Performance Criteria for Evaluation of Control Chart for Phase II Monitoring

Sonam Jaiswal and Nirpeksh Kumar*1

*DST-Center of interdisciplinary mathematical science, Banaras Hindu University, Varanasi

*Department of Statistics, Banaras Hindu University, Varanasi

Abstract

Several performance measures are considered in literature to design and evaluate the control charts when the parameters are known such as average run length, median run length etc. However, when parameters are unknown, they are customarily estimated from phase I samples and then phase II control limits are constructed by using estimators in place of their respective parameters. It is well known that these performance metrics are found inappropriate to design and evaluate the phase II control charts because they become random variables being the function of estimators. Therefore, several measures are suggested for the same. Few of them reflect the unconditional performance whereas few reflect conditional performance of the chart. In this study, we discuss various performance criteria (conditional and unconditional) which might be useful in designing and evaluating the charts with estimated control limits.

Keywords: Conditional and unconditional performance, control chart, $\bar{X}$-chart, lower prediction bound, probability ratio.

1. Introduction

Statistical process control (SPC) consists of statistical tools used for monitoring and controlling the process variability. There are several SPC tools, mainly consists of seven magnificent tools. The seven magnificent tools of SPC includes histogram or stem and leaf plot, check sheet, pareto chart, cause-and-effect diagram, defect concentration diagram, scatter diagram, control chart. Among them, control chart is the most popular SPC tool which is used to detect unnatural variation in the process and gives a signal when shift in the process mean or variance has been occurred. It is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. A typical control chart consists of three horizontal lines, namely, center line (CL), upper control limit (UCL) and lower control limit
The CL represents the average value of the quality characteristic corresponding to the in-control state whereas UCL and LCL are chosen so that if the process is in control, nearly all of the sample points will fall between them. As long as the points lie within the control limits without any non-random pattern, the process is assumed to be in control and no action is necessary. However, a point that plots outside of the control limits, it is interpreted as evidence that the process is out of control and investigation and corrective action are required to find and eliminate the assignable cause or causes responsible for this behaviour[1]. Let $Y$ be the charting statistic which is plotted on the control chart with control limits, UCL and LCL which are defined as follows.

\[ P[Y > UCL] = \frac{a_0}{2} \quad \text{and} \quad P[Y < LCL] = \frac{a_0}{2} \tag{1} \]

where $a_0$ is the nominal false alarm rate (FAR). Let $F_\theta(\cdot)$ denote the cumulative distribution function (CDF) of $Y$, indexed by the parameter $\theta$ (it may be scalar or vector), then we can express the control limits as follows.

\[ UCL = F_\theta^{-1}\left(1 - \frac{a_0}{2}\right) = \xi_1(\theta_0) \quad \text{and} \quad LCL = F_\theta^{-1}\left(\frac{a_0}{2}\right) = \xi_2(\theta_0) \tag{2} \]

where $\theta_0$ is the known value of parameter $\theta$ and $\xi_1(\cdot)$, $\xi_2(\cdot)$ are the real-valued function which depend on $\theta_0$ for specified $a_0$. In particular, if the quality characteristic $Y$ follows normal distribution with mean $\mu$ and variance $\sigma^2$, then the control limits of $\bar{X}$-chart for monitoring the average level of quality characteristic are given by

\[ UCL = \mu_0 + z_{a_0/2} \frac{\sigma_0}{\sqrt{n}} \quad \text{and} \quad LCL = \mu_0 - z_{a_0/2} \frac{\sigma_0}{\sqrt{n}} \tag{3} \]

where $\mu_0$ and $\sigma^2_0$ are the specified values of $\mu$ and $\sigma^2$ respectively and $z_{a_0/2}$ is the 100$(1 - a_0/2)$th percentile of standard normal distribution. Usually, we consider the value of $z_{a_0/2}$ equal to 3 and then the control limits are known as 3-$\sigma$ control limits which are given by

\[ UCL = \mu_0 + 3 \frac{\sigma_0}{\sqrt{n}} \quad \text{and} \quad LCL = \mu_0 - 3 \frac{\sigma_0}{\sqrt{n}} \tag{4} \]

