REVIEW OF RESEARCH

IMPACT FACTOR : 5.7631(UIF) UGC APPROVED JOURNAL NO. 48514 ISSN: 2249-894X

VOLUME - 8 | ISSUE - 6 | MARCH - 2019

NONPARAMETRIC CUSUM CONTROL CHART FOR PROCESS LOCATION

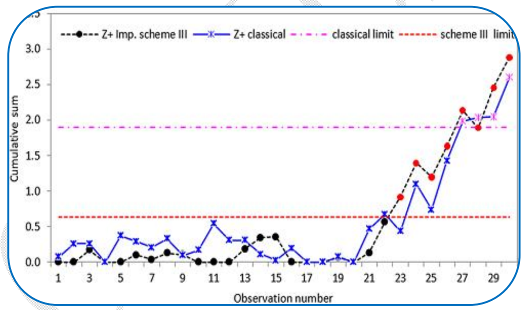
D. M. Zombade[1] and V. B. Ghute[2]

[1]Department of Statistics , Walchand College of Arts and Science, Solapur (MS), India.

[2]Department of Statistics, Solapur University, Solapur (MS), India.

ABSTRACT :

Control charts are useful and commonly used tools for monitoring the process parameters. It is well known that the CUSUM control charts are very effective in detecting small shifts in process parameters as compared to Shewhart-type charts. In this paper, nonparametric CUSUM control chart based on run statistic is developed for monitoring the known location of a continuous process. The average run length performance of the proposed CUSUM chart is investigated using a simulation study and is compared with parametric CUSUM chart under normal and non-normal process distributions. The study reveals that the proposed chart perform better than the parametric CUSUM chart for detecting small shifts in process location under normal and non-normal process distributions.



KEYWORDS : *Average run length, Run test, Control chart, Process location, CUSUM statistic.*

1. INTRODUCTION

Shewhart $\overline{X}$ chart is popular and commonly used control chart for monitoring process mean. The Shewhart $\overline{X}$ chart only uses the information from the last sample; hence it is insensitive to detect small and moderate shifts in the process mean. One alternative approach to address the detection of small shifts in process mean is to use the cumulative sum (CUSUM) chart. The CUSUM chart was introduced by Page (1954) and has been widely used for monitoring the mean of a quality characteristic of a production process. The CUSUM chart is designed such that it uses the past information along with the current information, which makes it more efficient than the Shewhart $\overline{X}$ chart in detecting small and moderate shifts in the process mean. Parametric Shewhart and CUSUM control charts often assumes that the process data come from some parametric distribution, most commonly the normal distribution. If underlying process distribution is unknown or not normal, these control charts may not be appropriate. In such situations, development and application of control charts that does not depend on particular distributional assumption is desirable. Nonparametric control charts can serve this purpose. The main advantage of nonparametric control chart is that it does not assume any probability distribution for the characteristic of interest. A formal definition of nonparametric control chart is given in terms of its in-control run-length distribution. The number of samples that need to be collected before the first out-of-control signal given by a chart is a random variable called the run-length; the probability distribution of the run-length is referred to as the run-length distribution. If the in-control run length distribution is same for every continuous distribution then the chart is called nonparametric (Chakraborti et al (2004)).

In the recent years, nonparametric control charts have been attracted much more attention from researchers. In literature, several nonparametric control charts are proposed for monitoring the parameter of the process. Chakraborti et al. (2001) presented an extension overview of the literature on univariate nonparametric control charts. Chakraborti and Graham (2007) provided an overview on nonparametric control charts and discussed their advantages. The location and scale of a process are two main parameters often monitored in nonparametric control charts. The problem of monitoring the location of a process is important in many applications. The location parameter could be the mean or the median or some percentiles of the distribution. For monitoring multivariate process location, some nonparametric control charts based on sign and signed-rank statistics are also available in literature. Das (2009) proposed multivariate nonparametric control chart based on bivariate sign test. Boone and Chakraborti (2011) proposed two Shewhart-type multivariate nonparametric control charts based on multivariate forms of the sign and signed-rank tests. Ghute and Shirke (2012a) developed nonparametric synthetic control chart based on bivariate signed-rank test to monitor changes in the location of a bivariate process. Ghute and Shirke (2012b) also developed nonparametric synthetic control chart based on bivariate sign test to monitor changes in the location of a bivariate process. For monitoring the location of a univariate continuous process, some nonparametric CUSUM charts have been developed. Bakir and Reynolds (1979) developed a nonparametric CUSUM chart to monitor process center based on within group signed ranks. McDonald (1990) proposed a CUSUM procedure based on sequential ranks. Amin et al. (1995) developed nonparametric CUSUM control chart for grouped data based on sign test statistic. Li et al. (2010) proposed a nonparametric CUSUM control chart based on well known Mann-Whitney test statistic for monitoring the unknown location of a process. Yang and Cheng (2011) have proposed a nonparametric CUSUM chart to monitor the possible small shifts in the process mean. Liu et al. (2015) proposed a sequential rank based nonparametric CUSUM control chart for detecting arbitrary magnitude of shifts in the location parameter. Zombade and Ghute (2018) developed Shewhart-type nonparametric control chart for process location.

