Optimal Design of the Double Sampling $\bar{X}$ Chart Based on Median Run Length

Wei Lin Teoh and Michael B. C. Khoo

Abstract—The double sampling (DS) $\bar{X}$ chart is effective in detecting small to moderate process mean shifts and reducing the sample size. The average run length (ARL) is an intuitively appealing and widely used optimization criterion in the design of a control chart. Nevertheless, the skewness of the run length distribution changes with the size of the process mean shift; thus, ARL is not necessarily a good representative of the run length distribution. In such a case, the median run length (MRL) provides a more reliable interpretation, for the in-control and out-of-control performances of a control chart. In this paper, an optimal design of the DS $\bar{X}$ chart based on MRL is proposed. New optimal parameters are provided for selected in-control median run lengths, MRL$_{\delta}$ $\in$ {250, 500} and in-control average sample sizes, ASS$_{\delta}$ $\in$ {3, 5, 7, 9}. These optimal parameters facilitate the implementation of the DS $\bar{X}$ chart in practice.

Index Terms—Average sample size, double sampling $\bar{X}$ chart, median run length, optimization.

I. INTRODUCTION

Contemporarily, the quality of the final products and customers’ satisfaction are vitally viewed in the manufacturing and service processes, such as in the chemical industries, food industries and health care services. Statistical Process Control (SPC) is a collection of statistical tools that provide improvement in yield and reduction in production costs. A control chart is one of the most valuable tools in SPC to reduce variability in key parameters and produce conforming products. Because of the operational simplicity, the Shewhart $\bar{X}$ chart is extensively used to detect large process mean shifts in industries. However, it is relatively insensitive in detecting small and moderate process mean shifts.

To overcome this problem, Daudin [1] applied the concept of double sampling plans to propose the DS $\bar{X}$ chart, which is viewed as a two-stage Shewhart $\bar{X}$ chart. The optimization model to minimize the in-control average sample size (ASS$_{\delta}$) is presented in Daudin’s [1] paper. From the statistical viewpoint, Irianto and Shinozaki [2] constructed an optimization model to minimize the out-of-control average run length (ARL$_{\delta}$). Daudin [1] and Costa [3] compared the performances of the DS $\bar{X}$ chart with the Shewhart, EWMA, CUSUM, variable sampling interval (VSI) and variable sample size (VSS) charts. They concluded that the DS $\bar{X}$ chart has a better performance in certain cases. In light of this advantage, recently, Torng et al. [4], Irianto and Juliani [5], Costa and Machado [6], Khoo et al. [7], and Lee et al. [8], to name a few, contributed to the area of DS control charts.

The sole dependence on the average run length (ARL) as a measure of a chart’s performance has been subjected to criticism in recent years (see [9]-[11]). Since the skewness of the run length distribution changes from highly skewed when the process is in-control to approximately symmetric when the process mean shift is large, interpretation based on ARL alone could be misleading. For instance, the DS $\bar{X}$ chart with (n$_1$, n$_2$, L$_1$ = 0.253, L = 5.046, L$_2$ = 3.067) has an in-control ARL (ARL$_{0}$) of 500, where the notations (n$_1$, n$_2$, L$_1$, L, L$_2$) of the DS $\bar{X}$ chart are defined in Section II. However, it is found that 50% of all the run lengths are less than 347 (i.e. the in-control median run length, MRL$_{\delta}$ = 347) and about 63% of all the run lengths are less than 500. In view of this disadvantage, Chakraborti [10] pointed out that the median run length (MRL) provides a more accurate measure of a chart’s performance. For example, the DS $\bar{X}$ chart with the parameters stated above has an out-of-control MRL (MRL$_{\delta}$) of 23 at shift $\delta$ = 0.5. This indicates that in 50% of the time, the out-of-control run length is less than or equal to 23 samples. Note that all the MRLs and the percentage point of the run length distribution are calculated by using the formulae shown in Section II. In short, interpretation based on MRL is more readily comprehended.