Note that when the parameters are not known, the estimated control limits are constructed by replacing the parameters by their respective estimates from phase I samples. It is well established that error in estimating parameters adversely affects the charts’ performance[2],[3]. Moreover, the performance of the chart cannot be predicted in the sense that if two practitioners use two different quantiles of the estimators to estimate the parameters, then both practitioners would have different control limits and hence may draw different inferences about the same process. For example, two practitioners may get different in-control (IC) average run length (will be defined in following section) values for their charts if they use different estimated values of the parameters[4].
It is worth noting that in earlier literature, the impact of estimation error on the performance of the charts were overlooked, however, in recent years, researches are more focused on examination of the effect of estimation error on the chart and for this purpose, several measures have been suggested in literature to evaluate and design the charts with estimated control limits, for example, measures based on conditional performance of the chart. The objective of this paper is to discuss several performance measures that are being used in literature to evaluate the estimated control charts and to make recommendations based on them.

Rest of the paper is organised as follows: In section 2, the performance measures for known parameter case are discussed. In Section 3, various performance measures are discussed for evaluation of phase II control limits and finally, concluding remarks are offered in Section 4.

2. Performance measures for known parameter case

The performance of a control chart is often judged in terms of certain characteristics associated with its run length distribution. The run length is a discrete random variable which may be defined as the number of charting statistics that must be plotted in order to give an out-of-control signal. One popular measure of the performance is the mean of run length distribution, called the average run length (ARL), which is defined as the average number of plotting points on the chart to give first OOC signal. Recall that the run length \( R_L \) follows a geometric distribution with parameter, say \( \beta \). We define

\[
\beta = P[\text{Signal}] = P[Y < LCL \text{ or } Y > UCL] = 1 + F_\theta(\xi_1(\theta_0)) - F_\theta(\xi_2(\theta_0))
\]

(5)

where \( Y \) is the charting statistic with distribution function \( F_\theta(\cdot) \), indexed by parameter \( \theta \). Thus, \( \beta \) depends on both \( \theta_0 \) and \( \theta \) and can be represented as \( \beta(\theta_0, \theta) \). If \( \theta = \theta_0 \), then the process is said to be IC and \( \beta \) is known as probability of type I error and when \( \theta \neq \theta_0 \), the process is OOC, \( 1 - \beta \) is known as probability of type II error. By definition, ARL is the expected value of the run length variable and is given by

\[
ARL = E(RL) = \frac{1}{\beta(\theta_0, \theta)}
\]

(6)

Let us denote IC ARL by \( ARL_{\text{in}} \) and OOC ARL by \( ARL_{\text{ooc}} \). It is desirable that a good control chart must have high \( ARL_{\text{in}} \) value and low \( ARL_{\text{ooc}} \) ARL values. In earlier studies, the standard deviation of run length distribution was also reported with ARL to reflect the variability in run lengths which is given by

\[
SD_{RL} = \sqrt{\frac{1-\beta(\theta_0, \theta)}{\beta^2(\theta_0, \theta)}}
\]

(7)

Because of the fact that the distribution of \( RL \) is significantly right-skewed and the skewness is more pronounced for small values of \( \beta \) typically encountered in SPC[2], ARL does not reflect...
the typical performance of the chart very well. Therefore, several researchers advocated the use of other measures as an alternative of ARL to assess the performance of the chart in more representative way[5],[6]. One such measure is a percentile of the run length distribution[7]. The percentile is an informative and robust measure of the performance. Note that the median run length (MRL) is 50% percentile which is being more preferred than the ARL in current literature to design the charts[6],[8],[9],[10],[11]. The 100 pth \( (0 < p < 1) \) percentile of the run length distribution, denoted by \( \kappa_p \) is defined as the smallest integer such that

\[
P[R \leq \kappa_p] \geq p
\]

(8)

we can express \( \kappa_p \) as follows.

\[
\kappa_p = \left\lfloor \frac{\ln(1-p)}{\ln(1-\beta(\theta_0, \theta))} \right\rfloor
\]

