An EWMA chart for monitoring the process standard deviation when parameters are estimated

Petros E. Maravelakis\textsuperscript{a,}\textsuperscript{*}, Philippe Castagliola\textsuperscript{b}

\textsuperscript{a} Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovassi, 83200, Samos, Greece
\textsuperscript{b} Université de Nantes & IRCCyN UMR CNRS 6597, Carquefou, France

ARTICLE INFO

Article history:
Received 17 July 2008
Received in revised form 6 January 2009
Accepted 9 January 2009
Available online 20 January 2009

ABSTRACT

The EWMA chart for the standard deviation is a useful tool for monitoring the variability of a process quality characteristic. The performance of this chart is usually evaluated under the assumption of known parameters. However, in practice, process parameters are estimated from an in-control Phase I data set. A modified EWMA control chart is proposed for monitoring the standard deviation when the parameters are estimated. The Run Length properties of this chart are studied and its performance is evaluated by comparing it with the same chart but with process parameters assumed known.

1. Introduction

Control charts are popular due to their ability to detect shifts in a process (see, e.g. Galeano (2007), Shu et al. (2008) and Wu et al. (2008)). The Exponentially Weighted Moving Average (EWMA) control chart is a well known tool for process monitoring. The EWMA chart was introduced by Roberts (1959) and it is used to detect persistent shifts in a process. The main advantage of this chart is that it is able to detect quickly small and moderate shifts. The EWMA chart for monitoring the variability has received less attention than its counterpart for the mean, although it is equally important. A number of publications about monitoring the process variance using an EWMA chart have appeared in the literature — see, e.g. Wortham and Ringer (1971), Sweet (1986), Ng and Case (1989), Domangue and Patch (1991), Crowder and Hamilton (1992), MacGregor and Harris (1993) and, more recently, Castagliola (2005).

An indispensable assumption for the development of (Phase II) control charts is that the process parameters are assumed known. In practice, the parameters are usually estimated from an in-control historical (Phase I) data set. When the parameters are estimated, the performance of the control chart differs from the known parameters case due to the variability of the estimators (see, e.g. Chen (1997, 1998)). The performance of the EWMA chart for the mean has been studied in the case of estimated parameters (Jones et al., 2001). In a review paper Jensen et al. (2006) emphasized the need for more research about the performance of control charts with estimated parameters for monitoring the variance. Jensen et al. (2006) argued that the EWMA chart for monitoring the variance has not been studied. In this paper we investigate the performance of an EWMA chart for the standard deviation with estimated parameters.

In Section 2, we present the EWMA chart for monitoring the standard deviation and we introduce a modified version used in the remainder of the paper. In Section 3, we compute the exact run length distribution of the proposed scheme along with its two first moments, using two different approaches: integral equations and Markov chain. The effect of estimating the parameters on the performance of the proposed scheme is presented in Section 4. An illustrative example is given in Section 5. Finally, in the concluding remarks we give some comments and recommendations.

\textsuperscript{*} Corresponding author. Tel.: +30 2273082354; fax: +30 2273082309.
E-mail address: maravel@aegean.gr (P.E. Maravelakis).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.01.004
2. An EWMA control chart for monitoring the process standard-deviation

In order to monitor the standard-deviation of a process, the unknown in-control process variance $\sigma_0^2$ has to be estimated from a Phase I data set composed of $m$ samples of size $n$, $X_{ij}, i = 1, \ldots, m$, $j = 1, \ldots, n$. If we assume that $X_{ij} \sim N(\mu_0, \sigma_0)$, where $\mu_0$ is the in-control process mean, $\sigma_0^2$ is the in-control process variance, and the samples are simple random samples themselves and are independent of each other, then an unbiased estimate $\hat{\sigma}_0^2$ of $\sigma_0^2$ is

$$\hat{\sigma}_0^2 = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2,$$

where $\bar{X}_i$ is the sample mean corresponding to the $i$th sample, i.e.

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}.$$

Now, let $Y_{i1}, \ldots, Y_{in}$ be Phase II samples obtained at times $i = 1, 2, \ldots$. Let us assume that $Y_{ij} \sim N(\mu_0, \sigma_0)$ where $\tau > 0$ is a constant reflecting the shift in dispersion. When $\tau = 1$, the process variability is in control; otherwise the process variability has changed. It is worth noting that when $\tau < 1$ the out-of-control process condition corresponds to a reduction in process variability, which is usually the outcome of a corrective action on the process itself. On the other hand, when $\tau > 1$ the out-of-control process condition may be the result of a deterioration in the process performance which increases the process variability. Let $\bar{Y}_i$ and $S_i^2$ be the sample mean and sample variance of $Y_{i1}, \ldots, Y_{in}$, i.e.

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_i)^2$$

with $\bar{Y}_i = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}$.

Crowder and Hamilton (1992) introduced an EWMA control chart for monitoring a process standard deviation using the scheme

$$T_0 = \ln \left( \sigma_0^2 \right),$$

$$T_i = \max(\lambda \ln S_i^2 + (1-\lambda)T_{i-1}, \ln(\sigma_0^2)), \quad i = 1, 2, \ldots \quad (1)$$

where $0 < \lambda \leq 1$ is a smoothing parameter. The Upper Control Limit (UCL) of this chart (i.e. for the $T_i$'s) in case of independent observations is given by

$$UCL = K\sigma = K \sqrt{\frac{(\lambda - 2)(\lambda)}{2 - \lambda} + \frac{2}{2(n-1)^2} + \frac{4}{3(n-1)^3} - \frac{16}{15(n-1)^5}}, \quad (2)$$

where $K$ is a constant chosen together with $\lambda$, so as to achieve the desired performance for the chart, and $\sigma$ is the asymptotic standard deviation of $T_i$ without the restart feature. The chart defined in Eq. (1) can be used only with subgrouped data (i.e. $n > 1$) therefore in this paper we confine ourselves to this case. Another feature of this chart is that it can be used to identify only upward shifts in the variability. This does not imply that decreases in the variability are not important. However, a decrease in the variability usually occurs only after a corrective action in the process. Therefore, the time of the change in the variability is known and a control chart is not needed to detect the shift.

