uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Jun Mao

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Mathematics)

GRADE / DEGREE

Department of Mathematics & Statistics

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Cumulative Sum Quality Chart Using Voronoi Depht

TITRE DE LA THÈSE / TITLE OF THESIS

D. MacDonald

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

M. Zarepour

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

M. Alvo

G. Ivanoff

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

CUMULATIVE SUM QUALITY CHART USING VORONOI DEPTH

Jun Mao

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
for the degree of
Master of Science in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The M.Sc. Program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Your file Votre référence
ISBN: 978-0-494-61319-1
Our file Notre référence
ISBN: 978-0-494-61319-1

Abstract

In this thesis, we discuss statistical quality control. A new nonparametric multivariate cusum based on Voronoi rank is proposed. We can reasonably predict on-target run lengths and the off-target run lengths are as short as any competing nonparametric procedure.

Acknowledgements

I wish to express my sincere gratitude to my supervisors, Professor David McDonald and Professor Mahmoud Zarepour. Their wide knowledge and logical way of thinking have been of great value to me. Their understanding, encouragement and personal guidance have provided a good basis for the present thesis, and their many useful suggestions and comments have been received with great appreciation.

Dedication

I wish to dedicate this thesis to my parents, Ximmin Mao and Yulan Li. Both of them give me strength and support, which have encouraged me to achieve my goals in life. This thesis is also dedicated to my beloved wife Dan Jin. None of this would be possible without your love and support.

Contents

List of Figures

Chapter 1

Introduction

Approximately 1.7 million years ago, our ancestors began manufacturing tools. Up until the Middle Ages, the manufacturing technique remained artisanal; every item was one of a kind. After the industrial revolution, quality control became a major concern in industry when parts became interchangeable. Quality improvement tools have been introduced which helped reduce problems in the production and distribution of manufactured goods. Improvement of quality did not extend only to products, but also included the process used for making products.

Statistical methods now play an important role in quality control. The concept had its greatest development during the 20th century. Statistical Process Control (SPC) was introduced by Walter Shewhart at Bell labs in the early 1920s. The work ensures reliability of telephone communication devices. He described "common cause" and "special cause" variation. Common cause variation cannot be eliminated easily from the process without fundamental changes. Attempts to remove this common cause variation were time consuming, costly, and wasteful. On the other hand, special cause variation is a difference in an outcome of a process that must be removed from the process to improve quality. The purpose of statistical process control is to detect special cause variation.

Statistical Process Control is an effective method of using statistical techniques to measure and analyze process variation. SPC monitors manufacturing processes in

order to keep the process close to a specified target. SPC is also a tool for improving product quality through the reduction of variability. When special causes of variation have been eliminated a process is said to be in a state of statistical control (in-control). If there are changes in the distribution over time, the process is said to be out-of-control.

1.1 Statistical Process Control chart

A primary tool used for SPC is the control chart, a graphical representation of the quality statistic for the manufacturing process. Shewhart [31] created the basis for the control chart and invented the concept of a state of statistical control. As the interest of statistical process control increased, many charts were proposed by researchers. These charts have been shown to be useful in monitoring many industrial processes. All control charts have two basic components:

(1) a descriptive statistic based on performance data is plotted over time.
(2) control lines define the limits of special cause variations.

The descriptive statistic is calculated from the samples. When the plotted statistic falls outside the control lines, an out-of-control signal is generated. When a signal is observed, engineers are expected to correct the manufacturing process.

There are two different phases in control charting practice. In Phase I, the focus is to test historical data in order to establish whether the process is in control or not. The use of control charts in Phase I is to determine appropriate control limits to use in Phase II. In Phase II, the aim is to monitor production data in order to decide whether the process is in control or not. Much work, process understanding, and process improvement is often required in the transition from Phase I to Phase II [36]. Most of the research on control charts is on Phase II performance. In the article, we only consider Phase II quality control charts.

A quality control procedure needs to meet certain statistical criteria. The number of samples until the signal occurs is called a run length. The run length is a random number, it can vary from run to run. The most common measure of control chart is the average run length (ARL). If the process is in control, the ARL should be large. If the process is out-of-control, the ARL should be small. The in-control ARL and out-of-control ARL distribution provides a comprehensive understanding of the control chart's performance.

1.2 Overview of the thesis

Many different control charts have been proposed by researchers. We will review the univariate parametric, univariate nonparametric, multivariate parametric and multivariate nonparametric charts in Chapter two.

A new multivariate nonparametric chart called Voronoi Cusum is developed in Chapter Three. Our chart is based on Voronoi tessellation, which was first formally introduced by R. Descartes (1850) in the 17th century. Simulations are conducted and the results show the new chart has better performance than existing multivariate nonparametric methods. Another advantage is that the chart is easily applied to high dimensional data.

