Comparing the Effects of Skewed Distributions on the $S$-chart and $S^2$-EWMA Chart Based on the Median Run Length Criterion

Ong Ker Hsin$^a$, Teh Sin Yin$^a$, Khoo Michael Boon Chong$^b$, Teoh Wei Lin$^c$, Soh Keng Lin$^a$

$^a$School of Management, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
$^b$School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
$^c$Department of Physical and Mathematical Science, Faculty of Science, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia

Abstract

The $S^2$-EWMA (called the $S$ square exponentially weighted moving average) control chart is effective in detecting small and moderate process variance shifts. Previously, the chart was designed based on the assumption that the distribution of the quality characteristic is normally distributed. This study designs the $S$-EWMA control chart for skewed distributions. The skewed distributions considered in this paper are the lognormal and gamma distributions. The performance of the $S^2$-EWMA control chart is compared with that of the traditional Shewhart $S$-chart, in terms of median run length (MRL), based on simulation using the Statistical Analysis System (SAS). The results show that regardless of the type of skewed distributions, sample size and skewness level, $\gamma$, in most of the cases, the $S^2$-EWMA chart outperforms the $S$-chart. Moreover, the findings reveal that the MRL performances of the $S$-chart and $S^2$-EWMA chart are significantly influenced by skewed distributions.

Keywords: Median run length, process variability, skewed distributions, $S$-chart, $S^2$-EWMA control chart

1.0 INTRODUCTION

Originally introduced by Walter Shewhart, the Shewhart control chart is a powerful tool commonly used for process monitoring purpose in Statistical Process Control [1]. The aim of a quality control chart is to reduce the process variability by monitoring and removing assignable causes of variation from the process [2]. Noting this, a control chart helps to prevent a process from delivering defective units and producing wastes. Generally, the Shewhart control chart is used when large process shifts are of interest while the exponentially weighted moving average (EWMA) control chart is one of the alternative charts when small process shifts are the major concern.

The median run length (MRL) is a performance measure that could be used in measuring control chart performances. It is the median number of samples plotted on a control chart before the first out-of-control sample is detected. The MRL criterion provides a more reliable interpretation of a chart performance compared with the average run length criterion [3].

Traditionally, a quality characteristic is assumed to be normally distributed. In other words, the data follow a normal or approximately normal distribution. This assumption is often valid, but in certain situations such as in semiconductor, chemical and cutting tool wear processes, this assumption is doubtful because these situations comprise processes that follow skewed distributions [4-6]. For a skewed population, as the skewness increases, the false alarm rate increases [5]. To date, there are a number of researchers who studied the control chart performances based on...
skewed distribution [5-7]. Previously, the control chart performances for both S-chart and \( S^2 \)-EWMA chart were investigated by assuming the distribution is normally distributed [8-9]. Thus, the objective of this paper is to study the effect of skewness on the performances of the S-chart and \( S^2 \)-EWMA chart based on MRL. The lognormal and gamma distributions are selected in this study as these two distributions provide a wide variety of shapes from nearly symmetric to highly skewed [7]. The organization of this paper is as follows: Section 2 reviews the S-chart and \( S^2 \)-EWMA chart. The simulation study in this work is presented in Section 3. Section 4 discusses the designs of the S-chart and \( S^2 \)-EWMA chart while Section 5 compares the performances of the S-chart and \( S^2 \)-EWMA chart, for skewed distributions based on the MRL. Lastly, conclusions are drawn in Section 6.

2.0 REVIEWS ON THE S-CHART AND \( S^2 \)-EWMA CHART

2.1 A Review on the S-chart

Equations (1) to (3) present the formulae for computing the upper control limit (UCL), centre line (CL) and lower control limit (LCL) of the S chart when the standard deviation, \( \sigma \) of a process is known [10].

\[
\begin{align*}
UCL &= c_4 \sigma + 3\sigma \sqrt{1 - c_4^2}, \\
CL &= c_4 \sigma, \quad \text{and} \\
LCL &= c_4 \sigma - 3\sigma \sqrt{1 - c_4^2},
\end{align*}
\]

where \( c_4 \) is a constant that depends on the sample size. The values of \( c_4 \) can be obtained from Appendix VI in [11]. When \( \sigma \) is unknown, it has to be estimated from past data. Assume that \( m \) preliminary samples, each of size \( n \), are taken and let \( S \) be the standard deviation of the \( k \)th sample. Then the following formula is used to calculate the average sample standard deviation [11]:

\[
\bar{S} = \frac{1}{m} \sum_{k=1}^{m} S_k, \quad m = 1, 2, \ldots.
\]

