1}{\lambda}[(j\omega + \omega/2) - (1-\lambda)i\omega],$$

$i, j = -t, -t{+}1, \cdots, t$, $2t{+}1$ is number of states in Markov chain, SCL_U and SCL_L represent the upper and lower Shewhart control limits, respectively. Similarly, for Shewhart with upper-sided CUSUM $C^+(S_0, k, h)$, X_L and X_U need to be changed into

$$X_L = j\omega - \omega/2 - i\omega + k, \;\; X_U = j\omega + \omega/2 - i\omega + k.$$

Then the one-step transition probability is

$$\mathbf{P} = \begin{pmatrix} \mathbf{R}^* & \mathbf{p}^* \\ \mathbf{0}^T & 1 \end{pmatrix},$$

where $\mathbf{p}^* = (\mathbf{I} - \mathbf{R}^*)\mathbf{1}$.

It is useful to consider a robust CUSUM or EWMA that can provide protection against occasional outliers in the data that might otherwise cause an OC signal. The combined Shewhart-CUSUM or Shewhart-EWMA should not be used in these situations because the addition of Shewhart limits causes the combined Shewhart-CUSUM or Shewhart-EWMA to be sensitive to the occurrence of outliers. [60] and [47] proposed robust CUSUM and robust EWMA, respectively. They considered the two-in-a-row rule, i.e., a single observation outside of the outlier limits (e.g., $\pm 4\sigma$) does not enter the CUSUM or EWMA, but two outliers in a row are considered to be an OC signal.

In [47],

> The run length properties of a robust CUSUM or EWMA using a two-in-a-row rule can be obtained by modifying the transition probability matrix of a combined Shewhart-CUSUM or Shewhart-EWMA. The IC transition probabilities will be identical to the IC transition probabilities for the combined Shewhart-CUSUM or Shewhart-EWMA, with the Shewhart limits representing outlier limits. When a single outlier is observed, the control statistic remains in the same state and a counter is set. If the next observation lies within the outlier limits, the counter is reset to 0; otherwise an OC signal is given.

The two-in-a-row rule requires a transition matrix which is twice the size of the transition matrix for the previous rules. The doubling of the size of the transition matrix is necessitated by an indicator, which keeps track of whether the previous observation was a suspected outlier or not.

[61] gave out a unified Markov chain approach for computing the run length distribution in control charts with simple or compound rules. [62] proposed combined CUSUM and Shewhart variance chart that provides simultaneous control of the process mean and process variance. Exact expressions for the moments of the combined CUSUM and Shewhart variance chart are given as integral equations. These integral equations can be solved by using numerical methods for the computation of ARL and higher moments of the combined CUSUM chart and Shewhart variance chart. [63] numerically compared the joint $\bar{X}$-S^2, two-sided CUSUM-S^2 and EWMA-S^2 control charts when process parameters are known and process parameters are estimated from retrospective data, respectively. In both cases, equations for the conditional and unconditional run length distributions are developed, and expressions for the ARL, the second moment of the run length (SMRL), and the standard deviation of the run lengths (SDRL) are derived for these charts.

5. Methods for Adaptive Control Charts

Extensive research in recent years has developed VSR control charts that vary the sampling rate as a function of current and prior sample results since [64] proposed $\bar{X}$ charts with VSI. [65] showed that one important area of SPC research continues to be the use of control charts with VSS and/or VSI. The widely used performance indicators for adaptive control charts are the ATS, AATS (SSATS) and ANSS. Noting that the computation methods for these indicators are similar to those for ARL, we will just give a brief review on these methods in this section.

5.1. *Methods for Adaptive Shewhart Control Charts*

The sampling scheme of adaptive control charts is to use a longer sampling interval or a smaller sample size as long as the monitoring statistic is close to the target so that there is no indication of process shifts. However, if the monitoring statistic is far from the target, but still within the control limits so that there is some indication of a process shift, then a shorter sampling interval or a larger sample size is used. If the monitoring statistic falls out of the control limits, then the process is considered to be OC.

[9] proposed a combined $\overline{X}$ control scheme with VSSI and [66] proposed $\overline{X}$ control chart with variable parameters (VP). The monitoring statistics are

$$Y_i = \frac{\overline{X}_i - \mu_0}{\sigma/\sqrt{n(Y_{i-1})}},$$

where $n(Y_{i-1})$ is the sample size of the ith sample determined by Y_{i-1}. Denoting the upper control limit by UCL, the lower control limit by LCL and the warning limit by w, the sampling schemes for the next sample are

$$(d(y), n(y)) = \begin{cases} (d_1, n_1), & \text{if} \quad Y_i \in [-w, w], \\ (d_2, n_2), & \text{if} \quad Y_i \in (LCL, -w) \cup (w, UCL), \end{cases}$$

where $d_2 < d < d_1$ and $n_1 < n < n_2$ with d and n being the sampling interval and sample size for an FSR sample scheme.

