A Comparison of the performance of mean, median, and precedence control charts for nonnormal data

Junghee Kim*, Sungim Lee†, Heon Jin Park‡, Jaecheol Lee*, Youngchul Jang*

Abstract

In this article, we will compare the performance of the mean control chart, the median control chart, the transformed mean control chart, the transformed median control chart, and the precedence control chart by simulation study. For control charts with transformed data, Yeo-Johnson transformation is used. Under the in-control condition, ARL's in all control charts coincide with the designed ARL in the normal distribution, but in the other distributions, only the precedence control chart provides the in-control ARL as designed. Under the out-of-control condition, the mean control chart is preferred in the normal distribution and the median control chart is preferred in the heavy-tailed distribution and the precedence control chart outperforms in the short-tailed distribution.

Key words: Average run length; control chart; median control chart; precedence control chart; Yeo-Johnson transformation.

1 Introduction

Statistical quality control charts are plots of statistics such as mean or median over time to distinguish between normal variation and unexpected true changes. The run length is the number of samples taken until an alarm is signaled by the control chart and the performance of control charts is usually evaluated by average run length. Hence, the control charts can be said it performs well if it has long ARL when the system has not changed or if it has short ARL when the process goes out of control.

In this article, we will investigate the performance of control charts for normal and nonnormal data in terms of ARL. Traditionally, if the data is mildly nonnormal, then often a simple transformation on each data point, such as the log or square root, will make them normally distributed. Or the data is severely nonnormal, nonparametric approach can be preferred. For details, refer to Gibbons and Chakraborti(2003) and references cited therein. The purpose of this article is to provide the performance of various control charts including

*System Engineering Team, Memory Division, Device Solution Network, Samsung Electronics, Hwasung, Kunggi-Do, Korea, 445-701

†Full-time lecturer, Department of Information statistics, Dankook University San 8, Hannam-Dong Yongsan-Ku, Seoul, Korea, 140-714, E-mail: silee@dankook.ac.kr

‡Associate Professor, Department of statistics, Inha University,253 Yongyun-Dong, Nam-Gu, Incheon, Korea, 402-751, E-mail: hjpark@inha.ac.kr

nonparametric control chart for nonnormal data via simulation. Also, the appropriate choice of control charts will be considered depending on the degrees of abnormality.

2 Review on control charts

In this section, we will review the mean, the median, and the precedence control charts with their control limits. All these control charts can be used to monitor the central tendency of a process.

The mean control chart is the most commonly used control chart to detect the lack of control. The control limits are given by

$$\bar{\bar{X}} \pm 3\frac{\bar{R}}{d_2\sqrt{n}}. \tag{1}$$

Here, $\bar{\bar{X}}$ is the mean of means of subgroups in the reference sample and $\bar{R}$ is the mean of ranges of subgroups in the reference sample.

The median control chart uses the median in the subgroup as a test statistic and the control limits are obtained from the range of the subgroup from the reference sample assuming normality.

$$\bar{\bar{X}} \pm 3m_3\frac{\bar{R}}{d_2\sqrt{n}} \tag{2}$$

These the mean and median control charts are well explored by Wheeler and Chambers(1992), and Montgomery(2001).

Chakraborti et al.(2004) provides a distribution-free control chart, called the precedence control chart. In the precedence control chart, the test statistic is the median. For the control limits for the median with subgroup size=n, Chakraborti et al.(2004) derive control limit from hyper-geometric distribution depending on the subgroup size and the number of the reference sample. For example, when the subgroup size is 5 and the number of subgroup for the reference sample is 200, from the table given in Chakraborti et al.(2004), the lower control limit is the 53-th smallest observation and the upper control limit is the 948-th smallest observation among 1000 observations.

Yeo and Johnson (2000) provides a power transformation maximizing likelihood. We use Yeo-Johnson transformation for fixing non-normality. That is, for data set, Yeo-Johnson transformation is performed and the mean control chart and the median control chart for the transformed data is adopted. We call the mean control chart and the median chart for transformed data the transformed control chart and the transformed median chart, respectively.

3 Simulation study

To compare the performance of control charts, we present simulation results for four kinds of distributions as follows.

(1) standard normal distribution with mean 0 and variance 1 (N(0, 1))

(2) χ^2 distribution with df=5 (χ^2_5)

(3) t distribution with df=4 (t_4)

(4) signed square-root standard normal distribution

$$Y = \begin{cases} -\sqrt{N(0,1)} & \text{if} \quad N(0,1) < 0 \\ \sqrt{N(0,1)} & \text{if} \quad N(0,1) \geq 0 \end{cases}$$

Note that χ^2_5 is skewed, t_4 is heavy-tailed, and the signed square-root standard normal distribution is short-tailed. For the simulation, random numbers were generated in SAS/IML. The subgroup size was five, which is most commonly-used, and 200 subgroups were used to obtain control limits under in-control condition. We designed $\alpha = 0.0027$ and ARL=370 with which control limits are 3σ in the normal distribution. We used Yeo-Johnson transformation (2000) for the reference sample, which provides the maximum likelihood estimate for the power of transformation. With the estimated power, the data were transformed and control limits with the transformed mean control chart and the transformed median control chart were acquired for in-control ARL. To obtain ARL, subgroups with size 5 were generated until test statistics were out of control limits. We simulated 1000 times in SAS/IML. ARL's under the in-control condition for the mean control chart, the median control chart, the transformed mean control chart, median control chart and the precedence control chart(Chkraborti et al.,2004) are in Table 1. Table 1 shows that empirical ARL's of all control charts coincide the in-control ARL of 370 as designed. Also the precedence control chart provides the designed ARL. In the χ^2_5 distribution, the mean, median, transformed median control charts provide smaller ARL than ARL as designed and the median control chart provides lager ARL. For t_4 distribution, the non-transformed control charts have the smaller ARL and the control charts with transformed data provide the larger ARL. For the signed square-root normal distribution, all control charts except the precedence control chart provide very large ARL. As a result, the precedence control chart outperforms over the other control charts under the in-control condition.