The statistic \( \frac{\bar{S}}{c_4} \) is an unbiased estimator of \( \sigma \). Thus, the limits of the S-chart are

\[
\begin{align*}
UCL &= \bar{S} + \frac{3\bar{S}}{c_4} \sqrt{1 - c_4^2}, \\
CL &= \bar{S}, \quad \text{and} \\
LCL &= \bar{S} - \frac{3\bar{S}}{c_4} \sqrt{1 - c_4^2}
\end{align*}
\]

Note that the sample standard deviation is

\[
S_k = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (X_{kj} - \bar{X}_k)^2}
\]

where \( X_{kj} \) is the \( j \)th observation in sample \( k \) and \( n \) is the sample size. Note that \( \bar{X}_k \) is the mean of sample \( k \).

2.2 A Review on the \( S^2 \)-EWMA Chart

The ARL measure for constructing the \( S^2 \)-EWMA chart for normal distribution was originally proposed by [8], [9] extended the study by introducing the MRL measure in designing the \( S^2 \)-EWMA chart. The \( S^2 \)-EWMA chart is a control chart used for monitoring the process variance. When the standard deviation shifts from \( \sigma_0 \) to \( \sigma_1 \) (where \( \sigma_0 \) is the nominal process standard deviation and \( \sigma_1 \) is the new process standard deviation), the process is considered as out-of-control. For this chart, the mean remains at its nominal value \( \mu \) while the magnitude of the standard deviation shift is determined using the parameter \( \tau = \frac{\sigma_1}{\sigma_0} \), where \( \sigma_0 \) is assumed to be known in this paper. Note that \( S_k^2 \) is the variance of sample \( k \), i.e.

\[
S_k^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_{kj} - \bar{X}_k)^2.
\]

[8] proposed the following transformation on \( S_k^2 \) in order to monitor the process dispersion, i.e.

\[
T_k = u + v \ln(S_k^2 + w),
\]

where \( u, v \) and \( w > 0 \) (to avoid problems arising with the logarithmic transformation) are three constants. Following this, the classical EWMA approach is applied on \( T_k \) statistic,

\[
Z_k = (1 - \lambda) Z_{k-1} + \lambda T_k,
\]

where the smoothing constant, \( \lambda \) has to satisfy \( 0 < \lambda \leq 1 \). This is to ensure that the distribution of \( T_k \) will be quasi-symmetrical. Besides, it is also to make sure that the distribution looks like a standard normal distribution if the constants \( u, v \) and \( w \) are properly selected.

Corresponding to the \( Z_k \) statistic, control limits for \( S^2 \)-EWMA chart are [8]

\[
\begin{align*}
LCL_{SE} &= E(T_k) - K \times \sqrt{\frac{\lambda}{2 - \lambda}} \times \sigma(T_k), \\
UCL_{SE} &= E(T_k) + K \times \sqrt{\frac{\lambda}{2 - \lambda}} \times \sigma(T_k).
\end{align*}
\]

Note that \( K > 0 \), \( E(T_k) \) and \( \sigma(T_k) \) are the theoretical mean and standard deviation for \( T_k \), respectively. The constants \( u, v \) and \( w \) are equal to [8]

\[
\begin{align*}
u &= V(n), \\
w &= W(n)\sigma_0^2.
\end{align*}
\]
and
\[ u = U(n) - 2V(n)\mu_0(\sigma_0) \]  
(16)
in order to calculate the value of \( T_e \). Note that the \( U(n) \), \( V(n) \) and \( W(n) \) are three functions that depend completely on the sample size, \( n \).