Therefore, MRL is suggested as an alternative performance criterion to design control charts. For related literatures, see Gan [12]-[13], Maravelakis et al. [14], Golosnoy and Schmid [15], Low et al. [16], and Khoo et al. [17].

This paper is structured as follows: First, Section II briefly introduces the DS $\bar{X}$ chart and its run length properties. Since MRL can increase the quality control engineers’ confidence in understanding a control chart, the main contributions of this paper are to propose an optimization
model for the DS $X$ chart by minimizing MRL$_i$ and provide specific optimal combinations for MRL$_{ij} \in \{250, 500\}$ and ASS$_i \in \{3, 5, 7, 9\}$, which are presented in Section III. Conclusions are drawn in Section IV.

| TABLE I: OPTIMAL $(n_1, n_2, L_1, L_2)$ COMBINATION (FIRST ROW OF EACH CELL) AND (MRL$_i$, ASS$_i$) VALUES (SECOND ROW OF EACH CELL) OF THE DS $X$ CHART WHEN MRL$_i \in \{250, 500\}$, ASS$_i \in \{3, 5, 7, 9\}$ and $\delta_{opt} = \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ |
|-----------------|-----------------|-----------------|-----------------|
| $\delta_{opt}$ | ASS$_i = 3$ | ASS$_i = 5$ | ASS$_i = 7$ | ASS$_i = 9$ |
|-----------------|-----------------|-----------------|-----------------|
| 0.2 | (1, 14, 1.465, 4.093, 2.527) | (3, 12, 1.383, 4.186, 2.775) | (5, 10, 1.282, 4.513, 2.904) | (4, 11, 0.748, 4.268, 2.947) |
| 0.4 | (1, 14, 1.464, 3.608, 2.565) | (17, 3.437) | (6, 9, 1.593, 4.025, 2.885) | (12, 9.258) |
| 0.6 | (1, 13, 1.425, 3.875, 2.567) | (2, 12, 1.150, 3.756, 2.803) | (3, 12, 0.967, 3.537, 2.928) | (2, 13, 0.615, 3.508, 2.914) |
| 0.8 | (1, 11, 1.335, 3.894, 2.635) | (2, 13, 1.020, 3.692, 2.763) | (1, 10, 0.524, 2.417, 2.900) | (2, 11, 0.748, 3.714, 2.966) |
| 1.0 | (1, 7, 1.066, 3.481, 2.853) | (2, 4.429) | (1, 10, 0.524, 2.417, 2.900) | (1, 12, 7.127) |
| 1.2 | (2, 6, 1.382, 3.748, 2.842) | (1, 5.626) | (5, 3, 0.430, 3.400, 3.039) | (6, 4, 0.318, 3.413, 3.039) |
| 1.4 | (1, 4, 0.674, 3.999, 2.934) | (1, 4.121) | (5, 3, 0.430, 3.400, 3.039) | (6, 4, 0.318, 3.413, 3.039) |
| 1.6 | (1, 3, 0.430, 3.400, 3.046) | (1, 3.593) | (5, 3, 0.430, 3.400, 3.039) | (5, 5, 0.253, 3.419, 3.047) |
| 1.8 | (1, 3, 0.430, 3.400, 3.046) | (1, 3.618) | (4, 4, 0.318, 3.413, 3.048) | (5, 5, 0.253, 3.419, 3.047) |
| 2.0 | (1, 3, 0.430, 3.400, 3.046) | (1, 3.606) | (4, 4, 0.318, 3.413, 3.048) | (4, 6, 0.210, 3.422, 3.052) |

