

# Bootstrap-T Technique for Minimax Multivariate Control Chart

Johnson A. Adewara<sup>1</sup>, Kayode S. Adekeye<sup>2</sup>

<sup>1</sup>Distance Learning Institute, University of Lagos, Lagos, Nigeria

<sup>2</sup>Department of Mathematical Sciences, Redeemer's University, Redemption City, Nigeria

Email: samadek\_2017@yahoo.co.uk

Received August 8, 2012; revised September 10, 2012; accepted September 22, 2012

## ABSTRACT

Bootstrap methods are considered in the application of statistical process control because they can deal with unknown distributions and are easy to calculate using a personal computer. In this study we propose the use of bootstrap-t multivariate control technique on the minimax control chart. The technique takes care of correlated variables as well as the requirement of the distributional assumptions needed for the operation of the minimax control chart. The bootstrap-t

technique provides the mean  $\hat{\theta}_B$  of all the bootstrap estimators  $\hat{\theta}_B = \frac{\sum_{i=1}^B \hat{\theta}_i}{B}$  where  $\hat{\theta}_i$  is the estimate using the  $i^{\text{th}}$

bootstrap sample and B is the number of bootstraps. The computation of the proposed bootstrap-t minimax statistic was performed on the values obtained from the bootstrap estimation. This method was used to determine the position of the four control limits of the minimax control chart. The bootstrap-t approach introduced to minimax multivariate control chart helps to detect shifts in the mean vector of a multivariate process and it overcomes the computational complexity of obtaining the distribution of multivariate data.

**Keywords:** Bootstrap; Minimax; Multivariate Data; Control Limits; Process Control

## 1. Introduction

The high rate of interest in multivariate control charts is now on increase after a volume of research in all areas of univariate control charts. Multivariate control charts have the advantage of being able to monitor multiple quality characteristics simultaneously for both changes in the mean vector and the correlation structure while maintaining a specified probability of Type I error ( $\alpha$ ). Multivariate control charts are at a disadvantage because it is difficult to identify which subset of the quality characteristics are responsible for a signal since any single characteristic or combination of characteristics could have experienced a shift in mean value, variance, or correlation.

Many industrial applications today have several correlated quality characteristics that needed to be monitored simultaneously. Each item in the sample is inspected and the values of each of its quality characteristics are determined. The information from the sample, contained in the sample statistics, is used to determine whether the process appears to be in control with reasonable variation. If the mean vector of one or more of the variables appears to have deviated from its desired value, or if the variation in one or more of the variables seems to have

increased, the process is considered out of control [1].

The development of multivariate control charts originates from the work by [2]. [3] provided a general framework on constructing control charts for both univariate and multivariate situations. [4] proposed a bootstrap control chart that can monitor both independent and dependent observations. [5] discussed the performance of techniques for constructing bootstrap control charts, [6] proposed a bootstrap control chart based on the Birnbaum-Saunders distribution, and [7] proposed median control charts whose control limits were determined by estimating the variance of the sample median via the bootstrap technique. Recent work by [8] proposed a bootstrap-based multivariate  $T^2$  control chart that can efficiently monitor a process when the distribution of observed data is non-normal or unknown. The concept of minimax control chart for multivariate quality control was introduced by [9]. The minimax control chart uses the joint probability distribution of the maximum and minimum standardized sample means to obtain the control limits for monitoring purposes. This chart operates under the assumptions of normality and uncorrelated variable. [10] uses a non-linear polynomial function approach to solve the problem of normality assumption needed to obtain the

control limits for the minimax control chart. In this paper, we introduce the concept of bootstrapping technique to address the issues of uncorrelated variables as well as the distributional assumptions of minimax control charts. Furthermore, the control limits for minimax control chart were determined for varying level of significance to enable process control operators to make a robust choice of control limits that will suite their purpose.