According to [8], the probability density function (pdf) \( f_\mu(T_e) \) of \( T_e \) that depends solely on \( n \), is as follows:
\[ f_\mu(T_e) = \frac{1}{V(n)} \exp\left( \frac{t - U(n)}{V(n)} \right) \exp\left( \frac{t - U(n)}{V(n)} \right) - W(n) \left| n^{-1/2} \right| \]
(17)
where \( f_\mu \) is the pdf of a gamma distribution with parameters \( \frac{n-1}{2} \) and \( \frac{2}{n-1} \). The pdf \( f_\mu(T_e) \) shows that the computation of the values of \( E(T_e) \) and \( \sigma(T_e) \) do not depend on the value of \( \sigma_0 \). The computation of \( E(T_e) \) and \( \sigma(T_e) \) were derived by [8] via the numerical quadrature. As the values of \( E(T_e) \) is close to zero, it is assumed that \( E(T_e) = 0 \). Note that the value of \( Z_0 \) can be obtained through
\[ Z_0 = U(n) + V(n)n[1 + W(n)] \]  
(18)
From Equation (18), it is obvious that \( Z_0 \) depends on \( n \) only but not on \( \sigma_0 \). The value of \( Z_0 \) can also be replaced with zero since the value of \( Z_0 \) found to be very close to zero. By referring to [8], the derivative of \( T_e \) has the distribution of the transformed random variable \( r^2S^2 \) with pdf
\[ f_r(t|n, \tau) = \frac{1}{V(n)} \exp\left( \frac{t - U(n)}{V(n)} \right) \exp\left( \frac{t - U(n)}{V(n)} \right) - W(n) \left| n^{-1/2} \right| \]
(19)
This is the reason why the distribution \( f_r(t|n, \tau) \) of \( T_e \) only depends on \( n \) and \( \tau \).

### 3.0 Simulation Study

Skewness measures the degree of asymmetry of a distribution. A data set or a distribution is symmetric when the median divides the right side and the left side into two identical regions. Sample skewness can be measured using the following equation [12]:
\[ \text{Skewness} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{(n-1)S^3}, \]  
(20)
where \( \bar{X} \) is the sample mean and \( n \) is the number of data points while \( S \) is the sample standard deviation. In general, the skewness for a symmetric distribution is zero. Positive values signify that the data are positively (right) skewed and negative values signify that the data are negatively (left) skewed.

This study considers two types of distributions. They are the lognormal and gamma distributions. These distributions are selected because they have high flexibility, they are able to represent a wide range of shapes, from nearly symmetric to highly skewed, by appropriately selecting the parameters.

For a lognormal distribution, its cdf is given by [13] as
\[ F(j) = \Phi\left( \frac{\log j - \omega}{\pi} \right), \]  
(21)
where \( \omega \) is the location parameter and \( \pi \) is the scale parameter. For this distribution, the location parameter of zero is chosen. By letting \( \omega = 0 \) and \( P_j = Pr(J \leq \mu) \) where \( \mu \) is the targeted mean value of \( J \), the equation becomes
\[ P_j = \Phi\left( \frac{\pi}{2} \right). \]  
(22)
The \( \pi \) values which satisfy the given \( P_j \) in Equations (22) can be uniquely obtained with a numerical method.

For a gamma distribution with location parameter \( \mu \) and scale parameter of one, its cdf is given by [14] as
\[ F(j) = \frac{\Gamma(j|\mu)}{\Gamma(\mu)}, \text{ for } j \geq 0, \]  
(23)
where \( \Gamma(j|\mu) = \int_{0}^{j} r^{\mu-1} e^{-r} dr \) and \( \Gamma(\mu) = \int_{0}^{\infty} r^{\mu-1} e^{-r} dr \). In this case,
\[ P_j = F(\mu) \]  
(24)
as \( \mu = \theta \). Note that \( \theta \) denotes the shape parameter of the gamma distribution.