The ATS and ANSS can be computed by the Markov chain approach for computing ARL ([22]), only with minor modifications. For control charts with VSI, it is shown that

$$ATS = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{d}, \quad ANSS = n(\mathbf{I} - \mathbf{R})^{-1}\mathbf{1},$$

where $\mathbf{d}$ is a vector consisted of sampling intervals corresponding to different states. For control charts with VSS, it is shown that

$$ATS = ARL = d(\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}, \quad ANSS = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{n},$$

where $\mathbf{n}$ is a vector consisted of sample sizes corresponding to different states. For control charts with VSSI, it is shown that

$$ATS = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{d}, \quad ANSS = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{n}.$$

5.2. *Methods for Adaptive CUSUM Control Charts with VSI*

[13] used Markov chain approach to evaluate properties such as the ATS and the ANSS and they derived

$$\mathbf{ANSS} = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{1} = \mathbf{M1}, \quad \mathbf{ATS} = \mathbf{Md}.$$

Recently, [67] proposed a new CUSUM control chart, which is based on both adaptive and VSI features, and developed a two-dimensional Markov chain model to evaluate its runtime performance. The monitoring statistics are

$$\begin{cases} S_0 = 0\,, \\ S_i = \max\left(\,0, S_{i-1}\,\right) + (X_i - k_i)/h(k_i), \\ k_i = \hat{\delta}_i/2\,, \end{cases}$$

where

$$\hat{\delta}_i = \max\left\{\,\delta_{\min}, (1-\lambda)\hat{\delta}_{i-1} + \lambda X_i\,\right\}\,,$$

$0 < \lambda \leq 1$ is the smoothing parameter and $\delta_{\min} > 0$ is a minimum magnitude of interest for early detection. The sampling scheme is

$$d(.) = \begin{cases} d_1, \text{ if } S_i < w\,, \\ d_2, \text{ if } w \leq S_i < h. \end{cases}$$

With a similar approach to [31], [67] modeled $(S_i, \hat{\delta}_i)^T$ as a two-dimensional Markov chain. Therefore, the proposed adaptive CUSUM with VSI can be viewed as an two-dimensional Markovian process with transition matrix $\mathbf{R}$, which is an $N \times N$ matrix

$$\mathbf{R}_{[(i-1)m_\delta + j,\,(m-1)m_\delta + n]} = f_{(i,j),(m,n)},$$

where $f_{(\,i,j\,),(\,m,n\,)}$ is the transition probability of $(\,S, \hat{\delta}\,)$ from state $(\,i,j\,)$ to state $(\,m,n\,)$ and m_δ is the number of states for $\hat{\delta}_i$.

5.3. *Methods for Control Charts with VSI at Fixed Time (VSIFT)*

Although control chart with VSI is able to detect most process changes substantially faster than that with fixed sampling interval. One disadvantage of control chart with VSI is that the prediction of time at which samples will be taken can not be done for more than the next sample. [16, 68] considered a modification of the VSI idea in which samples are always taken at specified, equally-spaced, fixed-time points with additional samples allowed between these fixed times when indicated by the process observations.

Specifically, samples will be taken using the sampling interval d_F as long as there is no indication of a problem with the process. However, if there is some indication of a problem with the process, the additional samples are allowed between the fixed times. Suppose that the interval d_F between two fixed times is divided into η subintervals of length d_1 such that the possible sampling times within d_F are $d_1, 2d_1, \ldots, (\eta-1)d_1$.

Take EWMA control chart for example. Supposing the first sample is sampled at time $t_1 = md_1$, $S_1 = s$, $S_0 = s_0$ and denoting $A(s,m)$ the average time from t_1 to signal and $f(s'|s)$ the transition probability of S from $s \in C$ to $s' \in C$, the ATS satisfies the following integral equation

$$ATS = t_1 + \int_{s \in C} A(s,m) f(s|s_0) ds,$$

where $A(s,m)$ satisfies

$$A(s,m) = d_1 + \int_{s' \in C} A(s', m+1) f(s'|s) ds', \quad s \in R_1, \; m = 0, 1, \ldots, \eta - 2.$$

For $s \in R_1$ (the warning region), $m = \eta - 1$ and

$$A(s, \eta - 1) = d_1 + \int_{s' \in C} A(s', 0) f(s'|s) ds'.$$

For $s \in R_2$ (the central region), $m = 0, 1, \ldots, \eta - 1$ and

$$A(s,m) = (\eta - m) d_1 + \int_{s' \in C} A(s', 0) f(s'|s) ds'.$$

6. Methods for Steady State ARL and Steady State ATS

In the previous sections, the process shift is assumed to occur as soon as the monitoring begins. However, in most practical cases, it is never known when the shift occurs. Note that control schemes are used to monitor a process and a shift often occurs after the process has been operating for some time. The monitoring statistics may not be zero when the shift occurs; in fact it has a distribution over its possible values (called its steady state distribution). [69] proposed the concept of SSARL, i.e. the weighted average of ARLs given the initial values of the monitoring statistics, using the steady state distribution of monitoring statistics values as the weights.

