Quality Technology &
Quantitative Management
Vol. 6, No. 2, pp. 155-169, 2009

QTQM

A New Multivariate Non-Parametric Control Chart Based on Sign Test

Nandini Das

SQC-OR Unit, Indian Statistical Institute, India
(*Received* April 2008, *accepted* August 2008)

Abstract: Multivariate statistical process control deserves particular attention in the recent scenario. Though, Hotelling T^2 control chart is quite popular and widely used technique in this field but its performance is deteriorated when the underlying distribution of the quality characteristics is not following multivariate normal distribution. Hence the need of developing a non-parametric multivariate control chart arises which does not require any underlying distribution. In this work a multivariate non-parametric control chart based on bivariate sign test is proposed. Its performance in both in control and out of control state was evaluated by simulating data from multivariate normal and multivariate t distribution and compared with those of existing multivariate parametric control chart.

Keywords: ARL, bivariate sign test, non-parametric control chart.

1. Introduction

Multivariate statistical process control (MSPC) is particularly important in contemporary industries where data are collected on more than one variable. Most quality characteristics to be controlled and monitored are not independent. One of the main reasons for this correlation is that most manufacturing systems are composed of many subsystems, which are highly interconnected. The excess of subsystems gives rise to the difficulties in monitoring the multiple variables since it is often misleading for the operators to monitor those correlated variables individually. Under such situations, how to monitor the variables and simultaneously to integrate them is crucial to success of MSPC. The situations demanding such analysis ranges from a rather simple plastic processing to more complex engine manufacturing processes. The use of multiple univariate control charts does not deliver a useful solution in this situation. The problems are that, the overall probability of signaling a false 'out-of-control' situation is not controlled and more seriously the correlation among the variables are ignored.

Extensive research has been performed in the field of multivariate control charts since the 1940's when Hotelling first recognized that the quality of product may depend on several correlated characteristics. One type of MSPC is a multivariate chart extended from univariate SPC methods, including Hotelling's T^2 chart, multivariate EWMA and multivariate CUSUM charts (Montgomery [13]). Another type of MSPC is based on the latent variable projection, such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) (Lowry and Montgomery [11], MacGregor and Kourti [12], Raich and Cinar [17]).

The most familiar multivariate process monitoring and control procedure is the Hotelling T^2 control chart for monitoring the mean vector of the process. This control chart is based on the assumption that the multiple quality characteristics have a joint probability density function, which is multivariate normal.

In real-life problems, the actual distribution is usually unknown and is also not easy to estimate accurately, especially when the number of samples is not large enough to approximate the asymptotic distribution (Polansky [15]). In such cases, the control charts may not perform as well as expected. Schilling and Nelson [19] and many other researchers have investigated the effects of non-normality on the control limits and charting performance. To alleviate such effects, some distribution-free or non-parametric control charts have been proposed.

The main advantages of the non-parametric control charts is the flexibility derived from not needing to assume any parametric probability distribution for the underlying process, at least as far as establishing and implementing control charts are concerned. Obviously, this is very beneficial in the field of process control, particularly in start-up situation where not much data is available to use a parametric procedure. Also the nonparametric charts are likely to share the robustness properties of nonparametric tests and confidence intervals and are, therefore far more likely to be less impacted by outliers. It should be noted that nonparametric methods can be somewhat less efficient than their parametric counterparts, provided of course that one has the complete knowledge of the underlying stochastic process for which the particular parametric method is specifically designed; however, the reality is that such information is seldom, if ever, available to the quality practitioner. Moreover, in today's computer based process monitoring and control, "less efficiency" can often be compensated for by more observations. Another perceived disadvantage of nonparametric charts is that for small sample sizes one needs special tables. Again, this should not be a problem given the ubiquitous presence of computers today.

Recent literature reveals the development of substantial number of non-parametric control charts where no underlying distribution is assumed on the process output. Woodall and Montgomery [20] foresaw an increasing role for non-parametric methods in control charting application. Chakraborti *et al.* [2] gave an overview and discussed the advantages of several non-parametric control charts over their normal theory counterparts. Bakir [1] complied and classified several non-parametric control charts according to the driving non-parametric idea behind each one of them. In last two decades a number of non-parametric control charts have been reported in several literature. But most of them concentrate on univariate non-parametric control chart.

The present study attempts to summarize different multivariate non-parametric control chart available in the existing literature and proposed a new multivariate non-parametric control chart for controlling location parameters based on bivariate sign test. Its performance with respect to in-control ARL and power to detect shift in location parameter has been evaluated and compared them with that of the most widely used multivariate parametric Hotelling T^2 or χ^2 control chart.

2. Literature Survey

Very few multivariate nonparametric control charts (MNCC) exist in the literature. Hayter and Tsui [6] proposed a Shewhart type MNCC for individual to monitor process location parameter. The chart is based on the M statistic, which is the maximum of deviation of the observations from their sample means. The calculation of control limits is based on the empirical distribution of an initial reference sample. Kapatou Reynolds [8, 9] proposed an EWMA type multivariate CCs for groups based on the signed rank statistics.

