Nonparametric Control Charts for Monitoring Process Variability

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

V. B. Ghute
Department of Statistics,
Solapur University,
Solapur-413255, (M.S.), India.

Abstract--**In this paper, two nonparametric control charts are developed for monitoring the process variability. The charts are Shewhart-type charts and are based on nonparametric two sample tests for testing equality of variance developed by Sukhatme and Mood. The performance of the proposed control charts is evaluated through average run length for the normal, double exponential and uniform distributions.**

Keywords: Average run length, Control chart, Process variability, Statistical process control, Nonparametric tests.

1. INTRODUCTION

Control charts are statistical process control tools that are widely used for controlling and monitoring a process. Shewhart R and S control charts are most popular control charts for monitoring process variability. Both of these control charts are designed and evaluated under the assumption that the underlying distribution of the quality characteristic is normal. In real applications, there are many situations in which the process data come from a non-normal distribution which need to be monitored by appropriate control charts. To monitor such type of data, development of control charts that do not depend on a particular distributional assumption is desirable. Nonparametric control charts can serve this purpose. The main advantage of a nonparametric control chart is that it does not assume any probability distribution for the characteristic of interest. A formal definition of nonparametric or distribution-free control chart is given in terms of its in-control run length distribution. The number of samples that needs 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 distribution-free or nonparametric (Chakraborti et al. (2004)).

In literature, several nonparametric control charts are proposed for monitoring location of a univariate process. Some of these are based on sign and/or rank statistics by assuming a known in-control target value for process location. Chakraborti et al. (2001) presented an extensive overview of the literature on univariate nonparametric control charts. Bakir (2004) developed a distribution-free Shewhart control chart for monitoring process center based on the signed-ranks of grouped observations. Bakir (2006) proposed Shewhart, CUSUM and EWMA control charts based on signed-rank-like statistics of grouped data for monitoring a process center when in-control target center was not specified and studied the robustness of the charts against outliers. There exist only few articles on nonparametric charts for monitoring process variability. Lehmann (1975) suggested using non-parametric tests for the equality of two variances for use as control statistics in nonparametric control charts for variability. Control charts using tests statistics for comparing two variances would require obtaining an initial sample (of size m) when the process is considered to be in-control. Then at each sample time i, a sample of size n is obtained from the process, and the pooled sample of size $(m + n)$ is obtained. The observations in the pooled sample then are ranked from smallest to largest, and some statistic based on the ranks of the observations is calculated. Das and Bhattacharya (2008) proposed a nonparametric control chart for monitoring process variability based on Conver's squared rank test for variance. Das (2008) developed two nonparametric control charts for monitoring process variability based on two nonparametric tests. Murakani and Matsuki (2010) proposed a nonparametric control chart for dispersion based on the rank sum statistic.

When the process distribution is normal, Shewhart R and S charts are appropriate control chart for monitoring the process variability. If underlying process distribution is non-normal, then the need of development of nonparametric control chart based on appropriate nonparametric test arises. In this paper, we introduce two Shewhart-type nonparametric control charts for monitoring process variability for the case that the location parameter is under control. The proposed nonparametric control charts are based on two sample nonparametric tests proposed by Sukhatme (1956) and Mood (1954). These are most powerful test statistics for detecting scale shifts. The performance of the proposed charts is assessed for both the in-control state and out-of-control state under different underlying distributions.

2. MATERIALS AND METHODS

2.1 Sukhatme test based nonparametric control chart

Sukhamte (1956) proposed a nonparametric test for two independent samples dispersion problem. Suppose we want to compare two independent random samples

$X = (X_1, X_2, ..., X_m)$ and $Y = (Y_1, Y_2, ..., Y_n)$ which are drawn from absolute continuous distributions and differ only in the scale parameters. Let σ_X and σ_Y be the arbitrary measures of dispersion of X and Y respectively then problem of testing of hypothesis is $H_0 : \sigma_X = \sigma_Y$ against $H_1 : \sigma_X \neq \sigma_Y$. The Sukhamte test statistic for testing null hypothesis is defined as,