In this paper, we propose a modified version of the scheme in Eq. (1). Specifically, the proposed chart is

$$Z_0 = z_0,$$

$$Z_i = \max(\lambda \ln(S_i^2/\hat{\sigma}_0^2) + (1-\lambda)Z_{i-1}, 0), \quad i = 1, 2, \ldots \quad (3)$$

where $0 \leq z_0 < UCL$ is a constant used as the initial value. If $z_0 > 0$ then the chart has a head start (see Lucas and Crosier (1982) and Lucas and Saccucci (1990)). The upper control limit in Eq. (2) was derived by Crowder and Hamilton (1992) using the fact that $\ln S_i^2$ has a log-gamma distribution and $\ln S_i^2 \sim \chi^2_{n-1} / 2 + \frac{2}{(n-1)^2} + \frac{4}{3(n-1)^3} - \frac{16}{15(n-1)^5}$. Treating $\hat{\sigma}_0^2$ as a constant, we have $V(\ln(S_i^2/\hat{\sigma}_0^2)) = V(\ln(S_i^2) - \ln(\hat{\sigma}_0^2)) = V(\ln(S_i^2))$ and we then deduce that the upper control limit UCL for our proposed chart is the same as the one in Eq. (2).

Let us define $Q_i = S_i^2/\hat{\sigma}_0^2$ and $W = \hat{\sigma}_0^2/\sigma_0^2$. It is well known (see e.g. Chen (1998)) that $Q_i$ and $W$ are both gamma random variables of respective parameters $(\frac{n-1}{2}, 2\sigma_0^2/n-1)$ and $(\frac{m(n-1)}{2}, \frac{2}{m(n-1)})$. Consequently, the probability density functions $f_Q(q)$ and $f_W(w)$ of $Q_i$ and $W$ are equal to

$$f_Q(q) = f_Y \left( q \left| \frac{n-1}{2}, \frac{2\sigma_0^2}{n-1} \right. \right),$$

$$f_W(w) = f_Y \left( w \left| \frac{m(n-1)}{2}, \frac{2}{m(n-1)} \right. \right).$$
where
\[ f_γ(x|a, b) = \frac{e^{-x/b}x^{a-1}}{b^a\Gamma(a)}. \]

It is also clear that \( Q_i/W = S_i^2/\hat{\sigma}_0^2 \) and Eq. (3) can thus be rewritten
\[ Z_0 = z_0, \]
\[ Z_i = \max(\lambda \ln(Q_i/W) + (1 - \lambda)Z_{i-1}, 0), \quad i = 1, 2, \ldots \]

Since we know the exact distribution function of the statistic plotted, we are able to compute the exact run length distribution of the proposed EWMA chart and its first two moments. This is carried out in the following section.

3. The run length properties of the EWMA Chart for the standard deviation with estimated parameters

Control charts are generally evaluated in terms of their Run Length (or RL in short) distribution. The RL is defined as the number of samples until the statistic plotted on a control chart crosses one of the control limits. Let \( L \) denote the random variable representing the RL of the proposed EWMA chart, i.e.
\[ L = \inf\{\ell \geq 1|Z_\ell > UCL\}, \]
and let
\[ f_{\ell}(\ell|w, z_0) = P(L = \ell|W = w, Z_0 = z_0), \]
be the (conditional) probability mass function of \( L \) given a specific value \( w \) for the random variable \( W \) and the starting value (constant) \( z_0 \). For simplicity, the term \( W = w \) will be removed in the rest of the paper when unnecessary. In this section, we will derive the main properties of the Run Length, i.e. probability mass function, cumulative distribution function, mean (denoted ARL) and standard-deviation (denoted SDRL). We will firstly present an integral equation approach and then a Markov chain approach, both approaches being numerically solvable.

3.1. Integral equation approach

Jones et al. (2001) evaluated the performance of the EWMA chart for the mean when the parameters are estimated. We will use a similar approach in order to evaluate the performance of our proposed EWMA chart for monitoring the variance. In Appendices A and B, we demonstrate that for \( \ell = 1 \), we have
\[ f_1(1|w, z_0) = 1 - F_Q \left( w \exp \left( \frac{UCL - (1 - \lambda)z_0}{\lambda} \right) \right), \]
and for \( \ell = 2, 3, \ldots \)
\[ f_\ell(\ell|w, z_0) = f_\ell(\ell - 1|w, 0)F_Q \left( w \exp \left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right) \]
\[ + \frac{w}{\lambda \int_0^{UCL} f_\ell(\ell - 1|w, z) \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) f_Q \left( w \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) \right) dz, \]
where \( F_Q \) is the cumulative distribution function of \( Q_i \). The unconditional probability mass function \( f_\ell(\ell|z_0) \) of \( L \) is given by
\[ f_\ell(\ell|z_0) = \int_0^{\infty} f_\ell(\ell|w, z_0)f_W(w)dw, \]
and the unconditional cumulative distribution function \( F_\ell(\ell|z_0) \) of \( L \) can immediately be derived from the previous formula
\[ F_\ell(\ell|z_0) = \sum_{i=1}^{\ell} f_i(i|z_0). \]