The purpose of this paper is to develop a nonparametric CUSUM control chart for monitoring the location of a process. The proposed chart is based on runs computed within samples and used in place of sample means in the parametric CUSUM chart. In this paper we focus on positive-sided CUSUM chart in which upward shifts in the process location are of interest. The rest of the paper is organized as follows. A parametric CUSUM chart for monitoring process mean based on $\overline{X}$ statistics is described in Section 2. A brief introduction of nonparametric run test for location parameter is discussed in Section 3. The proposed nonparametric CUSUM chart for monitoring process location based on run statistic is presented in Section 4. The performance of proposed nonparametric CUSUM chart is evaluated and compared with parametric CUSUM chart in Section 5. Some conclusions are given in Section 6.

2. PARAMETRIC CUSUM CHART FOR PROCESS MEAN

Let *X* denote the process variable being measured and suppose that *X* has a normal distribution with mean μ and standard deviation σ. Let μ_0 be in-control value for μ and σ_0 be in-control value for σ. Usually, two symmetric CUSUM charts are used to detect two-sided mean shifts. For detecting positive shifts in μ, a one-sided (upper) CUSUM consists in computing recursively the sequence C_i^+, $i \geq 1$, where

$$C_0^+ = 0\,;\; C_i^+ = \max[\,0,\; C_{i-1}^+ + (\overline{X}_i - \mu_0) - k\,], i \geq 1 \qquad (1)$$

Where *i* is subgroup number; $\overline{X}_i$ is the mean of study variable *X*, μ_0 is the target mean of the study variable *X* and *k* a positive constant is the reference value of the CUSUM scheme. The CUSUM signals a change in process mean as soon as C_i^+ exceeds a control limit $h > 0$, interpreting that process mean has shifted upward.

To control downward shifts in the process mean μ, a one-sided (lower) CUSUM consists in computing recursively the sequence C_i^-, $i \geq 1$, where

$$C_0^- = 0\,;\; C_i^- = \max[\,0,\; C_{i-1}^- - (\overline{X}_i - \mu_0) - k\,],\; i \geq 1 \tag{2}$$

The CUSUM chart using this statistic would signal whenever signals a change in process mean as soon as C_i^- exceeds $C_i^- < -h$, where $h > 0$, interpreting that process mean has shifted downward. For a two-sided CUSUM chart, the two charting statistics C_i^+ and C_i^- are plotted against a single control limit *h*. The starting value for both plotting statistics is usually taken as $C_0^+ = C_0^- = 0$. A signal is given if $C^+ > h$ or $C^- < -h$. We refer parametric CUSUM chart based on $\overline{X}$ as $CSM - \overline{X}$ chart.