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} D(X_i, Y_j), \qquad (1)$$

$$\text{where } D(X, Y) = 1 \text{ if either } 0 < X < Y \text{ or } Y < X < 0$$
$$= 0 \text{ otherwise}$$

We reject hypothesis if T is too large or too small. The mean and variance of the statistic T is given by, $E(T) = \frac{1}{4}$ and

$$\text{Var } (T) = \frac{(m + n + 7)}{48mn}$$

For a large sample,

$$Z = \frac{T - E(T)}{\sqrt{\text{Vat } (T)}} \qquad (2)$$

has a standard normal distribution and the test is performed on the basis of tabulated values of the standard normal distribution and the test is performed on the basis of tabulated values of the standard normal distribution.

We consider Z as the control chart statistic for the nonparametric control chart for monitoring process variability and the chart is referred as NP-S chart. We consider $X = (X_1, X_2, ..., X_m)$, as reference sample of size *m* from an in-control process and that $Y = (Y_1, Y_2, ..., Y_n)$ be an arbitrary test sample of size *n*. The sample statistics Z computed from independent observations from the process are plotted against an upper control limit UCL = 3 and LCL = -3. The process is considered out-of-control when a plotted point lies above UCL or below LCL.

2.2 Mood test based nonparametric control chart

Mood (1954) developed a nonparametric test for equality of variances.

Suppose, we have two independent random samples $X = (X_1, X_2, ..., X_m)$ and $Y = (Y_1, Y_2, ..., Y_n)$. We wish to test $H_0 : \sigma_X = \sigma_Y$ against $H_1 : \sigma_X \neq \sigma_Y$. Let $R_1 < R_2 < .... < R_m$ be the combined samples ranks of the X-values in increasing order of magnitude. The Mood test statistic for testing null hypothesis is defined as,

$$M = \sum_{i=1}^{n} \left(R_i - \frac{N+1}{2} \right)^2, \text{ where } N = m + n \qquad (3)$$

The mean and variance of the statistic M is given as

$$E(M) = \frac{m(N^2 - 1)}{12}$$

$$\text{and } \text{Var } (M) = \frac{mn(N+1)(N^2 - 4)}{180}$$

For N greater than or equal to 30, we may consider the normalized random variable W

$$W = \frac{M - E(M)}{\sqrt{\text{Var}(M)}} \qquad (4)$$

and perform the test on the basis of tabulated values of the standard normal distribution.

For N less than 30, it is not advisable to use directly the normal approximation. In that case, Laubscher recommended the use of a correction for continuity yielding the following test statistic:

$$W = \frac{M - E(M)}{\sqrt{\text{Var}(M) + \frac{1}{2\,\text{Var}(M)}}} \qquad (5)$$

and perform the test again based on the tabulated values of the standard normal distribution.

We consider W as the control chart statistic for the nonparametric control chart for monitoring process variability and the chart is referred as NP-M chart. We consider $X = (X_1, X_2, ..., X_m)$, as reference sample of size *m* from an in-control process and that $Y = (Y_1, Y_2, ..., Y_n)$ be an arbitrary test sample of size *n*. The sample statistics W computed from independent observations from the process are plotted against an upper control limit UCL = 3 and LCL = -3. The process is considered out-of-control when a plotted point lies above UCL or below LCL.

3. PERFORMANCE OF THE PROPOSED CONTROL CHARTS

To examine the ability of proposed NP-S and NP-M charts to detect variability shift in a process, we consider underlying process distributions as normal, double exponential and uniform with mean zero and variance one. The uniform distribution is considered as process distribution to see the effect of a light tailed distribution and double exponential distribution is considered to see the effect of heavy tailed distribution on the performance of proposed nonparametric control charts. Consider a process where quality characteristic of interest X is distributed with mean μ and standard deviation σ. Let μ_0 and σ_0 be the in-control values of μ and σ respectively. When a shift in process standard deviation occurs, we have change from the in-control value σ_0 to the out-of-control value $\sigma_1 = \delta\sigma_0$ $(0 < \delta \neq 1)$. Therefore, when control chart for variability is employed, the process shifts are measured through

$\delta = \frac{\sigma_1}{\sigma_0}$. When $\delta = 1$, the process is considered to be in-control . For $\delta > 1$ an increase in σ occurs and for $\delta < 1$, decrease in σ occurs. Computer programs written in C language are used to study the performance of the proposed control charts. The in-control and out-of-control ARL values of the proposed control charts are computed using 10000 simulations for sample size of n = 10, 15, 20 and 25.