Chapter 2

Statistical Process Control Charts

2.1 Univariate Parametric Process Control Charts

2.1.1 The Shewhart $\bar{X}$ Control Charts

The control chart was introduced by Walter Shewhart [31] in 1931. This chart has proven to be a simple and effective means of understanding data. Figure 2.1 gives a typical Shewhart $\bar{X}$-chart. The control statistic $\bar{X}$ is computed from a small sample of production data. The central line on the chart represents the average of the summary statistics when the process is in the state of control or perhaps some nominal value. The upper and lower control limits labeled as UCL and LCL, indicates the range of variation for the summary statistics when the process is in control. The control limits are usually set as three standard errors of the target. As long as the points stay inside the control limits, the process is deemed in-control. An out-of-control signal is given if $\bar{X}$ falls outside of control limits. This procedure has no memory, and so although it is very effective for detecting isolated special causes that lead to large shifts in the data, it is not very effective in detecting more moderate persistent shifts [12].

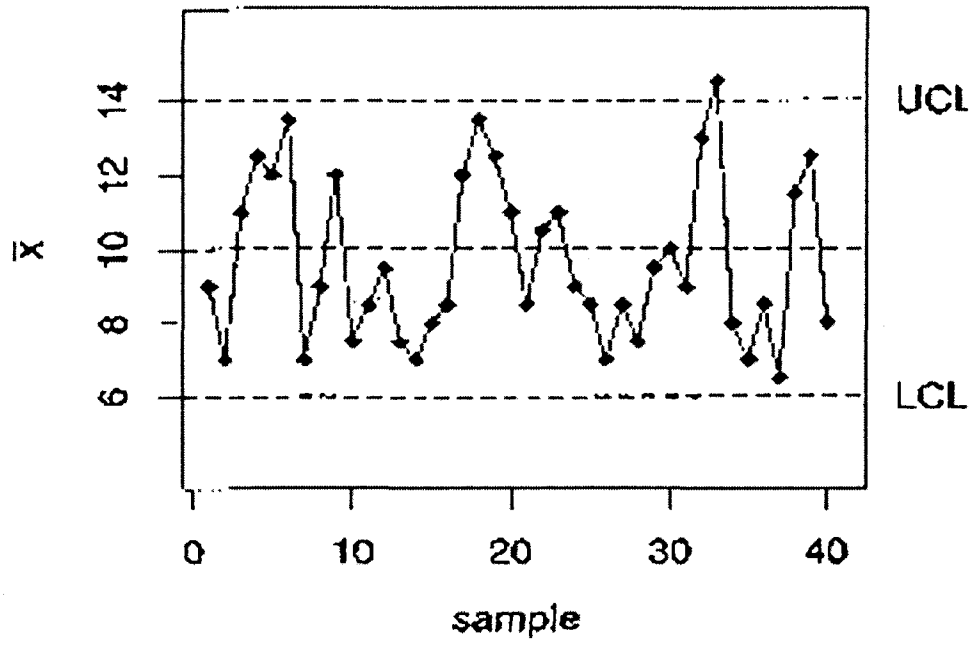

Figure 2.1: Shewhart $\bar{X}$ Chart.

2.1.2 Cumulative Sum Charts

A very effective alternative to the Shewhart $\bar{X}$ chart is the Cumulative Sum Chart (or **Cusum**). Page [27] introduced the idea in 1954, although not in the form used now. The Cusum has been well studied and developed by many statisticians. It has been shown to be more effective than the $\bar{X}$-chart for detecting small shifts in the mean of the process.

The Cusum chart works as follows: suppose that the sample size is m and $\overline{x}_j$ is the average of the jth sample. If μ_0 is the target value for the mean of the process, define the cumulative sum by

$$C_n = \sum_{j=1}^{n} (\overline{x}_j - \mu_0)$$

C_n is called the cumulative sum up to the nth sample. If the process in control,

the C_n should remain close to zero.

There are two general methods to detect the shift: the V-mask control scheme and the Decision Interval. The historical tool is the V-mask, the name comes from its shape. A typical V-mask is given in Figure 2.2. The mask looks like the letter V lying on its side. If all the Cusum C_n lie within the arms of V-mask, the process is deemed in control. However, if any of C_n is outside of the arms, it implies the process is out of control. For more detail on the V-mask, see [12].

