Percentile bootstrap control chart for monitoring process variability under non-normal processes

Nadia Saeed¹, Shahid Kamal² and Muhammad Aslam³*

¹College of Statistical and Actuarial Sciences, University of the Punjab Lahore-54000, Pakistan; Email: nadia.stat@pu.edu.pk
²GC University Faisalabad, Faisalabad, Pakistan; Email: kamal_shahid@hotmail.com
³Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21551, Saudi Arabia; Email: aslam_ravian@hotmail.com, Tel: 00966593329841

*Corresponding author

Abstract

In the recent years, another approach named as the bootstrap method is getting popular in statistical process control specifically when the underlying distribution of the process is unknown. The bootstrap estimators are getting popularity in statistical process control due to their remarkable properties for non-normal distribution. In this paper the bootstrap control chart is developed for monitoring process variability and robustness is discussed through simulation studies. It appears that the proposed control chart for monitoring process variability based on the bootstrap method is performing better to detect out-of-control signal in a case when data follow skewed distributions. Therefore, the proposed chart is more recommendable for industrial practitioners.

Keywords: Bootstrap; Control Chart; MAD; Non-Normal; Robust

Mathematics Subject Classification: 62F35, 62F40
1. Introduction

In statistical process control (SPC), the control chart is mostly used to detect the assignable causes of variation. The control chart for variation ($S$-chart), introduced by Shewhart, is widely accepted as a standard tool for monitoring univariate, independent and "nearly" normal processes (cf., [1]) but this is not well developed beyond these types of data (cf., [2]). The chart based on $S$ estimator is generally used to control the variability of the process. The traditional Shewhart-$S$ chart is constructed on the assumption that the process generator follows normal distribution but under non-normal processes, this chart may poorly perform (cf., [3]). In SPC, the scale estimators play a vital role to monitor the process variability. Although true process standard deviation $\sigma$ can be estimated through sample standard deviation $S$ but unfortunately $S$ is considered non-robust due to its slight departure from normality (cf., [4]). Hence for non-normal processes, the robust scale estimators are may perform well.

According to Mosteller and Tukey [5], a robust estimator has two properties; first is the resistance which means that the estimator does not cause a large change by change in the size of sample data, and second is the efficiency which means that it should be efficient in a variety of situations and free from distributional assumptions. Many scale estimators are available in the literature, one of them is Median Absolute Deviation (MAD) which was first introduced by Hampel [6]. The $MAD$ estimator is popular due to its 50% breakdown point which shows its resistance against outliers. Hence many researchers recommended $MAD$ as a robust scale estimator for normal and non-normal processes (cf. [7-8]). Moreover, a class of robust scale estimators is also proposed by Rousseeuw and Croux [9]. Most of them are based on the median.
In the recent years, another approach named as the bootstrap method is getting popular in statistical process control specifically when the underlying distribution of the process is unknown. The method was first developed by Efron [10]. The importance of bootstrap mean estimator for the construction of control chart is already discussed by Liu and Tang [2]. Based on the idea given by Liu and Tang [2], a control chart using a moving block bootstrap (MBB) method on dependent multivariate data was constructed by Liu et al. [11]. The bootstrap method was also applied to construct the control chart for Weibull percentiles by Nichols and Padgett [12]. After that Chatterjee and Qiu [13] developed CUSUM control charts using the bootstrap method. Wararit and Somchit [14] used a bootstrap approach for the construction of confidence intervals of difference between two process capability indices under half logistic distribution. Saeed and Kamal [15] proposed a bootstrap variability chart under the normal process. After that Wang and Hryniewicz [16] proposed a non-parametric Shewhart control chart based on fuzzy data using bootstrap method. Recently Hila et al. [17] extended the work by applying bootstrap methods in exponentially weighted moving average chart. The bootstrap methods are not only popular in variable control charts but these methods are also used for the construction of attribute control limits such as Zhao and Driscoll [18] used the bootstrap method for constructing control limits of c-chart. On the basis of average run lengths and false alarm rates, the bootstrap adjusted control limits showed better performance. Kashif et al. [19] proposed bootstrap confidence intervals of modified process capability index under lifetime distributions. The distribution free charts also getting popularity in the recent years. Marchant et al. [20] proposed robust multivariate control charts under generalized Birnbaum–Saunders distributions. In the next year, a comparative study was made by Ikpotokin and Siloko [21] for multivariate EWMA charts on the basis of bootstrap methods for the early detection of shifts. In the same year, Mutlu and Alakent [22] used reweighted robust standard deviation estimators to modify Shewhart $S$-
chart. The $MAD$, as robust scale estimator was also suggested by Koukouvinos and Lappa [23]. Based on the simulations, the performance of $MAD$ estimator was considered better as compared to standard deviation. The performance of confidence intervals under different sampling schemes was addressed by Mahdizadeh and Zamanzade [24-25]. Ajadi et al. [26] presented a review of dispersion control charts. Ugaz et al. [27] studied the adoptive EWMA chart for variance.