Table 1 to Table 4 provide the ARL values of the proposed nonparametric control charts when the underlying process data actually follows normal, double exponential and uniform distributions with sample sizes n =10, 15, 20 and 25 respectively.

Table-1: ARL values of NP-S and NP-M charts when n = 10

Shift δ	NP-S Chart			NP-M Chart		
	Normal	Double Exponential	Uniform	Normal	Double Exponential	Uniform
1.0	332.45	340.16	339.58	328.93	332.94	335.48
1.2	158.54	189.13	116.89	135.24	174.71	85.03
1.4	85.85	133.77	53.03	62.06	92.96	30.26
1.6	51.65	75.38	29.67	33.09	54.86	14.83
1.8	34.50	52.40	19.90	20.13	36.10	9.16
2.0	24.85	40.09	14.01	13.66	24.69	6.43
2.2	18.59	30.24	10.71	9.91	18.15	5.01
2.4	15.00	24.80	8.70	6.82	12.63	3.71
2.6	12.24	20.71	7.44	6.13	11.67	3.44
2.8	10.04	17.16	6.38	5.13	9.51	2.99
3.0	8.84	14.98	5.60	4.39	7.96	2.69
3.5	6.40	10.73	4.45	3.34	5.79	2.25
4.0	5.24	8.57	3.71	2.71	4.45	1.95
4.5	4.36	7.05	3.26	2.34	3.67	1.79
5.0	3.82	5.88	2.92	2.10	3.16	1.67

Table-2: ARL values of NP-S and NP-M charts when n = 15

Shift δ	NP-S Chart			NP-M Chart		
	Normal	Double Exponential	Uniform	Normal	Double Exponential	Uniform
1.0	414.70	419.74	415.25	420.16	415.82	411.57
1.2	164.85	213.42	105.95	137.21	184.70	66.75
1.4	68.36	103.25	35.08	46.02	78.83	17.62
1.6	34.13	58.25	16.18	20.50	39.69	7.49
1.8	19.71	34.98	9.46	11.25	23.02	4.53
2.0	12.61	23.20	6.41	7.10	14.85	3.19
2.2	8.94	16.49	4.80	5.00	10.45	2.44
2.4	6.56	12.65	3.76	3.59	6.75	1.91
2.6	5.29	9.76	3.14	3.05	5.95	1.80
2.8	4.30	8.03	2.69	2.59	4.83	1.64
3.0	3.72	6.65	2.41	2.23	4.05	1.53
3.5	2.67	4.57	1.93	1.74	2.85	1.32
4.0	2.20	3.53	1.67	1.49	2.26	1.23
4.5	1.86	2.83	1.52	1.36	1.92	1.18
5.0	1.66	2.43	1.42	1.28	1.71	1.16

Table-3: ARL values of NP-S and NP-M charts when n = 20

Shift δ	NP-S Chart			NP-M Chart		
	Normal	Double Exponential	Uniform	Normal	Double Exponential	Uniform
1.0	332.45	340.16	339.58	418.01	414.41	416.31

1.2	158.54	189.13	116.89	109.29	166.97	45.11
1.4	85.85	133.77	53.03	31.19	58.02	10.23
1.6	51.65	75.38	29.67	12.73	26.67	4.42
1.8	34.50	52.40	19.90	6.61	14.51	2.65
2.0	24.85	40.09	14.01	4.09	8.92	1.95
2.2	18.59	30.24	10.71	2.93	6.11	1.58
2.4	15.00	24.80	8.70	2.28	4.50	1.42
2.6	12.24	20.71	7.44	1.87	3.46	1.28
2.8	10.04	17.16	6.38	1.66	2.87	1.20
3.0	8.84	14.98	5.60	1.47	2.44	1.16
3.5	6.40	10.73	4.45	1.25	1.81	1.09
4.0	5.24	8.57	3.71	1.15	1.51	1.05
4.5	4.36	7.05	3.26	1.09	1.34	1.04
5.0	3.82	5.88	2.92	1.07	1.24	1.03