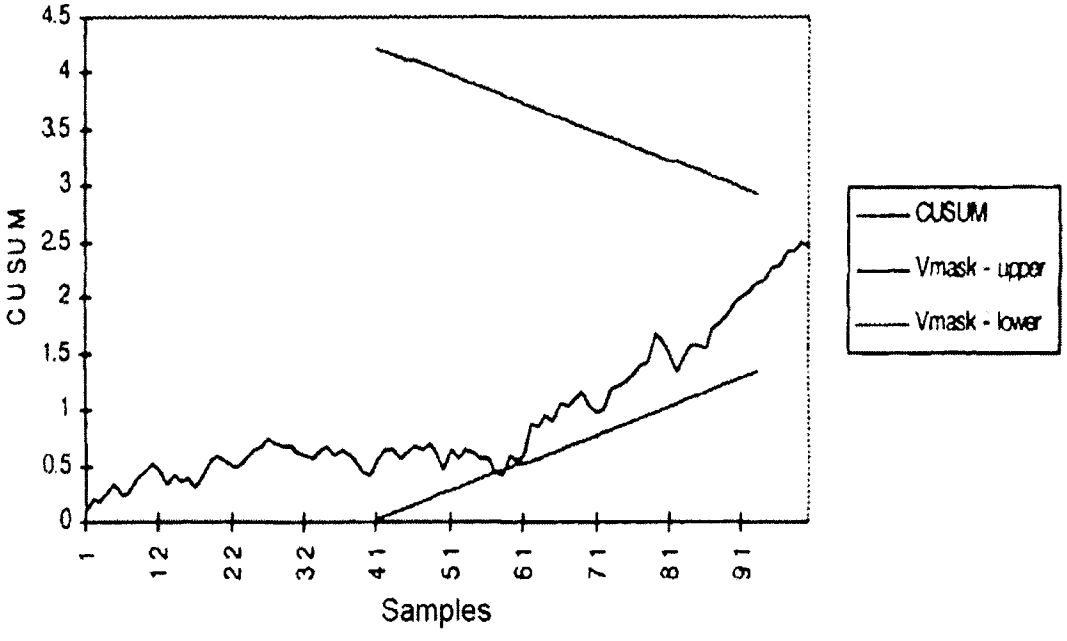

Figure 2.2: Cusum V-Mask.

The Decision Interval is more popular and is used in this thesis. We need to define the Cusum as follows;

Let $\overline{x}_j$ denote the average of a small sample of quality measurements then

$$C_0 = 0$$

$$C_n = max(0, C_{n-1} + \overline{x}_j - \mu_0 - k).$$

A signal occurs if

$$C_n > h$$

In the equation, k is called the reference value and is chosen to produce an optimal response to a shift of a specified size δ [12]. The parameter h is called the decision interval. If the Cusum C_n exceeds h, the process is deemed out-of-control. The in-control ARL depends on the values of k and h. The table with values of k and h and in-control ARL for Cusum of standardized normal data is provided in Section 3.

2.1.3 Exponentially Weighted Average Control Charts

Another type of primary control chart is called the Exponentially Weighted Average Control (or **EWMA**) Chart. An average of the recent data is used to smooth the noise in the data to generate a better estimate of the process mean than that obtained from the most recent observation [26]. The smoothing factor is denoted by λ, where $\lambda \in [0, 1]$. The mean $\bar{x}_t$ represents the average value of the samples at time t. The exponentially weighted moving average is defined by:

$$Z_t = \sum_{i=0}^{t} \lambda(1-\lambda)^i \bar{x}_{t-i} + (1-\lambda)^t Z_0$$

with the first value Z_0 in this sum generally being set to the mean of former observations. The smoothing factor λ is chosen by the user to determine the weighted average of all previous data. The EWMA chart is good in detecting the smaller shift (from 0.5δ to 2δ). However it is not very efficient in detecting large shifts in the process.

2.2 Univariate Nonparametric Process Control Charts

Nonparametric control charts are designed for data that does not follow the normal density or any other specific parametric distribution assumption. Chakraborti [5]

gave the following definition for the nonparametric control chart: if the in-control run length distribution is the same for every continuous distribution, then the chart is called nonparametric or distribution-free. Chakraborti [5] also discussed the advantages of nonparametric control charts: (i) their simplicity; (ii) the lack of a need to assume a particular distribution for the underlying process; (iii) an in-control run length distribution that is the same for all continuous distributions; (iv) their greater robustness and insensitivity to outliers; (v) their greater efficiency in detecting changes when the true distribution is markedly non-normal, particularly with heavier tails; (vi) the lack of a need to estimate the variance to set up charts for the location parameter.