6.1. *Methods for Steady State ARL and Steady State ATS*

[47] derived a computing procedure for SSARL for EWMA control charts. [47] showed

> that an exact steady state probability vector does not exist because the transition probability matrix is not ergodic. The steady state probability vector that best models the way control schemes are used is a cyclical steady state probability vector that is obtained by altering the transition probability matrix so that the control statistic is reset to state 0 whenever it goes into the OC state; that is
>
> $$\mathbf{P}^* = \begin{pmatrix} \mathbf{R} & (\mathbf{I} - \mathbf{R})\mathbf{1} \\ 0 \cdots 1 \cdots 0 & 0 \end{pmatrix}.$$
>
> This transition probability matrix is ergodic. The steady state probability vector, $\mathbf{p}_{ss}$, is found by solving $\mathbf{p} = (\mathbf{P}^*)^T \mathbf{p}$ subject to $\mathbf{1}^T \mathbf{p} = 1$. Then, $\mathbf{p}_{ss} = (\mathbf{1}^T \mathbf{q})^{-1} \mathbf{q}$, where $\mathbf{q}$ is a vector of length t obtained from $\mathbf{p}$ by deleting the entry corresponding to the absorbing state; that is, $\mathbf{p}_{ss}$, is the probability vector obtained from $\mathbf{p}$ by deleting the entry corresponding to the absorbing state and normalizing so that the probabilities sum to 1.

Then the SSARL can be computed by

$$SSARL = \mathbf{p}_{ss}^T (\mathbf{I} - \mathbf{R})^{-1} \mathbf{1}.$$

In computing the SSATS, the following assumptions are usually made ([13]).

(i) The probability of the shift falling in an interval of a particular length is proportional to the product of this interval length and the conditional stationary probability of using this interval given that there is no signal.
(ii) When the shift occurs within an interval, the position of the shift within the interval is uniformly distributed over the interval.

6.2. *Methods for Control Chart with VSSIFT for Correlated Data*

[70] proposed "variable sampling at fixed times" (VSFT), which includes both VSIFT and VSSFT schemes, using a Markov chain model and integral equations for the autocorrelated process.

The monitoring statistics in [70] are $\bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ki}$, where the ith observation at sampling point t_k, X_{ki}, can be represented as

$$X_{ki} = \mu_k + \zeta_{ki} \quad k = 1, 2, \cdots,$$

where μ_k is the random process mean at sampling time k and ζ_{ki}'s are independent normal random variables with mean 0 and variance σ_ζ^2. For the case in which $0 < \phi < 1$, it is assumed that if the process mean is constant and samples $k-1$ and k are $t > 0$ units apart, then μ_k can be expressed in terms of μ_{k-1} as

$$\mu_k = (1-\phi^t)\xi + \phi^t \mu_{k-1} + \alpha_k \quad k = 1, 2, \ldots,$$

where ξ is the process mean, α_k is a random shock and assumed to be independent normal random variable with mean 0 and variance σ_α^2. It is assumed that the starting value μ_0 follows a normal distribution with mean ξ and variance $\sigma_\mu^2 = \sigma_\alpha^2/(1-\phi^2)$.

In order to construct the VSFT control scheme for the autocorrelated AR(1) plus error model, $\bar{X}_k$ is standardized to

$$Z_k = \frac{\bar{X}_k - \xi_0}{\sqrt{\sigma_\mu^2 + \sigma_\zeta^2/n(k)}},$$

where $n(k)$ is the sample size of the kth sample and ξ_0 is the IC process mean. Denoting $\psi = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{1}{n_0}\sigma_\zeta^2}$ as a measure index which indicates the proportion of the variance σ_X^2 that is due to the AR(1) process, where n_0 is the average IC sample size. Then Z_k can be rewritten as

$$Z_k = \frac{\bar{X}_k - \xi_0}{\sigma_\zeta \sqrt{\frac{1}{n_0}\left(\frac{\psi}{1-\psi} + \frac{n_0}{n(k)}\right)}}.$$

Let Y_k be the state at sample point k for $k = 1, 2, \ldots$ and the state transition probability

$$p(j|i,v) = Pr(Y_{k+1} = j | \mu_{k+1} = v, Y_k = i).$$

Table AI of [70] showed the state transition probability $p(j|i,v)$ given $\mu_{k+1} = v$ without the elements associated with the absorbing state.