The skewness coefficient, \( \gamma \), is unique for a given value of \( \pi \) or \( \theta \). The shape parameters, \( \pi \) for the lognormal distribution and \( \mu \) for the gamma distribution, are computed in the case that the skewness coefficient, \( \gamma \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\} \). The skewness coefficient \( \gamma = 0.0 \) represents symmetric distribution; \( \gamma = 0.5 \) and 1.0 represent low levels of skewness; \( \gamma = 1.5 \) and 2.0 represent moderate levels of skewness while \( \gamma = 2.5 \) and 3.0 represent high levels of skewness. A shift in the mean is represented by \( \mu_1 = \mu_0 + \delta \sigma_0 \), in which \( \delta > 0 \) is the magnitude of a shift, in terms of the number of standard deviation units, while \( \mu_0 \) and \( \sigma_0 \) represent the in-control mean and in-control standard deviation respectively. Note that this paper only considers the in-control process, i.e. when \( \delta = 0.0 \). For a random variable, \( J \), from the lognormal and gamma distributions, their in-control means are
\[ \mu_0 = e^{\bar{X}}, \]  
(25)
and
\[ \mu_1 = \theta, \]  
(26)
respectively, whereas their in-control standard deviations are
\[ \sigma_1 = \sqrt{e^{2\bar{X}} - 1}, \]  
(27)
and 
\[ \sigma_{j,0} = \sqrt{\theta} \]  
respectively [15].

4.0 DESIGNS OF S-CHART AND S²-EWMA CHART

The steps in designing the S-chart based on MRL are outlined as follows:

Step 1: Determine the value of \( MRL_0 \) and \( n \).

Step 2: Let \( p = 0.5 \) as the 50\(^{th}\) percentile or median of the run length distribution is considered. Calculate the Type-I error, \( \alpha \) by solving Equation (29).

\[ MRL \geq \frac{\ln(1-p)}{\ln(1-\alpha)} \]  

(29)

Step 3: Divide the level of significance, \( \alpha \), by 2 because the two-tailed test is used, corresponding to both tails of the probability density function.

Step 4: Compute the nominal value of the process standard deviation, \( \sigma_0 \). In this study, the value of \( \sigma_0 \) corresponds to the standard deviation of the skewed distribution which is influenced by the shape parameter.

Step 5: Compute the value of \( LCL \) so that the tail probability of the density function of \( S \) is closest to \( \alpha /2 \), where \( NIntegrate \) gives a numerical approximation to the integral; \( s \) is the value of the standard deviation; and \( Gamma \) represents a gamma function [10]. This is to ensure that the \( MRL_0 \) value specified in Step 1 is attained. Note that the \( LCL \) is computed via the following code using Mathematica 9.0.

\[ LCL = NIntegrate\left\{ \frac{2^{3-\nu} \Gamma(n-\nu) \Gamma(\frac{1}{2}) \Gamma(\frac{n}{2})}{\Gamma(n-1) \Gamma(n-\nu) \sigma_0^2} \exp\left( -\frac{\text{MRL} - \bar{\text{UCL}}}{\sigma_0} \right) \right\} \{0, \text{LCL}\} \]

(30)

Step 6: Compute the \( \text{UCL} \) so that the tail probability of the density function of \( S \) is closest to \( \alpha /2 \) in order to attain the \( MRL_0 \) specified in Step 1. Note that the \( \text{UCL} \) is computed via the following code using Mathematica 9.0.

\[ \text{UCL} = NIntegrate\left\{ \frac{2^{3-\nu} \Gamma(n-\nu) \Gamma(\frac{1}{2}) \Gamma(\frac{n}{2})}{\Gamma(n-1) \Gamma(n-\nu) \sigma_0^2} \exp\left( -\frac{\text{MRL} - \bar{\text{LCL}}}{\sigma_0} \right) \right\} \{0, \text{UCL}\} \]

(31)

Step 7: Calculate the S statistic via Equation (8) in Section 2.1.

Steps 1 to 7 are repeated for 10,000 simulation trials to compute the \( MRL_0 \). The SAS program is available upon request from the corresponding author. As for the \( S^2 - \) EWMA chart, please refer to [9] for the design of this chart.