Table-4: ARL values of NP-S and NP-M charts when n = 25

Shift	NP-S Chart			NP-M Chart		
δ	Normal	Double Exponential	Uniform	Normal	Double Exponential	Uniform
1.0	416.96	420.62	420.04	418.44	424.64	426.27
1.2	108.62	154.37	55.64	90.68	139.41	33.15
1.4	31.47	55.39	13.23	22.46	44.64	6.83
1.6	12.92	24.67	5.44	8.39	19.05	2.96
1.8	6.82	13.41	3.19	4.42	9.86	1.88
2.0	4.14	8.32	2.25	2.83	5.81	1.46
2.2	2.96	5.61	1.78	2.05	4.10	1.27
2.4	2.27	4.15	1.51	1.65	3.05	1.16
2.6	1.88	3.26	1.36	1.44	2.41	1.11
2.8	1.64	2.62	1.26	1.29	2.02	1.07
3.0	1.48	2.28	1.19	1.20	1.76	1.05
3.5	1.24	1.69	1.10	1.09	1.39	1.02
4.0	1.14	1.42	1.05	1.04	1.21	1.01
4.5	1.08	1.26	1.03	1.02	1.13	1.01
5.0	1.05	1.17	1.02	1.02	1.08	1.00

Examinations of Table 1 to Table 4 lead to the following findings:

- In-control ARL values of the proposed NP-S and NP-M control charts for different process distributions are approximately same.
- Out-of-control ARL values of NP-M chart are smaller than that of the NP-S chart. Therefore, NP-M chart is more efficient than NP-S chart for normal, light tailed uniform and heavy tailed double exponential distributions.
- For normally distributed data, both NP-S and NP-M charts performs better than double exponential data.
- For uniformly distributed data, both NP-S and NP-M charts perform better than normally and doubly exponential data.

4. CONCLUSIONS

In this paper, two nonparametric control charts are developed for monitoring process variability. The performance of the proposed control charts is studied by simulation under normal, light tailed and heavy tailed distributions. Our simulation study indicates that the NP-M control chart is more efficient than NP-S control chart for detecting shifts in process variability for different process distributions. Both NP-M and NP-S control charts perform better when underlying process distribution is light tailed.

REFERENCES

[1] Bakir, S. T. (2004). A distribution-free Shewhart quality control chart based on signed-ranks, *Quality Engineering*, 16, 613-623.

[2] Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics, *Communications in Statistics- Theory and Methods*, 35, 743-757.

[3] Chakraborti, S., Van der Laan, P. and Bakir, S. T. (2001). Nonparametric control charts: An overview and some results, *Journal of Quality Technology*, 3, 304-315.

[4] Chakraborti, S. Van der Laan, P. and Van de Wiel, M. A. (2004). A class of distribution- free control charts, *Journal of Royal Statistical Society, Series C: Applied Statistics*, 53(3), 443-462.

[5] Das, N. (2008). Non-parametric control chart for controlling variability based on rank test, *Economic Quality Control*, 2, 227-242.

[6] Das, N. and Bhattacharya, A. (2008). A new nonparametric control chart for controlling variability, *Quality Technology and Quantitative Management*, 5(4), 351-361.

[7] Lehmann, E.L. (1975). *Nonparametric Statistical Methods Based on Ranks.* Holden-Day, San Fransisco, California.
[8] Mood, A.M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests, *Annals of Mathematical Statistics*, 25, 514-522.
[9] Murakami, H. and Matsuki, T. (2010). A nonparametric control chart based on the Mood statistic for dispersion, *International Journal of Advanced Manufacturing Technology*, 49, 757-763.
[10] Sukhamte, B.V. (1956). On certain two-sample nonparametric tests for variances, *The Annals of Mathematical Statistics,* 28(1), 188-194.