Let \( ARL(w, z_0) = E(L|w, z_0) \) be the (conditional) Average Run Length of this chart. In Appendix C, we demonstrate that \( ARL(w, z_0) \) can be obtained by solving the following integral equation:
\[ ARL(w, z_0) = 1 + ARL(w, 0)F_Q \left( w \exp \left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right) \]
\[ + \frac{w}{\lambda \int_0^{UCL} ARL(w, z) \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) f_Q \left( w \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) \right) dz. \]
Let us also define $E2RL(w, z_0) = E(L^2|w, z_0)$ to be the (conditional) 2nd non-central moment of the run length $L$. In Appendix D, we demonstrate that $E2RL(w, z_0)$ can be obtained by solving the following integral equation:

$$E2RL(w, z_0) = 1 + 2 \left[ ARL(w, 0) f_Q \left( w \exp \left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right) \right. + \frac{w}{\lambda} \int_0^{UCL} ARL(w, z) \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) f_Q \left( w \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) \right) dz \bigg] + E2RL(w, 0) f_Q \left( w \exp \left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right) + \frac{w}{\lambda} \int_0^{UCL} E2RL(w, z) \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) f_Q \left( w \exp \left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) \right) dz.$$

The unconditional Average Run Length $ARL(z_0)$ and 2nd non-central moment $E2RL(z_0)$ of $L$ are equal to

$$ARL(z_0) = \int_0^{\infty} ARL(w, z_0) f_W(w) dw,$$

$$E2RL(z_0) = \int_0^{\infty} E2RL(w, z_0) f_W(w) dw,$$

and the Standard Deviation Run Length $SDRL(z_0)$ of this chart is

$$SDRL(z_0) = \sqrt{E2RL(z_0) - (ARL(z_0))^2}.$$

The exact solutions of the integral equations for $ARL(w, z_0)$ and $E2RL(w, z_0)$ can be approximated using numerical integration for specific value of $W$ (see, for example, the methodology presented in Hamilton and Crowder (1992)). Then the equations for $ARL(z_0)$ and $E2RL(z_0)$ are solved numerically using a Gaussian quadrature for variable $W$.

3.2. Markov chain approach

Let us assume that we have a discrete-time Markov chain with $p + 2$ states, where states 0, 1, ..., $p$ are transient and state $p + 1$ is an absorbing one. The transition probability matrix $P$ of this discrete-time Markov chain is

$$P = \begin{pmatrix} Q & r \\ 0^T & 1 \end{pmatrix},$$

where $Q$ is the $(p + 1, p + 1)$ matrix of transient probabilities, where $0^T = (0, 0, \ldots, 0)$ and where the $(p + 1, 1)$ vector $r$ satisfies $r = 1 - Q1$ (i.e. row probabilities must sum to 1) with $1^T = (1, 1, \ldots, 1)$. Let $\mathbf{q}$ be the $(p + 1, 1)$ vector of initial probabilities associated with the $p + 1$ transient states, i.e. $\mathbf{q}^T = (q_0, q_1, \ldots, q_p)$.

The number of steps $L$ until the process reaches the absorbing state is a Discrete Phase-type (or DPH) random variable of parameters $(Q, \mathbf{q})$, see for instance Neuts (1981) or Latouche and Ramaswami (1999). In control chart terminology, the random variable $L$ is also called the Run Length. The probability mass function $f_{DPH}(\ell|Q, \mathbf{q})$ and the cumulative distribution function $F_{DPH}(\ell|Q, \mathbf{q})$ of a DPH random variable $L$ are defined for $\ell = 1, 2, \ldots, \infty$ and equal to

$$f_{DPH}(\ell|Q, \mathbf{q}) = \mathbf{q}^T Q^{\ell-1} r,$$

$$F_{DPH}(\ell|Q, \mathbf{q}) = 1 - \mathbf{q}^T Q^{\ell} 1.$$

If there is no simple formula for the central moment $\mu_i = E((L - E(L))^i|Q, \mathbf{q})$ of order $i \geq 1$ of a DPH random variable $L$, there is nevertheless a simple formula for the factorial moment $v_i = E(L(L-1)\cdots(L-i+1)|Q, \mathbf{q})$ of order $i \geq 1$, equal to $v_i = i! \mathbf{q}^T (1 - Q)^{-1} Q^{i-1} 1$.

We have, in particular,

$$v_1 = \mathbf{q}^T (1 - Q)^{-1} 1,$$

$$v_2 = 2 \mathbf{q}^T (1 - Q)^{-2} Q 1,$$

and we easily deduce the mean $E(L)$ and standard deviation $\sigma(L)$ of $L$

$$E(L) = v_1,$$

$$\sigma(L) = \sqrt{v_2 - v_1^2} + v_1.$$

The properties (probability mass function $f_{|w, z_0}$, cumulative distribution function $F_{|w, z_0}$, $ARL(w, z_0) = E(L)$, $SDRL(w, z_0) = \sigma(L), \ldots)$ of the Run Length $L$ of our chart can be numerically evaluated using the formulas presented above. This flexible and relatively easy to use procedure, originally proposed by Brook and Evans (1972) for the $ARL$, involves
dividing the interval between $LCL = 0$ and $UCL$ (see Fig. 1) into $p$ subintervals of width $2\delta$, where $\delta = UCL/(2p)$. By definition, $H_j, j = 1, \ldots, p$, represents the midpoint of the $j$th subinterval and $H_0 = 0$ corresponds to the “restart state” feature of our chart. When the number $p$ of subintervals is sufficiently large, this finite approach provides an effective method that allows the Run Length properties to be accurately evaluated. In our particular case, the generic element $Q_{ij}, i = 0, 1, \ldots, p$, of the matrix $Q$ of transient probabilities is equal to (see Appendix E)