(9)

where \([x] \) denotes the greatest integer contained in \( x \). The IC percentiles can be obtained by substituting \( \theta = \theta_0 \) in (9). We can obtain MRL using (9) by letting \( p = 1/2 \), thus, we have

\[
MRL = \left\lfloor -\frac{0.69347}{\ln(1-\beta(\theta_0, \theta))} \right\rfloor
\]

(10)

In particular, for the \( \bar{X} \) chart with control limits in (4), the probability of signal \( \beta(\theta_0, \theta) \) is obtained as

\[
\beta(\delta) = 1 + \Phi(-3 - \delta \sqrt{n}) - \Phi(3 - \delta \sqrt{n})
\]

where \( \Phi(\cdot) \) denotes the CDF of standard normal variable and we assume that the characteristic follows normal distribution with mean \( \mu_1 = \mu_0 + \delta \sigma_0 \) and variance \( \sigma_0^2 \). The \( \delta \) represents the amount of shift and when \( \delta = 0 \), the process is IC, otherwise OOC. Next, we calculate various performance metrics which are discussed above for the \( \bar{X} \)-chart and report them in Table 1.

**Table 1: Percentiles and the average run length of the Shewhart \( \bar{X} \)-chart when process mean and variance are known with subgroup size \( n=5 \)**

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>ARL</th>
<th>Percentiles of the run length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>0</td>
<td>370.37</td>
<td>38</td>
</tr>
<tr>
<td>0.25</td>
<td>133.15</td>
<td>13</td>
</tr>
<tr>
<td>0.50</td>
<td>33.40</td>
<td>3</td>
</tr>
<tr>
<td>0.75</td>
<td>10.76</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4.50</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 shows that the IC MRL is 256 which is 45% smaller than the IC ARL when \( \delta = 0 \) whereas it is 64% smaller than ARL when \( \delta = 1 \). This implies that the shape of run length distribution depends heavily on \( \beta \) which is more skewed for higher values of \( \beta \) as shift increases. It is now well accepted in literature that there is an advantage to use MRL(or percentile) that
50% (or 100 \( p \)% of the run lengths will be less than the MRL (percentile) value irrespective of change in the shape of run length distribution\([6],[7]\). On the other hand, ARL only provides information about the expected run length. The entire run length distribution gives important information about the chart’s performance and a number of selected percentiles should be reported in addition to ARL to summarize this information follow the reference\([7]\).

3. Performance measures when the parameters are unknown

As stated earlier, when the parameter(s) are unknown, the ‘plug-in’ control limits are constructed by replacing the parameters by their suitable estimators. The parameters are estimated form a phase I samples from IC process and the resulting control limits are used for Phase II monitoring. It is well accepted in literature\([2],[12],[13],[14],[15]\) that whenever these estimated control limits are implemented, the properties of the control chart are fundamentally changed and the chart’s performance is significantly deviated from the chart with known parameters.

Recall that the run length \( R \) follows geometric distribution when the process parameters are known, however, when the plug-in control limits are used, the distribution of \( R \) is no longer geometric because the signalling events are dependent\([2]\). Traditionally, the unconditional run length (URL) distribution and associated characteristics are used to evaluate the chart’s performance when the parameters are unknown. Suppose the parameter \( \theta \) is estimated by \( \hat{\theta} \) by using \( m \) phase I samples, then the estimated control limits are obtained by replacing \( \theta \) by \( \hat{\theta} \) in \((2)\) as follows.

\[
\bar{UCL} = \xi_1(\hat{\theta}) \quad \text{and} \quad \bar{LCL} = \xi_2(\hat{\theta})
\]

(11)

It is worth to noting that conditionally on \( \hat{\theta} \), the run length variable follows a geometric distribution with probability of signal equal to \( \hat{\beta} \) where

\[
\hat{\beta} = P[Y < \xi_1(\hat{\theta}) \text{ or } Y > \xi_2(\hat{\theta})]
\]

(12)

Therefore, the distribution of URL, denoted by \( R_{uc} \) is given by

\[
P[R_{uc} \leq a] = 1 - \int_{\theta} \left(1 - \hat{\beta}\right)^a f(\theta) d\theta
\]

(13)

Thus, the average and standard deviation of \( R_{uc} \) are given by

\[
ARL_{uc} = \int_{\theta} \frac{1}{\hat{\beta}} f(\theta) d\theta
\]