In recent years, the bootstrap control charts are considered best using Bayes estimator in high quality processes for monitoring the fraction nonconforming (cf., [28]). The performance of different types of robust estimators is also addressed by Moheghi et al. [29], Dizicheh et al. [30], Ahmed et al. [31], Raza et al. [32], Ugaz et al. [27] and Abu-Shawiesh et al. [33]. Under remarkable properties of bootstrap methods, the bootstrap-$S$ chart is proposed for non-normal processes and its robustness is also being discussed in this research.

2. $S$ and $MAD$ charts for variability

2.1 The Shewhart-$S$ control limits

It is customary to know that the sample standard deviation ($S$) is a biased estimator of the population standard deviation and under the normal process, the $S$ estimator is considered as an unbiased estimator of $c_4\sigma$, where $c_4 \approx \frac{4(n-1)}{4n-3}$ is a bias adjusting constant and its value depends upon subgroup size $n$. Furthermore the standard deviation of $S$ estimator is

$$\sigma\sqrt{1-c_4^2}$$

The three-sigma control limits for Shewhart-$S$ chart with $\sigma$ known are:

$$UCL = c_4\sigma + 3\sigma\sqrt{1-c_4^2}$$
\[ CL = c_4 \sigma \]

\[ LCL = c_4 \sigma - 3\sigma \sqrt{1-c_4^2} \]

where Upper Control Limit (UCL) and Lower Control Limit (LCL) are the parameters of the control chart while Control Line (CL) is the central line of the control chart. It is customary to define the two constants

\[ B_5 = c_4 - 3\sqrt{1-c_4^2} \]

\[ B_6 = c_4 + 3\sqrt{1-c_4^2} \]

Consequently, the control limits become

\[ UCL = B_6 \sigma \quad (1) \]

\[ CL = c_4 \sigma \quad (2) \]

\[ LCL = B_5 \sigma \quad (3) \]

In situation when \( \sigma \) is not known, it must be estimated. Suppose that \( m \) preliminary samples are available, each of size \( n \) and let \( S_i \) be the standard deviation of the \( i \)th sample. The average of \( m \) standard deviations is \( \overline{S} \). Since \( \frac{\overline{S}}{c_4} \) is an unbiased estimator of process standard deviation \( \sigma \), therefore the control limits for Shewhart-\( S \) chart would be:

\[ UCL = \overline{S} + 3\frac{\overline{S}}{c_4} \sqrt{1-c_4^2} \]

\[ CL = \overline{S} \]
We usually define the constants

\[ B_3 = 1 - \frac{3}{c_4} \sqrt{1 - c_4^2} \]

\[ B_4 = 1 + \frac{3}{c_4} \sqrt{1 - c_4^2} \]

Finally the control parameters may be written as

\[ UCL = B_4 \bar{S} \] \hspace{1cm} (4)

\[ CL = \bar{S} \] \hspace{1cm} (5)

\[ LCL = B_3 \bar{S} \] \hspace{1cm} (6)

For Shewhart charts, further details are available in Montgomery [34].

**2.2 The control limits using MAD estimator**

The estimator based on median absolute deviation (MAD) taken from medians was considered one of the robust scale estimator due to its simple formula, bounded influence function and 50% break down point (cf., [9]). The *MAD* estimator is defined as:

\[ MAD = 1.4826 MD \left\{ \left| X_i - MD \right| \right\}, \hspace{1cm} i = 1, 2, \ldots, n \]

where *MD* is the sample median.

The transformed control limits for *S* chart based on *MAD* estimator developed by Abu-Shawiesh [7] are:
\[ UCL = c_n b_n \overline{MAD} + 3b_n \overline{MAD} \sqrt{1 - c^2_4} \]  
\[ LCL = c_n b_n \overline{MAD} - 3b_n \overline{MAD} \sqrt{1 - c^2_4} \]

where \( b_n \) is the correction factor. The values of \( b_n \) under different subgroup sizes are calculated by Abu-Shawiesh [7].