- if $j = 0$
  \[ Q_{00} = F_Q \left( w \exp \left( -\frac{(1-\lambda)H_0}{\lambda} \right) \right). \]

- if $j = 1, 2, \ldots, p$
  \[ Q_{ij} = F_Q \left( w \exp \left( \frac{H_j + \delta - (1-\lambda)H_i}{\lambda} \right) \right) - F_Q \left( w \exp \left( \frac{H_j - \delta - (1-\lambda)H_i}{\lambda} \right) \right). \]

and the generic element $q_j$ of vector $q$ of initial probabilities is equal to

\[ q_j = \begin{cases} 
1 & \text{if } H_i - \delta < z_0 < H_j + \delta, \\
0 & \text{otherwise.} 
\end{cases} \]

The unconditional probability mass function $f_1(\ell|z_0)$, cumulative distribution function $F_1(\ell|z_0)$, $ARL(z_0)$ and $SDRL(z_0)$ can be obtained from the conditional ones by applying the same formulas presented in the previous subsection.

4. The performance of the EWMA chart for the standard deviation with estimated parameters

The mean ($ARL$) and the standard deviation ($SDRL$) of the RL distribution are the usual measures to evaluate the performance of control charts. When a process is in-control, we expect, at some point, the control chart to signal due to the inherent variability (false alarm). The average RL in such a case will be denoted $ARL_0$ and the standard deviation of the RL, $SDRL_0$. On the other hand, the average RL when a process is out-of-control will be denoted $ARL_1$ and the standard deviation of the RL, $SDRL_1$. Apparently, a control chart is considered better than its competitors if it has the smaller $ARL_1$ value for a specific shift $\tau$ when $ARL_0$ is the same for all the charts (fixed).

Now we are ready to present several numerical results that clarify the effect of estimating the nominal standard-deviation $\sigma_0^2$ on the EWMA chart for the Standard Deviation. In Table 1 we present the $ARL$ and $SDRL$ values for the out-of-control cases ($ARL_1$ and $SDRL_1$ values respectively). These values are calculated for both the known variance case ($m = \infty$) and the estimated variance case ($m = 10, 20, 40, 80$) for different sample size values $n = 3, 5, 7, 9$. The $ARL_0$ value is 370.4 for all the cases considered in the table and $z_0$ was set equal to 0. The $ARL_1$ and $SDRL_1$ values have been computed using the Markov Chain approach presented in Section 3.2 (the number of Markov chain states was set to $p = 400$ when $\tau = 1.1$ and $p = 100$ when $\tau \geq 1.2$). We have to note that using either the integral equations method or the Markov chain method to compute the $ARL_1$ and $SDRL_1$ values would lead us to the same equation. This statement is attributed to the fact that in both cases the equations for $ARL(w, z_0)$ and $E2RL(w, z_0)$ are numerically evaluated by dividing the interval between 0 and $UCL$ (either by using Gaussian quadrature or the Markov chain methodology described in Section 3.2).

The first column of Table 1 displays the shift in dispersion $\tau = 1.1, 1.2, 1.4, 1.6, 1.8$. In the second column, we provide the optimal EWMA parameters $(\lambda^*, K^*)$ for Crowder and Hamilton’s (1992) approach (i.e. $\sigma_0$ known case) corresponding to the specific shift $\tau$. For instance, if $n = 5$ and $\tau = 1.2$, the optimal EWMA parameters for Crowder and Hamilton’s (1992) approach are $(\lambda^* = 0.013, K^* = 0.705)$. In this case, the optimality property means that the specific choice of $(\lambda^*, K^*)$ gives the lowest $ARL_1$ value, for the $\sigma_0$ known case, among all possible combinations $(\lambda, K)$, for the specific shift $\tau$ and when
Table 1
ARL₁ and SDRL₁ when n = 3, 5, 7, 9 and m = 10, 20, 40, 80, ∞.

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>τ = 3</th>
<th>m = ∞</th>
<th>m = 10</th>
<th>m = 20</th>
<th>m = 40</th>
<th>m = 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.10</td>
<td>(0.010, 0.446)</td>
<td>(96.9, 91.4)</td>
<td>(183,844.1, 183,838.8)</td>
<td>(597.8, 592.4)</td>
<td>(184.6, 179.1)</td>
<td>(128.3, 122.8)</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
<td>(0.010, 0.446)</td>
<td>(40.9, 35.9)</td>
<td>(1798.1, 1793.3)</td>
<td>(97.2, 92.3)</td>
<td>(56.8, 51.8)</td>
<td>(47.4, 42.4)</td>
</tr>
<tr>
<td>7</td>
<td>1.10</td>
<td>(0.147, 1.211)</td>
<td>(15.1, 12.4)</td>
<td>(31.3, 28.6)</td>
<td>(19.6, 16.9)</td>
<td>(17.0, 14.2)</td>
<td>(16.0, 13.2)</td>
</tr>
<tr>
<td>9</td>
<td>1.10</td>
<td>(0.474, 1.408)</td>
<td>(8.2, 6.8)</td>
<td>(11.2, 9.8)</td>
<td>(9.4, 8.0)</td>
<td>(8.7, 7.4)</td>
<td>(8.5, 7.1)</td>
</tr>
</tbody>
</table>

The ARL₀ value is fixed at 370.4. It must be noted that for computational reasons, the value of λ must not be too small (in this case both the integral and the Markov Chain approaches give unreliable results). This fact has been quoted in many papers and, for this reason, we have chosen to constrain λ* ≥ 0.01.