3. NONPARAMETRIC RUN TEST FOR LOCATION

In this Section, we briefly review the run test described by Varon (2010). Let $X_1, X_2, \ldots, X_n$ be a subgroup sample of size *n* > 1 from a distribution with location θ and standard deviation σ. It is assumed that these observations are independent and have a continuous distribution symmetric about location (median or median) θ. Let θ_0 denote the target known value of the process location. Without loss of generality, we assume that $\theta_0 = 0$ and $\sigma_0 = 1$. A test for the hypothesis $\mathrm{H}_0 : \theta = 0$ versus $\mathrm{H}_1 : \theta > 0$, based on runs has been discussed by Varon (2010). A run is defined as a succession of two or more identical symbols which are followed and preceded by different symbols or no symbol at all. At each inspection point, a nonparametric run statistic R is computed using a subgroup sample $X_1, X_2, \ldots, X_n$. For the construction of runs, the variable η_j is defined as

$$\eta_j = S(X_{Dj}) = \begin{cases} 1\,, & \text{if } X_{Dj} > 0\,,\; j = 1,2,3,\ldots.,n \\ 0\,, & \text{otherwise} \end{cases}, \tag{3}$$

where D_j is the antirank of $|X|_{(j)}$ such that $|D_j| = |X|_{(j)}$. Hence D_j labels the X which corresponds to the j^{th} order absolute value. Then the sequence $\eta_1, \eta_2, \ldots, \eta_n$ is a dichotomized sequence. The changes in the dichotomized succession are identified with the following indicators:

Define $I_1 = 1$

$$I_j = \begin{cases} 1\,, & \text{if } \eta_{j-1} \neq \eta_j,\; j = 2,3,\ldots.,n. \\ 0\,, & \text{if } \eta_{j-1} = \eta_j \end{cases} \tag{4}$$

The number of runs until the j^{th} element of the dichotomized succession is obtained through the following partial sums:

$$r_i = \sum_{j=1}^{i} I_j \;, i = 1,2,\ldots.,n. \tag{5}$$

Naturally $r_i \le r_j$ for $i < j$ and r_n is the total number of runs in the sequence. Test statistic based on runs is given as

$$R = \frac{1}{r_n}\sum_{j=1}^{n}\delta_j\, r_j \qquad (6)$$

Where $$\delta_j = \begin{cases} 1\ , & if\ \eta_j = 1, \\ -1\ , & if\ \eta_j = 0 \end{cases} \qquad j = 1,\ 2, 3, \ldots, n. \qquad (7)$$

Note that R includes the number of runs until every element of the dichotomized succession, increasing their value when $\eta_j = 1\ (\delta_j = 1, \text{runs of ones})$ and decreasing when $\eta_j = 0\ (\delta_j = -1, \text{runs of zeros})$ the large value of R indicate greater number of runs of ones and it is an indication that $\theta > 0$. Additionally the inverse of total number of runs $\frac{1}{r_n}$ is used as a factor of standardization. It should be noted that the statistic R takes values between $-n$ and n. Large values of R indicate a positive shift where as small value indicate a negative shift. For $\theta > 0$, it is expected that R takes large positive values. Accordingly H_0 is rejected for large values of R.

4. NONPARAMETRIC CUSUM CHART FOR LOCATION

In this Section, we develop a nonparametric CUSUM chart for monitoring location of a process. Let X denote the process variable being measured and suppose that X has a continuous symmetric distribution with location parameter θ. Here we have to monitor location parameter θ through control charting. Let θ_0 be in-control or target value of θ. The location parameter of a distribution under study is usually unknown in practice and need to be estimated from the analysis of the preliminary samples taken when the process is assumed to be in-control. The proposed nonparametric CUSUM charting technique for detecting a change in location θ from in-control value θ_0 to some out-of-control value θ_1 is based on first transforming the observed data into a nonparametric run statistic R and then applying the CUSUM chart on the transformed statistic (the chart is referred as *CSM-R* chart). The proposed *CSM-R* chart is constructed by accumulating the statistics $R_1, R_2, \ldots$ sequentially from each sample subgroup $X_1, X_2, \ldots, X_n$ of size n. When detection of shift in location θ (from its specified value θ_0) in only one direction (up or down) is of interest, a one-sided CUSUM chart is desirable. When the objective is to detect increase in θ an upper one-sided CUSUM uses the plotting statistic

$$S_i^+ = \max[\,0,\ \ S_{i-1}^+ + (R_i - \theta_0) - k] \qquad (8)$$

where starting value of plotting statistic is $S_0^+ = 0$ and k is the reference value of the CUSUM scheme. The statistic S_i^+ is plotted on the chart along with control limit h. If any S_i^+ plots on or outside the control limit h, the process is declared out-of-control and search for assignable causes is started, otherwise, the process is considered in-control and control procedure continues. When the objective is to detect decrease in θ lower one-sided CUSUM uses the plotting statistic

$$S_i^- = \max[\,0,\ \ S_{i-1}^- - (R_i - \theta_0) - k] \qquad (9)$$

where starting value of plotting statistic is $S_0^- = 0$ and k is the reference value of the CUSUM scheme. The statistic S_i^- is plotted on the chart along with control limit *h*. If any S_i^- plots on or outside the control limit -*h*, the process is declared out-of-control and search for assignable causes is started.