Let $\sigma_i^2 = \psi(1-\phi^{2l_i})\sigma_X^2$ be the variance of α_{k+1}, where $l_i, i = 1, 2, \cdots, \eta$ is the interval between t_k and t_{k+1} which is determined by Y_k. Define the transition density $f(v|u,i)$ at $\mu_{k+1} = v$ given that $\mu_k = u$ and $Y_k = i$ as follows

$$f(v|u,i) = f_N\Big[\frac{v-\xi}{\sigma_i} - \phi^{l_i}\Big(\frac{u-\xi}{\sigma_i}\Big)\Big]\frac{1}{\sigma_i},$$

where $f_N(\cdot)$ is the standard normal PDF. Then the joint transition density at $\mu_{k+1} = v$ and $y_{k+1} = j$ given that $\mu_k = u$ and $Y_k = i$ can be expressed as $p(j|i,v)f(v|u,i)$.

For any given sampling time t_k, let $A(u,i)$ be the expected time from t_k to the time that the chart signals, given that $\mu_k = u$ and $Y_k = i$, where $i = 1, 2, \ldots, 2\eta$. It follows that $A(u,i)$ satisfies the integral equation

$$A(u,i) = l_i + \sum_{j=1}^{2\eta} \int_{-\infty}^{\infty} A(v,j)p(j|i,v)f(v|u,i)dv.$$

Thus, the ATS is

$$\text{ATS} = t_1 + \sum_{j=1}^{2\eta} \int_{-\infty}^{\infty} A(v,j)p(j|1,v)f_1(v)dv,$$

where $f_1(v)$ is the marginal density of μ_1.

7. Conclusions and Extensions

Control charts are widely used in industries to monitor a process for quality improvement and play an important role in the area of SPC. An inevitable problem for a practitioner is to choose one from many possible control charts. In the literature, ARL and ATS are the key indexes for comparing the performance of static and adaptive control charts, respectively.

In this paper, we review the Markov chain and integral equation methods for computing ARL, ATS and the run length distribution for CUSUM, EWMA, and combined control charts, respectively. [30] compared Markov chain approach and integral equation approach and showed that these two methods are actually equivalent when used with conventional CUSUM and EWMA. Some references have been given in this paper that should be helpful to those interested in control charts or in their practical applications.

From the review in this paper, it seems that the state space of the monitoring statistics should be partitioned into intervals ([2]), rectangles ([31]) or concentric spherical shells ([71]), it is, however, not necessarily so. Interested readers are referred to [72, 73].

Although this paper is the result of an extensive literature review of the most recent developments in the area of computation of ARL, ATS and related indexes, there are still some aspects not involved. First, we only consider computation of ATS, AATS and ANSS for adaptive control charts. Other comparison indicators, such as average number of switches (ANSW) proposed in [74], are not reviewed due to the limited application. Second, besides Markov chain approach and integral equation approach, there are some numerical methods ([26, 75]), which we believe are still effective complements to Markov chain approach and integral equation approach. Third, we focus on univariate process monitoring, [76–78], however, pointed out that multivariate control charts are one of the most rapidly developing areas of SPC and suggested that basic and applied research is needed on methods for monitoring multiple parameters that arise in models for the cases of single or multiple process variables. For the IC situation, [79] approximated the multivariate CUSUM (MCUSUM) ([80]) by using a discrete Markov chain model. [71] extended the advantages of the Markov chain approximation to multivariate EWMA (MEWMA) ([81]) and showed that this analysis can be applied whenever the multivariate control statistic can be modeled as a Markov chain and the run length performance depends on the off-target mean only through the noncentrality parameter.

Acknowledgements

The authors are grateful to the Editor and the anonymous referee for their valuable comments that have improved this paper vastly. This paper is supported by the National Natural Science Foundation of China Grants 11001138, 11071128, 11131002, 11101306, 11201246, the RFDP of China Grant 20110031110002 and the Fundamental Research Funds for the Central Universities 65010731 and 65012231.