They are not truly non-parametric since some elements of covariance need to be estimated. Liu [10] combined the idea of reference sample with the concept of data depth and proposed a new type of MNCC. Sun and Tsung [18] proposed a multivariate control chart based on the kernel distance, which is a measure of the distance between the 'kernel centre' and the incoming new sample to be monitored. The kernel distance can be calculated using support vector methods. This chart makes use of information extracted from in-control preliminary samples. Hamurkarouglu *et al.* [5] proposed a nonparametric control chart based on Mahalanobis depth. The chart was constructed with respect to the rank of Mahalanobis depth. Hur [7] developed a wavelet-based Control Charts for general multivariate processes.

Hotelling T^2 control chart:

This control chart is based on the assumption that the multiple quality characteristics have a joint probability density function, which is multivariate normal, which is of the following form:

$$f(\boldsymbol{X}) = (2\pi)^{-p/2} \Sigma^{-1/2} \exp\{-(1/2)(\boldsymbol{X}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{X}-\boldsymbol{\mu})\};$$

$p =$ number of variables, $\boldsymbol{\mu} =$ mean vector, $\Sigma =$ variance-covariance matrix.

Suppose that two quality characteristics x_1 and x_2 are jointly distributed according to the bivariate normal distribution. Let μ_1 and μ_2 be the mean values of the quality characteristics and let σ_1 and σ_2 be the standard deviations of x_1 and x_2 respectively. The covariance between x_1 and x_2 is σ_{12}. We assume that they are known.

If $\boldsymbol{x} = (x_1\ x_2)'$ is the sample mean vector and Σ is the variance-covariance matrix, then the statistic,

$$n(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}); \text{ where } n = \text{sample size},$$

follows Chi-square distribution with 2 degrees of freedom. This equation can be used as the basis of a control chart for the process means μ_1 and μ_2. If the process means remain at the values μ_1 and μ_2, the values of the statistic should be less than the upper control limit $\text{UCL} = \chi^2_{\alpha,\,2}$, where $\chi^2_{\alpha,\,2}$ is the upper α percentage point of the chi-square distribution with 2 degrees of freedom. If at least one of the means shift to some new value, the probability of the statistic exceeding upper control limit increases.

In practice, however, it is necessary to estimate $\boldsymbol{\mu}$ and Σ from the analysis of the preliminary samples taken when the process is assumed to be in control. Suppose, $\boldsymbol{X}$ and $\boldsymbol{S}$ are the unbiased estimates of the population mean ($\boldsymbol{\mu}$) and covariance matrix (Σ) respectively. Then the test statistic becomes

$$T^2 = n(\boldsymbol{x}-\boldsymbol{X})' \boldsymbol{S}^{-1} (\boldsymbol{x}-\boldsymbol{X}),$$

which follows Hotelling T^2 distribution. For **phase 1** (when the objective is to establish the control limits by obtaining the in-control set of observations), the upper control limit is

$$\text{UCL} = \{p(m-1)(n-1)/(mn-m-p+1)\} F_{\alpha,\,p,\,mn-m-p+1},$$

where $m =$ number of initial samples.

In **Phase 2**, when the chart is in use, the control limit is

$$\text{UCL} = \{p(m-1)(n-1)/(mn-m-p+1)\} F_{\alpha,\,p,\,mn-m-p+1}.$$

If $\boldsymbol{\mu}$ and Σ are estimated from a large number of preliminary samples (≥20), it is customary to use $\mathrm{UCL} = \chi^2_{\alpha,\,2}$ as the upper control limit in both phase 1 and phase 2.

3. Proposed Method

Theory: A bivariate sign test for location

In this section the basic theory of bivariate sign test is described. In this context see Puri and Sen [16].

Let $X_j = (X_{1j} X_{2j})^T$, $j = 1, \ldots, n$, be n independent stochastic vectors having continuous cumulative distribution functions $F_1(x), \ldots, F_n(x)$, $x \in R^2$, respectively. The problem is to test whether $F_1, \ldots, F_n$ have n specified pairs of (marginal) medians (assumed to be uniquely defined). By suitably choosing the origins, we may assume that the pair of hypothetical medians for each X_j is $0 = (0\ 0)^T$, $j = 1, \ldots, n$. Then the null hypothesis is

$$H_0 : F_j(0, \infty) = F_j(-\infty, 0) = \frac{1}{2}, \quad \text{for all} \quad j = 1, \ldots, n; \tag{1}$$

when $F_1, \ldots, F_n$ are otherwise arbitrary. For each X_j, the events $(X_{1j} \le 0,\, X_{2j} \le 0)$ and $(X_{1j} \ge 0,\, X_{2j} \ge 0)$ are called concordance of first and second kind, and the events $(X_{1j} \le 0,\, X_{2j} \ge 0)$ and $(X_{1j} \ge 0,\, X_{2j} \le 0)$ as discordance of the first and second kinds respectively, $j = 1, \ldots, n$. Also, let γ_j be the probability of concordance of $(X_{1j},\, X_{2j})$, and assume that

$$0 < \gamma_j < 1 \quad \text{for all} \quad j = 1, \ldots, n; \tag{2}$$

Finally, let us denote for X_j the conditional probability of a concordance (discordance) of the first kind given concordance (discordance) by $\theta_j(\tau_j)$ for $j = 1, \ldots, n$. Then, (1) can be written as

$$H_0 : \theta_j = \tau_j = \frac{1}{2} \quad \text{for all} \quad j = 1, \ldots, n; \tag{3}$$

The sign test to be considered below is based on the following principle. Among the n observations X_j, $j = 1, \ldots, n$, let $C_i(D_i)$ be the number of concordances (discordances) of the i^{th} kind, $i = 1, 2$. Also, let $C = C_1 + C_2$ and $D = D_1 + D_2$. Then $C + D = n$. Under (3), C_1 and C_2 (as well as D_1 and D_2) should be stochastically equal, whereas if (1) does not hold (3) can not hold, and hence, $C_1 - C_2$ or $D_1 - D_2$ will be stochastically different from zero. So, it is suggested that the test may be based on $C_1 - C_2$ and $D_1 - D_2$. However, even when (3) holds the joint distribution of (C_1, C_2, D_1, D_2) depends on F_j $j = 1, \ldots, n$ through the unknown values of $F_j(0, 0)$, $j = 1, \ldots, n$. Therefore, the following conditional probability law is used to construct a test which is conditionally distribution-free. We consider the conditional distribution of C_1, C_2, D_1, D_2 given which pairs among $(X_{1j},\, X_{2j})^T$ are concordant and which one is discordant. Let $c (0 \le c \le n)$ be any integer. We consider any partition

$$(i_1, \ldots, i_c),\ (i_{c+1}, \ldots, i_n) \quad \text{where} \quad i_1 < \ldots < i_c,\ i_{c+1} < \ldots < i_n, \tag{4}$$

of the set of numbers $1, \ldots, n$ into two disjoint subsets containing c and $n - c$ numbers respectively. (If $c = 0$ or n, then the first or second subset is of course empty.). Let $\varepsilon_{i1}, \ldots, \varepsilon_{ic}$ be the event that among $X_1, \ldots, X_n$, the $i_1^{th}, \ldots, i_c^{th}$ pairs are concordant and the rest are discordant. Then

$$P\{\varepsilon_{i1}, \ldots, \varepsilon_{ic}\} = \gamma_{i1} \ldots \gamma_{ic} (1 - \gamma_{i\,c+1}) \ldots (1 - \gamma_{in}). \tag{5}$$

The probability that $\varepsilon_{i1\ldots ic}$ will occur and there will be just c_1 concordances of the first kind and just d_1 discordances of the first kind $(0 \le c_1 \le c,\, 0 \le d_1 \le n-c)$ is

$$\gamma_{i1}\ldots \gamma_{ic}(1-\gamma_{i\,c+1})\ldots(1-\gamma_{in})\cdot\{(1-\theta_{i1})\ldots(1-\theta_{ic})$$
$$\textstyle\sum_1 \prod_1 [\theta_{ij}/(1-\theta_{ij})]\}\cdot\{(1-\tau_{i\,c+1})\ldots(1-\tau_{in})\sum_2 \prod_2 [\tau_{ij}/(1-\tau_{ij})]\}, \tag{6}$$

where $\prod_1$ denotes the product over a subset of c_1 of the values $1, 2, \ldots, c$ of j and $\sum_1$ denotes the sum over all the $(c\ c_1)^T$ such subsets; similarly, $\prod_2$ denotes the product over a subset of d_1 of the values $c+1, \ldots, n$ and so on. From (5) and (6) the required conditional distribution of C_1 and D_1 is given by

$$P\{C_1 = c_1,\, D_1 = d_1 | \varepsilon_{i1,\ldots,ic}\} = \{(1-\theta_{i1})\ldots(1-\theta_{ic})$$
$$\textstyle\sum_1 \prod_1 [\theta_{ij}/(1-\theta_{ij})]\}\cdot\{(1-\tau_{i\,c+1})\ldots(1-\tau_{in})\sum_2 \prod_2 [\tau_{ij}/(1-\tau_{ij})]\}; \tag{7}$$
$$(0 \le c_1 \le c,\, 0 \le d_1 \le n-c).$$

Under H_0 given by (3), whatever be $(F_1, \ldots, F_n)$, (7) gives

$$P\{C_1 = c_1,\, D_1 = d_1 | \varepsilon_{i1,\ldots,ic},\, H_0\} = P\{C_1 = c_1,\, D_1 = d_1 | C = c,\, H_0\} = {}^cC_{c1}\,{}^{n-c}C_{d1}2^{-n};$$
$$(0 \le c_1 \le c,\, 0 \le d_1 \le n-c). \tag{8}$$

Thus, under H_0, given $C = c$, C_1 and D_1 are independently distributed as binomial random variables with parameters $(c, 1/2)$ and $(n-c, 1/2)$ respectively. Hence for testing H_0, it seems reasonable to use the statistic

$$T = (4/C)(C_1 - C/2)^2 + [4/(n-C)][D_1 - (n-C)/2]^2. \tag{9}$$

For $C = 0$ or n, one of the terms in T is absent. Given $C = c$, the conditional distribution of T under H_0 would be clearly distribution-free. Let $F_n(t|c)$ denote the c. d. f of this distribution as obtained by summing (8) over those combinations (c_1, d_1) for which the value of T does not exceed t. For any $\alpha (0 < \alpha < 1)$ let $t_{n,\alpha}(c)$ be the value of t for which $1 - F_n(t|c) \le \alpha < 1 - F_n(t-0|c)$. When we define a critical function $\Phi(t, c)$ which assumes the value 0 for $t < t_{n,\alpha}(c)$ and 1 for $t > t_{n,\alpha}(c)$. For $t = t_{n,\alpha}(c)$ we take $\Phi(t, c) = a_{n,\alpha}(c)$ where $0 \le a_{n,\alpha}(c) < 1$ is chosen such that

$$E\{\Phi(T, C) | C = c,\, H_0\} = \alpha. \tag{10}$$

It follows that $\Phi(t, c)$ is a strictly size α test for H_0 in (3). This is a conditional test and is randomized in nature. A non-randomized test may be as follows:

$$\text{reject } H_0 \text{ if } T > t_{n,\alpha}(c), \text{ otherwise accept } H_0, \tag{11}$$

whatever be $F_1, \ldots, F_n$, the size of the test $\le \alpha$.

For large n, we may approximate $F_n(t|c)$ by the c. d. f $F_{\chi^2}(t)$ for the χ^2 distribution with 2 degrees of freedom and use the upper α point $\chi^2_{2,\alpha}$ of the latter from $t_{n,\alpha}(c)$.

Therefore for large n, the randomized as well as non-randomized test reduces to the test:

$$\text{reject } H_0 \text{ if } T > \chi^2_{2,\alpha}, \text{ otherwise accept } H_0. \tag{12}$$

The test is shown to be consistent and unbiased against certain alternatives by Chatterjee [3].

4. Steps for the Proposed Multivariate Non-Parametric Control Chart

1. Collect at least 20 samples of size n (preferably ≥ 10).
2. Calculate statistic T as described in previous section for each sample.
3. Choose an α and set up upper control limit as $\chi^2_{2,\alpha}$.
4. Plot T in the chart.
5. If any point goes beyond the limit take control action.

5. Performance Comparison of the Proposed Method with the Existing One

A popular measure of chart performance is the expected value of the run length (the number of samples or subgroups that need to be collected, before the first out of control signal is given by a chart, is a random variable called the run length.) distribution, called the average run length (*ARL*). It is desirable (often stipulated) that the *ARL* of a chart be large when the process is in control. Larger the value of the in-control *ARL* better the performance of the chart with respect to false alarm.

By definition, the run length is a positive integer valued random variable, so the *ARL* looses much of its attractiveness as a typical summary if the distribution is skewed (as is often the case). As a result misleading conclusions can be drawn based on *ARL*. Hence, the use of *ARL* has been criticized, owing to the skewness of the run length distribution in the out-of-control case and its non-normality. In such cases it is recommended to characterize run length distribution by Median run length (*MRL*) and standard deviation of run length (*sdRL*). Since, it becomes inconvenient to compare different situations with three measures (*viz*, *ARL*, *MRL*, *sdRL*) it has been decided to compare different methods vis-à-vis to their powers in terms of fraction of correct classification.

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. A second measure as fraction of correct classification to detect different shifts in location may serve the above purpose. Larger the value of the fraction correct classification for a particular shift greater the efficiency of the chart to detect the shift. Some of researchers used this measure to evaluate the performance of a control chart. In this context see Niaki and Abbasi [14] and Das and Prakash [4]. This measure is related to probability of type-II error (β) or power of a test ($1-\beta$). The type-II error is the error that occurs when the chart does not give out of control signal when actually the process is out of control. From the established theory of test of hypothesis it is known that both the errors (Type-I and Type-II) can not be reduced simultaneously for a particular test. Hence test of hypothesis is designed by keeping one error in a small value thereby reducing the other error. Since Control chart is nothing but test of hypothesis the same theory will be applicable. The relationship between Power of a control chart and *ARL* is as follows.

$$
\begin{aligned}
\beta &= \text{Probability of type II error.} \\
&= P(\text{A point lies within thecontrol limits but assignable cause has occurred}) \\
&= P(\text{control chart fails to detect the shift})
\end{aligned}
$$

Power of a chart = P(A point lies beyond the control limits when assignable cause has occurred)

= P(control chart succeeds todetect the shift)

$= 1 - \beta$

Out of control $ARL = 1/(1-b)$.

Power can be expressed in terms of probability also. This probability value represents the fraction of cases in which the chart is able to detect a shift considering. This is called fraction correct classification. The performance of the proposed bivariate nonparametric control chart was evaluated by simulating data from both Multivariate Normal and multivariate t distribution.

First, consider multivariate normal case. In-control state was assumed to be characterized by $N_2(0, 0, 1, 1, \rho)$. The value of correlation coefficient (ρ) was assumed as 0.5. In-control state performance was measured by estimating in-control ARL. Samples of size n were simulated from $N_2(0, 0, 1, 1, \rho)$. Both Hotelling T^2 statistic and nonparametric statistic were calculated and plotted in the respective control charts. This exercise was carried out till the value exceeded the UCL fixing a value for α. The number of samples needed to exceed the limit was counted which was nothing but run length. 10000 such run length were estimated for both the chart and then its average was calculated and reported in Table 1 as in control ARL. The value of n was taken as 10, 15 and 20. The value of α was taken as 0.05, 0.01 and 0.001.

Table 1. In-Control ARL performance of existing Hotelling T^2 control Chart & proposed nonparametric control chart when data simulated from bivariate normal.

Sample Size	α	In control ARL (Hotelling T^2)	In control ARL (Nonparametric)
10	0.050	17.56	23.5700
	0.010	107.06	258.2300
	0.001	1102	>10000
15	0.050	19.65	28.2900
	0.010	107.44	286.7800
	0.001	934	>10000
20	0.050	18.3	26.6500
	0.010	89.35	141.9100
	0.001	1126.7	>10000

The out of control performance of the proposed control chart was compared with that of the existing Hotelling T^2 chart by estimating its power to detect the shift in location parameters in terms of fraction correct classification. 10000 samples of size n were simulated from bivariate normal population with shift in mean vector and number of cases where a particular chart was able to detect the shift was noted. From this number an estimate of power in terms of fraction correct classification was obtained and reported in Table 2. The shift was taken in two variables as ±0.25, ±0.5, ±1.0, ±1.5, ±2.0, ±2.5, and ±3.0. The value of n was taken as 10, 15, and 20. The value of α was taken as 0.05.

Table 2. Out of Control performance in terms of fraction correct classification for existing Hotelling T^2 control Chart & proposed nonparametric control chart for different shifts in location parameter when data simulated from bivariate normal ($\alpha = 0.05$).

Shift in First variable mean	Shift in Second variable mean	$n = 20$		$n = 15$		$n = 10$	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
0.25	0.25	0.165	0.110	0.164	0.098	0.118	0.087
	0.50	0.508	0.324	0.378	0.230	0.278	0.188
	1.00	0.990	0.861	0.959	0.679	0.844	0.543
	1.50	1	0.992	1	0.953	0.995	0.858
	2.00	1	1	1	0.996	1	0.978
	2.50	1	1	1	1	1	0.998
	3.00	1	1	1	1	1	1
	-0.25	0.489	0.216	0.397	0.128	0.274	0.106
	-0.50	0.866	0.428	0.780	0.338	0.570	0.182
	-1.00	1	0.921	1	0.815	0.968	0.552
	-1.50	1	0.999	1	0.980	1	0.888
	-2.00	1	1	1	1	1	0.987
	-2.50	1	1	1	1	1	0.999
	-3.00	1	1	1	1	1	1
0.50	0.50	0.632	0.404	0.498	0.336	0.359	0.262
	1.00	0.988	0.851	0.943	0.679	0.804	0.573
	1.50	1	0.995	1	0.952	0.999	0.887
	2.00	1	1	1	0.998	1	0.977
	2.50	1	1	1	1	1	1
	3.00	1	1	1	1	1	1
	-0.25	0.867	0.464	0.764	0.344	0.602	0.203
	-0.50	0.985	0.698	0.931	0.521	0.828	0.317
	-1.00	1	0.970	1	0.901	0.995	0.650
	-1.50	1	1	1	0.990	1	0.919
	-2.00	1	1	1	1	1	0.981
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
1.00	1.00	0.996	0.965	0.992	0.867	0.928	0.764
	1.50	1	0.996	1	0.974	0.992	0.929
	2.00	1	1	1	1	1	0.983
	2.50	1	1	1	1	1	0.998
	3.00	1	1	1	1	1	1
	-0.25	1	0.925	0.998	0.816	0.968	0.593
	-0.50	1	0.968	1	0.897	0.995	0.652
	-1.00	1	0.999	1	0.993	1	0.865
	-1.50	1	1	1	1	1	0.976
	-2.00	1	1	1	1	1	0.997
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1

Table 2. (Continued).

Shift in First variable mean	Shift in Second variable mean	$n = 20$		$n = 15$		$n = 10$	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
1.5	1.50	1	1	1	0.995	1	0.962
	2.00	1	1	1	0.999	1	0.992
	2.50	1	1	1	1	1	0.999
	3.00	1	1	1	1	1	1
	-0.25	1	0.999	1	0.987	1	0.885
	-0.50	1	1	1	0.992	1	0.911
	-1.00	1	1	1	1	1	0.970
	-1.50	1	1	1	1	1	0.998
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
2.0	2.00	1	1	1	1	1	0.999
	2.50	1	1	1	1	1	1
	3.00	1	1	1	1	1	1
	-0.25	1	1	1	1	1	0.980
	-0.50	1	1	1	0.999	1	0.984
	-1.00	1	1	1	1	1	0.998
	-1.50	1	1	1	1	1	1
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
2.5	2.50	1	1	1	1	1	1
	3.00	1	1	1	1	1	1
	-0.25	1	1	1	1	1	0.999
	-0.50	1	1	1	1	1	1
	-1.00	1	1	1	1	1	0.998
	-1.50	1	1	1	1	1	1
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
3.0	3.00	1	1	1	1	1	1
	-0.25	1	1	1	1	1	1
	-0.50	1	1	1	1	1	1
	-1.00	1	1	1	1	1	1
	-1.50	1	1	1	1	1	1
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
-0.25	-0.25	0.205	0.134	0.156	0.105	0.118	0.090
	-0.50	0.496	0.302	0.391	0.222	0.257	0.170
	-1.00	0.993	0.838	0.959	0.671	0.853	0.510
	-1.50	1	0.997	1	0.960	0.999	0.874
	-2.00	1	1	1	0.999	1	0.985
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1

Table 2. (Continued).

Shift in First variable mean	Shift in Second variable mean	$n = 20$		$n = 15$		$n = 10$	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
-0.50	-0.50	0.611	0.399	0.486	0.285	0.367	0.248
	-1.00	0.986	0.831	0.947	0.703	0.817	0.581
	-1.50	1	0.990	1	0.964	0.995	0.870
	-2.00	1	1	1	0.994	1	0.977
	-2.50	1	1	1	1	1	0.998
	-3.00	1	1	1	1	1	1
-1.00	-1.00	1	0.997	1	0.978	0.996	0.900
	-1.50	1	1	1	0.994	0.998	0.969
	-2.00	1	1	1	1	1	0.990
	-2.50	1	1	1	1	1	0.997
	-3.00	1	1	1	1	1	1
-1.50	-1.50	1	1	1	0.999	1	0.991
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
-2.00	-1.50	1	1	1	1	1	1
	-2.00	1	1	1	1	1	1
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
-2.50	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
-3.00	-3.00	1	1	1	1	1	1

The above exercise was repeated by simulating data from a non-normal distribution *i.e.* multivariate t distribution where in control state was characterized by mean vector $(0, 0)$, $\sum$ matrix with diagonal elements as 1 and off-diagonal elements as $\rho(=0.5)$ and degree of freedom as 3. The corresponding in control *ARL* was reported in Table 3 and out of control performance in terms of fraction correct classification was reported in Table 4. Programs were written using MATLAB 7.0.1 to prepare the results presented in the these tables.

Table 3. In-Control *ARL* performance of existing Hotelling T^2 control Chart & proposed nonparametric control chart when data simulated from bivariate t distribution.

Sample Size	α	In control *ARL* (Hotelling T^2)	In control *ARL* (Nonparametric)
10	0.050	3.46	23.44
	0.010	6.19	258.18
	0.001	11.45	7845
15	0.050	3.19	23.9
	0.010	5.46	314.61
	0.001	11.07	8736
20	0.050	3.47	30.71
	0.010	5.84	150.81
	0.001	11.03	6149

Table 4. Out of Control performance in terms of fraction correct classification for existing Hotelling T^2 control Chart & proposed nonparametric control chart for different shifts in location parameter when data simulated from bivariate t distribution ($\alpha = 0.05$).

Shift in First variable mean	Shift in Second variable mean	$n = 20$		$n = 15$		$n = 10$	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
0.25	0.25	0.414	0.114	0.386	0.102	0.344	0.085
	0.50	0.589	0.238	0.525	0.175	0.480	0.168
	1.00	0.953	0.722	0.894	0.540	0.817	0.408
	1.50	0.997	0.948	0.986	0.846	0.977	0.689
	2.00	0.999	0.997	1	0.947	0.993	0.864
	2.50	1	0.998	1	0.984	1	0.939
	3.00	1	1	1	0.995	0.999	0.957
	-0.25	0.593	0.158	0.523	0.134	0.455	0.105
	-0.50	0.808	0.368	0.740	0.267	0.640	0.151
	-1.00	0.986	0.802	0.976	0.629	0.906	0.414
	-1.50	1	0.960	0.995	0.904	0.995	0.718
	-2.00	1	0.992	1	0.968	0.998	0.873
	-2.50	0.999	1	1	0.993	1	0.947
	-3.00	1	1	1	0.996	1	0.978
0.50	0.50	0.666	0.349	0.588	0.236	0.445	0.203
	1.00	0.938	0.722	0.890	0.548	0.778	0.447
	1.50	0.998	0.948	0.985	0.834	0.957	0.682
	2.00	1	0.993	0.999	0.957	0.996	0.891
	2.50	1	1	1	0.988	0.998	0.924
	3.00	1	0.999	1	0.992	1	0.967
	-0.25	0.828	0.407	0.748	0.288	0.645	0.156
	-0.50	0.944	0.573	0.885	0.451	0.785	0.225
	-1.00	0.995	0.895	0.989	0.759	0.949	0.530
	-1.50	1	0.987	0.999	0.932	0.994	0.774
	-2.00	1	0.998	1	0.974	0.999	0.883
	-2.50	1	0.999	1	0.988	1	0.951
	-3.00	1	1	1	0.996	1	0.983
1.00	1.00	0.972	0.890	0.943	0.720	0.867	0.590
	1.50	0.996	0.970	0.991	0.870	0.964	0.760
	2.00	0.998	0.990	0.998	0.960	0.994	0.890
	2.50	1	1	1	0.990	1	0.940
	3.00	1	1	1	1	0.998	0.970
	-0.25	0.984	0.790	0.975	0.700	0.918	0.440
	-0.50	0.995	0.900	0.995	0.750	0.964	0.510
	-1.00	1	0.980	0.998	0.940	0.993	0.720
	-1.50	1	1	1	0.980	0.999	0.880
	-2.00	1	1	1	0.990	0.999	0.950
	-2.50	1	1	1	1	0.999	0.980
	-3.00	1	1	1	1	0.999	0.990

Table 4. (Continued).

Shift in First variable mean	Shift in Second variable mean	n = 20		n = 15		n =10	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
1.5	1.50	0.999	0.987	0.998	0.939	0.979	0.843
	2.00	0.999	0.995	0.998	0.974	0.993	0.900
	2.50	1	1	1	0.991	1	0.965
	3.00	0.999	1	1	0.998	1	0.979
	-0.25	1	0.975	0.998	0.898	0.995	0.760
	-0.50	0.999	0.984	0.999	0.921	0.994	0.783
	-1.00	1	0.998	1	0.987	0.998	0.885
	-1.50	1	1	1	0.994	1	0.950
	-2.00	1	1	1	0.999	1	0.974
	-2.50	1	1	1	1	1	0.991
	-3.00	1	1	1	1	1	0.997
2.0	2.00	1	1	1	0.990	0.997	0.940
	2.50	1	1	1	1	0.997	0.960
	3.00	1	1	1	1	1	0.990
	-0.25	1	0.990	1	0.980	1	0.870
	-0.50	1	1	1	0.990	1	0.920
	-1.00	1	1	1	1	1	0.950
	-1.50	1	1	1	1	1	0.980
	-2.00	1	1	1	1	1	0.980
	-2.50	1	1	1	1	1	1
	-3.00	1	1	1	1	1	1
2.5	2.50	1	1	1	0.998	1	0.978
	3.00	1	1	1	1	1	0.985
	-0.25	1	1	1	0.994	1	0.936
	-0.50	1	1	1	0.994	1	0.960
	-1.00	0.999	1	1	0.999	1	0.971
	-1.50	1	1	1	1	1	0.988
	-2.00	1	1	1	1	1	0.997
	-2.50	1	1	1	1	1	0.998
3.0	3.00	1	1	1	1	1	0.992
	-0.25	1	0.999	1	0.996	0.999	0.976
	-0.50	1	1	1	0.996	1	0.986
	-1.00	1	1	1	1	1	0.991
	-1.50	1	1	1	1	1	0.994
	-2.00	1	1	1	1	1	0.995
	-2.50	1	1	1	1	1	0.998
	-3.00	1	1	1	1	1	0.999
-0.25	-0.25	0.413	0.108	0.364	0.098	0.333	0.088
	-0.50	0.608	0.253	0.527	0.171	0.447	0.185
	-1.00	0.947	0.704	0.905	0.535	0.821	0.415
	-1.50	0.995	0.933	0.991	0.833	0.971	0.678
	-2.00	0.999	0.994	0.999	0.968	0.998	0.858
	-2.50	0.999	1	1	0.983	0.999	0.947
	-3.00	1	0.999	1	0.997	1	0.973

Table 4. (Continued).

Shift in First variable mean	Shift in Second variable mean	$n = 20$		$n = 15$		$n = 10$	
		Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric	Hotelling T^2	Non-Parametric
-0.50	-0.50	0.669	0.365	0.614	0.262	0.520	0.237
	-1.00	0.944	0.731	0.911	0.555	0.802	0.464
	-1.50	0.997	0.954	0.984	0.817	0.966	0.685
	-2.00	1	0.994	0.998	0.949	0.994	0.859
	-2.50	1	0.999	0.999	0.985	0.999	0.935
	-3.00	1	1	1	0.995	1	0.976
-1.00	-1.00	0.962	0.854	0.959	0.730	0.850	0.585
	-1.50	0.995	0.968	0.988	0.879	0.958	0.755
	-2.00	1	0.995	1	0.962	0.993	0.884
	-2.50	1	1	1	0.981	0.999	0.931
	-3.00	1	1	1	1	0.999	0.974
-1.50	-1.50	0.999	0.993	0.996	0.948	0.987	0.845
	-2.00	1	0.995	0.997	0.977	0.998	0.911
	-2.50	1	1	1	0.995	1	0.957
	-3.00	1	1	1	0.997	1	0.976
-2.00	-2.00	1	0.999	1	0.988	0.990	0.955
	-2.50	1	1	1	0.991	1	0.980
	-3.00	1	1	1	0.998	1	0.990
-2.50	-2.50	1	1	1	0.999	1	0.983
	-3.00	1	1	1	0.997	1	0.992
-3.00	-3.00	1	1	1	0.999	1	0.997

6. Conclusion

- As far as in control *ARL* is concerned performance of the proposed method is better than that of existing Hotelling T^2 chart. With the increase of α in control *ARL* decreases.
- As far as out of control performance is concerned though performance of the proposed method is slightly poorer than the existing method but it is quite satisfactory. Especially for detecting higher shift (>1.5) the performance of the proposed method is as good as the existing one.
- With the increase of sample size the performance of the chart is improving to detect a particular shift.
- In both type control charts we have noticed a different picture in 'in control' and 'out of control' criteria. This is expected since for any test of hypothesis if we try to increase probability of Type-I error (α) then probability of Type-II error will decrease and vice versa.

6.1. Advantage and Disadvantage of the Proposed Control Chart

- It is easy to calculate the test statistic for the proposed method.
- It does not require any distributional assumptions.
- Estimate of $\sum$ is needed for existing methodology but for the proposed method it is not needed.
- Its limitation is it cannot be recommended for sample size less than 10.
- It can be applied only for bivariate cases. When number of variables is more than 2, multivariate Generalization of the bivariate Sign Test for location can be applied.

6.2. Scope for Further Study

The above study can be extended by simulating data from other non-normal multivariate distribution like multivariate gamma. Multivariate Generalization of the bivariate Sign test for location can also be applied taking data for more than 2 variables assuming different distributions.

References

1. Bakir, S. T. (2001). Classification of distribution free control charts. *Proceedings of Annual Meeting of American Statistical Association*, *Section on Quality and Productivity*. www.amstat.org/sections/SRMS/Proceedings/y2001/Proceed/00088.pdf
2. Chakrabarti, S., Vander Der, Laan, P. And Bakir, S. T. (2001). Nonparametric control charts: an overview and some results. *Journal of Quality Technology*, 3, 304-315.
3. Chatterjee, S. K. (1966). A bivariate sign test for location. *The Annals of Mathematical Statistics*, 37, 1771-1782.
4. Das, N. and Prakash, V. (2008). Interpreting the out of control signal in multivariate control chart – a comparative study. *International Journal of Advanced Manufacturing Technology*, 37, 966-979.
5. Hamurkarouglu, C., Mert, M. and Sayken, Y. (2004). Nonparametric control charts based on Mahalanobis depth. *Hacettepe Journal of Mathematics and Statistics*, 33, 57-67.
6. Hayter, A. J. and Tsui, K. L. (1994). Identification and quantification in multivariate quality control procedure. *Journal of Quality Technology*, 26, 197-208.
7. Hur, Y. (2007). *Wavelet-based Control Charts for General Multivariate Processes, Seminar*. Department of Applied Mathematics and Statistics. The John Hopkins University. www.mts.jhu.edu/~seminar/seminar/20071206hur.pdf
8. Kapatou, A. and Reynolds, M. R. Jr. (1994). Multivariate nonparametric control chart using small samples. *Proceedings of Annual Meeting of American Statistical Association*, *Section on Quality and Productivity*, 241-246.
9. Kapatou, A. and Reynolds, M. R. Jr. (1998). Multivariate nonparametric control chart for the case of known Σ. *Proceedings of Annual Meeting of American Statistical Association*, *Section on Quality and Productivity*, 77-82.
10. Liu, R. (1995). Control charts for multivariate processes. *Journal of the American Statistical Association*, 90, 1380-1387.
11. Lowrey, C. A. and Montgomery, D. C. (1995). A review of the multivariate control charts. *IIE Transactions*, 27, 800-810.
12. Macgregor, J. F. and Kourti, T. (1995). Statistical quality control of multivariate processes. *Control Engineering Practice*, 3, 403-414.
13. Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, 4^{th} edition. Wiley, New York.
14. Niaki, S. T. A. and Abbasi, B. (2005). Fault diagnosis in multivariate control charts using artificial neural networks. *Quality Reliability Engineering International*, 21(8), 825-840.
15. Polansky, A. M. (2001). A smooth nonparametric approach to multivariate process capability. *Technometrics*, 43, 199-211.
16. Puri, M. L. and Sen, P. K. (1976). *Nonparametric Methods in Multivariate Analysis.* John Wiley & Sons Inc, New York.

17. Raich, A. and Cinar, A. (1996). Statistical process monitoring and disturbance diagnosis in multivariate continuous processes. *American Institute of Chemical Engineers Journal*, 42, 995-1009.

18. Sun, R. and Tsung, F. (2003). A kernel distance based multivariate control chart using support vector methods. *International Journal of Production Research*, 41 (13), 2975-2989.

19. Shilling, A. G. and Nelson, P. R. (1976). The effect of non-normality on the control limits of X charts. *Journal of Quality Technology*, 8, 183-188.

20. Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31, 376-386.

Author's Biographies:

Nandini Das did her Master degree in Statistics and M-Tech in Quality, Reliability and Operations research from Indian Statistical Institute. Presently she is working as a faculty in Statistical Quality Control and Operations Research division of Indian Statistical Institute. Her area of interest is statistical process control.