When the objective is to detect both increase and decrease in θ from θ_0, a two-sided *CSM-R* chart uses both of the statistics S_i^+ and S_i^- simultaneously. A signal is given if $S_i^+ > h$ or $S_i^- < -h$. We assume that the objective of monitoring the process is to detect any special cause that changes location θ from θ_0. Although detecting decrease in location θ may be of interest in some applications, here we focus on more important problem of detecting increase in the process location θ. The upper one-sided *CSM-R* chart can be constructed as follows:

Step 1. Collect a subgroup sample $X_i = (X_{i1}, X_{i2}, \ldots, X_{in})$, $i = 1, 2, \ldots, n$ of size *n* from a process.

Step 2. Compute run statistic R_i from the subgroup sample X_i, *i* = 1, 2, ..., *n*.

Step 3. Construct the CUSUM statistic as $S_i^+ = Max\,\{0,\ S_{i-1}^+ + (R_i - \theta_0) - k\}$, where $(k \geq 0)$ is reference parameter of CUSUM scheme and S_i^+ will detect upward location shift.

Step 4. Plot S_i^+ against control limit *h*.

Step 5. If S_i^+ exceeds *h*, process is declared to be out-of-control at the *i*th sample otherwise the process is considered to be in-control and monitoring continues to the next sample.

5. PERFORMANCE COMPARISONS

The performance of a control chart is usually measured by ARL, which is the average number of samples required to signal an out-of-control case. The in-control ARL is denoted by ARL_0 and out-of-control ARL is denoted by ARL_1. The performance of the proposed *CSM-R* chart is compared with the performance of the parametric $CSM - \overline{X}$ chart which is known to have very good properties when underlying process distribution is normal. A computer programme developed in C language is used to simulate ARL. Three process distributions are considered in simulation study namely, normal distribution with location θ and scale σ, Laplace distribution with location θ and scale λ, which is symmetric and having heavier tails than normal distribution and the uniform distribution with location θ and scale λ, which is symmetric and having lighter tails than the normal distribution. Equation (10), (11) and (12) respectively gives probability density functions of normal distribution with location θ and scale σ, Laplace distribution with location θ and scale λ and uniform distribution with location θ and scale λ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},\ -\infty < x < \infty \text{ and } \sigma > 0 \qquad (10)$$

$$f(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x-\theta|}{\lambda}\right),\ -\infty < x < \infty \text{ and } \lambda > 0 \qquad (11)$$

$$f(x) = \frac{1}{2\lambda}\ ,\ \theta - \lambda < x < \theta + \lambda \text{ and } \lambda > 0 \qquad (12)$$

The distributions have been shifted and scaled such that they all have an expected value of 0 and standard deviation 1, so simulation results are easily comparable. To achieve standard deviation of

1, we choose $\sigma = 1$ for normal distribution, $\lambda = \frac{1}{\sqrt{2}}$ for Laplace distribution and $\lambda = \sqrt{3}$ for uniform distribution.

Consider a process where quality characteristic of interest X is distributed with location θ and standard deviation σ. Let θ_0 and σ_0 be the in-control values of θ and σ respectively. When a shift in process location occurs, we have change from the in-control value θ_0 to the out-of-control value $\theta_1 = \theta_0 + \delta\,\sigma_0$, $(\delta > 0)$. Therefore, when control chart for location is employed, the process shifts are measured through $\delta = \frac{|\theta_1 - \theta_0|}{\sigma_0}$, where θ_1 is the shifted location and θ_0 is in-control location. The amount of a shift in the location is taken over the range $\delta = 0\ (0.2)1.2$. When $\delta = 0$, the process is in-control. The ARL values of the $CSM - \overline{X}$ and proposed *CSM-R* charts are computed using 10000 simulations. The optimal reference value *k* is taken to be $\frac{\delta}{2}$ if a CUSUM chart is to be able to detect a standardized location shift of size δ. Once *k* is determined, parameter *h* is usually chosen to achieve specified ARL_0. In simulation study we choose k = 0.5 and $ARL_0 = 370$.