2.2.1 Sequential Rank Cusum

McDonald [24] proposed the univariate Cusum procedure based on sequential ranks. For a sequence of independent identically distributed random variables $X_1, \ldots, X_n$, the sequential ranks are defined as follows [24],

$$R_i = 1 + \sum_{k=1}^{i-1} I(X_k - X_i)$$

$I(\cdot)$ is the indicator function

$$I(\cdot) = \begin{cases} 1 & \text{if } X_k - X_i \leq 0 \\ 0 & \text{otherwise,} \end{cases}$$

The statistic U_i is defined as

$$U_i := \frac{R_i}{(i+1)}$$

U_i is independent of U_j for $j \neq i$ and is uniformly distributed on

$$\left\{ \frac{1}{i+1}, \frac{2}{i+1}, \ldots, \frac{i}{i+1}, \right\}$$

Based on the U_i, the Cusum procedure can be constructed as follows

$$T_0 = 0, \quad T_i = max\{T_{i-1} + U_i - k, 0\}, \quad i = 1, 2, 3, \ldots$$

where $k > 0$ (the reference value) and $h > 0$ (the signal value). The procedure signals an out-of-control alarm when $T_i \geq h$.

When the process is in control, the ARL depends on the value of k and h and does not depend on the on-target distribution F. For a given reference value k, one can fix the ARL by the corresponding signal level h.

In order to study the relationship between k, h and ARL, consider the hypotheses in the following,

$$H_0 : U_i \sim \text{uniform on } [0,1] \text{ and independent}$$

$$H_1 : U_i \sim G \circ F^{-1}, \quad \text{where } F \sim N(0,\sigma^2),\ G \sim N(\mu\sigma,\sigma^2).$$

then

$$\text{reject } H_0 \ if \ \sum_{i=1}^{N} \Phi^{-1}(\bar{U}_i) \geq \frac{1}{\mu} \ln \frac{1}{\alpha}$$

where α is the type I error probability. For a process in control, the U_i's are approximately independent uniforms $\bar{U}_i$.

The associated lower arm of the V-mask is determined by

$$k = \frac{\mu}{2}, \qquad h = \frac{1}{\mu} \ln \frac{1}{\alpha}$$

Note that $\Phi^{-1}(U_i)$ is approximately a standard normal distribution if the process is in control. To determine h, one should specify the on-target ARL with the appropriate reference value k. Table 2.1 represents the optimal choice of h and k for a cusum based on sequential ranks.

If the procedure is in control, the U_i is approximately uniform on $[0,1]$. After the change, U_i is dependent and not uniform but the first few are almost independent with approximate distribution GF^{*-1}, where F^* is the limit of the empirical distributions F^n calculated from the observation $\{X_i\}$ [24].

McDonald [24] mentioned a censoring problem for nonparametric procedures. In the first case, the simple case where a false alarm does not imply any recalibration.

Table 2.1: Optimal choice of h and k for the sequential rank cusum for different on-target ARL and μ

μ	ARL_{On}	k	h
2.0	1000	0.7993	0.456
1.0	1000	0.6428	1.382
0.5	1000	0.5706	2.507
0.1	1000	0.5143	5.710
2.0	500	0.7993	0.395
1.0	500	0.6425	1.2031
0.5	500	0.5706	2.125
0.1	500	0.5143	4.455
2.0	100	0.7993	0.276
1.0	100	0.6428	0.798
0.5	100	0.5706	1.306
0.1	100	0.5143	2.181

The sampling distribution is uncensored; restart the cusum at 0 and keep the empirical distribution. In the second case, a false alarm does imply the recalibration. The sampling distribution is censored and restart the calculation of the cusum and the empirical distribution. There are two more different methods in the analysis of the change point. The worst-case analysis assumes that when cusum is 0 the change point has occured. The steady-state analysis suggests that the location of the cusum at the change point is distributed like the long-run distribution of the on-target cusum. Based on the above scenarios, four different simulations of sequential rank cusums are given in [24].

1. No recalibration after false alarms with the worst-case analysis.
2. Recalibration after false alarms with the steady-state analysis.

3. The worst case analysis of the optimal parametric cusum.
4. The steady-state analysis of the optimal parametric cusum.

The sequential rank cusum does very well in all four cases.

2.3 Multivariate Parametric Process Control Charts

A single variable is not enough to describe quality in the modern manufacturing plant. One product contains many different properties and one needs to monitor them simultaneously. For example, in the semiconductor industry, each chip is subject to a large number of quality measurements.

Jackson [10] pointed that any multivariate process control procedure should fulfill four conditions: a) an answer to the question: "Is the process in control?" must be available; b) an overall probability for the event "Procedure diagnoses an out-of-control state erroneously" must be specified; c) the relationships among the variables should be taken into account; d) an answer to the question: "If the process is out-of-control, what is the problem?" should be available. Many scholars have developed multivariate statistical process control charts. Woodall and Montgomery [37] state that the multivariate process control chart is one of the most rapidly developing areas of statistical process control chart. Unless explicitly stated otherwise, all control charts in this section follow a multivariate normal distribution.