## 2. Method

The minimax control chart developed by [9] is similar to the charts proposed by [11] and [12]. The minimax control chart also uses the minimum and maximum standardized sample means to make the decision if the process should be considered in control or out of control. However, the minimax chart uses both lower and upper control limits on both the maximum and minimum standardized sample means. This is facilitated by the development of the capability to determine the value of the joint density function of the maximum and minimum standardized sample means. This not only facilitates a method for setting the control limits, but also allows for the comparison of the performance of the minimax chart relative to other charts through computation of the out-of-control average run length.

The minimax chart operates by controlling the minimum and maximum standardized sample means. The correlation between the quality characteristic variables, which is assumed to be known, is incorporated into the construction of the control limits for the minimum and maximum standardized sample mean. In addition, the chart operates at a relatively low and fixed probability of Type I error ( $\alpha$ ). There are several advantages of using the minimax control chart. Although the construction of the chart is fairly complex, the application of the chart is extremely straight-forward. The computations involve the standardization of each of the mean quality characteristic values. The maximum and minimum of these values are then plotted, each against its upper and lower control limits.

The minimax chart monitors  $p$  quality characteristics simultaneously. At fixed time intervals, a sample of size  $n$  is taken and the value of each of the quality characteristics is determined. Let  $X_{ij}$  represent the value of quality characteristic  $i$  of item  $j$  for  $i = \{1, 2, \dots, p\}$  and  $j = \{1, 2, \dots, n\}$ , and let  $\bar{X} = \{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p\}^T$  contains the sample mean value for each quality characteristic. Although the observations may be correlated within each sample, the samples are assumed independent of each other. In addition, it is assumed that  $\bar{X}$  is distributed with a multivariate normal distribution with known process mean vector  $\mu = \{\mu_1, \mu_2, \dots, \mu_p\}^T$  and covariance matrix  $\Sigma = \sigma_{ij}^2$ , where  $\sigma_{ii}^2$  is the variance of the  $i^{th}$

quality characteristic and  $\sigma_{ii}^2$  is the covariance between the  $i^{th}$  quality characteristic and the  $j^{th}$  quality characteristic. The covariance structure is assumed to remain stable over time leaving only a shift in the mean vector to cause an out-of-control situation.

The minimax chart requires the minimum and maximum standardized sample means to fall between their respective control limits. In order to obtain the minimum standardized sample mean  $Z[1]$  and the maximum standardized sample mean  $Z[p]$ , each of the elements of the sample mean vector,  $\bar{X}$  is standardize using

$$Z_i = \frac{\sqrt{n}(\bar{X}_i - \mu_i)}{\sigma_{ii}} \quad (1)$$

Note that  $Z_i$  is simply the standardized sample mean for the  $i$ th quality measure and  $Z_{[i]}$  is an order statistic defining the  $i$ th smallest value of the  $p$  standardized sample means. Therefore  $Z_{[1]}$  and  $Z_{[p]}$  are determined from  $Z = \{Z_1, Z_2, \dots, Z_p\}$ . There are four control limits for the minimax chart. These are  $UCL_{[1]}$  and  $LCL_{[1]}$  *i.e.* the respective upper and lower control limits for  $Z_{[1]}$  and  $UCL_{[p]}$  and  $LCL_{[p]}$  are the respective upper and lower control limits for  $Z_{[p]}$ . It should be noted that the minimax chart is essentially two dependent charts. One chart monitors and controls  $Z_{[1]}$  and the other chart monitors and controls  $Z_{[p]}$ . The charts are dependent because the joint distribution of  $Z_{[1]}$  and  $Z_{[p]}$  is used to construct the control limits. The probability of a Type I error,  $\alpha$ , is the sum of the probabilities of a Type I error in the  $Z_{[1]}$  chart,  $\alpha_{[1]}$ , and a Type I error in the  $Z_{[p]}$  chart,  $\alpha_{[p]}$ .