The third column of Table 1 displays the ARL₁ and SDRL₁ values for the σ₀ known case. The remaining columns (4–7) present the results for the ARL₁ and SDRL₁ values when the variance is estimated. It must be stressed that the results in columns 4–7 are computed for the specific combination of \((λ^*, K^*)\) listed in the second column. As it can be noticed in Table 1, the values of ARL₁ and SDRL₁ are very large in the case τ = 1.1 and m = 10 and are actually really different from the values of ARL₁ and SDRL₁ in the case m = ∞ (i.e., σ₀ known case). Of course, as intuitively expected, the difference between ARL₁ values corresponding to the estimated and the known σ₀ case is decreasing as m increases for each specific shift. A similar result holds for the SDRL₁ values. Note that for large shifts (τ ≥ 1.8), the difference becomes negligible, even for small values of m and n. It is also obvious that the difference decreases as n increases. A notable comment according to the results of Table 1 is that as the shift increases for a specific n, so does the optimal \(λ^*\), keeping in mind that the optimal \(λ^*\) and K* displayed are associated with the chart, with known parameters. This result validates the well-known conclusion that small λ values must be used to detect small to moderate shifts and larger λ values must be used for large shifts when the process parameters are known.

In Fig. 2, we present the cumulative distribution function \(F_L(ℓ)\) of the RL L for n = 5, τ = 1.2, 1.6, for both the known case (m = ∞) and the estimated case m = 10, 20, 40, 80. The cumulative distribution function \(F_L(ℓ)\) has also been computed using the Markov Chain approach presented in Section 3.2. According to the results, when the variability increases, estimating the variance results at the beginning in a higher probability of a signal and, afterwards, it gives a lower probability for a signal than in the known variance case. Apparently, at least m = 80 samples of size n = 5 each are needed for a Phase I data set so that the EWMA chart for the Standard Deviation approximatively matches the performance of the known variance case. For other sample sizes the number of samples needed is properly modified (for smaller n larger m is needed and vice-versa).

Table 2 displays the ARL₁, SDRL₁ values for m = 10, 20, 40, 80, n = 3, 5, 7, 9 and τ = 1.2, 1.4, 1.6, 1.8, 2.0 along with the optimal \(λ^*\) and K* values for each case considered. For example when n = 5, m = 40 and τ = 1.4 we observe that the optimal values are \(λ^* = 0.152, K^* = 1.416\) and for these values we find that ARL₁ = 7.4 and SDRL₁ = 5.0. Note that the ARL₀ value is 370.4 for all the entries in the table and z₀ is set equal to 0. From the results we observe that λ* increases as long as m increases, apart from the case τ = 1.2 where λ* is almost everywhere equal to 0.01 due to the constraint we imposed (λ* ≥ 0.01). Additionally, λ* is getting larger as the shift increases except for combinations with small values for n, m and τ. K* resembles the behavior of λ* since it increases as m increases. Furthermore, as τ increases so does K* apart from a few limiting cases for τ = 2.

As an aid to practitioners, we may argue that for small to moderate shifts (τ ≤ 1.4) λ* should be less than 0.35 and if n ≤ 5 then λ* must be below 0.2. Value of λ* less than 0.1 should be chosen for small shifts. For all the other cases λ* can
Fig. 2. CDF for $n = 5$, $m = 10, 20, 40, 80, \infty$ and $\tau = 1.2, 1.6$.

Values below unity should be used only for small shifts or for moderate shifts with a small number of samples ($m$).
Table 2