To compare the performance under the out-of-control condition, mean shift varied from 0.1 to 3. The performance under the out-of-control condition depends on the ARL under the in-control condition. That is, the smaller ARL under the in-control condition tends to provide the smaller ARL under the out-of-control condition. To compare the performance under the out-of-control condition, length from the mean or median to the control limit was adjusted in order for the in-control ARL to be 370. ARL's with the adjusted control limits under the out-of-control condition were shown in Figure 1.

In the standard normal distribution, the mean control chart is better than other control charts and ARL of the transformed mean control chart is close to that of the mean control chart. The median control chart, the transformed median control chart and the precedence control chart provide the similar ARL in the normal distribution. In χ^2_5 distribution, the transformed mean control chart has the best performance and the transformed median control chart does the second. The precedence control chart works poorly, especially in small mean shift with larger ARL than the designed ARL=370 in χ^2_5 distribution. In t_4 distribution, the median control chart, the transformed median control chart and the precedence control chart work better than the mean control chart and the transformed mean control chart. Even though the median control chart, the transformed median control chart and the precedence control chart performs closely, the median control chart is the best in t_4 distribution. In the signed square-root distribution, the precedence control chart provides

the best performance, and the mean control chart and the transformed mean control chart have worse performance than others. In the signed square-root distribution, ARL's for the median control chart and the transformed median are close to that for the mean control chart for the small shift and move to that for the precedence control chart as shift gets larger.

Table 1. Results of the in-control ARL for control charts in several distributions

Control chart	Distribution			
	N(0,1)	χ_5	t_4	Y
Mean	380.4	130.8	92.5	1121.3
Transformed mean	365.2	461.1	85.8	1128.1
Median	348.4	131.4	433.4	2580.8
Transformed median	333.1	311.0	420.1	2826.7
Precedence	371.7	363.9	341.4	371.7

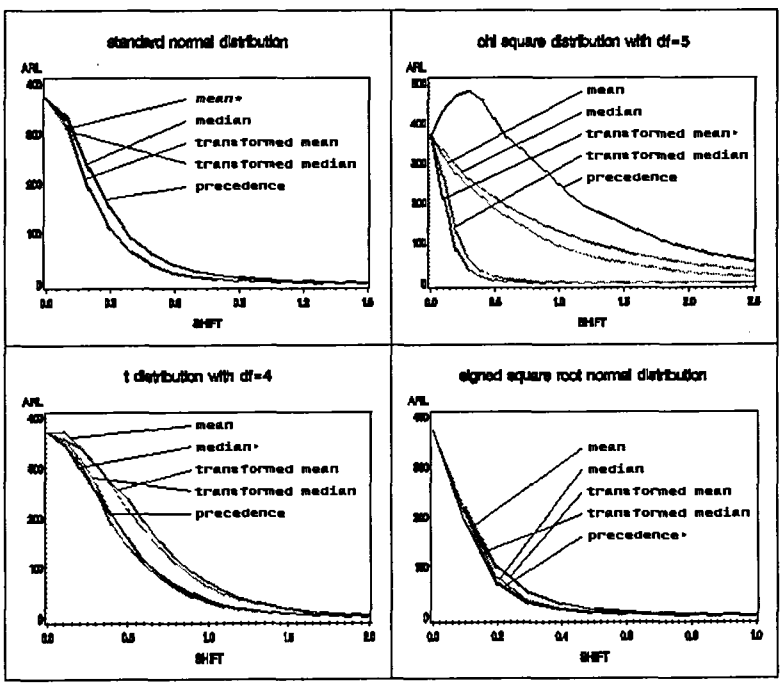

Figure 1: Out-of-control ARL for several distribution. The control chart with * stands for chart with the smallest ARL.

4 Conclusions

Findings in this article suggest that the appropriate control charts for the nonnormal data may differ from its shape. In the presence of skewness, the transformed mean control charts performs better than the other control charts. Except the mean control chart for normally distributed process data, we need to make adjustment about control limits in all control charts for given in-control ARL. For the heavy-tailed distribution, the median control chart outperforms compared to other charts. For the short-tailed distribution, the precedence control chart performs best. As a result, the skewed data need transformation to be symmetric. After skewness is fixed, if data are normally distributed, the mean control chart is the best choice. For the transformed data with heavy tail, the median control chart with adjusted control limit performs well and the precedence control chart is recommended for the transformed data with short tail.

Our findings emphasize the importance of the transformation of the data in dealing with the nonnormal data and also the choice of appropriate control charts depending on degrees of abnormality of measurement data. The control limit adjustment via reference sample for given in-control ARL needs to be considered in future study.

References

Chkraborti, S. and van der Laan, P. and van de Wiel, M. A. (2004). A class of distribution-free control charts.*Appl. Statist.* **53**, 443-462.

Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric Statistical Inference 4th edn. *New York: Dekker.*

Montgomery, D. C. (2001). Introduction to Statistical Quality Control. *New York: Wiley.*

Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika*, **87**, 954-959.

Wheeler D.J. and Chambers D.S.(1992). Understading Statistical Process Control. *SPC Press: Knoxville, TN.*