3. Proposed bootstrap-S chart

The performance of traditional control charts depends on the distribution of process data. In the construction of all traditional charts, the normal distribution is assumed. Therefore their robustness to this assumption has long been an issue in SPC (cf., [34]). Since non-normality of the process can adversely affect the performance of the control chart, some authors have suggested bootstrap methods in the construction of control limits which are completely non-parametric and free from distributional assumptions.

3.1 Control limits using bootstrap-S estimator

Using Jacknife approach by Liu and Tang [2], the control limits based on IID observations for \( B \) bootstrap samples can be constructed using histogram of variate \( \sqrt{n}(S_n - \bar{S}_m) \) where \( n \) represents subgroup size, \( S_n \) is subgroup standard deviation and \( \bar{S}_m \) is the average standard deviation over \( m \) samples.

Hence the control limits for bootstrap-S chart are:

\[ LCL = \bar{S}_m + \frac{q_{a/2}}{\sqrt{n}} \]  
\[ UCL = \bar{S}_m + \frac{q_{1-a/2}}{\sqrt{n}} \]
where $q_{a/2}$ and $q_{1-a/2}$ are used as estimated $(\frac{\alpha}{2})^{th}$ and $(1-\frac{\alpha}{2})^{th}$ quantiles of variate
\[\sqrt{n}(S_n - \bar{S}_m)\] respectively.

4. **Simulation study**

The simulation study is carried out for the construction of Shewhart-$S$, $MAD$ and bootstrap-$S$ control limits. The random numbers for thirty samples each of size 5 and 10 are simulated from three non-normal distributions such as Exponential, Cauchy and Logistic distributions. The exponential and logistic are lifetime distributions which are commonly used in quality and life testing problems while Cauchy is heavy-tailed skewed distribution. The control limits of Shewhart-$S$, $MAD$ and bootstrap-$S$ charts are calculated for each distribution. For the construction of bootstrap limits, one thousand bootstrap samples are considered and the histogram of the variate $\sqrt{n}(S_n - \bar{S}_m)$ is constructed. The out-of-control points for above mentioned three distributions are also calculated so that the comparison could be made. The specific algorithm is as follows:

**Step 1** The random numbers from Exponential (2) distribution are generated for thirty samples ($m = 30$) each with the subgroup size $n = 5$.

**Step 2** The control limits of Shewhart-$S$ (using eqs. 4-6) and $MAD$ (using eqs. 7-8) are constructed.

**Step 3** For each of one thousand bootstrap samples ($B = 1000$), the differences of each sample wise standard deviation and the overall standard deviation is calculated. After multiplying each difference with $\sqrt{n}$, the histogram is constructed.

**Step 4** The bootstrap-$S$ control limits (using eqs. 9-10) are calculated.

**Step 5** The number of out-of-control points is also calculated for each of the three charts.
Step 6 The overall process is repeated for subgroup size \( n = 10 \).

Step 7 Steps 1–6 are repeated for other non-normal distributions such as Cauchy(0,1) and Logistic(0,1) distributions.

5. Results discussion

The findings of Table 1 show the control limits and interval widths (IW) for Shewhart-\( S \), \( MAD \) and bootstrap-\( S \) control charts using subgroup size 5 and under Exponential(2), Cauchy(0,1) and Logistic(0,1) distributions. The results show that the bootstrap control chart has a shorter interval width as compared to Shewhart-\( S \) and robust \( MAD \) charts for exponential and logistic processes. Hence it can be concluded that bootstrap control limits are more robust for detecting an out-of-control signal for both non-normal processes. For heavy-tailed distribution such as Cauchy distribution, the performance of \( MAD \) chart is better than the traditional Shewhart-\( S \) and proposed bootstrap-\( S \) charts in terms of having a tighter interval width.

The supporting evidence is shown in Figure 1 through which the comparison of control charts can be observed. The different panels of Figure 1 represent Shewhart-\( S \), \( MAD \) and bootstrap-\( S \) control charts with the subgroup size 5 and 10 using Exponential, Cauchy and Logistic distributions. By comparing these figures, it can clearly be observed that the control limits for bootstrap-\( S \) chart are more resistant and hence they can detect out-of-control signals more quickly.