2.3.1 Hotelling T^2

Hotelling [14] is the first statistician to study multivariate quality control. Hotelling's T^2 control chart is optimal for detecting a change in the process mean when the distribution is multivariate normal. Let X_i be the p-dimensional observations where $i = 1, 2, 3, \ldots, m$. The average of the sample is $\bar{X}$ and the target vector is μ_0. If S^{-1}

is the inverse of the sample covariance matrix, then the statistic T^2 is the following

$$T^2 = m(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0)$$

The Upper Control Limit (UCL) is $\chi^2_{\alpha,p}$, where p is the number of variables and α is the upper α percentage point of the chi-square distribution.

An extension of the subgroup procedure is more useful in practice. Set observations into k subgroups, each of size n, then compute a statistic T_s^2 as follows

$$T_s^2 = n(\bar{x}_i - \mu_0) S_p^{-1} (\bar{x}_i - \mu_0), i = 1, 2, \ldots, k.$$

where $S_p = S/k$ is the pooled covariance matrix obtained by averaging the k subgroup of covariance matrix.

The Upper Control Limit (UCL) is given by

$$UCL \sim \frac{p(k-1)(n-1)}{k(n-1)+1-p} F^{\alpha}_{p,k(n-1)+1-p}$$

where p is the number of variables, n is the size of the subgroup, and k is the number of subgroups in the sample.

The main advantage of the Hotelling's T^2 is that the calculations are simple. It was recognized as a major tool to analyze multivariate statistics quality procedures. Most statistical software has a package that implements Hotelling's T^2. Mason and Young [22] show the basic steps for the implementation of multivariate statistical process control using the Hotelling's T^2.

However, Hotelling's T^2 only uses the current sample information. It is very hard to detect a small shift in the mean vector. CUSUM and EWMA charts have already been proven to be more powerful in detecting a change point in the univariate case. Scholars also extended them to multivariate statistics quality control. MCUSUM and MEWMA control charts have been developed to overcome the problem.

2.3.2 Multivariate Cumulative Sum Control Chart

Woodall and Ncube [38] proposed extending the univariate CUSUM procedure to the multivariate case. Their idea is that the sample vector is a collection of unconnected

univariate data. The CUSUM analysis is based on the individual component. It is used to detect p-dimension mean vector shift, operate p one-sided or two-sided CUSUM charts simultaneously and evaluate the collection.

Let $X_n = (X_{1n}, X_{2n}, \ldots, X_{pn})^T$, $n = 1, 2, \ldots$, be p-variate independent normal random vectors. For an MCUSUM method either a one-sided or two-sided CUSUM procedure is applied to each random variable $X_{in}, i = 1, \ldots, p$. The one-sided CUSUM statistics are the following

$$S_{i,n} = \max\{0, S_{i,n-1} + X_{in} - k_i\}$$

$$S_{i,0} = 0 \quad \text{and} \quad k_i \geq 0$$

The run length of a one-sided CUSUM procedure for the ith variable is ARL(i)=min$\{n : S_{i,n} \geq h_i\}$, where $h_i > 0$. The two-sided CUSUM procedure also uses the statistics

$$T_{i,n} = \min\{0, T_{i,n-1} + X_{in} - k_i\}$$

$$T_{i,0} = 0$$

The run length of a two-sided CUSUM procedure is ARL(i)=min$\{ n : S_{i,n} \geq h_i$ or $T_{i,n} \leq -h_i\}$.

The reference value k_i depends individually on the ith variable which is considered to be out of control. Woodall and Ncube [38] pointed out ARL performance in a MCUSUM procedure is in relation to independent and dependent quality characteristics.

Healy [13] shows that the CUSUM procedure is related to the sequential probability ratio test. Let X_i be the ith observation from the multivariate normal with mean vector μ_0 and covariance matrix Σ. Define the noncentrality parameter λ

$$\lambda = [(\mu_1 - \mu_0)' \Sigma^{-1} (\mu_1 - \mu_0)]^{1/2}$$

where μ_1 is the out-of-control mean vector.

The one-sided CUSUM for detecting a positive shift in the mean of multivariate normal is

$$S_i = \max\{S_{i-1} + a'(X_i - \mu_0) - 0.5\lambda, 0\}$$

where

$$a' = \frac{(\mu_1 - \mu_0)'\Sigma^{-1}}{\lambda}$$

When $S_i > h$, the CUSUM generates an out-of-control signal. The variable $a'X_i$ is a linear combination of normal random variables, so $a'X_i$ is also normal. Then $a'(X_i - \mu_0)$ has a standard univariate normal distribution when X_i has mean equal to μ_0; and when X_i has mean equal to μ_1, $a'(X_i - \mu_0)$ has a univariate normal distribution with mean λ and variance 1 [13]. The multivariate CUSUM procedure reduces to a univariate normal procedure. The variables h and S_0 for a univariate normal CUSUM can also be applied to Healy's multivariate CUSUM. The ARL performance does not change as p increases and it is the same as a one-sided univariate normal CUSUM.