II. THE DS $X$ CHART

It is assumed that the observations of the quality characteristic $X$ are independently and identically distributed (iid) normal random variables, with an in-control mean $\mu_0$ and an in-control variance $\sigma_i^2$. Let $L_i$ and $L_2$ be the warning and control limits of the first sample, respectively; whereas $L_i$ be the control limit of the combined samples. The regions in Fig. 1 are represented as $I_i = [-L_i, L_i]$, $I_2 = [-L_i - L_i] \cup L_i$, $I_3 = (-\infty, -L) \cup (L, +\infty)$, and $I_4 = [-L_2, L_2]$. By referring to Fig. 1, the design procedure of the DS $X$ chart demonstrated by Daudin [1] is as follows:

1. Take a first sample of size $n_1$ and compute the sample mean $X_{i_1} = \sum_{j=1}^{n_1} X_{i_1,j} / n_1$, where $X_{i_1,j}$ with $j = 1, 2, \ldots, n_1$ is the $j$th observation in the $i$th sampling time of the first sample.
2. If $Z_{i_1} = \left[ (X_{i_1} - \mu_0) / \sigma_i \right] / \sigma_i \in I_1$, the process is considered as in-control.
3. If $Z_{i_1} \in I_2$, the process is concluded as out-of-control.
4. If $Z_{i_1} \in I_3$, take a second sample of size $n_2$ and calculate the sample mean $X_{i_2} = \sum_{j=1}^{n_2} X_{i_2,j} / n_2$, where $X_{i_2,j}$ is the $j$th observation in the $i$th sampling time of the second sample.

For $n_1, n_2 \in \{250, 500\}$, ASS$_i \in \{3, 5, 7, 9\}$ and $\delta_{opt} = \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$, the optimal combinations and control limits are given in Table I.

with \( j = 1, 2, \ldots, n \) is the \( j \)th observation in the \( i \)th sampling time of the second sample.

5) At the \( i \)th sampling time, calculate the combined-sample mean
\[
X_i = \left( n_1 \bar{X}_i + n_2 \bar{X}_{i2} \right) / (n_1 + n_2),
\]
where \( \bar{X}_i \) is obtained as
\[
\bar{X}_i = \left[ (\bar{X}_i - \mu_0) / \sigma_0 \right] / \sigma_0 \in L_i,
\]
the process is declared as in-control; otherwise, it is deemed as out-of-control.

Note that the \( i \)th sampling time refers to the \( i \)th time when either the first sample of size \( n_1 \) only or the combined samples of size \( n_1 + n_2 \), respectively are collected.

It is further assumed that \( \mu_0 \) and \( \sigma_0^2 \) is known. Let \( P_{in} \) be the probability that the process is regarded as in-control at stage \( k \) with \( k \in \{1, 2\} \). Then \( P = P_{in} + P_{out} \) is the probability of the in-control process, with (see [1])
\[
P_{in} = P(Z_i \in I_1) = \Phi \left( L_1 + \delta \sqrt{\frac{n_1}{n_2}} \right) - \Phi \left( -L_1 + \delta \sqrt{\frac{n_1}{n_2}} \right),
\]
and
\[
P_{out} = P(Z_i \in I_4) = \int_{z \in I_4} \left[ \Phi \left( cL_2 + \sigma_0(z) \sqrt{\frac{n_1}{n_2}} \right) - \Phi \left( -cL_2 + \sigma_0(z) \sqrt{\frac{n_1}{n_2}} \right) \right] \phi(z) dz,
\]
where \( \Phi(.) \) and \( \phi(.) \) are the cumulative distribution function (cdf) and probability density function (pdf) of the standard normal random variable, respectively. Here, \( \delta = |\mu - \mu_0| / \sigma_0 \) is the magnitude of the standardized mean shift with the out-of-control mean \( \mu_0 \), \( r = \sqrt{n_1 + n_2} \), \( c = r / \sqrt{n_2} \) and \( I_1 = \left[ -L_1 + \delta \sqrt{\frac{n_1}{n_2}}, -L_1 + \delta \sqrt{\frac{n_1}{n_2}} \right] \). Note that a run length (RL) is the number of samples taken before the chart signals the first out-of-control condition. Montgomery [18] pointed out that the run length distribution of a Shewhart chart is the geometric distribution. It is known that the DS \( \bar{X} \) chart is a Shewhart type chart; thus, all the run length properties of the DS \( \bar{X} \) chart can be characterized by the geometric distribution. Hence, the cdf of RL, i.e. \( F_{RL}(\ell) \) is obtained as
\[
F_{RL}(\ell) = P (RL \leq \ell) = 1 - P_{\ell}'
\]
where \( \ell \in \{1, 2, 3, \ldots\} \). Then the MRL of the DS \( \bar{X} \) chart is equal to (see [12] and [13])
\[
P(\text{RL} \leq \text{MRL} - 1) \leq 0.5 \quad \text{and} \quad P(\text{RL} \leq \text{MRL}) > 0.5.
\]
Daudin [1] also showed that the average sample size (ASS) at each sampling time is equal to
\[
\text{ASS} = n_1 + n_2 P(Z_i \in I_2),
\]
where
\[
P(Z_i \in I_2) = \Phi \left( L_1 + \delta \sqrt{\frac{n_1}{n_2}} \right) - \Phi \left( -L_1 + \delta \sqrt{\frac{n_1}{n_2}} \right) + \Phi \left( -L_1 - \delta \sqrt{\frac{n_1}{n_2}} \right) - \Phi \left( -L_1 - \delta \sqrt{\frac{n_1}{n_2}} \right).
\]