Moreover a large number of out-of-control points show that irrespective of the wider interval width than \( MAD \) chart for heavy tailed distribution, the bootstrap-\( S \) chart has always precise control limits.
Although the out-of-control points may be used as one of the indicators of non-normality of population distribution but the clear evidence can be observed by constructing the histogram based on $\sqrt{n}$ times the differences of bootstrapped subgroup standard deviations $S_n$ by overall sample standard deviation $\bar{S}_m$.

In case of unknown population distribution, the bootstrap histogram can be helpful to estimate it. Hence the bootstrap histogram constructed in Figure 2 strongly resembles the non-normal distribution of the process. Although the simulated samples are extracted from non-normal processes but if it is supposed that the distribution is unknown as justified in most practical data sets, the bootstrap histogram can be evident for its abnormality.

### 6. Comparative study

This section comprises the performance comparison of proposed chart with traditional Shewhart-$S$ and robust MAD charts developed by Abu-Shawiesh [7]. The performance of these charts is evaluated on the basis of simulated out-of-control points under non-normal processes. It is observed that for heavy tailed distributions such as Cauchy distribution, although the MAD chart showed tighter interval width (Table 1) but the proposed bootstrap-$S$ chart has relatively precise limits which can also be verified by calculating out-of-control points (Table 2).

On the basis of a large number of out-of-control points (Table 2), it is concluded that if underlying process is not normal, the bootstrap-$S$ control limits are more resistant to show the out-of-control position as compared to traditional Shewhart-$S$ and robust MAD charts. Our findings also agree with the study of Liu and Tang [2] in which bootstrap control limits showed better performance as compared to the standard charts under non-normal processes.
In recent literature, the variability control limits based on bootstrap estimator are found to be robust under normal process by Saeed and Kamal [15].

7. An industrial application
The data related to melting index of an extrusion grade polyethylene compound, used by Elamir [35] and Saeed and Kamal [36], are provided in Table 3. The seven days data set is measured on twenty consecutive shifts with the subgroup size $n = 4$.

The control limits of Shewhart-$S$ and $MAD$ charts are calculated for twenty samples each with the subgroup size $n = 4$. Over one thousand bootstrap samples ($B = 1000$), the bootstrap-$S$ control limits are constructed.

Table 4 signifies the control limits, interval widths and out-of-control points using Shewhart-$S$, $MAD$ and bootstrap-$S$ charts respectively for monitoring process variability. The shortest interval width of bootstrap-$S$ chart indicating its robustness to detect out-of-control points more quickly which also resulting from a large number of out-of-control points. The similar results are shown through the control charts in Figure 3 (panel i-ii).

Moreover the histogram of $\sqrt{n}(S_n - S_m)$ differences indicates the non-normality of population distribution (panel-iii). Hence it is observed that the traditional Shewhart chart for variability based on the assumption of normality does not provide a clear indication if the process shows a departure from normality. Similarly, the chart based on $MAD$ estimator shows no out-of-control points while bootstrap-$S$ control chart indicates more out-of-control points when process distribution is non-normal (Table 4).

8. Conclusion
The bootstrap methods are popular due to their good theoretical properties. In statistical process control, traditional Shewhart-$S$ chart is useful only if data follow the normal
distribution. In real life, mostly data sets do not follow the normal distribution. Due to the reason, the bootstrap- $S$ control chart based percentile bootstrap method is proposed. Under the good properties of bootstrap methods, the proposed bootstrap- $S$ chart has performed well under non-normal distribution i.e. exponential, Cauchy and logistic distributions. Based on the similar approach as proposed by Liu and Tang [2], the proposed bootstrap- $S$ chart showed better performance than the traditional Shewhart- $S$ and $MAD$ charts due to the detection of a large number of out-of-control points under non-normal processes. Moreover, the bootstrap methods are also helpful to find the population distribution using the sampling distribution of bootstrap statistic. Since in real life, the distribution of the process data does not necessarily follow the normal distribution, our proposed control chart may be recommended as it does not require any distributional assumption. Furthermore, the study can be extended on fair grounds for other types of bootstrap methods, i.e., parametric or moving block bootstrap methods.

**Acknowledgement:** The authors are thankful to the editor and reviewers for their valuable comments to improve the quality of this manuscript.