(14)

\[
SD_{uc} = \sqrt{\left(\int_{\theta} \frac{2 - \hat{\beta}}{\hat{\beta}^2} f(\theta) d\theta - ARL_{uc}^2\right)}
\]

(15)

In like manner, the 100\( p \)-percentile of the unconditional run length distribution can be obtained as
\[ \int_0^\infty (1 - \hat{\theta})^\kappa \frac{e^{-\theta}}{\theta} d\theta \leq 1 - p \]  
(16)

Consider a Shewhart \( \bar{X} \)-chart for monitoring the average assuming that the characteristic \( X \) follows normal distribution with unknown mean \( \mu \) and unknown variance \( \sigma^2 \). Thus, the phase II 3-\( \sigma \) control limits of \( \bar{X} \)-chart can be obtained from the control limits in (4) by replacing \( \mu \) by \( \bar{\bar{X}} \) and \( \sigma \) by \( S \) calculated from \( m \) phase I samples of size \( n \) and defined below in (17). Therefore, the phase II control limits of \( \bar{X} \)-chart are given by

\[
\begin{align*}
UCL &= \bar{\bar{X}} + 3 \frac{S}{\sqrt{n}} \\
CL &= \bar{\bar{X}} \\
LCL &= \bar{\bar{X}} - 3 \frac{S}{\sqrt{n}}
\end{align*}
\]

(17)

where \( \bar{\bar{X}} = \frac{1}{m} \sum_{j=1}^{m} X_j \), \( S^2 = \frac{1}{m} \sum_{j=1}^{m} S^2_j \). The \( \bar{X}_j \) and \( S^2_j \) are the sample mean and variance of \( j \)th sample of size \( n \) respectively.

Conditionally on \( \bar{\bar{X}} \), the probability of signalling event \( \hat{\beta} \) for \( \bar{X} \)-chart with control limits in (17) can be written as

\[
\hat{\beta} = 1 + \Phi \left( -\delta \sqrt{n} + \frac{\sqrt{n}(\bar{\bar{X}} - \mu_1)}{\sigma} - z_{\alpha_0/2} \frac{S}{\sigma} \right) - \Phi \left( -\delta \sqrt{n} + \frac{\sqrt{n}(\bar{\bar{X}} - \mu_1)}{\sigma} + z_{\alpha_0/2} \frac{S}{\sigma} \right)
\]

where \( \mu_1 = \mu_0 + \delta \sigma_0 \). Using the facts that (i) \( Z = \frac{\sqrt{n}(\bar{\bar{X}} - \mu_1)}{\sigma} \) follows a standard normal distribution and (ii) \( W = \frac{\nu S^2}{\sigma^2} \) follows a chi-square distribution with \( \nu = n - 1 \) degrees of freedom and (iii) both \( Z \) and \( W \) are independent, \( \hat{\beta} \) can be re-expressed as

\[
\hat{\beta} = 1 + \Phi \left( -\delta \sqrt{n} + \frac{Z}{\sqrt{m}} - z_{\alpha_0/2} \frac{W}{\sqrt{\nu}} \right) - \Phi \left( -\delta \sqrt{n} + \frac{Z}{\sqrt{m}} + z_{\alpha_0/2} \frac{W}{\sqrt{\nu}} \right)
\]

(18)

Using formulae (13)-(16) and (18), the metrics for the \( \bar{X} \)-chart are obtained when the parameters are unknown based on \( m \) samples of size \( n \) which are presented in Table 2 for \( \alpha_0 = 0.0027 \).

It is clear from Table 2 that when small number of Phase I samples are used to estimate parameters, the ARL values are deviated significantly from 370.4 (Known parameter case). For example, when \( m = 20 \), the ARL value is 422 which is about 14\% larger than 370.4 and comes close only for \( m \geq 200 \). In like manner, all the percentiles are much different from the respective values for the known parameter case.