In this paper, the control limits proposed by [9] for the minimax control chart were adopted. These control limits are given in Equation (2)

$$\begin{aligned} UCL_{[p]} &= \left\{ \tau : \int_{-\infty}^{\tau} \dots \int_{-\infty}^{\tau} f(Z) dz = 1 - \alpha \right\} \\ LCL_{[p]} &= \left\{ \tau : \int_{-\infty}^{\tau} \dots \int_{-\infty}^{\tau} f(Z) dz = \alpha \right\} \\ UCL_{[1]} &= \left\{ \tau : 1 - \int_{\tau}^{\infty} \dots \int_{\tau}^{\infty} f(Z) dz = 1 - \alpha \right\} \\ LCL_{[1]} &= \left\{ \tau : 1 - \int_{\tau}^{\infty} \dots \int_{\tau}^{\infty} f(Z) dz = \alpha \right\} \end{aligned} \quad (2)$$

These control limits are function of the probability of Type I error ( $\alpha$ ),  $\tau$  and the distribution of the transformed variable. It is worthy to note that the control limits depends heavily on the value of  $\alpha$  chosen. Therefore, we consider varying the value of  $\alpha$  to obtain the four control limits of the minimax control chart.

### Bootstrap-T Method

The bootstrap method, introduced by [13], is a powerful tool for estimating the sampling distribution of a given statistic. The bootstrap estimate of the sampling distribution is generally better than the normal approximation

based on the central limit theorem (cf. [14] and [15]), even if the statistic is not standardized (cf. [16] and [17]). Let  $\{X_1, \dots, X_N\}$  be an iid sample following the distribution  $F$  with mean,  $\mu$  and variance  $\sigma^2$ . The standard bootstrap procedure is to draw with replacement a random sample of size  $N$  from  $\{X_1, \dots, X_N\}$ . Denote the bootstrap sample by  $\{X_1^*, \dots, X_N^*\}$  and denote their mean and standard deviation by  $\bar{X}_N^*$  and  $S_N^*$ . Suppose  $F_N$  indicate the empirical distribution of  $\{X_1, \dots, X_N\}$ , then the sampling distribution of  $(\bar{X}_N^* - \bar{X}_N)$  under  $F_N$  is the bootstrap approximation of the sampling distribution of  $(\bar{X}_N - \mu)$  under  $F$ . Its approximation error is shown to be negligible by the proposition derived by [14] and [15]). The bootstrap technique provides the mean  $\hat{\theta}_B$  of all the bootstrap estimators

$$\hat{\theta}_B = \frac{\sum_{i=1}^B \hat{\theta}_i}{B} \tag{3}$$

where  $\hat{\theta}_i$  is the estimate using the  $i^{th}$  bootstrap sample and  $B$  is the number of bootstraps.

The bootstrap handles cases where a standard distribution cannot be assumed. The control limit is estimated by resampling the observed data to estimate the distribution of the observed variable. For many problems in statistics, we are interested in the distribution of values from a random sample of the population. If the underlying distribution from which the values are drawn is known, we can use developed theory to generate the sampling distribution. [13] proposed the use of bootstrapping when there is little or no significant information about the underlying distribution. The idea behind the bootstrap is very simple, namely that (in the absence of any other information), the sample itself offers the best guide of the sampling distribution. By resampling with replacement from the original sample, we can create a bootstrap sample, and use the empirical distribution of our estimator in a large number of such bootstrapped samples to construct confidence intervals and tests for significance. In this paper, the bootstrap-t method is employed.

The bootstrap-t method procedure directly estimates the standard normal distribution ( $Z$ ) from the data, much like the use of the normal and  $t$  tables. We generate  $B$  bootstrap samples  $x^{*1}, x^{*2}, \dots, x^{*B}$  and for each we compute