**Conflict of interest:** The authors declare no conflict of interests regarding the publication of this paper.

**References**


3. Bonett, D.G. “Confidence interval for a coefficient of quartile variation”, 


https://doi.org/10.1007/s00362-020-01197-x


Table 1. Control limits and interval widths (in brackets) under non-normal distributions

<table>
<thead>
<tr>
<th>Control Charts</th>
<th>Exponential(2)</th>
<th>Cauchy(0,1)</th>
<th>Logistic(0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 5</td>
<td>n = 10</td>
<td>n = 5</td>
</tr>
<tr>
<td>S</td>
<td>0–0.992</td>
<td>0.121–0.733</td>
<td>0–15.426</td>
</tr>
<tr>
<td></td>
<td>(0.992)</td>
<td>(0.612)</td>
<td>(15.426)</td>
</tr>
<tr>
<td>MAD</td>
<td>0–0.695</td>
<td>0.085–0.516</td>
<td>0–4.618</td>
</tr>
<tr>
<td></td>
<td>(0.695)</td>
<td>(0.431)</td>
<td>(4.618)</td>
</tr>
<tr>
<td>Bootstrap-S</td>
<td>0.338–0.596</td>
<td>0.340–0.546</td>
<td>3.410–12.881</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
<td>(0.206)</td>
<td>(9.471)</td>
</tr>
</tbody>
</table>

Figure 1. Shewhart-S, MAD and bootstrap-S control charts
Figure 2. The histogram of $\sqrt{n}(S_n - \overline{S}_m)$ differences of bootstrap samples

(i) $\text{Exp}(2), n=5$

(ii) $\text{Cauchy}(0,1), n=5$

(iii) $\text{Logistic}(0,1), n=5$

(iv) $\text{Exp}(2), n=10$

(v) $\text{Cauchy}(0,1), n=10$

(vi) $\text{Logistic}(0,1), n=10$

<table>
<thead>
<tr>
<th>Table 2. Number of out-of-control points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Charts</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$S$ (Shewhart 1931)</td>
</tr>
<tr>
<td>$MAD$ (Shawiesh 2008)</td>
</tr>
<tr>
<td>Proposed (Bootstrap-$S$)</td>
</tr>
</tbody>
</table>

Table 3. The data set of melt index measurements used by Elamir [35]

<table>
<thead>
<tr>
<th>Day</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shift</td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>I</td>
</tr>
<tr>
<td>218</td>
<td>228</td>
<td>280</td>
<td>210</td>
<td>243</td>
</tr>
<tr>
<td>224</td>
<td>236</td>
<td>228</td>
<td>249</td>
<td>240</td>
</tr>
<tr>
<td>220</td>
<td>247</td>
<td>228</td>
<td>241</td>
<td>230</td>
</tr>
</tbody>
</table>
Table 4. Control limits and out-of-control points of melt index data set

<table>
<thead>
<tr>
<th>Control Charts</th>
<th>UCL</th>
<th>CL</th>
<th>LCL</th>
<th>IW</th>
<th>Out-of-Control Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shewhart-S</td>
<td>18.939</td>
<td>8.358</td>
<td>0</td>
<td>18.939</td>
<td>1</td>
</tr>
<tr>
<td>MAD</td>
<td>14.874</td>
<td>6.564</td>
<td>0</td>
<td>14.874</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. The control charts and bootstrap histogram of melt index data set

(i) Shewhart-S and Bootstrap-S charts

(ii) MAD chart

(iii) Histogram

Biographies

Nadia Saeed is an Assistant Professor at College of Statistical and Actuarial Sciences, University of the Punjab, Lahore. Dr. Nadia Saeed did her PhD from University of the Punjab, Lahore-Pakistan. Nadia's publications have appeared in national and international...
journals. The areas of her research interest include Statistical Quality Control, Robust Methods and Statistical Inference.

**Shahid Kamal** is a Professor. He served University of the Punjab since 1985 and currently performing the duties of Vice Chancellor, GC University Faisalabad. Dr. Shahid Kamal did his PhD from Exeter University, Exeter, UK and awarded research scholarship. He has tremendously contributed for the uplift of quality education with current era need. His current research interests include statistical process control, Regression and robust methods.

**Muhammad Aslam** did PhD (statistics) from NCBA & E, Pakistan. Prof. Muhammad Aslam is the founder of Neutrosophic Inferential Statistics (NIS), Neutrosophic Circular Statistics (NCS), Neutrosophic Applied Statistics (NAS), and Neutrosophic Statistical Quality Control (NSQC). He is the author of three books. He is listed in, top 2% of scientists of the world in the list released by Standford University, USA, and at rank 35/93 among the King Abdulaziz University scientists.