ASM Sc. J., 12, Special Issue 5, 2019 for ICoAIMS2019, 124-129

# Nonparametric CUSUM Control Chart based on Wilcoxon Signed-Rank Statistics and Hodges Lehmann Estimator

Ainaa Salfariena Razalee[1], Nazihah Mohamed Ali[2] and Ow Su Shing[3]

[1,2,3]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.*

Nowadays, statistical process control (SPC) is widely used in the industry for measuring and controlling the quality for the manufacturing process. Cumulative sum control chart (CUSUM) is effective in detecting a small shift of a process. By using classical CUSUM, the normality assumption of the data is needed. However, the normality assumption always cannot be reached in real situation. To overcome these problems, the nonparametric statistics is used since the nonparametric statistics does not require normality assumption. Hence, Hodges-Lehmann (HL) estimator and Wilcoxon Signed-Rank (SR) statistics are used to integrate with the CUSUM control chart and formulate nonparametric CUSUM control chart which are HL-CUSUM and SR-CUSUM control chart. Average run length ($ARL$) is used to measure the performance of each CUSUM control chart. Our simulation study showed that the HL-CUSUM perform best in the small skewed data while the SR-CUSUM has the best performance when the skewness of the data increase.

**Keywords:** CUSUM control chart; normality; nonparametric statistics; average run length; skewness

## I. INTRODUCTION

Statistical process control (SPC) is widely used in the industry nowadays. SPC is a method to visualize and control the process performed by reducing product variability and improve production efficiency (Parkash *et al.*, 2013). Control charts are used to monitor for this SPC. Control charts can be classified into two categories which are memory-less control chart (Shewhart's type) and memory-type control chart. Memory-less control charts like $\bar{X}$ and $R$ chart use current information only while memory-type charts like cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) chart consider past information together with current information which makes this control chart effective in detecting small shifts quickly (Montgomery, 2007).

CUSUM control chart was first proposed by Page in 1954 while the EWMA control chart was proposed by Roberts in 1959. These two are the commonly used memory-type control charts. According to Montgomery (2007), CUSUM and EWMA are particularly effective in detecting small sustained shift quickly and have the similar performance. The parameter of $h = 4$ or $h = 5$ and $k = 0.5$ were suggested to use in order to provide a CUSUM with good ($ARL$). Koshti (2011) found that CUSUM is more efficient in detecting small shifts in the mean of a process than Shewhart charts although CUSUM is not simple to operate.

Normality assumption is needed in order to produce an efficient control chart. If the normality assumption cannot be reached or lack of information of the data, the control charts will be less practical to be used and the efficiency of the control chart will reduce (Das & Bhattacharya, 2008). To solve this problem, nonparametric is more desirable to use (Yang & Cheng, 2011). The advantage of nonparametric statistics are incorrect conclusions can be avoided since the assumption of the data is unnecessary and it can be used for data with small sample size (Nahm, 2016). Besides that, nonparametric statistics is more suitable to use when the skewness of the data is large (Fitzgerald, 2001).

A new nonparametric CUSUM mean chart was proposed by (Yang & Cheng, 2011). The proposed nonparametric two-sided CUSUM mean chart provides a good alternative to $\bar{X}$ control chart and classical CUSUM control chart when the distribution in not normal or unknown. Kaur and Kumar (2015) studied the parametric and nonparametric

*Corresponding author's e-mail: nazihanma@upm.edu.my

tests. From their study, nonparametric tests give a better performance compare to parametric test when the normality assumption cannot be met. According to Chakraborti and Van de Wiel (2008) nonparametric control charts are more preferable for skewed distribution because it is less sensitive to the outlier.

Some authors have discussed several nonparametric statistics on control charts. Bakir (2004) had proposed a distribution-free Shewhart control chart based on Wilcoxon Signed-Rank statistics where he used the data with uniform distribution, double exponential distribution and Cauchy distribution. Chakraborti and Eryilmaz (2007) had proposed a Shewhart-type distribution-free control chart based on Wilcoxon Signed-Rank statistics. This control chart provide more desirable false alarm rates and in-control average run length ($ARL$) . Graham *et al.* (2011) suggested the nonparametric EWMA control chart using the Wilcoxon Signed-Rank statistics.

Abu-Shawiesh and Abdullah (2001) studied the performance of new Shewhart control chart by using Hodges-Lehmann (HL) and Shamos-Bickel-Lehmann (SBL) estimators. Hodges-Lehmann estimator are more suitable when the special cause variations and outlier that present in Shewhart-type control charts and memory control charts (Nazir *et al.*, 2013). Subgroup size, $m = 5$ and $m = 10$ were used in their study. The process is simulated for 10000 times and the average out of these run length is considered to compare the performance of the control charts. From their study, HL is suitable when the data is not normally distributed. Hodges-Lehmann control chart had been studied by (Pongpullponsak & Jayathavaj, 2014). In their study, Weibull distribution with 11 skewness was used.

Average run length ($ARL$) is defined as the average number of samples taken before an out of control (OOC) signal is obtained. This shows that control chart with smaller ARL has a better performance since a smaller sample is required before an out of control signal is obtained. Hence, ($ARL$) can be used to determine the performance of the control chart (Montgomery, 2007). In addition, many study such as Pongpullponsak *et al.* (2006) and Tapang (2016) used ($ARL$) to compare the performance of the control charts.

Before continue to the implementation of nonparametric CUSUM control chart, we prepare the basic structure of standardized parametric CUSUM control chart in next section.

## II. THE STANDARDIZED CUSUM CONTROL CHART

In this study, data with a rational subgroup of sample size, $m > 1$ is used. The two advantages of standardized CUSUM control chart are the choices of parameter $h$ and $k$ do not depend on $\sigma$ and it leads naturally to CUSUM for controlling variability (Montgomery, 2007). The standardized two-sided CUSUM are defined as follows:

$$C_i^+ = \max[0, C_{i-1}^+ + y_i - k] \quad (1)$$

$$C_i^- = \max[0, C_{i-1}^- - y_i - k] \quad (2)$$

where $i$ is the subgroup number, $y_i$ is the standardized mean of each subgroup define as:

$$y_i = \frac{\bar{x}_i}{\sigma_{\bar{x}}} \quad (3)$$

$C_i^+$ and $C_i^-$ are upper and lower CUSUM respectively and $k$ is the reference parameter. The initial value of $C_i^+$ and $C_i^-$ are usually taken as equal to 0. A process is said to be out of control if the plotting of $C_i^+$ and $C_i^-$ exceed the decision interval, $h$.

## III. DESCRIPTION OF NONPARAMETRIC CUSUM CONTROL CHART

In this section, we shortly describe the nonparametric CUSUM control chart using Hodges Lehmann estimator (HL) and Wilcoxon Signed-Rank statistics (SR). Let $x_1, x_2, x_3, \dots, x_m$ be a random sample obtain from a distribution with size $m$.

### *A. Hodges Lehmann (HL) CUSUM*

Chakraborti *et al.* (2001) showed the usage of Hodges Lehmann estimator on Shewhart-type control chart. Here, we implemented the method on CUSUM control chart. The steps to obtain the Hodges Lehmann estimator are as follows:

i. Compute $M = \frac{m(m+1)}{2}$ (4)

ii. Compute the Walsh average, $W_r = \frac{x_i + x_j}{2}$ (5)

Where $r = 1,2,3, \dots, M$

$i = 1,2,3, \dots, m$

$j = 1,2,3, \dots, m$

iii. Reorder the Walsh average in ascending order

$$W_{(1)} \leq W_{(2)} \leq W_{(3)} \leq \cdots \leq W_{(M)}$$

iv. $HL_i = median\{W_{(1)}, W_{(2)}, W_{(3)}, \dots, W_{(M)}\}$ (6)

The $HL_i$ of each subgroup will be obtained and standardized by substituting the value of $\bar{x}_i$ from equation (3) with $HL_i$.

### B. Wilcoxon Signed-Rank (SR) CUSUM

Suppose that $X_{ij}$, $j = 1,2,3, \dots, n$ and $i = 1,2,3, \dots, m$ denote $j_{th}$ observation in the $i_{th}$ rational subgroup of size $m > 1$. Let $R_{ij}^{+}$ be the rank of the $|X_{ij}|$ among $|X_{i1}, \dots, X_{in}|$, $j = 1,2,3, \dots, n$ within the $i_{th}$ subgroup and $\theta$ is the target value or the mean. The Wilcoxon Signed-Rank is defined as:

$$SR_i = \sum_{j=1}^{m} sign(X_{ij} - \theta) R_{ij}^{+} \quad (7)$$

For $i = 1,2,3, \dots, m$, where

$$sign(X_{ij} - \theta) = \begin{cases} 1, \ if(X_{ij} - \theta) > 0 \\ 0, \ if(X_{ij} - \theta) = 0 \\ -1, \ if(X_{ij} - \theta) < 0 \end{cases} \quad (8)$$

Then, the $SR_i$ will be obtained and replace the $\bar{x}_i$ from equation (3).

## IV. CONTROL CHART PERFORMANCE

From the past research, a most popular way to compare the efficiency of the control chart is by using average run length ($ARL$). $ARL$ is the number of samples or subgroup needed before the first out of control signal is obtained for a control chart. The value of $ARL$ can be calculated by

$$ARL = \frac{1}{\alpha} \quad (9)$$

where $\alpha$ is the probability that any point exceeds the control limits.

The $ARL$ of the two-sided CUSUM from the $ARL^{+}$ and $ARL^{-}$ is calculated as

$$\frac{1}{ARL} = \frac{1}{ARL^{+}} + \frac{1}{ARL^{-}} \quad (10)$$

R code is used to generate the suitable parameter of $h$ with respective $h$ which the $ARL$ is equal to 370 as shown in Table 1. According to Montgomery (2007), using $h = 4$ or $h = 5$ and $k = 0.5$ will provide a CUSUM that has good $ARL$ properties. Hence, $h = 5$ and $k = 0.5$ will be used for each of the control chart in this study.

Table 1. Parameter of $h$ with respective $k$ for $ARL = 370$

| | $k$ | **0.25** | **0.50** | **0.75** | **1.00** | **1.50** | **2.00** |
|---|---|---|---|---|---|---|---|
| $h$ | CUSUM | 9.84 | 5.16 | 3.34 | 2.39 | 1.40 | 0.80 |
| | HL-CUSUM | 10.57 | 5.56 | 3.60 | 2.58 | 1.54 | 0.92 |
| | SR-CUSUM | 9.79 | 4.96 | 3.00 | 1.95 | 0.70 | 0.02 |

For this study, the performance of CUSUM, HL-CUSUM and SR-CUSUM are determined based on their$ARL$.Weibull distribution and Lognormal distribution are used for the simulation study. For the said purpose, 2000 observations from each distribution with $m = 5$ and $m = 10$ are generated. The skewness coefficient is introduced in the process to evaluate the $ARL$. The respective sample number when out-of-control signal occur in noted. This process is repeated 1000 times so that we can average out the run length to get the $ARL$ value. The parameter of each distribution will be adjusted to obtain various skewness coefficients. Different skewness coefficient Weibull and Lognormal distribution are used. The skewness coefficient with the respective parameter of each distribution is shown in Table 2 (Pongpullponsak *et al.*, 2006).

Table 2. Skewness coefficient with the respective parameter for Weibull and Lognormal distributions

| Skewness coefficient | Weibull distribution | | Lognormal distribution | |
|---|---|---|---|---|
| | Scale parameter, $\lambda$ | Shape parameter, $k$ | Scale parameter, $\mu$ | Shape parameter, $\sigma$ |
| **0.1** | 0.1 | 3.2219 | 0.1 | 0.0334 |
| **0.5** | 0.5 | 2.2110 | 0.5 | 0.1641 |
| **1.0** | 1.0 | 1.5630 | 1.0 | 0.3142 |
| **2.0** | 2.0 | 1.0000 | 2.0 | 0.5513 |
| **3.0** | 3.0 | 0.7686 | 3.0 | 0.7156 |
| **4.0** | 4.0 | 0.6478 | 4.0 | 0.8326 |
| **5.0** | 5.0 | 0.5737 | 5.0 | 0.9202 |

## V. RESULT AND DISCUSSION

Table 3 shows the result of simulation by using data from Weibull distribution with different skewness. This data is analyse by using CUSUM, HL-CUSUM and SR-CUSUM control charts and the values of $ARL$ is generated. From table 3, $ARL$ values decreasing with the increase of skewness of the data. $ARL$ values of CUSUM control chart decreasing slowly for sample size $m = 5$ and $m = 10$. For HL-CUSUM control chart, $ARL$ values decrease slowly when $m = 5$ and decrease drastically when $m = 10$ while $ARL$ values of SR-CUSUM control chart decrease drastically for both sample size.

Figure 1 graphically shows the $ARL$ values for those three control chart for sample size $m = 5$ and $m = 10$. The figure shows that $ARL$ values for CUSUM and SR-CUSUM control charts are almost the same for small skewed data while HL-CUSUM control chart has the smallest $ARL$ value for both sample size. This shows that HL-CUSUM control chart has the best performance for small skewed data. When the skewness increase, SR-CUSUM control chart has the smallest $ARL$ values compare to other two control chart for both sample size. This shows that SR-CUSUM control chart has the best performance when $m = 5$. All three control charts have comparable performance for data with large skewness when $m = 10$.

Table 3. Simulation result of $ARL$ for Weibull distribution with different skewness

| Control chart | | $m = 5$ CUSUM | HL-CUSUM | SR-CUSUM |
|---|---|---|---|---|
| Skewness coefficient | 0.1 | 149.21 | 107.39 | 152.62 |
| | 0.5 | 137.97 | 94.04 | 54.27 |
| | 1.0 | 94.44 | 75.56 | 4.07 |
| | 2.0 | 31.41 | 42.06 | 0.04 |
| | 3.0 | 5.93 | 30.97 | 0.00 |
| | 4.0 | 1.14 | 32.47 | 0.00 |
| | 5.0 | 0.13 | 42.41 | 0.00 |
| Control chart | | $m = 10$ CUSUM | HL-CUSUM | SR-CUSUM |
| Skewness coefficient | 0.1 | 148.22 | 100.58 | 147.61 |
| | 0.5 | 130.36 | 52.88 | 26.20 |
| | 1.0 | 120.05 | 0.92 | 0.99 |
| | 2.0 | 73.19 | 0.92 | 0.00 |
| | 3.0 | 26.71 | 0.25 | 0.00 |
| | 4.0 | 11.34 | 0.19 | 0.00 |
| | 5.0 | 3.97 | 0.14 | 0.00 |

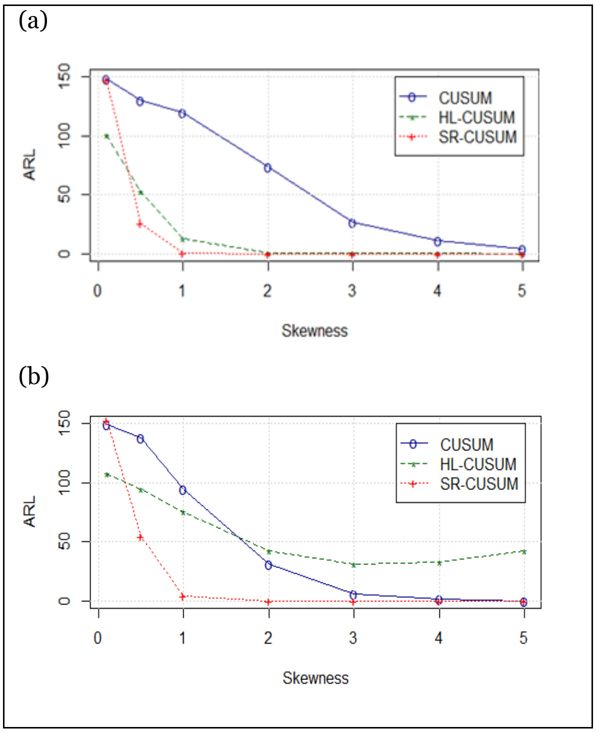


Figure 1. $ARL$ values for values for Weibull distribution with (a)$m = 5$ and (b)$m = 10$

Table 4 show the result of simulation by using data from Lognormal distribution with different skewness. $ARL$ values decreasing with the increase of the skewness of the data. $ARL$ values of CUSUM control chart decrease slowly for $m = 5$ and $m = 10$. For HL-CUSUM control chart,$ARL$ values decrease slowly when $m = 5$ and decrease drastically when $m = 10$ while $ARL$ values for SR-CUSUM control chart decrease drastically for both sample size.

Figure 2 shows the $ARL$ values for the three control chart graphically. The figure shows that $ARL$ values for CUSUM and SR-CUSUM control charts are almost the same for small skewed data while HL-CUSUM control chart has the smallest $ARL$ value for both sample size. This shows that HL-CUSUM control chart has the best performance for small skewed data. When the skewness increase, SR-CUSUM control chart has the smallest $ARL$ values compare to other two control chart for both sample size. This shows that SR-CUSUM control chart has the best performance when $m = 5$. All three control charts have comparable performance for data with large skewness when $m = 10$.

Table 4. Simulation result of *ARL* for Lognormal distribution with different skewness

| Control chart | | $m = 5$ | | |
|---|---|---|---|---|
| | | CUSUM | HL-CUSUM | SR-CUSUM |
| Skewness coefficient | 0.1 | 142.66 | 121.48 | 143.37 |
| | 0.5 | 136.35 | 119.61 | 56.36 |
| | 1.0 | 105.89 | 101.95 | 6.07 |
| | 2.0 | 38.24 | 89.47 | 0.17 |
| | 3.0 | 10.08 | 91.02 | 0.01 |
| | 4.0 | 3.77 | 96.05 | 0.00 |
| | 5.0 | 0.17 | 101.27 | 0.00 |
| Control chart | | $m = 10$ | | |
| | | CUSUM | HL-CUSUM | SR-CUSUM |
| Skewness coefficient | 0.1 | 151.02 | 125.87 | 147.76 |
| | 0.5 | 140.57 | 65.79 | 31.07 |
| | 1.0 | 130.52 | 20.11 | 1.55 |
| | 2.0 | 67.02 | 1.94 | 0.01 |
| | 3.0 | 32.10 | 0.42 | 0.00 |
| | 4.0 | 12.33 | 0.17 | 0.00 |
| | 5.0 | 6.73 | 0.12 | 0.00 |

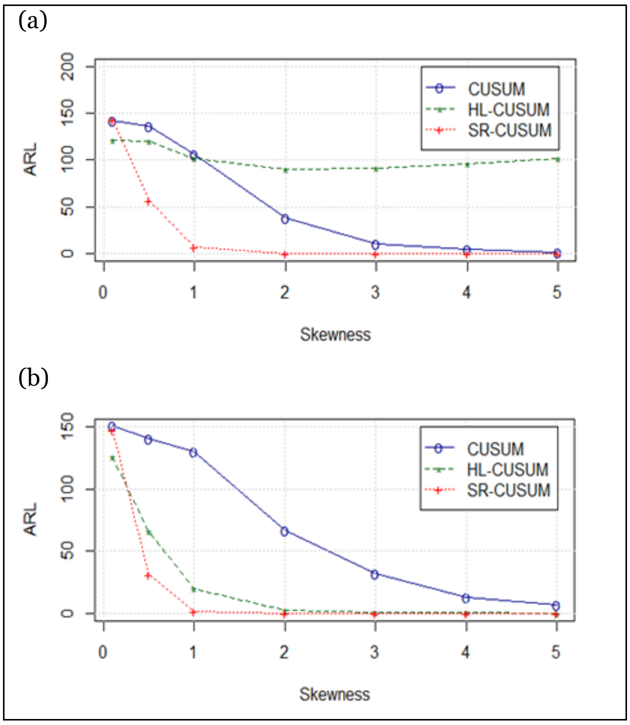


Figure 2. *ARL* values for values for Lognormal distribution with (a)$m = 5$ and (b)$m = 10$

## VI. CONCLUSION

To apply and check the performance of control chart, the normality assumption of the data is very important. If the normality assumption of the data is violated, misleading conclusion of that control chart can occur. Therefore, nonparametric statistics are used to overcome these problems. Hodges Lehmann estimator and Wilcoxon Signed-Rank statistics are used by integrating it with the classical CUSUM control chart to formulate nonparametric control charts which are HL-CUSUM and SR-CUSUM control chart.

From the simulation study, the *ARL* values of each control chart is computed and compared. Data from Weibull distribution and Lognormal distribution with different skewness are used by adjusting the parameter of each distribution. The parameter of CUSUM control chart $h = 5$ and $k = 0.5$ are used. The subgroup with sample size 5 and 10 are included in the study.

As the conclusion, SR-CUSUM control chart has the best performance among these three control charts when the data is non-normally distributed. Besides that, HL-CUSUM control chart can be used when the skewness of the data is small.

## VII. REFEREENCES

Abu-Shawiesh, M O & Abdullah, M B 2001. 'A new robust bivariate control chart for location'. *Communications in Statistics-Simulation and Computation*, vol. 30, no. 3, pp. 513-529.

Bakir, S T 2004. 'A distribution-free Shewhart quality control chart based on signed-ranks'. *Quality Engineering*, vol. *16*, no. 4, pp. 613-623.

Chakraborti, S & Eryilmaz, S 2007. 'A nonparametric Shewhart-type signed-rank control chart based on runs'. *Communications in Statistics—Simulation and Computation®*, vol. *36*, no. 2, pp. 335-356.

Chakraborti, S & Van de Wiel, M A 2008. *A nonparametric control chart based on the Mann-Whitney statistic* (pp. 156-172). Institute of Mathematical Statistics.

Chakraborti, S, Van der Laan, P & Bakir, S T 2001. 'Nonparametric control charts: an overview and some results'. *Journal of Quality Technology*, vol. *33*, no. 3, pp. 304-315.

Das, N & Bhattacharya, A 2008. 'A new non-parametric control chart for controlling variability'. *Quality Technology & Quantitative Management*, vol. *5*, no. 4, pp.351-361.

Fitzgerald, S, Dimitrov, D & Rumrill, P 2001. 'The basics of nonparametric statistics'. *Work*, vol. *16*, no. 3, pp. 287-292.

Graham, M A, Chakraborti, S & Human, S W 2011. 'A nonparametric exponentially weighted moving average signed-rank chart for monitoring location'. *Computational Statistics & Data Analysis*, vol. *55*, no.8, pp. 2490-2503.

Kaur, A & Kumar, R 2015. 'Comparative analysis of Parametric and Non-parametric Test'. *Journal of Computer and Mathematical Sciences*, vol. *6*, no. 6, pp. 336-342.

Koshti, V V 2011. 'Cumulative sum control chart'. *International Journal of Physics and Mathematical Sciences*, vol. *1*, no. 1, pp. 28-32.

Montgomery, D C 2007. *Introduction to statistical quality control*. John Wiley & Sons.

Nahm, F S 2016. 'Nonparametric statistical tests for the continuous data: the basic concept and the practical use'. *Korean journal of anesthesiology*, vol. *69*, no. 1, pp. 8.

Nazir, H Z, Riaz, M, Does, R J & Abbas, N 2013. 'Robust CUSUM control charting'. *Quality Engineering*, vol. *25*, no. 3, pp. 211-224.

Page, E S 1954. 'Continuous inspection schemes'. *Biometrika*, vol. *41*, no. 1/2, pp. 100-115.

Parkash, V, Kumar, D & Rajoria, R 2013. 'Statistical Process Control'. *Int. J. Res. Eng. Tech*, vol. *2*, pp. 70-72.

Pongpullponsak, A & Jayathavaj, V 2014. 'The new Hodges-Lehmann estimator control charting technique for the known process distributions'. In *Proceedings of International Conference on Applied Statistics (ICAS 2014)* (pp. 47-58).

Pongpullponsak, A, Suracherkiiati, W & Intaramo, R 2006. 'The Comparison of Efficiency of Control Chart by Weighted Variance Method, Scaled Weighted Variance Method, Empirical Quantiles Method and Extreme-Value Theory for Skewed Populations'. *CURRENT APPLIED SCIENCE AND TECHNOLOGY*, vol. *6*, no. 2a, pp. 456-465.

Roberts, S W 1959. 'Control chart tests based on geometric moving averages'. *Technometrics*, vol. *1*, no. 3, pp. 239-250.

Tapang, W, Pongpullponsak, A & Sarikavanij, S 2016. 'Three non-parametric control charts based on ranked set sampling'. *Chiang Mai Journal of Science*, vol. *43*, no. 4, pp. 914-929.

Yang, S F & Cheng, S W 2011. 'A new non-parametric CUSUM mean chart'. *Quality and Reliability Engineering International*, vol. *27,* no. 7, pp.867-875.