References

[1] W.A. Shewhart, *The Application of Statistics as an Aid in Maintaining Quality of a Manufactured Product*, J. Amer. Statist. Assoc. 20 (1925), pp. 546-548.
[2] E.S. Page, *Continuous Inspection Schemes*, Biometrika. 41 (1954), pp. 100-115.
[3] S.W. Robert, *Control Chart Test Based on Geometric Moving Averages*, Technometrics. 1 (1959), pp. 239-250.
[4] D.C. Montgomery, *Introduction to Statistical Quality Control*, 6th ed, John Wiley & Sons, New York, 2009.
[5] J.S. Oakland, *Statistical Process Control*, 5th ed, Butterworth-Heinemann, Oxford, 2003.
[6] T.P. Ryan, *Statistical Methods for Quality Improvement*, 2nd ed, John Wiley & Sons, New York, 2000.
[7] D.M. Hawkins and D.H. Owell, *Cumulative Sum Charts and Charting for Quality Improvement*, Springer-Verlag, New York, 1998.
[8] G.B. Wetherill and D.W. Brown, *Statistical Process Control*, Chapman and Hall, 1991.
[9] S.S. Prabhu, D.C. Montgomery and G.C. Runger, *A Combined Adaptive Sample Size and Sampling Interval $\overline{X}$ Control Schemes*, J. Qual. Tech. 26:3 (1994), pp. 164-176.
[10] A.F.B. Costa, *Joint $\overline{X}$ and R Charts with Variable Parameters*, IIE Trans. 30 (1998), pp. 505-514.
[11] Y.K. Chen and K.C. Chiou, *Optimal Design of VSI $\overline{X}$ Control Charts for Monitoring Correlated Samples*, Qual. Rel. Eng. Int. 21 (2005), pp. 757-768.
[12] G. Celano, A.F.B. Costa and S. Fichera, *Statistical Design of Variable Sample Size and Sampling Interval $\overline{X}$ Control Charts with Run Rules*, Int. J. Adv. Manuf. Technol. 28 (2006), pp. 966-977.
[13] M.R.Jr. Reynolds, R.W. Amin and J.C. Arnold, *CUSUM Charts with Variable Sampling Intervals*, Technometrics. 32 (1990), pp. 371-384.
[14] S. Zhang and Z. Wu, *A CUSUM Scheme with Variable Sample Sizes for Monitoring Process Shifts*, Int. J. Adv. Manuf. Technol. 33 (2007), pp. 977-987.
[15] Z. Wu, S. Zhang and P. Wang, *A CUSUM Scheme with Variable Sample Sizes and Sampling Intervals for Monitoring the Process Mean and Variance*, Qual. Rel. Eng. Int. 23 (2007), pp. 157-170.
[16] M.R.Jr. Reynolds, *Shewhart and EWMA Variable Sampling Interval Control Charts with Sampling at Fixed Times*, J. Qual. Tech. 28:2 (1996a), pp. 199-212.
[17] M.R.Jr. Reynolds and J.C. Arnold, *EWMA Control Charts with Variable Sample Sizes and Variable Sampling Intervals*, IIE Trans. 33 (2001), pp. 511-530.
[18] Z. Li, Y. Luo and Z. Wang, *Cusum of Q Chart with Variable Sampling Intervals for Monitoring the Process Mean*, Int. J. Prod. Res. 48:16 (2010), pp. 4861-4876.
[19] Z. Li and Z. Wang, *An Exponentially Weighted Moving Average Scheme with Variable Sampling Intervals for Monitoring Linear Profiles*, Comput. Ind. Eng. 59:4 (2010a;), pp. 630-637.
[20] J.Y. Liu, M. Xie, T.N. Goh, Q.H. Liu and Z.H. Yang, *Cumulative Count of Conforming Chart with Variable Sampling Intervals* Int. J. Prod. Econ. 101 (2006), pp. 286-297.
[21] F. Aparisi and L.C. Haro, *Hotelling's T^2 Control Chart with Variable Sampling Intervals*, Int. J. Prod. Res. 39:14 (2001), pp. 3127-3140.
[22] D. Brook and D.A. Evans, *An Approach to the Probability Distribution of CUSUM Run Length*, Biometrika. 59 (1972), pp. 539-549.
[23] S.V. Crowder, *A Simple Method for Studying Run Length Distribution of Exponentially Weighted Moving Average Control Charts*, Technometrics. 29 (1987), pp. 401-407.
[24] D. Han and F. Tsung, *A Generalized EWMA Control Chart and Its Comparison with the Optimal EWMA, CUSUM and GLR Schemes*, Ann. Statist. 32:1 (2004), pp. 316-339.