III. OPTIMAL DESIGN OF THE DS \( \bar{X} \) CHART BASED ON MRL

In this paper, the performance of the DS \( \bar{X} \) chart is evaluated, in terms of the MRL and ASS. When the process is in-control, the MRL is denoted as MRL; while the out-of-control MRL is denoted as MRL. Similarly, two ASSs are usually of interest, namely the in-control ASS, ASS and the out-of control ASS, ASS. In this section, an optimal design of the DS \( \bar{X} \) chart for minimizing MRL is developed. Here, \( \delta_{opt} \) represents the size of a process mean shift, for which a quick detection is required. It is noted that when MRL, ASS and \( \delta \) are fixed, a control chart is considered superior to its competitors if it has the smallest MRL value. Similarly, a DS \( \bar{X} \) chart, for optimally detecting a desired shift is obtained when the optimal parameters giving the lowest MRL, are identified from all the possible \( (n_1, n_2, L_1, L_2) \) combinations. The proposed optimization model is illustrated as follows:
\[
\text{Min} \quad \text{MRL} \left( \delta_{opt} \right),
\]
Subject to
\[
\text{MRL} = \tau,
\]
where \( \tau \) is the expected in-control median run length;
\[
\text{ASS} = n,
\]
where \( n \) is the expected in-control average sample size;
\[
1 \leq n_1 < n < n_1 + n_2 \leq n_{\text{max}},
\]
where \( n_{\text{max}} \) is the upper bound of \( n_1 + n_2 \).

A nonlinear optimization algorithm is employed to determine the optimal \( (n_1, n_2, L_1, L_2) \) combination for each case considered in this paper. Practically, either a small or a moderate sample size is adopted in industry; thus, \( n_{\text{max}} = 15 \) is set in this paper. It must be emphasized that MRL is an integer due to the discrete property of RL; hence, there may exist several optimal \( (n_1, n_2, L_1, L_2) \) combinations for which the MRL value is minimum at a specific \( \delta_{opt} \neq 0 \). In such a case, the DS \( \bar{X} \) chart with the optimal \( (n_1, n_2, L_1, L_2) \) combination having the smallest ASS, is preferred.