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\hat{S}_e^*(b)} \tag{4}$$

where  $\hat{\theta}^*(b)$  is the value of  $\hat{\theta}$  for the bootstrap sample  $x^{*b}$ , and  $\hat{S}_e^*(b)$  is the estimated standard error of  $\hat{\theta}^*(b)$  for the bootstrap sample  $x^{*b}$ . The  $\alpha^{th}$  percentile of  $Z^*(b)$  is estimated by the value  $\hat{i}(\alpha)$  such that

$$\#\{Z^*(b) \leq \hat{i}(\alpha)\} / B = \alpha \tag{5}$$

where  $\#\{\cdot\}$  is the  $B \cdot \alpha^{th}$  value in the ordered list of the smallest value of the  $Z^*(b)$ s in interval  $\{\cdot\}$ . For instance, if  $B = 1000$ , the estimate of the 90% point is the 900<sup>th</sup> smallest value of the  $Z^*(b)$ s. The ‘‘bootstrap-t’’ confidence interval is

$$\left[ \hat{\theta}^* - \hat{i}(1-\alpha) \hat{S}_e^*(b), \hat{\theta}^* - \hat{i}(\alpha) \hat{S}_e^*(b) \right]$$

In this study we adopt the bootstrap-t method given in Equation (4) to generate the data set that we used to obtain the control limits for the minimax control chart developed by [9]. The vector  $Z^*(b) = [Z_i]$ ,  $i = 1, 2, \dots, p$  is now defined as the standardized bootstrap sample mean vector. The maximum sample mean  $(Z^*(b)_p)$  is defined as the maximum of the elements of the vector  $Z^*(b)$  that is,  $Z^*(b)_{[p]} = \max(Z^*(b)_i)$ . Also, the minimum standardized sample mean  $(Z^*(b)_{[1]})$  is defined as the minimum of the elements of the bootstrap vector  $Z^*(b)$  as given by  $Z^*(b)_{[1]} = \min(Z^*(b)_i)$ . These minimum and maximum statistics are then implemented on the control limits in Equation (2).

### 3. Results

The data used in this study for illustrative purpose were collected from the production line of a soft drinks manufacturing company in Nigeria. We considered five important quality characteristics of the soft drink. These are:  $X_1 =$  Contents in ml,  $X_2 =$  Brix,  $X_3 =$  pressure,  $X_4 =$  CO<sub>2</sub>, and  $X_5 =$  Temperature. The data was first tested for autocorrelation and it was discovered that the data are uncorrelated. Thereafter, the data was bootstrapped using the approach described in Section 2. The result of the bootstrap analysis is presented in **Table 1**.

To obtain the control limits in Equation (2), an algorithm was developed and implemented on C language using different levels of significance ( $\alpha$ ) from 1% to 50%. The Control limits for five variables case considered in this study were determined for varying level of significance from 0.01 to 0.5. The obtained control limits are presented in **Table 2**

### 4. Discussion of Results

The results in **Table 1** represent the summary of the descriptive statistics of the bootstrap data and the original data from the production line, where  $\bar{X}^*(b)$ ,  $SD^*(b)$ , and  $Z^*(b)$  are the mean, standard deviation, and the standardized values of the bootstrapped data. The bootstrap statistics shows the same feature with the real data. The Control limits in **Table 2** are obtained for five vari-

**Table 1. Summary of result for the bootstrap data for five variables.**

	Variable				
	1	2	3	4	5
$\bar{X}^*(b)$	349.090	10.199	44.075	69.657	3.730
$SD^*(b)$	0.5128	0.0086	0.3132	0.3767	0.0080
$Z^*(b)$	-0.0757	-0.1287	0.0479	0.0313	-0.0741
$\bar{x}$	348.8153	10.1875	44.225	69.775	3.7238