Crosier [7] presented two multivariate CUSUM quality-control procedures. The first method is called a COT scheme. The CUSUM is given by

$$S_n = \max\{0, S_{n-1} + T_n - k\}$$

where $T_n = \sqrt{(X_n - \mu_0)'\Sigma^{-1}(X_n - \mu_0)}$, $S_0 = 0$ and $k > 0$. X_n is the nth observation and μ_0 is the on target mean of a multivariate normal distribution with known covariance matrix Σ. The procedure will signal if $S_n > h$. The optimal value of k is the square root of the dimension.

The ARL's of COT schemes depend on the mean vector and covariance matrix, because the statistic is based on the square root of Hotelling's T^2. The COT schemes have several advantages compared to multivariate Shewhart charts: 1) the COT schemes easily detects the shift in the process mean; 2) the COT schemes contains the fast initial response feature; 3) quick detection of off-target conditions is achieved just by starting with S_0 equal to $h/2$; 4) the COT combines the property of CUSUM and multivariate Shewhart charts simultaneously.

The second procedure is developed by Crosier [7]: the vector-based CUSUM scheme is extended by replacing the scalar quantities of the univariate CUSUM by a vector. The multivariate CUSUM schemes is expressed as follows:

$$C_n = [(S_{n-1} + X_n - \mu_0)' \Sigma^{-1} (S_{n-1} + X_n - \mu_0)]^{1/2}.$$

then

$$S_n = 0 \qquad \text{for } C_n \leq k$$
$$S_n = (S_{n-1} + X_n - \mu_0)(1 - k/C_n) \quad \text{for } C_n > k$$

where we define $S_0 = 0$ and $k > 0$; when $[S_n' \Sigma^{-1} S_n]^{1/2} > h$, the multivariate CUSUM schemes will generate a signal.

Table 2.2: ARL for Bivariate Cusum Schemes with $k = .5, 1$ and 1.5 [7].

	$h = 5.50$	$h = 2.99$	$h = 1.87$
λ	$k = 0.5$	$k = 1$	$k = 1.5$
0	200	200	200
.5	28.8	48	78.7
1.0	9.35	11.0	18.4
1.5	5.94	5.08	7.14
2.0	4.20	3.48	3.72
2.5	3.26	2.51	2.36
3.0	2.78	2.08	1.69

Table 2.2 compares the ARL performance of Crosier's multivariate CUSUM for choices of $k = 0.5, 1$ and 1.5. From the table, it shows that the choice of $k = \lambda/2$ minimize the ARL. There are details of a method for selecting the interval value h in [7].

The ARL of COT schemes and multivariate CUSUM schemes depends on the noncentrality parameter λ. Both COT schemes and multivariate CUSUM schemes detect small shifts in the mean vector faster than multivariate Shewhart charts. Because multivariate CUSUM schemes allow observations in opposite directions from

the target mean to cancel each other, multivariate CUSUM gives faster detection than COT schemes.

Crosier[7] also compared the multivariate CUSUM schemes to the Woodall and Ncube method. For the same value of λ, the ARL's of the Woodall and Ncube method are usually larger than the vector-based CUSUM schemes. The difference between the two procedures is proportional to the standard error of the ARL of the vector-based CUSUM schemes.

A CUSUM procedure (MC2) similar to COT schemes is proposed by Pignatiello and Runger [28]. Define S_n as

$$S_n = \max\{0, S_{n-1} + T_n^2 - k\}$$

and

$$T_n^2 = (X_n - \mu_0)'\Sigma^{-1}(X_n - \mu_0)$$

where $S_0 = 0$ and k is chosen to be $p + 0.5\lambda^2$. Note that p is the number of variables and λ^2 is the measure of the distance from μ_1 to μ_0. The process is out of control if S_n exceeds an upper control limit h.

Pignatiello and Runger [28] provided the other version of the multivariate CUSUM (MC1) chart. The cumulative sums are:

$$S_i = \sum_{j=i-n_i+1}^{i} X_j$$

and

$$MC_i = \max\{0, (S_i'\Sigma^{-1}S_i)^{1/2} - kn_i\}$$

$$n_i = \begin{cases} n_{i-1} + 1 & \text{if } MC_{i-1} > 0 \\ 1 & \text{otherwise,} \end{cases}$$

with $MC_0 = 0$ and $k = 0.5\lambda$. An out-of-control signal will be generated when $MC_i > h$.