Noting that the performance analysis in terms of URL distribution is known as unconditional performance analysis which provides only a viewpoint of ‘on an average’ performance of a typical control chart and does not reveal the complete picture of the performance. Most of the
research on the effects of parameter estimation on the performance of control charts has been focused on the marginal (or unconditional) run length distribution and some associated characteristics[16]. In the works of Chen[17] and Chakraborti[2], the unconditional run length distribution provides a compact and convenient way to study chart performance but it does not give any impression of a particular control chart conditionally on estimated values of process parameters. The performance of the resulting chart would depend on the particular value of the parameter estimates obtained from this reference sample. Thus, two practitioners using two reference samples, from the same IC performance can then have different parameter estimates and resulting control charts with very different chart properties. For example, suppose two practitioners are using respectively first and third quartiles of $\bar{X}$ as estimates of $\mu$. Figure 1 shows the boxplots of run lengths of their respective control charts conditioning on their estimated value $\bar{X}$ for IC process.

Table 2. IC performance metrics for the Shewhart $\bar{X}$-chart when the mean and variance are unknown based on m samples of size 5 and $\alpha_0=0.0027$

<table>
<thead>
<tr>
<th>$m$</th>
<th>ARL</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>422.31</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>398.77</td>
<td>29</td>
</tr>
<tr>
<td>50</td>
<td>384.19</td>
<td>32</td>
</tr>
<tr>
<td>100</td>
<td>375.91</td>
<td>36</td>
</tr>
<tr>
<td>200</td>
<td>372.75</td>
<td>38</td>
</tr>
<tr>
<td>300</td>
<td>371.86</td>
<td>38</td>
</tr>
<tr>
<td>500</td>
<td>371.22</td>
<td>39</td>
</tr>
<tr>
<td>1000</td>
<td>370.78</td>
<td>39</td>
</tr>
<tr>
<td>2000</td>
<td>370.57</td>
<td>39</td>
</tr>
<tr>
<td>5000</td>
<td>370.45</td>
<td>39</td>
</tr>
</tbody>
</table>

The recent paper by Saleh[18] and Epprecht[19] highlighted the important issues involved in dealing with the so-called practitioner-to-practitioner variability and recommended that a study of the conditional RL distribution and some associated characteristics given the reference sample is more informative. Kumar & Chakraborti [3] elaborates the impact of parameter estimation (from the phase I reference sample) on the performance of a phase II $t_r$-chart is investigated based on the conditional false alarm rate (CFAR), the conditional average run length (CARL) and the number of phase I observations required to have a desired level of IC average run length (ARL) value.
Figure 1. Boxplots of conditional run length distributions when $\mu$ is estimated by first quartile of $\bar{X}$ (on the top) and third quartile of $\bar{X}$ (on the bottom) respectively.

Note that when the control limits are estimated, the conditional average run length (CARL) becomes random variable which plays a vital role in evaluating the chart in case U. Usually, average and standard deviation of $CARL$ are used in literature as conditional performance measures which will be denoted as $AARL$ and $SD_{CARL}$ respectively. Former provides information that a practitioner can get on an average his CARL value equal to AARL whereas later reflects the variability among the CARL values of practitioners. Steiner called it practitioner-to-practitioner variability and its low value is desirable. These metrics are defined as

$$AARL = \int_{\hat{\theta}}^{\hat{\theta}} \frac{1}{\hat{\beta}} f(\hat{\theta}) d\hat{\theta}$$

(16)

$$SD_{CARL} = \sqrt{\left(\int_{\hat{\theta}}^{\hat{\theta}} \frac{1}{\hat{\beta}^2} f(\hat{\theta}) d\hat{\theta} - AARL^2\right)}$$

(17)

Note that the measures based on first two moments i.e. $AARL$ and $SD_{CARL}$ do not provide full information about the $CARL$ distribution and hence other metrics based on $CARL$ distribution should also be considered in [3] e.g. LPB and PR value. The LPB is defined as the smallest IC $CARL$(denoted by $CARL_{in}$) value that can be attained by the chart with some specified probability[3]. On the other hand, PR is the probability that a user would get his IC $CARL$ value will be at least nominal value of ARL($ARL_0$). Recently, the metric LPB is suggested by several authors including[20],[21],[22] to adjust the control chart limits to guarantee the nominal $ARL_0$ value in the phase II stage. Now we define100(1 $\gamma$)% LPB (0 < $\gamma$ < 1) and $PR$ as follows.