**Table 2. The minimax control limits for varying level of significance ( $\alpha$ ).**

Level of significance ( $\alpha$ )	Maximum			Minimum		
	UCL <sub>[5]</sub>	LCL <sub>[1]</sub>	Limit Interval	UCL <sub>[5]</sub>	LCL <sub>[1]</sub>	Limit Interval
0.01	2.4423	-0.8624	3.305	3.0422	2.7027	0.340
0.02	2.4364	-1.033	3.469	3.0393	2.7069	0.332
0.03	2.4305	-1.1489	3.579	3.0364	2.7111	0.325
0.04	2.4245	-1.2394	3.664	3.0335	2.7153	0.318
0.05	2.4185	-1.3148	3.733	3.0306	2.7195	0.311
0.06	2.4124	-1.38	3.792	3.0276	2.7236	0.304
0.07	2.4063	-1.4379	3.844	3.0247	2.7277	0.297
0.08	2.4001	-1.4901	3.890	3.0218	2.7318	0.290
0.09	2.3939	-1.5378	3.932	3.0188	2.7359	0.283
0.10	2.3876	-1.5819	3.970	3.0158	2.74	0.276
0.11	2.3813	-1.623	4.004	3.0128	2.744	0.269
0.12	2.3749	-1.6615	4.036	3.0099	2.748	0.262
0.13	2.3684	-1.6978	4.066	3.0069	2.752	0.255
0.14	2.3619	-1.7321	4.094	3.0038	2.756	0.248
0.15	2.3554	-1.7648	4.120	3.0008	2.7600	0.241
0.16	2.3487	-1.796	4.145	2.9978	2.7639	0.234
0.17	2.3421	-1.8258	4.168	2.9948	2.7679	0.227
0.18	2.3353	-1.8544	4.190	2.9917	2.7718	0.220
0.19	2.3285	-1.8819	4.210	2.9887	2.7757	0.213
0.20	2.3216	-1.9084	4.230	2.9856	2.7795	0.206
0.21	2.3147	-1.934	4.249	2.9825	2.7834	0.199
0.22	2.3077	-1.9587	4.266	2.9794	2.7872	0.192
0.23	2.3006	-1.9827	4.283	2.9763	2.7911	0.185
0.24	2.2935	-2.0059	4.299	2.9732	2.7949	0.178
0.25	2.2863	-2.0285	4.315	2.9701	2.7987	0.171
0.26	2.2790	-2.0505	4.330	2.967	2.8024	0.165
0.27	2.2716	-2.0719	4.344	2.9638	2.8062	0.158
0.28	2.2642	-2.0927	4.357	2.9607	2.8099	0.151
0.29	2.2567	-2.113	4.370	2.9575	2.8137	0.144
0.30	2.2491	-2.1328	4.382	2.9544	2.8174	0.137
0.40	2.1681	-2.3096	4.478	2.922	2.8536	0.068
0.50	2.0759	-2.4578	4.534	2.8885	2.8585	0.03

ables scenario. It should be noted that the control limits for situation for more than five variables can be obtained by an extension of the program used in this study. From the bootstrap data, the value of the minimum  $Z^*b(1) = -0.076$  and the value of the maximum  $Z^*b(5) = 0.048$ . If the management is interested in taking decision based on 1% level of significance ( $\alpha = 0.01$ ), then the control limits from **Table 2** will be  $UCL(5) = 2.442$ ,  $LCL(5) = -0.862$ ,  $UCL(1) = 3.042$  and  $LCL(1) = 2.703$ . Therefore, the maximum value  $Z^*b(5)$  will be within the control limits whereas, the minimum value  $Z^*b(1)$  will be lying outside the lower control limit. However, if the management is interested in taking decision based on 5% level of significance ( $\alpha = 0.05$ ), then the control limits from **Table 2** will be  $UCL(5) = 2.419$ ,  $LCL(5) = -1.315$ ,  $UCL(1) = 3.031$  and  $LCL(1) = 2.72$ . This will give the same conclusion as that of when  $\alpha = 0.01$  is used. Furthermore, the results show that the control limit interval increases as the level of significance increases. For instance, the control limit interval when  $\alpha = 0.01$  are 3.304 for the maximum and 0.339 for the minimum. When  $\alpha = 0.05$ , the control limit intervals are 3.734 for the maximum and 0.311 for the minimum.