[25] D. Han and F. Tsung, *A Reference-Free Cuscore Chart for Dynamic Mean Change Detection and a Unified Framework for Charting Performance Comparison*, J. Amer. Statist. Assoc. 101 (2006), pp. 368-386.
[26] M.R.Jr. Reynolds, *An approximations to the Average Run Length in Cumulative Sum Control Charts*, Technometrics. 17:1 (1975), pp. 65-71.
[27] P.B. Robinson and T.Y. Ho, *Average Run Length of Geometric Moving Average Chats by Numerical Methods*, Technometrics. 20 (1978), pp. 85-93.
[28] W.H. Woodall, *The Distribution of the Run Length of One-sided CUSUM Procedures for Continuous Random Variables* Technometrics. 25:3 (1983), pp. 295-301.
[29] M.S. Gold, *The Geometric Approximation to the Cusum Run Length Distribution*, Biometrika. 76 (1989), pp. 725-733.
[30] C.W. Champ and S.E. Rigdon, *A Comparison of the Markov Chain and the Integral Equation Approach For Evaluating the Run Length Distribution of Quality Control Charts*, J. Stat. Comput. Sim. 20:1 (1991), pp. 191-204.
[31] L. Shu and W. Jiang, *A Markov Chain Model for the Adaptive CUSUM Control Chart*, J. Qual. Tech. 38:2 (2006), pp. 135-147.
[32] D.M. Hawkins, *Evaluation of The Average Run Length of Cumulative Sum Charts for an Arbitrary Data Distribution*, Commun. Stat.-Simul. C. 21 (1992), pp. 1001-1020.
[33] G. Lorden, *Procedures for Reacting to a Change in Distribution*, Ann. Math. Statist. 42 (1971), pp. 1897-1908.
[34] M. Pollak, *Optimal Detection of a Change in Distribution*, Ann. Statist. 13 (1985), pp. 206-227.
[35] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Spring-Verlag, New York, 1985.
[36] G.V. Moustakides, *Optimal Stopping Times for Detecting Changes in Distributions*, Ann. Statist. 14:4 (1986), pp. 1379-1387.
[37] R.S. Sparks, *CUSUM Charts for Signalling Varying Locations shifts*, J. Qual. Tech. 32 (2000), pp. 157-171.
[38] Z. Li and Z. Wang, *Adaptive CUSUM of the Q Chart*, Int. J. Prod. Res. 48:5 (2010b), pp. 1287-1301.
[39] C.P. Quesenberry, *SPC Q Charts for Start-Up Processes and Short or Long Runs*, J. Qual. Tech. 23:3 (1991), pp. 213C224.
[40] C.S. Van Dobben de Bruyn, *Cumulative Sum Test: Theory and Practice*, Griffin, London, 1968.
[41] J.M. Lucas and R.B. Crosier, *Fast Initial Response for CUSUM Quality Control Schemes: Give Your CUSUM a Head Start*, Technometrics. 24 (1982a), pp. 199-205.
[42] E. Yashchin, *On a Unified Approach to the Analysis of Two-sided Cumulative Sum Control Schemes with Headstarts*, Adv. Appl. Probab. 17:3 (1985), pp. 562-593.
[43] W.H. Woodall, *On the Markov Chain Approach to the Two-Sided CUSUM Procedure*, Technometrics. 26:1 (1984), pp. 41-46.
[44] F.F. Gan, *Exact Run Length Distribution for One-sided Exponential CUSUM Scheme*, Statist. Sinica. 2 (1992), pp. 297-312.
[45] L.A. Jones, C.W. Champ and S.E. Rigdon, *The Run Length Distribution of the CUSUM with Estimated Parameters*, J. Qual. Tech. 36 (2004), pp. 95-108.
[46] A. Luc eno and J. Puig-pey, *Evaluation of the Run-Length Probability Distribution for CUSUM charts: Assessing Chart Performance*, Technometrics. 42:4 (2000), pp. 411-416.
[47] J.M. Lucas and M.S. Saccucci, *Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements*, Technometrics. 32:1 (1990), pp. 1-29.
[48] M.S. Saccucci and J.M. Lucas, *Average Run Length for Exponentially Weighted Moving Average Control Schemes Using the Markov Chain Approach*, J. Qual. Tech. 22:2 (1990), pp. 154-162.
[49] G. Capizzi and G. Masarotto, *An Adaptive Exponentially Weighted Moving Average Control Chart*, Technometrics. 45 (2003), pp. 199-203.
[50] Y. Zhao, F. Tsung and Z. Wang, *Dual CUSUM Scheme for Detecting the Range Shift in the Mean*, IIE Trans. 37 (2005), pp. 1047-1058.
[51] E.H. Aarts, P.J.M. Van Laarhoven, *Simulated Annealing: an Introduction*, Statist. Neerlandica. 43 (1989), pp. 31C52.
[52] N.A. Saleh, M.A. Mahmoud and A.-S.G. Abdel-Salam, *The Performance of the Adaptive Exponentially Weighted Moving Average Control Chart with Estimated Parameters*, Qual. Rel. Eng. Int. DOI: 10.1002/qre.1408.
[53] F.F. Gan, *Exponentially Weighted Moving Average Control Charts with Reflecting Boundaries*, J. Stat. Comput. Sim. 46 (1993), pp. 45-67.
[54] F.F. Gan, *Designs of One- and Two-Sided Exponential EWMA Charts*, J. Qual. Tech. 30 (1998), pp. 55-69.
[55] Z. Li, Z. Wang and Z. Wu, *Necessary and Sufficient Conditions for Non-Interaction of a Pair of One-Sided EWMA Schemes with Reflecting Boundaries*, Statist. Probab. Lett. 79 (2009), pp. 368-374.
[56] L.A. Jones, C.W. Champ and S.E. Rigdon, *The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters*, Technometrics. 43 (2001), pp. 156-167.
[57] L.A. Jones, *The Statistical Design of EWMA Control Charts with Estimated Parameters*, J. Qual. Tech. 34:3 (2002), pp. 277-288.
[58] C.W. Champ and W.H. Woodall, *Exact Result for Shewhart Control Charts with Supplementary Run Rules*, Technometrics. 29:4 (1987), pp. 393-399.
[59] J.M. Lucas, *Combined Shewhart-CUSUM Quality Control Schemes*, J. Qual. Tech. 12 (1982), pp. 51-59.
[60] J.M. Lucas and R.B. Crosier, *Robust CUSUM: a Robustness Study for CUSUM Quality Control Schemes*, Comm. Statist.-Theory Methods. 11 (1982b), pp. 2669-2687.
[61] J.C. Fu, G. Shmueli and Y.M. Chang, *A Unified Markov Chain Approach for Computing the Run Length Distribution in Control Charts with Simple or Compound Rules*, Statist. Probab. Lett. 65 (2003), pp. 457-466.