The optimal \( (n_1, n_2, L_1, L_2) \) combinations and their corresponding \( (\text{MRL}, \text{ASS}) \) values for different combinations of MRL, ASS and \( \delta_{opt} \), are presented in the first and second rows of each cell in Table I, respectively. The above proposed optimization model (7)-(10) is applied
here to obtain the optimal \((n_1, n_2, L_1, L, L_2)\) combinations. It must be accentuated that \(MRL_0 = \tau \in \{250, 500\}\) (constraint (8)) and \(ASS_0 = n \in \{3, 5, 7, 9\}\) (constraint (9)) are attained for each case shown in Table I. The results in Table I have been verified with simulation. These new optimal \((n_1, n_2, L_1, L, L_2)\) combinations facilitate the implementation of the DS \(\bar{X}\) chart in practice. For example, consider a continuous chemical process in which a quick detection is desired at a shift with magnitude \(\delta_{opt} = 1.0\). If \(MRL_0 = 250\) and \(ASS_0 = 5\) are selected, Table I suggests using \((n_1 = 3, n_2 = 7, L_1 = 1.066, L = 3.481, L_2 = 2.976)\) as the optimal parameters to detect such a shift. Additionally, Table I tells us that 50% of the time, a shift with magnitude \(\delta = 1.0\) is detected by the first sample (i.e., \(MRL_1 = 1\)). Also, the chart needs 7.967 observations on the average to detect such a shift.

The choice of \(MRL_0\) primarily depends on the rate of production and the rate of sampling. Accordingly, two \(MRL_0\)s are considered in this paper, i.e. \(MRL_0 \in \{250, 500\}\). Examination of Table I reveals that a large \(MRL_0\) issues false alarms less frequently than a small \(MRL_0\); however, the latter responds quicker to an out-of-control condition, especially for a small \(\delta_{opt}\). As expected, the sensitivity of the DS \(\bar{X}\) chart in detecting a particular shift increases as \(ASS_0\) increases. This improvement is more pronounced for small \(\delta_{opt}\). It is interesting to note that for moderate and large shifts \((\delta_{opt} \geq 1.2)\), the DS \(\bar{X}\) chart has the same minimum \(MRL_1\) value regardless of the \(ASS_0\) used. Daudin [1] stated that the DS chart is designed, where the sample size should be increased if there is a higher chance of inferior quality. Therefore, it is observed that for most of the cases in Table I, the \(ASS_0\) taken to detect an out-of-control situation is higher than the corresponding \(ASS_0\). Apparently, some of the optimal \((n_1, n_2, L_1, L, L_2)\) combinations are optimally detecting a range of shifts. For instance, when \(MRL_0 = 500\) and \(ASS_0 = 7\), the optimal \((n_1 = 5, n_2 = 3, L_1 = 0.430, L = 3.585, L_2 = 3.246)\) combination is optimally detecting the shifts in the range \(1.2 \leq \delta_{opt} \leq 1.8\). This is a favorable property of the DS \(\bar{X}\) chart, optimized based on MRL, as it is sometimes difficult to predict the actual process mean shift in practice.

IV. CONCLUSION

A good understanding of the control charts used is crucial to engineers and shop floor personnel. Since the run length distribution changes in accordance to the magnitude of mean shifts, the MRL is more readily interpreted than the ARL. Moreover, the MRL provides more credible information for practitioners. This paper demonstrates the optimization design of the DS \(\bar{X}\) chart, based on MRL and thus, complements the work of Irianto and Shinozaki [2], who proposed the DS \(\bar{X}\) chart, optimized based on the ARL. Furthermore, specific optimal charting parameters, based on MRL are provided to aid practitioners in implementing the DS \(\bar{X}\) chart instantaneously.

REFERENCES


Wei Lin Teoh is a Ph.D. student in the School of Mathematical Sciences, Universiti Sains Malaysia (USM). She received her Bachelor of Science with Education (Honours) in Mathematics from USM in 2010.

Michael B.C. Khoo is a Professor in the School of Mathematical Sciences, Universiti Sains Malaysia (USM). He received his Ph.D. in Applied Statistics in 2001 from USM. He specializes in Statistical Process Control. He is a member of the American Society for Quality and serves as a member of the editorial boards of several international journals.