ARL₁ and SDRL₁ when $n = 3, 5, 7, 9$ and $m = 10, 20, 40, 80.$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$m = 10$</th>
<th>$m = 20$</th>
<th>$m = 40$</th>
<th>$m = 80$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>(0.010, 0.233)</td>
<td>(0.010, 0.310)</td>
<td>(0.010, 0.367)</td>
<td>(0.010, 0.404)</td>
</tr>
<tr>
<td>1.4</td>
<td>(0.010, 0.233)</td>
<td>(0.010, 0.310)</td>
<td>(0.061, 0.808)</td>
<td>(0.103, 1.035)</td>
</tr>
<tr>
<td>1.6</td>
<td>(0.257, 0.917)</td>
<td>(0.365, 1.176)</td>
<td>(0.420, 1.297)</td>
<td>(0.447, 1.353)</td>
</tr>
<tr>
<td>1.8</td>
<td>(0.358, 1.000)</td>
<td>(0.469, 1.218)</td>
<td>(0.519, 1.319)</td>
<td>(0.543, 1.367)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.420, 1.037)</td>
<td>(0.533, 1.237)</td>
<td>(0.583, 1.329)</td>
<td>(0.606, 1.373)</td>
</tr>
<tr>
<td>$n = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>(0.010, 0.316)</td>
<td>(0.010, 0.422)</td>
<td>(0.010, 0.508)</td>
<td>(0.010, 0.564)</td>
</tr>
<tr>
<td>1.4</td>
<td>(0.010, 0.316)</td>
<td>(0.085, 1.053)</td>
<td>(0.152, 1.416)</td>
<td>(0.186, 1.559)</td>
</tr>
<tr>
<td>1.6</td>
<td>(0.292, 1.259)</td>
<td>(0.395, 1.542)</td>
<td>(0.443, 1.667)</td>
<td>(0.465, 1.725)</td>
</tr>
<tr>
<td>1.8</td>
<td>(0.423, 1.371)</td>
<td>(0.529, 1.595)</td>
<td>(0.574, 1.693)</td>
<td>(0.595, 1.740)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.519, 1.428)</td>
<td>(0.622, 1.618)</td>
<td>(0.666, 1.703)</td>
<td>(0.687, 1.743)</td>
</tr>
<tr>
<td>$n = 7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>(0.010, 0.361)</td>
<td>(0.010, 0.482)</td>
<td>(0.010, 0.585)</td>
<td>(0.023, 0.943)</td>
</tr>
<tr>
<td>1.4</td>
<td>(0.078, 0.916)</td>
<td>(0.192, 1.510)</td>
<td>(0.249, 1.737)</td>
<td>(0.277, 1.836)</td>
</tr>
<tr>
<td>1.6</td>
<td>(0.170, 1.501)</td>
<td>(0.295, 1.754)</td>
<td>(0.512, 1.864)</td>
<td>(0.533, 1.916)</td>
</tr>
<tr>
<td>1.8</td>
<td>(0.516, 1.603)</td>
<td>(0.613, 1.797)</td>
<td>(0.653, 1.883)</td>
<td>(0.673, 1.924)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.633, 1.655)</td>
<td>(0.722, 1.816)</td>
<td>(0.760, 1.888)</td>
<td>(0.777, 1.922)</td>
</tr>
<tr>
<td>$n = 9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>(0.010, 0.390)</td>
<td>(0.010, 0.522)</td>
<td>(0.027, 0.980)</td>
<td>(0.059, 1.422)</td>
</tr>
<tr>
<td>1.4</td>
<td>(0.188, 1.351)</td>
<td>(0.284, 1.744)</td>
<td>(0.332, 1.913)</td>
<td>(0.354, 1.990)</td>
</tr>
<tr>
<td>1.6</td>
<td>(0.441, 1.666)</td>
<td>(0.535, 1.889)</td>
<td>(0.576, 1.987)</td>
<td>(0.595, 2.033)</td>
</tr>
<tr>
<td>1.8</td>
<td>(0.605, 1.757)</td>
<td>(0.690, 1.924)</td>
<td>(0.727, 1.999)</td>
<td>(0.744, 2.035)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.737, 1.798)</td>
<td>(0.812, 1.935)</td>
<td>(0.845, 1.997)</td>
<td>(0.861, 2.026)</td>
</tr>
</tbody>
</table>

5. An example

In order to illustrate the use of the EWMA $\ln(S^2)$ control chart, let us consider a procedure to be designed with the aim of monitoring the variability within a yogurt cup filling process: a critical to quality parameter of the filling process is the deviation of each cup filling from a specified target $\mu_0 = 125g$. We want to monitor the process dispersion by means of an EWMA $\ln(S^2)$ control chart. We also suggest to monitor the mean of the process using a traditional EWMA $\bar{X}$ control chart. The Phase I data set used in this example consists of $m = 10$ subgroups of size $n = 5$ plotted in the left part of Fig. 3(a) with “o”. From this Phase I data set we deduce $\bar{\mu}_0 = 124.982$ and $\hat{\sigma}_0^2 = 0.0766$. If $\hat{\sigma}_0^2$ is known to be an unbiased estimator for $\sigma_0^2$, $\hat{\sigma}_0$ is unfortunately a biased estimator for $\sigma_0$. An unbiased estimator for $\sigma_0$ is known to be $\hat{\sigma}_0/c_4(n)$, where the constant $c_4(n)$ is equal to

$$c_4(n) = \sqrt{\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}} \sqrt{2} \left( \frac{n}{n-1} \right).$$

In our case, we have $c_4(5) = 0.94$ and an unbiased estimator for $\sigma_0$ is $\frac{1}{\sqrt{2}} \left( \frac{0.0766}{0.94} \right)^{1/2} = 0.294$. Concerning the EWMA $\bar{X}$, we have chosen to select the optimal parameters $\lambda^* = 0.166$ and $L^* = 0.380$ optimally designed to detect a standardized mean shift $|\mu_0 - \mu_1|/\sigma_0 = 0.5$ when the in-control ARL = 370.4. This yields the following control limits for the EWMA $\bar{X}$
(a) Phase I and II dataset. (b) EWMA $\bar{X}$.

(c) EWMA $\ln(S^2)$, $\sigma_0^2$ estimated case. (d) EWMA $\ln(S^2)$, $\sigma_0^2$ known case.