Table 1 and Table 2 provide the ARL performance of the $CSM - \overline{X}$ and proposed *CSM-R* charts when underlying process data actually follows normal, Laplace and uniform distributions with sample sizes *n* = 10 and 15.

Table 1. ARL comparison for *n* = 10.

Shift δ	Normal distribution		Laplace distribution		Uniform distribution	
	CSM- $\overline{X}$ *h* = 0.40 *k* = 0.5	*CSM-R* *h* = 16.25 *k* = 0.5	*CSM-* $\overline{X}$ *h* = 0.4515 *k* = 0.5	*CSM-R* *h* = 16.25 *k* = 0.5	*CSM-* $\overline{X}$ *h* = 0.376 *k* = 0.5	*CSM-R* *h* = 16.25 *k* = 0.5
0.0	370.21	371.39	370.5	368.5	370.16	371.93
0.2	51.56	20.23	63.27	15.07	45.04	21.93
0.4	10.52	9.25	12.71	7.68	9.68	10.19
0.6	3.76	6.32	4.22	5.57	3.60	6.94
0.8	2.08	5.00	2.24	4.63	2.02	5.39
1.0	1.47	4.28	1.55	4.11	1.42	4.54
1.2	1.18	3.88	1.22	3.78	1.16	4.01
2.0	1.00	3.09	1.00	3.18	1.00	3.00

Table 2. ARL comparison for *n* = 15.

Shift δ	Normal distribution		Laplace distribution		Uniform distribution	
	CSM- $\overline{X}$ *h* = 0.2223 *k* = 0.5	*CSM-R* *h* = 19.85 *k* = 0.5	*CSM-* $\overline{X}$ *h* = 0.2515 *k* = 0.5	*CSM-R* *h* = 19.85 *k* = 0.5	*CSM-* $\overline{X}$ *h* = 0.211 *k* = 0.5	*CSM-R* *h* = 19.85 *k* = 0.5
0.0	369.0	371.61	371.36	371.92	370.14	369.91

0.2	40.97	15.93	49.24	12.41	37.51	16.75
0.4	7.71	7.80	8.81	6.68	7.22	8.26
0.6	2.70	5.45	2.93	4.94	2.61	5.77
0.8	1.53	4.37	1.58	4.14	1.50	4.61
1.0	1.15	3.78	1.18	3.66	1.15	3.95
1.2	1.03	3.40	1.05	3.38	1.03	3.47
2.0	1.00	3.00	1.00	3.02	1.00	3.00

Examination of Table 1 and Table 2 leads to the following findings:

- For monitoring a process operating under normal distribution, it is observed that, the proposed *CSM-R* chart is more efficient than the parametric $CSM-\overline{X}$ chart for detecting small shifts in the process location of size $\delta \leq 0.4$. For example, for a shift of size $\delta = 0.20$ under normal distribution with *n* = 10, the proposed *CSM-R* chart has $ARL_1 = 20.23$ which is smaller than the $ARL_1 = 51.56$ of parametric $CSM-\overline{X}$ chart.
- The proposed *CSM-R* chart is more efficient than the parametric $CSM-\overline{X}$ chart for detecting small shifts in the process location of size $\delta \leq 0.4$ when underlying distribution has tails heavier than the normal distribution. For example, for a shift of size $\delta = 0.20$ under Laplace distribution with *n* = 10, the proposed *CSM-R* chart has $ARL_1 = 15.07$ which is much smaller than $ARL_1 = 63.27$ of parametric $CSM-\overline{X}$ chart.
- The proposed *CSM-R* chart is more efficient than the parametric $CSM-\overline{X}$ chart for detecting small shifts in the process location of size $\delta \leq 0.4$ when process operates under light tail distribution such as uniform distribution. For example, for a shift of size $\delta = 0.20$ under uniform distribution with *n* = 10, the proposed CSM-R chart has $ARL_1 = 21.93$ which is much smaller than $ARL_1 = 45.04$ of parametric $CSM-\overline{X}$ chart.