[62] F.F. Gan, *Combined Cumulative Sum and Shewhart Variance Charts*, J. Stat. Comput. Sim. 32:3 (1989), pp. 149-163.
[63] M.E. Calzada and S.M. Scariano, *Joint Monitoring of the Mean and Variance of Combined Control Charts with Estimated Parameters*, Commun. Stat.-Simul. C. 36 (2007), pp. 1115-1134.
[64] M.R.Jr. Reynolds, R.W. Amin, J.C. Arnold and J.A. Nachlas, *$\bar{X}$ Charts with Variable Sampling Intervals*, Technometrics. 30 (1988), pp. 181-192.
[65] D.C. Montgomery, *SPC Research–Current Trends*, Qual. Rel. Eng. Int. 23 (2007), pp. 515-516.
[66] A.F.B. Costa, *X-bar Chart with Variable Parameters*, J. Qual. Tech. 31 (1999), pp. 408-416.
[67] Y. Luo, Z. Li and Z. Wang, *Adaptive CUSUM Control Chart with Variable Sampling Intervals*, Comput. Statist. Data Anal. 53:7 (2009), pp. 2693-2701.
[68] M.R.Jr. Reynolds, *Variable-sampling-interval Control Charts with Sampling at Fixed Times*, IIE Trans. 28 (1996b), pp. 497-510.
[69] R.B. Crosier, *A New Two-sided Cumulative Sum Quality Control Scheme*, Technometrics. 28 (1986), pp. 187-194.
[70] C. Zou, Z. Wang and F. Tsung, *Monitoring an Autocorrelated Processes Using Variable Sampling Schemes at Fixed-Times*, Qual. Rel. Eng. Int. 28 (2008), pp. 55-69.
[71] G.C. Runger and S.S. Prabhu, *A Markov Chain Model for the Multivariate Exponentially Weighted Moving Average Control Chart*, J. Amer. Statist. Assoc. 91:436 (1996), pp. 1701-1706.
[72] J. Zhang, C. Zou and Z. Wang, *A Control Chart Based on Likelihood Ratio for Monitoring Process Mean and Variability*, Qual. Rel. Eng. Int. 26 (2010), pp. 63-73.
[73] Z. Li, J. Zhang and Z. Wang, *Self-Starting Control Chart for Simultaneously Monitoring Process Mean and Variance*, Int. J. Prod. Res. 48:15 (2010), pp. 4537-4553.
[74] R.W. Amin and W. Letsinger, *Improved Switching Rules in Control Procedures Using Variable Sampling Intervals*, Commun. Stat.-Simul. C. 20 (1991), pp. 205-230.
[75] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.
[76] W.H. Woodall and D.C. Montgomery, *Research Issues and Ideas in Statistical Process Control*, J. Qual. Tech. 31 (1999), pp. 376-386.
[77] Z.G. Stoumbos, M.R.Jr. Reynolds, T.P. Ryan and W.H. Woodall, *The State of Statistical Process Control as We Proceed into the 21st Century*, J. Amer. Statist. Assoc. 95 (2000), pp. 992-998.
[78] S. Bersimis, S. Psarakis and J. Panaretos, *Multivariate Statistical Process Control Charts: an Overview*, Qual. Rel. Eng. Int. 23 (2007), pp. 517-543.
[79] Y. Dai, Y. Luo, Z. Li and Z. Wang, *A New Adaptive CUSUM Control Chart for Detecting the Multivariate Process Mean*, Qual. Rel. Eng. Int. 27:7 (2011), pp. 877-884.
[80] R.B. Croisier, *Multivariate Generalizations of Cumulative Sum Quality-Control Schemes*, Technometrics. 30 (1988), pp. 243-251.
[81] C.A. Lowry, W.H. Woodall, C.W. Champ and S.E. Rigdon, *Multivariate Exponentially Weighted Moving Average Control Chart*, Technometrics. 34 (1992), pp. 46-53.
[82] A.L. Goel and S.M. Wu, *Determination of ARL and a Contour Nomogram for Cusum Charts to Control Normal Mean*, Technometrics. 13:2 (1971), pp. 221-236.