Fig. 3. Phase I and II datasets, EWMA $\bar{X}$ and EWMA $\ln(S^2)$ control charts.

chart: $LCL = 124.982 - 0.380 \times 0.294 = 124.87$ and $UCL = 124.982 + 0.380 \times 0.294 = 125.094$. Concerning the EWMA $\ln(S^2)$ control chart, we have chosen to select:

- the optimal parameters $\lambda^* = 0.292$ and $K^* = 1.259$ optimally designed to detect a shift $\tau = 1.6$ when $\sigma_0$ is estimated from a Phase I data set consisting of $m = 10$ subgroups of size $n = 5$ (see Table 2). This yields the following upper limit for the EWMA $\ln(S^2)$ chart: $UCL = 0.418$.
- the optimal parameters $\lambda^* = 0.487$ and $K^* = 1.782$ optimally designed to detect a shift $\tau = 1.6$ when $\sigma_0$ is assumed known (see Table 1). This yields the following upper limit for the EWMA $\ln(S^2)$ chart: $UCL = 0.812$.

The Phase II data set used in this example consists of $m = 30$ subgroups of size $n = 5$ plotted in the right part of Fig. 3(a) with “•”. The first 15 subgroups are supposed to be in-control while the last 15 subgroups are supposed to have a larger variability and to be out-of-control. In Fig. 3(b) we plotted the EWMA $\bar{X}$ control chart corresponding to the Phase II data set. As we can see, there is no point outside the control limits and, in terms of the mean, the process seems to be in-control. In both Fig. 3(c) and (d) we plotted the EWMA $\ln(S^2)$ control chart corresponding to the Phase II data set. In Fig. 3(c) we used the optimal parameters $\lambda^* = 0.292$, $K^* = 1.259$ and control limits $UCL = 0.418$ defined above in the $\sigma_0^2$ estimated case, while in Fig. 3(d), we used the optimal parameters $\lambda^* = 0.487$, $K^* = 1.782$ and control limits $UCL = 0.812$ defined above in the $\sigma_0^2$ known case. In both cases, these control charts signal several out-of-control situations (the first one for the 20th subgroup). But, in this example, it can be noticed that the EWMA $\ln(S^2)$ control chart in Fig. 3(c) detects more systematically the out-of-control situations than the EWMA $\ln(S^2)$ control chart in Fig. 3(d), which leaves some points below the upper limits, not detecting these points as potential out-of-control situations.

6. Conclusions

The estimation of the parameters is known to reduce the ability of control charts to detect process shifts. In this paper we have presented the exact RL distribution and its mean and standard deviation of an EWMA chart for monitoring the standard deviation when the in-control variance $\sigma_0^2$ is estimated, using both integral equations and Markov chain approaches. The results show that the chart with estimated $\sigma^2$ needs more data on average in order to detect an out-of-control situation than the one with known $\sigma^2$. 
Appendix B

If \( Z_0 = 0 \), then the next value \( Z_1 \) will be

- either \( Z_1 = 0 \) \( \iff \) \( \lambda \ln(Q_1/w) + (1 - \lambda)z_0 \leq 0 \),
- or \( Z_1 = \lambda \ln(q/w) + (1 - \lambda)z_0 \) where \( q \in S \) is a particular value of \( Q_1 \) and where \( S = \{ q \in \mathbb{R} | 0 < \lambda \ln(q/w) + (1 - \lambda)z_0 < UCL \} \).

Then, for \( \ell = 2, 3, \ldots \), we can split \( f_\ell(\ell|w, z_0) = P(L = \ell|Z_0 = z_0) \) into two parts

\[
f_\ell(\ell|w, z_0) = P(L = \ell - 1|Z_0 = 0) P(\lambda \ln(Q_1/w) + (1 - \lambda)z_0 \leq 0) \]
\[
+ \int_S P(L = \ell - 1|Z_0 = \lambda \ln(q/w) + (1 - \lambda)z_0) f_q(q) dq.
\]

Isolating \( Q_1 \) in \( P(\lambda \ln(Q_1/w) + (1 - \lambda)z_0 \leq 0) \) yields

\[
P(\lambda \ln(Q_1/w) + (1 - \lambda)z_0 \leq 0) = P\left( Q_1 \leq w \exp\left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right),
\]
\[
= F_Q\left( w \exp\left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right).
\]

If we introduce the variable \( z = \lambda \ln(q/w) + (1 - \lambda)z_0 \) then \( q = w \exp(\frac{z - (1 - \lambda)z_0}{\lambda}) \), \( dq = \frac{w}{\lambda} \exp(\frac{z - (1 - \lambda)z_0}{\lambda}) dz \) and we finally deduce the following integral equation, for \( \ell = 2, 3, \ldots \)

\[
f_\ell(\ell|w, z_0) = f_\ell(\ell - 1|w, 0) F_Q\left( w \exp\left( \frac{-(1 - \lambda)z_0}{\lambda} \right) \right)
\]
\[
+ \frac{w}{\lambda} \int_0^{UCL} f_\ell(\ell - 1|w, z) \exp\left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) F_Q\left( w \exp\left( \frac{z - (1 - \lambda)z_0}{\lambda} \right) \right) dz.
\]

Appendix C

By definition we have

\[
ARL(w, z_0) = \sum_{\ell=1}^{\infty} \ell f_\ell(\ell|w, z_0).
\]
Separating the case $\ell = 1$, rearranging the remaining terms from $\ell = 1$ and expanding $\ell + 1$ gives

$$\text{ARL}(w, z_0) = f_1(1|w, z_0) + \sum_{\ell=2}^{\infty} \ell f_1(\ell|w, z_0),$$

$$= f_1(1|w, z_0) + \sum_{\ell=1}^{\infty} (\ell + 1) f_1(\ell + 1|w, z_0),$$

$$= f_1(1|w, z_0) + \sum_{\ell=1}^{\infty} f_1(\ell + 1|w, z_0) + \sum_{\ell=1}^{\infty} \ell f_1(\ell + 1|w, z_0).$$