5.0 A COMPARISON OF THE PERFORMANCES FOR THE S-CHART AND S²-EWMA CHART

The desired sample sizes, \( n = 3, 5, 7 \) and 9 with the in-control median run length of 200 are considered in this simulation study. The optimal \( (\lambda, K) \) combinations for the \( S^2 - EWMA \) chart are obtained via SAS Monte Carlo simulations. The distributions considered in this study are the lognormal and gamma distributions. Table 1 shows the \( MRL_0 \) for sample sizes, \( n = 3, 5, 7 \) and 9 for both S-chart and \( S^2 - EWMA \) chart. The closer the in-control MRL (\( MRL_0 \)) is to 200, the better is the chart performance.
Table 1 In-control run lengths for $S$-chart and $S^2$-EWMA chart when the underlying distributions are skewed, based on $MRL_0$ of 200 and $n \in \{3, 5, 7, 9\}$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>$\gamma$</th>
<th>$n = 3$</th>
<th>$n = 5$</th>
<th>$n = 7$</th>
<th>$n = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S$-chart</td>
<td>$S^2$-EWMA</td>
<td>$S$-chart</td>
<td>$S^2$-EWMA</td>
</tr>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td>0.010</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1656</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3170</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\pi$</td>
<td>0.4484</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5593</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6525</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.7315</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Gamma</td>
<td></td>
<td></td>
<td>38000</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.4</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.913</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\theta$</td>
<td>1.788</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.983</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.648</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.442</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Generally, irrespective of the type of skewed distribution, sample size and level of skewness, it is noticeable in Table 1 that most of the $MRL_0$ values (around 94.64%) of the $S^2$-EWMA chart are closer to 200 compared with the corresponding ones of the $S$-chart. For example, for the lognormal distribution, when $n = 9$ and $\pi = 0.3170$, the $MRL_0$ values for the $S$-chart and $S^2$-EWMA chart are 37 and 76, respectively. As for the gamma distribution, however, when $n = 9$ and $\theta = 3.913$, the corresponding $MRL_0$ values for the $S$-chart and $S^2$-EWMA chart are 42 and 84 respectively. Although there are a few exceptional cases (about 5.36%) for the gamma distribution, where the $MRL_0$ value of the $S$-chart is closer to 200 compared with the corresponding value of the $S^2$-EWMA chart, the $MRL_0$ values of the $S$-chart and $S^2$-EWMA chart do not differ much. For instance, when $n = 5$ and $\theta = 38000$, the $MRL_0$ values are 205 for the $S$-chart and 206 for the $S^2$-EWMA chart.

Other than that, it is also worthy to note that the $MRL_0$ decreases when the sample size increases for the case where the level of skewness, $\gamma$ remains unchanged. This is true for the $S$-chart and $S^2$-EWMA chart. This indicates that the $S$-chart and $S^2$-EWMA chart are less effective when the sample size increases.

When the skewness level, $\gamma$, has the value of zero, the distribution of the data is symmetric. Thus, the $MRL_0$'s under both skewed distributions would approach the $MRL_0$ value of the normal distribution, i.e., 200. Moreover, regardless of the type of distribution (lognormal or gamma distribution) and sample size, when the level of positive skewness, $\gamma$, increases, the $MRL_0$ decreases for both charts. A reduction in $MRL_0$ implies a higher incidence of false alarms. The $MRL_0$ value of the $S$-chart decreases drastically compared with that of the $S^2$-EWMA chart. Thus, the performance of the $S$-chart is more severely
affected by skewed distributions compared with the \( S^2 \)-EWMA chart.

6.0 CONCLUSIONS

For skewed populations, regardless of the level of skewness, \( \gamma \), type of skewed distribution and sample size, the \( S^2 \)-EWMA chart is found to have a more favorable \( \text{MRL}_0 \) performance than the \( S \)-chart. This is because the \( S^2 \)-EWMA chart gives \( \text{MRL}_0 \) values that are closer or equal to 200 than the \( S \)-chart in most of the cases, under both skewed populations. In short, the performance of the \( S^2 \)-EWMA chart is better than the \( S \)-chart not only when the underlying distribution is normal, but also when the underlying distribution is skewed, as shown by the \( \text{MRL}_0 \)s in Section 5. Note that this study has not taken into consideration the autocorrelated process. Process data are always assumed to be statistically independent. This may not be the case in some industries. Besides, the heavy tailed distributions are not considered in this study too. Therefore, some noteworthy future studies related to this topic are as follows:

(i) To evaluate the performance of the \( S^2 \)-EWMA chart based on an autocorrelated process, i.e. when the independence assumption is violated.

(ii) To investigate the performance of the \( S^2 \)-EWMA chart based on heavy tailed distributions, such as the Student-\( t \) or Cauchy distribution.

Acknowledgement

The authors would like to acknowledge the work that has produced this paper is funded by the Universiti Sains Malaysia (USM) Short Term Grant, number 304/PMGT/6312129.

References