Appendix A. Gaussian Quadrature Method for Numerical Computation of Integral Equation

For integral equations as in Eq. (4), the Gaussian quadrature method can be used to obtain numerical solutions. Supposing the Gaussian roots being $z_1, \dots, z_m$, and corresponding weights being $c_1, \dots, c_m$, Eq. (4) can be approximately written as,

$$L(\mu) = 1 + L(0)F(k-\mu) + \sum_{i=1}^{m} c_i L(z_i) f(z_i + k - \mu). \tag{A1}$$

Let

$$\begin{aligned}
&\delta_2 = (F(k-z_1), \dots, F(k-z_m))', \\
&\boldsymbol{\Omega} = (\omega_{ij}),\ \omega_{ij} = c_j f(z_j + k - z_i), \\
&\mathbf{A} = (\delta_\mathbf{2} | \boldsymbol{\Omega} - \mathbf{I})\,, \\
&\mathbf{L}' = (L(0), L(z_1), \dots, L(z_m)), \\
&\mathbf{c} = \begin{pmatrix} -1 \\ -1 \\ \dots \\ -1 \end{pmatrix}.
\end{aligned}$$

If μ in Eq. (A1) are set $\mu = z_1, \ldots, z_m$, respectively, we can have

$$A\mathbf{L} = \mathbf{c}.$$

Note that A and $\mathbf{c}$ are both given, then $\mathbf{L}$ can be obtained as well as the ARL $L(\mu)$ with the obtained $\mathbf{L}$.

As is known, this Gaussian quadrature method greatly depends on the number of Gaussian roots and the specific selection of these Gaussian roots and weights.

[82] used this method to compute ARL and the run length distribution when observations follow the normal distribution.

Appendix B. Gauss-Legendre Method for Numerical Computation of Integral Equation

For the distribution of ARL of upper-sided CUSUM control chart defined by Eq. (5) and Eq. (6), [46] gave out the following Gauss-Legendre approximation.

Let

$$\begin{aligned}
\delta_1 &= (1 - F(h+k-z_1), \ldots, 1 - F(h+k-z_m))',\\
\delta_2 &= (F(k-z_1), \ldots, F(k-z_m))',\\
\mathbf{p}_n &= (p(n,z_1), \ldots, p(n,z_m))',\\
\mathbf{w} &= (c_1 f(z_1+k), \ldots, c_m f(z_m+k))',\\
\mathbf{\Omega} &= (\omega_{ij}),\ \omega_{ij} = c_j f(z_j + k - z_i).
\end{aligned}$$

Eq. (5) and Eq. (6) can be approximately written as

$$\begin{cases} \mathbf{p}_n = p(n-1,0)\delta_2 + \mathbf{\Omega}\mathbf{p_{n-1}}, \\ \mathbf{p}_1 = \delta_1. \end{cases} \tag{B1}$$

When initial value is 0, we note that

$$\begin{cases} p(1,0) = 1 - F(h+k), \\ p(n,0) = p(n-1,0)F(k) + \mathbf{w}'\mathbf{p}_{n-1}. \end{cases} \tag{B2}$$

Due to the recursive form of Eq. (B2), substituting $\mathbf{p}_n$ with Eq. (B1), we can have

$$\begin{aligned}
&p(n,0)\\
&= F(k)p(n-1,0) + \mathbf{w}'[p(n-2,0)\delta_2 + \mathbf{w}'\mathbf{p}_{n-2}]\\
&= F(k)p(n-1,0) + \mathbf{w}'\mathbf{\Omega}^0\delta_2 p(n-2,0) + \cdots + \mathbf{w}'\mathbf{\Omega}^{n-3}\delta_2 p(1,0) + \mathbf{w}'\mathbf{\Omega}^{n-2}\delta_1\\
&= F(k)p(n-1,0) + \sum_{j=0}^{n-3} K_{2j}p(n-2-j,0) + K_{1\ n-2},
\end{aligned} \tag{B3}$$

where $K_{1j} = \mathbf{w}'\mathbf{\Omega}^j\delta_1, K_{2j} = \mathbf{w}'\mathbf{\Omega}^j\delta_2,\ j = 0,1,2,\ldots$.

If all the absolute values of the eigenvalues of $\mathbf{\Omega}$ are less than 1, for sufficiently large n, we can find a positive integer N large enough such that (B3) can be approximately written as

$$p(n,0) = F(k)p(n-1,0) + \sum_{j=0}^{N} K_{2j}p(n-2-j,0) + K_{1\ n-2}. \tag{B4}$$

The ARL with initial value 0 can be computed by Eq. (B3) and Eq. (B4) and the ARL with non-zero initial value by Eq. (B1) and Eq. (B4). [46], moreover, gave out the following numerical expression to compute ARL,

$$ARL(z_j) = r_{1j} + r_{2j}ARL(0), \tag{B5}$$

where $\mathbf{r}_i = (r_{i1}, \ldots, r_{im})', i = 1, 2$ satisfies integral equations

$$(\mathbf{I} - \mathbf{\Omega})\mathbf{r}_1 = \mathbf{1},\ (\mathbf{I} - \mathbf{\Omega})\mathbf{r}_2 = \delta_2.$$

Note that the computing speed of the approximation in Eq. (B5) is more fast because all of the values of $\mathbf{\Omega}, \delta_{\mathbf{1}}, \delta_{\mathbf{2}}, \mathbf{w}$ are independent of n so as to need to be computed and only computed for one time.