The first two terms sum to 1 and we have

$$\text{ARL}(w, z_0) = 1 + \sum_{\ell=1}^{\infty} \ell f_1(\ell + 1|w, z_0).$$

Because we have

$$f_1(\ell + 1|w, z_0) = f_1(\ell|w, 0) F_Q \left( w \exp \left( \frac{-(1 - \lambda) z_0}{\lambda} \right) \right)$$

$$+ \frac{w}{\lambda} \int_0^{\text{UCL}} f_1(\ell|w, z) \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) f_0 \left( w \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) \right) \, dz,$$

we then deduce

$$\text{ARL}(w, z_0) = 1 + \text{ARL}(w, 0) F_Q \left( w \exp \left( \frac{-(1 - \lambda) z_0}{\lambda} \right) \right)$$

$$+ \frac{w}{\lambda} \int_0^{\text{UCL}} \text{ARL}(w, z) \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) f_0 \left( w \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) \right) \, dz.$$

**Appendix D**

The demonstration for $E2RL(w, z_0) = E(L^2|w, z_0)$ is very similar to the demonstration of $\text{ARL}(w, z_0)$. By definition we have

$$E2RL(w, z_0) = \sum_{\ell=1}^{\infty} \ell^2 f_1(\ell|w, z_0).$$

Separating the case $\ell = 1$, rearranging the remaining terms from $\ell = 1$ and expanding $(\ell + 1)^2$ gives

$$E2RL(w, z_0) = f_1(1|w, z_0) + \sum_{\ell=2}^{\infty} \ell^2 f_1(\ell|w, z_0),$$

$$= f_1(1|w, z_0) + \sum_{\ell=1}^{\infty} (\ell + 1)^2 f_1(\ell + 1|w, z_0),$$

$$= f_1(1|w, z_0) + \sum_{\ell=1}^{\infty} f_1(\ell + 1|w, z_0) + 2 \sum_{\ell=1}^{\infty} \ell f_1(\ell + 1|w, z_0) + \sum_{\ell=1}^{\infty} \ell^2 f_1(\ell + 1|w, z_0).$$

The first two terms sum again to 1 and we have

$$E2RL(w, z_0) = 1 + 2 \sum_{\ell=1}^{\infty} \ell f_1(\ell + 1|w, z_0) + \sum_{\ell=1}^{\infty} \ell^2 f_1(\ell + 1|w, z_0).$$

Finally, if we expand $f_1(\ell + 1|w, z_0)$, we obtain

$$E2RL(w, z_0) = 1 + 2 \left[ \text{ARL}(w, 0) F_Q \left( w \exp \left( \frac{-(1 - \lambda) z_0}{\lambda} \right) \right) \right.$$

$$+ \frac{w}{\lambda} \int_0^{\text{UCL}} \text{ARL}(w, z) \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) f_0 \left( w \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) \right) \, dz \left. \right]$$

$$+ E2RL(w, 0) F_Q \left( w \exp \left( \frac{-(1 - \lambda) z_0}{\lambda} \right) \right)$$

$$+ \frac{w}{\lambda} \int_0^{\text{UCL}} E2RL(w, z) \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) f_0 \left( w \exp \left( \frac{z - (1 - \lambda) z_0}{\lambda} \right) \right) \, dz.$$
Appendix E

By definition, $Q_{i,j}$ is the probability that the Markov chain goes from state $i$ to state $j$ in one step. The probability $Q_{i,0}$ that the Markov chain goes from state $i = 0, 1, \ldots, p$ (i.e. starts from $H_i$) to state $j = 0$ is

$$Q_{i,0} = P(\lambda \ln(Q/w) + (1 - \lambda)H_i \leq 0).$$

Isolating $Q$ gives

$$Q_{i,0} = P \left( Q \leq w \exp \left( \frac{-(1 - \lambda)H_i}{\lambda} \right) \right).$$

$$= F_Q \left( w \exp \left( \frac{-(1 - \lambda)H_i}{\lambda} \right) \right).$$

The probability $Q_{i,j}$ that the Markov chain goes from state $i = 0, 1, \ldots, p$ (i.e. starts from $H_i$) to state $j = 1, 2, \ldots, p$ is

$$Q_{i,j} = P(H_j - \delta < \lambda \ln(Q/w) + (1 - \lambda)H_i < H_j + \delta).$$

Splitting the probability into two parts yields

$$Q_{i,j} = P(\lambda \ln(Q/w) + (1 - \lambda)H_i < H_j + \delta) - P(\lambda \ln(Q/w) + (1 - \lambda)H_i < H_j - \delta).$$

Isolating $Q$ again finally gives

$$Q_{i,j} = P \left( Q \leq w \exp \left( \frac{H_j + \delta - (1 - \lambda)H_i}{\lambda} \right) \right) - P \left( Q \leq w \exp \left( \frac{H_j - \delta - (1 - \lambda)H_i}{\lambda} \right) \right).$$

$$= F_Q \left( w \exp \left( \frac{H_j + \delta - (1 - \lambda)H_i}{\lambda} \right) \right) - F_Q \left( w \exp \left( \frac{H_j - \delta - (1 - \lambda)H_i}{\lambda} \right) \right).$$

References


Chen, G., 1997. The mean and standard deviation of the run length distribution of $\bar{X}$ charts when control limits are estimated. Statistica Sinica 7, 789–798.