## 5. Conclusion

This work used the bootstrap techniques to set up the control limits for a minimax control chart. The control limits for minimax control chart were determined for varying level of significance to enable process control operators to make a robust choice of control limits that will suite their purpose. Furthermore, the derived control limits for a five variable scnerio showed that the smaller the control limit interval, the higher the false alarm rate and hence high Type I error probability. Therefore, the obtained value of control limits in **Table 2** will help the management to take decision on the level of significance to use for monitoring purpose. An extension or review of the used algorithm will provide the control limits for multivariate problem with less or greater than five variables.

## REFERENCES

- [1] J. Rehmert, "A Performance Analysis of the Minimax Multivariate Quality Control Chart," Master of Science Dissertation, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, 1997
- [2] H. Hotelling, "Multivariable Quality Control—Illustrated by the Air Testing of Sample Bombsight," In: C. Eisenhart, M. W. Hastay and W. A. Wallis, Eds., *Techniques of Statistical Analysis*, McGraw Hill, New York, 1947, pp. 110-122.
- [3] A. M. Polansky, "A General Framework for Constructing Control Charts," *Quality and Reliability Engineering International*, Vol. 21, No. 6, 2005, pp. 633-653.
- [4] R. Y. Liu and J. Tang, "Control Charts for Dependent and Independent Measurements Based on the Bootstrap," *Journal of the American Statistical Association*, Vol. 91, No. 436, 1996, pp. 1694-1700.
- [5] I. A. Jones and W. H. Woodall, "The Performance of Bootstrap Control Charts," *Journal of Quality Technology*, Vol. 30, No. 4, 1998, pp. 362-375.
- [6] Y. L. Lio and C. Park, "A Bootstrap Control Chart for Birnbaum-Saunders Percentiles," *Quality and Reliability Engineering International*, Vol. 24, No. 5, 2008, pp. 585-600.
- [7] H. I. Park, "Median Control Charts Based on Bootstrap Method," *Communications in Statistics—Simulation and Computation*, Vol. 38, No. 1, 2009, pp. 558-570.
- [8] P. Phaladiganon, S. B. Kim, V. C. P. Chen, J. G. Baek and S. K. Park, "Bootstrap-Based  $T^2$  Multivariate Control Charts," *Communications in Statistics—Simulation and Computation*, Vol. 40, No. 5, 2011, pp. 645-662. [doi:10.1080/03610918.2010.549989](https://doi.org/10.1080/03610918.2010.549989)
- [9] A. Sepulveda, "The Minimax Control Chart for Multivariate Quality Control," Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering, Blacksburg, 1996.
- [10] J. A. Adewara, K. S. Adekeye, O. E. Asiribo and S. B. Adejuyigbe, "Minimax Multivariate Control Chart Using a Polynomial Function," *Applied Mathematics*, Vol. 2, No. 12, 2011, pp. 1539-1545. [doi:4236/am.2011.212219](https://doi.org/10.4236/am.2011.212219)
- [11] A. J. Hayter and K. L. Tsui, "Identification and Quantification in Multivariate Quality Control Problems," *Journal of Quality Technology*, Vol. 26, No. 3, 1994, pp. 197-208.
- [12] N. H. Timm, "Multivariate Quality Control Using Finite Intersection Tests," *Journal of Quality Technology*, Vol. 28, No. 2, 1996, pp. 233-243.
- [13] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, Vol. 7, No. 1, 1979, pp. 1-26. [doi:10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552)
- [14] P. J. Bickel and D. A. Freedman, "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, Vol. 9, No. 6, 1981, pp. 1196-1217.
- [15] K. Singh, "On the Asymptotic Accuracy of Efron's Bootstrap," *The Annals of Statistics*, Vol. 9, No. 6, 1981, pp. 1187-1195.
- [16] R. J. Beran, "Estimated Sampling Distributions: The Bootstrap and Competitors," *The Annals of Statistics*, Vol. 10, No. 1, 1982, pp. 212-215.
- [17] R. Liu and K. Singh, "On a Partial Correction by the Bootstrap," *The Annals of Statistics*, Vol. 15, No. 4, 1987, pp. 1713-1718.