It is natural to expect that parametric CUSUM outperform nonparametric CUSUM for normal process distribution. Surprisingly, the proposed *CSM-R* chart performs better than parametric $CSM-\overline{X}$ chart for detecting small shifts in the process location under normal and non-normal process distributions while the parametric $CSM-\overline{X}$ chart performs better only for larger shifts in process location. The effectiveness (speed of detection) of the proposed $CSM-R$ chart varies depending on the underlying process distribution.

6. CONCLUSIONS

In this paper, nonparametric CUSUM control chart based on run statistic is developed for monitoring the location parameter of a continuous symmetric process distribution. The proposed chart requires simple calculations and it is straightforward to implement. The performance of the proposed control chart is studied by simulation under normal, light tailed and heavy tailed process distributions. Simulation study indicates that the proposed CSM-R has the ability to detect small shifts of size $\delta \leq 0.4$ more quickly than the parametric $CSM-\overline{X}$ chart under normal, heavy tailed and light tailed distributions.

REFERENCES

1. R. W. Amin, M. R. Reynolds Jr., S. T. Bakir (1995), Nonparametric quality control charts based on the sign statistic, Communications in Statistics-Theory and Methods, 24(6), 1597-1623.
2. S.T. Bakir, M. R. Reynolds Jr. (1979), A nonparametric procedure for process control based on within group ranking, Technimetrics, 21, 175-183.
3. J. M. Boone, S. Chakraborti (2012), Two simple Shewhart-type multivariate nonparametric control charts, Applied Stochastic Models in Business and Industry 28: 130-140.
4. S. Chakraborti, M. A. Graham (2007), Nonparametric control charts, Encyclopedia of Statistics in Quality and Reliability, John Wiley, New York,.
5. S. Chakraborti, P. Van der Lann, M. Van de Wiel (2004), A class of distribution-free control charts, Applied Statistics, 53, 443-462.
6. S. Chakraborti, P. Van der Lann, S. T. Bakir (2001), Nonparametric control charts: An overview and some results, Journal of Quality Technology, 33, 304-315.
7. N. Das (2009), A multivariate nonparametric control chart based on sign test, Quality Technology and Quantitative Management 6(2), 155-169.
8. V. B. Ghute and D. T. Shirke (2012a), A nonparametric signed-rank control chart for bivariate process location, Quality Technology and Quantitative Management 9(4), 317-328.
9. V. B. Ghute, D. T. Shirke (2012b), Bivariate nonparametric synthetic control chart based on sign test, Journal of Industrial and system Engineering 6(2), 108-121.
10. S. Y. Li, L. C. Tang, S. H. Ng (2010), Nonparametric CUSUM and EWMA control charts for detecting mean shifts, Journal of Quality Technology, 42(2), 209-226.
11. L. Liu, J. Zhang, X. Zi; (2015), Dual nonparametric CUSUM control chart based on ranks, Communications in Statistics- Theory and Methods, 44, 756-772.
12. D. McDonald (1990), CUSUM procedure based on sequential ranks, Naval Research Logistics, 37(5), 627-646.
13. C. A. McGilchrist, K. D. Woodyer (1975), A note on a distribution-free CUSUM technique , Technimetrics, 17(3), 321-325.
14. E. S. Page; (1954), Continuous inspection schemes, Biometrika, 41, 100-115.
15. M. J. R. Varon (2010),Ph. D. Dissertation on Nonparametric test based on runs for a single sample location problem. L: http://kops.ub.uni-konstanz.de/volltexte/2010/11634.
16. S. Yang, S. W. Cheng (2011), A new nonparametric CUSUM mean chart, Quality and Reliability Engineering International, 27, 867-875.
17. D.M. Zombade and V.B. Ghute (2018), Shewhart-type nonparametric control chart for process location , Communications in Statistics - Theory and Methods, 0,1-14,
Link: https://doi.org/10.1080/03610926.2018.1435811

D. M. Zombade
Department of Statistics , Walchand College of Arts and Science, Solapur (MS), India.