From the simulation in [28], MC1 shows superior performance for detecting a variety of shifts in the mean. The ARL performance of the MC1 scheme depends only

on the noncentrality parameter. The reference value k is chosen to be the average of the distance from the on-target state to the off-target state. For alternate choices of k, the performance of off-target will change.

2.3.3 Multivariate Exponentially Weighted Moving Average Control Chart

Like the MCUSUM, the MEWMA chart can also be extended from the univariate version to a multivariate process. Lowry, Woodall, Champ and Rigdom [20] presented the multivariate case of EWMA's

$$Z_i = RX_i + (I - R)Z_{i-1}$$

where $Z_0 = 0$ and $R = diag(r_1, r_2, ..., r_p)$, $0 < r_j \leq 1$, $j = 1, 2, \ldots, p$. The MEWMA gives an out-of-control signal when

$$T_i^2 = Z_i' \Sigma_{Z_i}^{-1} Z_i > h$$

where h is the in-control limit, and Σ_{Z_i} is the covariance matrix of Z_i.

$$\Sigma_{Z_i} = \sum_{j=1}^{i} var[R(I - R)^{i-j} X_j]$$

If $r_1 = r_2 = \cdots = r$, then the MEWMA vector can be modified via

$$Z_i = rX_i + (1 - r)Z_{i-1}$$

and then $\Sigma_{Z_i}^{\star}$ is given by

$$\Sigma_{Z_i}^{\star} = r\Sigma[1 - (1 - r)^{2i}]/(2 - r)$$

For the derivation of the covariance matrix for Z_i, see [20]. The MEWMA chart is equivalent to Hotelling's T^2 when $r = 1$.

The ARL of the MEWMA chart depends only on the noncentrality parameter. If the unequal weighted parameter r is applied, the ARL of performance only depends

on the direction of the shift. Smaller values of r are more powerful in detecting small shifts in the mean vector. This is analogous to the univariate case. The optimal value of r does not change too much as the number of variables changes. The value of h is obtained from simulation. Lowry [20] verifies that the logarithm of the in-control ARL is very close to a linear function of h.

In [20], they compare the ARL performance of Hotelling's T^2, Crosier's vector-based CUSUM [7], Pignatiello and Runger's multivariate CUSUM(MC1) [28] and the MEWMA chart with covariance matrix $\Sigma^{\star}_{Z_i}$. Table 2.3, 2.4 and 2.5 show ARL comparisons for p=2, 3 and 10.

There are two implicit assumptions in the ARL comparisons of these multivariate control charts. First, it is assumed that any shift from the in-control mean is to be detected as quickly as possible. Second, it is assumed that the ARL is a function of the noncentrality parameter λ.

From the tables, the on-target ARL is designed to be around 200. The comparison shows that Hotelling's T^2 has poor performance in detecting the shift of the mean unless the shift is large ($\lambda > 2$). The MEWMA chart is more effective in detecting the shift than the other three procedures.

In reality, most of the multivariate data does not follow the multivariate normal distribution. When the sample size is sufficiently large, the sample mean $\bar{X}$ is approximately multivariate normal. Researchers want to explore the performance of non-normality on the existing statistics quality control chart. Stoumbos and Sullivan [33] investigate whether the MEWMA chart can be designed to have excellent performance in cases where normality does not hold.

Let $X_1, X_2, ...$ be a sequence of p-variate process observations. The mean vector of a process observation is $\bar{X}$ and the constant covariance matrix is Σ_X. The on-target mean is μ_0, the noncentrality parameter λ is defined in Section 2.2. The accumulation vector has a different format compared to the original MEWMA chart. We have

$$Z_i = r(X_i - \mu_0) + (1 - r)Z_{i-1}$$

Table 2.3: ARL Comparisons for p=2.

	T^2	*Crosier*	*MC*1	*MEWMA*
λ	$h = 10.6$	$h = 5.50$	$h = 4.75$	$h = 8.79$
		$k = .50$	$k = .50$	$r = .10$
0	200	200	203	200
.5	116	28.8	31.3	25.3
1.0	42.0	9.35	9.28	7.76
1.5	15.8	5.94	5.23	4.07
2.0	6.9	4.2	3.69	2.59
2.5	3.5	3.26	2.91	1.89
3.0	2.2	2.78	2.40	1.50

where r is called the smoothing parameter and $0 < r \leq 1$. The initial value of Z_i is 0. Then the control statistic of T^2 is same as in MEWMA,

$$T_i^2 = Z_i' \Sigma_{Z_i}^{-1} Z_i$$

where the covariance matrix Σ_Z is as follows

$$\Sigma_Z = \frac{r}{2-r} \Sigma_X$$

The MEWMA chart generates a signal when $T_i^2 > h$. A large value of r makes the MEWMA chart detect a large shift effectively and the chart is effective at detecting a small shift when r is small. For a shift size, there is a unique optimal r value for the corresponding MEWMA chart. When $r = 1$, the MEWMA chart reduces to Hotelling's T^2 chart. For the choice of control limit h, Stoumbos and Sullivan [33] made simulations based on 100000 replications. Table 2.6 presents the h value for in-control $ARL = 200$ under p-variate normality.