$$P[CARL_{in} \geq LPB] > 1 - \gamma$$

(18)

and

$$P[CARL_{in} > ARL_0] = \gamma$$

(19)

The following table 3 shows the in-control AARL, $SD_{CARL}$, LPB and PR values for the $\bar{X}$-chart for different sample size m. The values of Table 3 are found using equations (16),(17),(18) and (19).
Table 3 shows that in-control AARL is equal to the average of URL distribution. Note that in literature, about 200 phase I samples are recommended based on unconditional performance analysis to get good performance of the chart i.e. $AARL_{in}$ equal to 370.4. However, Table 3 shows that however, IC $AARL$ value is close to 370.4 but even then there is large variability among the $CARL_{in}$ values, for example, when $m = 200$, the $SD_{CARL}$ value is 95 which is significantly large. Zhang et al, 2014 suggested to keep $SD_{CARL}$ below 10% of $AARL_{in}$ value which requires more than 1000 Phase I samples. Moreover, $LPB$ is very low specially for small number of Phase I samples, for example, when $m = 200$, the $LPB$ is 263 which 29% smaller than 370.4 whereas when $m = 500$, it 300, still 19% smaller than 370.4. In like manner, the PR value is only 0.44 when $m = 200$ which implies that there is only 44% chance that $CARL_{in}$ will greater than 370.4 which is not desirable. The PR value is not more than 50% even for large number of Phase I sample, say, 5000. Thus, the attempt is to be made to consider these metrics for designing and evaluating the control charts. Recently, the metric $LPB$ is recommended by several authors including [20],[21],[22] to design the control chart limits to guarantee the nominal ARL value in the Phase II monitoring.

Table 3. The in-control AARL, $SD_{CARL}$, $LPB$ and PR values of the $\bar{X}$-chart for the unknown parameter case for different sample size $m$

<table>
<thead>
<tr>
<th>$m$</th>
<th>AARL</th>
<th>$SD_{CARL}$</th>
<th>$LPB$</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>422.31</td>
<td>459.33</td>
<td>111.25</td>
<td>0.38</td>
</tr>
<tr>
<td>30</td>
<td>398.77</td>
<td>313.13</td>
<td>142.18</td>
<td>0.40</td>
</tr>
<tr>
<td>50</td>
<td>384.19</td>
<td>214.00</td>
<td>180.18</td>
<td>0.42</td>
</tr>
<tr>
<td>100</td>
<td>375.91</td>
<td>139.17</td>
<td>225.26</td>
<td>0.44</td>
</tr>
<tr>
<td>200</td>
<td>372.75</td>
<td>94.66</td>
<td>263.74</td>
<td>0.45</td>
</tr>
<tr>
<td>300</td>
<td>371.86</td>
<td>76.32</td>
<td>281.75</td>
<td>0.46</td>
</tr>
<tr>
<td>500</td>
<td>371.22</td>
<td>58.53</td>
<td>300.46</td>
<td>0.47</td>
</tr>
<tr>
<td>1000</td>
<td>370.78</td>
<td>41.09</td>
<td>320.04</td>
<td>0.48</td>
</tr>
<tr>
<td>2000</td>
<td>370.57</td>
<td>28.95</td>
<td>334.15</td>
<td>0.48</td>
</tr>
<tr>
<td>5000</td>
<td>370.45</td>
<td>18.27</td>
<td>347.36</td>
<td>0.49</td>
</tr>
</tbody>
</table>

4. Conclusion

Control charts play a crucial role in controlling and improving the quality by giving an alarm when the assignable causes occur in the process. Therefore, to examine their performance has its own right in literature. In this study, we discussed various performance measures to evaluate the charts in terms of run length distribution when the process parameters are known and unknown as well. Because unconditional performance analysis does not provide a full picture of the
performance of the estimated control chart, conditional analysis must be carried out. Usually, conditional analysis is focused on the average and standard deviation of conditional run length distribution; however, it is recommended that other criteria such as $LPB$ and $PR$ can be considered in designing and evaluating the estimated control limits.

References:


