Volume 66, Issue 4, 2022
Journal of Scientific Research
of
The Banaras Hindu University

Nonparametric Control Charts to Jointly Monitor Scale and Location Parameters

M. P. Shah*[1], Sejal Desai[2]

*[1]Department of Statistics, Veer Narmad South Gujarat University, Surat. monikashah7@gmail.com
[2]J.Z.Shah Arts and H.P. Desai Commerce college, Amroli, Surat. sejal.desai6@gmail.com

Abstract: **Conventional Shewhart control chart works efficiently on many assumptions. The most important one is data should be normally distributed. But in many processes, we observed, process outcomes do not follow normal distribution. On another side nonparametric statistics works on very few assumptions. It does not require any knowledge regarding probability distribution of population under study. Also, to check process is under statistical control or not we need to draw two separate charts, chart for mean and chart for dispersion. But many statistics had been developed by researchers to jointly monitor location and scale parameters. To jointly monitor location and scale parameters without assuming distribution of parent population, in this article different non-parametric control charts based on Wilcoxon-Mood, Wilcoxon-Ansari Bradley, Wilcoxon-Siegel-Tukey, Wilcoxon-Conover Signed rank have been discussed. The performance of these charts under different distributions and sample sizes to control both parameters simultaneously, Monte Carlo simulations had been used.**

Index Terms: **Distribution free, Wilcoxon Signed Rank test, Mood, Ansari Bradley, Siegel-Tukey, simulation, Average Run Length (ARL)**

I. INTRODUCTION

Every product is made for some specific use, if it serves it completely then it is called of good quality. In different field definition of quality is different, but the concept is same. Conceptually a quality is working on few dimensions like, performance, reliability, durability, serviceability, aesthetics, features, perceived quality, and conformance to the standard.

In other way quality is inversely proportional to variability. Of course, some basic variation will be always there we can't control it exactly to the specified standard. We can tolerate variation up to some extent then after we have to identify the reasons which increase the variability in the product. So, the causes which degraded the quality of product arise due to materialistic error during the production process. To identify that the quality of variable is under control or not we use conventional Shewhart control chart. In Shewhart's control chart to check process is under control or not we use $\bar{X}$ with s or $\bar{X}$ with R chart under assumption of normality.

But if process data does not follow normal distribution the use of conventional Shewhart chart is suspicious. Also, to check process is under control or not we need to draw two separate charts, first chart to check process variability is under statistical control or not and second to check, specific standard is achieved or not. So basically, we draw two separate charts, one for dispersion and one for location parameter. But, can we simultaneously control on both parameters? Yes, many researchers suggested and proved the efficient applications of joint statistic to control process statistically.

1.1 Objective of the study:

To check the process is constructing standard/quality characteristics of the product or not we generally draw Shewhart's control-chart. To check process, for variable outcome, is under statistical control or not we draw two separate charts, first we draw s or R for dispersion and then we draw $\bar{X}$ for location. Many research had been done on simultaneous control of scale and location parameters. Again, Shewhart control chart works on assumption of normality, but in many cases process outcomes does not follow normal distribution but have skewed distributions. So here we have two main concerns of this study, first is joint monitoring of scale and location parameters and another is, that joint statistic should be a distribution-free.

1.2 Review of Literature

Some researchers did the deep research for the utilities of joint statistics and Nonparametric concepts in the SPC.

Chen and Cheng (1998) studied Max chart which combines two normalized statistics, one for the mean and one for

* Corresponding Author

the variance, by taking the maximum of the absolute values of two statistics. **A. F. B. Costa & M. A. Rahim (2004)** proposed a single chart that is based on the non-central chi-square statistic, which is more effective than the joint $\bar{X}$ and R charts in detecting assignable cause(s) that change the process mean and/or increase variability. They also shown that the EWMA control chart based on a non-central chi-square statistic is more effective in detecting both increases and decreases in mean and/or variability.

Gordan J.Ross, Dimitris K.Tasoulis, Niall M.Adams (2011) showed the shift may occur either in scale or in location parameter, hence it affects the whole model. For monitoring simultaneous change of both type of shifts they used Lepage type hypothesis test, $L=U^2+M^2$, where U is Mann Whitney test statistic for location and M is Mood test statistic for scale. Another Lepage type statistic used by **Mukherjee and Chakraborti (2012).** They proposed a single distribution-free control chart for joint monitoring of location and scale. The statistic is a combination of the Wilcoxon rank sum location statistic and Ansari-Bradely scale statistic to monitor process. **Chowdhury et al. (2014)** proposed distribution-free chart based on Cucconi statistic, for joint monitoring of location and scale parameters of continuous distribution. Performance of the chart is examined in a simulation study on the basis of the ARL, the standard deviation, the median and some percentiles of the run length distribution. **Wan-Chen Lee (2014)** introduced the procedures of deriving the distribution of the rank statistic considering the shift parameter, scale parameter individually and simultaneously by using finite Markov chain imbedding approach. And its power function using ARL had been discussed. **Pedro Carlos Oprime, et. al. (2016)** modified Shewhart Lepage control chart, to monitor location and scale parameters, proposed by Mukherjee and Chakraborti in 2012. To derive control limits, they used multiple linear regression model. **Jiujun Zhang, Erjie Li, Zhonghua Li (2017)** proposed ECvM chart, combining Exponentially weighted Moving Average (EWMA) and Cramer-von Mises test for joint monitoring of scale and location parameters. They compared performance of proposed chart with Shewhart Cucconi chart and Shewhart Lepage chart using ARL for various distributions.

ORIENTATION TOWARDS RESEARCH TOPIC CONSIDERED UNDER STUDY

II a) Brief introduction to statistics, which are going to be used

Wilcoxon (1945) proposed a test to identify significant change in location parameter of two samples, say X and Y with sample size m and n respectively. The statistic is very easy in computation and robust in its results. We need to assign ranks to combined sample of size N=m+n in ascending order. Compute W= Sum of the ranks assigned to any one sample. Too small or too large value of W indicates significance change in two location parameters. Alternatively, W could be expressed as

$$W = \sum_{i=1}^{N} iZ_i \qquad \text{………………. (1)}$$

Where,

$$Z_i = \begin{cases} 1, & when\ i^{th}\ order\ statistic\ of\ combined\ sample\ is\ a\ X \\ 0, & when\ i^{th}\ order\ statistic\ of\ combined\ sample\ is\ a\ Y \end{cases}$$

$$Z = \frac{W-\mu_W}{\sigma_W} \sim N(0,1) \qquad \text{………………. (2)}$$

where E(W) and V (W) can be found using

$$E(W) = \frac{m(N+1)}{2} \text{ and } V(W) = \frac{mn(N+1)}{12} \qquad \text{………………. (3)}$$

Mood (1954) developed a nonparametric test for equality of variances. For two independent random samples X and Y of size m and n respectively. In the combined sample of N=m+n observations with no ties, the average rank is the mean of first N integers (N+1)/2. The amount of deviations of the ranks of the observation about this mean is an indication of relative spread. Mood (1954) presented a test that is based on the sum of squares of the deviations of ranks of Y's from the average combined rank. The Mood statistic for scale is

$$M_N = \sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right)^2 Z_i \qquad \text{………………. (4)}$$

$$Z_i = \begin{cases} 1, & when\ i^{th}\ order\ statistics\ of\ the\ combined\ N\ observations\ is\ a\ Y. \\ 0, & when\ i^{th}\ order\ statistics\ of\ the\ combined\ N\ observations\ is\ a\ X. \end{cases}$$

A large value of M_N would imply that Y's are more widely dispersed, and M_N small implies the opposite conclusion. Mood (1954) derived the mean and variance of his test statistic as,

$$E(M_N) = \frac{n(N^2-1)}{12} \text{ and}$$

$$V(M_N) = \frac{mn(N+1)(N^2-4)}{180} \qquad \text{…………….. (5)}$$

Freund and Ansari (1957) later with some variations added by David and Bartlon (1958) and Ansari Bradley (1960) developed a nonparametric test for equality of variances. We consider two independent random samples X and Y of size m and n respectively. In the combined sample of N = m+n observations with no ties, the average rank is the mean of first N integers (N+1)/2. The amount of deviations of the ranks of the observation about this mean is an indication of relative spread. The Freund-Ansari-Bradley statistic for scale is

$$F_N = \frac{m(N+1)}{2} - \sum_{i=1}^{N} \left| i - \frac{N+1}{2} \right| Z_i \quad \text{……………………. (6)}$$

$$Z_i = \begin{cases} 1, when\ i^{th}\ order\ statistics\ of\ the\ combined\ N\ observations\ is\ a\ Y. \\ 0, when\ i^{th}\ order\ statistics\ of\ the\ combined\ N\ observations\ is\ a\ X. \end{cases}$$

A large value of F_N would imply that Y's are more widely dispersed, and M_N small implies the opposite conclusion. The mean and variance of this test statistic as,

$$E(F_N) = \frac{m(N+1)^2}{4N} \quad for\ \ N\ odd$$

and

$$E(F_N) = \frac{m(N+2)}{4} \quad for\ \ N\ even$$

$$V(F_N) = \frac{mn(N+1)(N^2+3)}{48N^2} \quad for\ N\ \ odd$$

$$V(F_N) = \frac{mn(N^2-4)}{48(N-1)} \quad for\ N\ \ even \quad \text{……….. (7)}$$

Siegel & Tukey (1960) suggested test for dispersion. The basic idea behind the test is that if two samples for two random variables differing in their variance, the sample from the population with greater variance will be more spread out. If there is a location difference the population should be aligned by shifting the median (mean) of one sample to coincide with that of the other sample. This requires an appropriate addition to or subtraction of a constant for all observations in one sample. The sample variance is unaltered by this change. Then the combined sample is arranged in order of their values and the rank 1 is assigned to the smallest variable, 2 to the largest, 3 to the next largest, 4 and 5 to the next two smallest, 6 and 7 to the next two largest and so on.

i.e. the rearrangement is:

i	1	2	3	…	N/2	…	N-2	N-1	N
ai	1	4	5	…	N	…	6	3	2

For N even, and if N is odd, the middle observation is removed from the calculation.

$$S_N = \sum_{i=1}^{N} a_i Z_i \quad \text{……………….. (8)}$$

$$a_i = \begin{cases} 2i & for\ i\ even, 1 < i \le \frac{N}{2} \\ 2i-1 & for\ i\ odd, 1 \le i \le \frac{N}{2} \\ 2(N-i)+2 & for\ i\ even, \frac{N}{2} < i \le N \\ 2(N-i)+1 & for\ i\ odd, \frac{N}{2} < i < N \end{cases}$$

The sum of the ranks of the sample for the random variable having the greater variance should be smaller than if there were no difference in variance. The smaller sum of ranks associated with one sample is calculated and tested for significance.

Large Sample approximation

$$E(S_N) = \frac{m(N+1)}{2} \ \ \&\ V(S_N) = \frac{mn(N+1)}{12} \quad \text{………………….. (9)}$$

Cucconi (1968) is very firstly developed the concept of jointly monitor location and scale parameters. It is easier to be computed because it requires only the ranks of one sample in the combined sample.

The Cucconi test is based on the following statistic:

$$C = \frac{U^2+V^2-2\rho UV}{2(1-\rho^2)} \quad \text{……………….. (10)}$$

where U is based on the standardized sum of squared ranks of the first sample elements in the pooled sample, and V is based on the standardized sum of squared contrary-ranks of the first sample elements in the pooled sample. ρ is the correlation coefficient between U and V. The test statistic rejects for large values.

Yves Lepage (1971) developed Concept of "Combination of Wilcoxon's and Ansari-Bradley's Statistics". And suggested the below statistic

$$L_1 = \left(\frac{W-E(W)}{\sqrt{V(W)}}\right)^2 + \left(\frac{F_N-E(F_N)}{\sqrt{V(F_N)}}\right)^2 \quad \text{…………….. (11)}$$

where W is Wilcoxon rank sum statistic for shift in location and F_N is Ansari-Bradley rank statistic for shift in scale.

Pettitt (1976) suggested Combination of Wilcoxon's and Mood test statistic for simultaneously monitoring the location and the scale,

$$L_2 = \left(\frac{W-E(W)}{\sqrt{V(W)}}\right)^2 + \left(\frac{M_N-E(M_N)}{\sqrt{V(M_N)}}\right)^2 \quad \text{…………….. (12)}$$

where W is Wilcoxon rank sum statistic for shift in location and M_N is Mood statistic for shift in scale.

II b) PROCEDURES TO CONSTRUCT CHART

Mukherjee and Chakraborti (2012) proposed a single nonparametric control chart based on Lepage (1971) test statistic for simultaneously monitoring the location and the scale parameters of a continuous process. The single plotting statistic for the joint monitoring of location and scale is given by L_1 in equation (11) and chart is called Shewhart-Lepage (SL) chart. They considered X= (X_1, X_2, ..., X_m), as reference sample of size m from an in-control process and that Y = (Y_1, Y_2, ..., Y_n) an arbitrary test sample of size n. As the value of charting statistic $L_1 \geq 0$, lower control limit (LCL) of the chart is 0.

Chowdhury et al. (2013) proposed a single nonparametric control chart based on Cucconi (1971) test statistic for simultaneously monitoring the location and the scale parameters of any continuous process. The single plotting statistic for the joint monitoring of location and scale is given C in equation (10) and chart is called Shewhart-Cucconi (SC) chart. They considered X= (X_1, X_2, ..., X_m), as reference sample of size m from an in-control process and that Y= (Y_1, Y_2, ..., Y_n) an arbitrary test sample of size n. As the value of charting statistic $C \geq 0$, LCL of the chart is 0.

Zombade Digambar Mahadev, Dr. Ghute V.B. (2015) proposed a single nonparametric chart based on Pettitt test statistic for simultaneously monitoring the location and the scale parameters of any continuous process. The single plotting statistic for the joint monitoring of location and scale is given L_2 in equation (12) and chart is called Shewhart-Pettitt (SP) chart. They considered X= (X_1, X_2, ..., X_m), as reference sample of size m from an in-control process and that Y= (Y_1, Y_2, ..., Y_n) an arbitrary test sample of size n. As the value of charting statistic $L_2 \geq 0$, LCL of the chart is 0.

We develop a single nonparametric chart based on Wilcoxon signed rank test for location and Siegel-Tukey test for scale. Here we consider location parameter(mean) is known and scale parameter (dispersion) is unknown. To draw control chart with plotting statistics we need to follow steps given below.

Step-1: Collect reference sample X from in-control process of size m and known mean μ_0.

Step-2: Assuming a subgroup Yi's of size n, compute the W statistic and Siegel-Tukey statistic from equations (1) and (8) for Yi against the reference sample X and denote them as Wi and S_{Ni} respectively.

Step-3: Standardize Wi and Si with their corresponding means and variances.

Step-4: Calculate the Wilcoxon-Seigel Tukey plotting statistic $T_i = T_{1i}^2 + T_{2i}^2$, i=1, 2, … where T_{1i} is standardized statistic of Wi, using eq.(2) & (3), and T_{2i} is standardized statistic of S_{Ni} using eq.(8) & (9).

Step 5. Plot Ti against an UCL > 0. The LCL is zero. Note that Ti ≥ 0 and larger value of Ti suggest an out-of-control process.

Step 6. If Ti exceeds UCL the process is declared out-of-control at the i^{th} test sample. If not, the process is thought to be in-control.

II c) DERIVE UCL FOR DIFFERENT SAMPLE SIZES

To test process is under statistical control or not, we need to derive control limits. Here LCL is 0, and deriving UCL is not an easy task, as distribution of data is not known. To derive UCL we used simulation. Performance of the chart can be measured by Average Run length distribution (ARL), expected value of the run length distribution.

$$\text{ARL}_0 = E\left(\frac{1}{P(T_i > UCL|X}\right)$$

Larger the value of the in-control ARL better the performance of the chart with respect to false alarm. The main task of a control chart is to detect the change in the process as quickly as possible and give an out-of-control signal. Clearly the quicker the detection and the signal, the more efficient the chart is. In conventional Shewhart control chart for 3σ limits ARL is 370. To derive UCL we choose ARL_0 as 250, 370, and 500.

For given pair (m, n) a search is conducted for different values of UCL, and that value of UCL is obtained for different values of in-control ARL. Computer programs written in R is used to derive UCL of the control charts under different distributions: Normal, Weibull, and beta using a simulation based on 10000 runs for sample size of n=5, 10 and 15, and size of reference sample, m=30.

Table 1.
UCL of different control charts for different sample sizes for N (5,1) (Symmetric)

ARL	Chart	n=5	n=10	n=15
250	Shewhart Cucconi	5.51	4.73	4.46
	Shewhart Lepage	16.28	18.06	34.84
	Shewhart Pettitt	8.75	21.11	28.06
	Shewhart Wilcoxon-Siegel	9.41	23.36	27.88
370	Shewhart Cucconi	5.89	5.14	4.74
	Shewhart Lepage	17.22	19.15	36.21
	Shewhart Pettitt	9.27	22.07	29.41
	Shewhart Wilcoxon-Siegel	9.71	24.34	29.98
500	Shewhart Cucconi	6.19	5.43	4.91
	Shewhart Lepage	17.37	19.55	37.32
	Shewhart Pettitt	9.52	22.39	30.32
	Shewhart Wilcoxon-Siegel	10.19	24.78	30.31

Table 2.
UCL of different control charts for different sample sizes for Weibull (5,3) (skewed)

ARL	Chart	n=5	n=10	n=15
250	Shewhart Cucconi	5.06	4.39	4.16
	Shewhart Lepage	11.31	16.86	27.75
	Shewhart Pettitt	12.19	16.8	29.87
	Shewhart Wilcoxon-Siegel	12.52	23.44	29.55
370	Shewhart Cucconi	5.46	4.61	4.34
	Shewhart Lepage	11.9	17.8	28.97
	Shewhart Pettitt	12.84	18.05	31
	Shewhart Wilcoxon-Siegel	13.02	24.56	30.51
500	Shewhart Cucconi	5.73	4.75	4.56
	Shewhart Lepage	12.22	18.12	29.65
	Shewhart Pettitt	12.94	18.58	32.02
	Shewhart Wilcoxon-Siegel	13.52	25.07	31.43

Table 3.
UCL of different control charts for different sample sizes for Beta (7, 3) (skewed)

ARL	Chart	n=5	n=10	n=15
250	Shewhart Cucconi	4.87	4.13	4.33
	Shewhart Lepage	11.31	17.51	29.21
	Shewhart Pettitt	9.12	19.77	27.62
	Shewhart Wilcoxon-Siegel	13.45	18.81	36.4
370	Shewhart Cucconi	5.34	4.31	4.69
	Shewhart Lepage	11.8	18.56	30.31
	Shewhart Pettitt	9.92	20.54	28.81
	Shewhart Wilcoxon-Siegel	14.18	19.58	37.9
500	Shewhart Cucconi	5.65	4.5	4.9
	Shewhart Lepage	12.26	19.04	31.1
	Shewhart Pettitt	10.34	20.94	29.24
	Shewhart Wilcoxon-Siegel	14.51	20.27	38.34

III FINDINGS

Observing each table, we can conclude that for each kind of distribution, as sample size-n is increasing, UCL is also increasing, for Shewhart Lepage chart, Shewhart Pettitt chart and for Shewhart Wilcoxon Siegel chat. While for Shewhart Cucconi chart we found inverse relation between sample size-n and UCL.

For Shewhart- Cucconi chart and Shewhart Lepage chart, UCLs for each sample size is largest for symmetric distribution. That is, we can say, these charts perform better in symmetric data compare to skewed data.

For Shewhart Pettitt chart we observe that UCLs for each sample size is highest for Weibull distribution. That is, for Weibull distribution Shewhart-Pettitt perform better.

For Shewhart Wilcoxon-Siegel chart we observe that UCLs for each sample size is greater for beta distribution. That is, for beta distribution Wilcoxon-Siegel perform better.

IV CONCLUSIONS

Every process works differently in different circumstances, each of these process outcomes may not follow normal distribution, hence use of conventional Shewhart control chart is inadequate. In such cases nonparametric process control charts are best alternatives. Also, to monitor process variability and process location parameters, we used to draw two separate charts. But here we discussed four different nonparametric control charts. For different distributions charts are performing differently, and results of simulation is discussed in FINDINGS.

REFERENCES

1) Chen & Cheng (1998): Max Chart: Combining X-Bar Chart and S Chart, *Statistica Sinica* 8, pp.263-271
2) Cucconi O.(1968): Un nuovo test non parametrico per il confronto tra due gruppi campionari. *Giornale degli Economisti*; XXVII:225–248. 2.
3) Yves Lepage (1971): A Combination of Wilcoxon's and Ansari-Bradley's Statistics. *Biometrika*, Vol. 58, No. 1, pp. 213-217
4) Pettitt, A. N. (1979): A non-parametric approach to the change point problem. Journal of the Royal Statistical Society Series C, *Applied Statistics* 28, 126-135.
5) A. F. B. Costa & M. A. Rahim (2004): Monitoring Process Mean and Variability with One Non-central Chi-square Chart, *Journal of Applied Statistics*, Vol. 31, No. 10, pp.1171–1183
6) Saad.T. Bakir (2004): A Distribution-Free Shewhart Quality Control Chart Based on Signed-Ranks. *Quality Engineering*, Vol.16, No.4, pp:613-623
7) S. Chakraborti1 and S. Eryilmaz (2007): A Nonparametric Shewhart-Type Signed-Rank Control Chart Based on Runs. *Communications in Statistics—Simulation and Computation*, 36:2, pp:335-356
8) Nandini Das (2009): A comparison study of three non-parametric control charts to detect shift in location parameters, *Int J Adv Manuf Technol* (2009) 41: pp:799–807
9) Chunguang Zhou, Changliang Zou, Yujuan Zhang, Zhaojun Wang (2009): Nonparametric control chart

based on change point model. *Stat-papers, Springer-Verlag*, pp:13-28

10) S.K.Khilare, D.T.Shirke (2010): A Nonparametric Synthetic Control Chart using Sign Statistic, *Communications in Statistics-Theory and Methods*, Vol 39, No. 18, pp:3282-3293.
11) Gordan J.Ross, Dimitris K.Tasoulis, Niall M. Adams (2011): Nonparametric monitoring of data steams for changes in location and scale, *Technometrics*, Vol 53, No.4, pp.379-389.
12) Mukherjee, A. and Chakraborti, S. (2012). A nonparametric phase II control chart for simultaneous monitoring of location and scale. *Quality and Reliability Engineering International*, 28(3), 335-352.
13) J. Ross, Niall M. Adams (2012): Two Nonparametric control charts for detecting arbitrary distribution changes, *Jounal of Quality Technology*, Vol 44, No.2, pp:102-116
14) S.K.Khilare and D.T. Shirke (2012): Nonparametric synthetic control charts for process variation. *Quality and reliability engineering International*, Vol.28, No.2, pp:193-202.
15) M.P.Shah, S.N.Desai (2018): An Empirical study on application of Wilcoxon signed rank test in Statistical Process control when process target is fixed. *International Journal of Management, IT & Engineering* Vol. 8 Issue 8, pp.130-136
16) A. K. McCracken and S. Chakraborti (2013): Control Charts for Joint Monitoring of Mean and Variance: An Overview. *Quality Technology & Quantitative Management* Vol. 10, No. 1, pp. 17-36
17) Chowdhury, S. Mukherjee A. and Chakraborti, S. (2014): A new distribution free control chart for joint monitoring of unknown location and scale parameters of continuous distributions, *Quality and Reliability Engineering International*, 30(2), pp.191-204
18) Wan-Chen Lee (2014): Joint distribution of rank statistics considering the location and scale parameters and its power study, *Journal of Statistical Distributions and Applications*, pp.1:16
19) Pedro Carlos Oprime, José Carlos de Toledo, Mario Orestes Aguirre González, S. Chakraborti (2016): Method for determining the control limits of nonparametric charts for monitoring location and scale, Gest. Prod., v. 23, n. 1, pp. 146-164.
20) Jiujun Zhang, Erjie Li, Zhonghua Li (2017): A Cram´er-von Mises Test-based Distribution-free Control Chart for Joint Monitoring of Location and Scale, Computers & Industrial Engineering
21) Wei-Heng Huang (2021): Control Charts for Joint Monitoring of the Lognormal Mean and Standard Deviation Symmetry 2021, 13, pp-549
22) Montgomery, D.C. Statistical Quality Control, 8th ed.; John Wiley & Sons (Asia) Pte. Ltd.: Hoboken, NJ, USA, 2019.

***