Stoumbos and Sullivan [33] applied the MEWMA chart on multivariate t and multivariate gamma distribution, with the number of variables p equal to 2, 5 and 10. For multivariate t distribution, the degrees of freedom used were $v = 3, 20, 40$

Table 2.4: ARL Comparisons for p=3.

	T^2	*Crosier*	$MC1$	$MEWMA$
λ	$h = 12.85$	$h = 6.88$	$h = 5.48$	$h = 10.97$
		$k = .50$	$k = .50$	$r = .10$
0	200	200	203	202
.5	130	32.7	33.5	28.5
1.0	52.6	11.2	10.1	8.64
1.5	20.5	6.69	5.66	4.47
2.0	8.8	4.7	4.00	2.84
2.5	4.4	3.83	3.17	2.08
3.0	2.6	3.17	2.63	1.62

and 100; and for multivariate gamma distribution, the shape parameters used were $\alpha = 1, 16, 256, 1024$ with scale parameter $\beta = 1$.

For bivariate observations, using r in the range of $[0.022, 0.046]$ gives ARL performance for the bivariate t and gamma distributions similar to that for the bivariate normal. For applications in excess of 5 dimensions, a value of r smaller than 0.02 is similar to multivariate normal data. The choice of r value is critical for the MWEMA chart to detect a shift in a variety of distributions.

The MEWMA chart's performance is very similar to its performance under the multivariate normal, regardless of (i) the distribution of the process data, (ii) the size of the subgroups, (iii) the magnitude of the shift in μ, (iv) the direction of the shift in μ [33].

Table 2.5: ARL Comparisons for p=10.

	T^2	*Crosier*	$MC1$	$MEWMA$
λ	$h = 10.6$	$h = 5.50$	$h = 4.75$	$h = 8.79$
		$k = .50$	$k = .50$	$r = .10$
0	200	200	202	201
.5	162	43.2	43.9	44.3
1.0	92.8	18.6	12.6	12.8
1.5	44.7	11.8	7.66	6.28
2.0	20.6	8.79	5.66	3.96
2.5	9.9	3.26	4.55	2.79
3.0	5.2	2.78	3.84	2.14

2.4 Multivariate Nonparametric Process Control Charts

2.4.1 Data Depth

For most multivariate data sets from the statistical experiment we do not know if they follow a certain distribution. Applying the charts mentioned above may fail to detect the real change in the process. The use of nonparametric analysis becomes attractive in the context of quality control. Multivariate nonparametric process control has been studied by many researchers recently. One approach is to try to transform the sample distribution to a multivariate normal and then apply the existing methods. However, it is often hard to find the transformation. Data depth plays an important role in nonparametric methods for multivariate statistics.

Liu [17] investigated the basic properties of the simplicial depth and developed three multivariate nonparametric control charts. Burr [3] proposed a revised definition of simplicial depth. Based on the simplicial depth, Messaoud [25] used the rMEWMA

Table 2.6: h value for in-control $ARL = 200$ under p-variate Normality.

r	$p = 1$	$p = 2$	$p = 3$	$p = 5$	$p = 10$
1.0	7.8825	10.5966	12.8335	16.7447	25.1763
0.4641	7.6756	10.4021	12.6437	16.5496	24.9774
0.2154	7.0300	9.7373	11.9583	15.8310	24.1644
0.1	6.0218	8.6345	10.7748	14.5397	22.6501
0.04641	4.7840	7.1971	9.1990	12.7507	20.4592
0.02154	3.4645	5.5498	7.3170	10.5173	17.5800
0.01	2.2484	3.8722	5.3035	7.9511	13.9680
0.004641	1.3049	2.4019	3.4137	5.3398	9.9079
0.002154	0.6905	1.3349	1.9503	3.1509	6.0950
0.001	0.3425	0.6842	1.0174	1.6716	3.3252

Control chart to detect the multivariate sequence change.

A data depth is a measure of how deep or how central a given point $x \in R^d$ is with respect to a multivariate distribution F [19]. Let $X = \{X_1, \ldots, X_m\}$ be a random sample from F. There are several definitions of data depth, such as Mahalanobis's depth, Simplicial depth, Tukey's depth, and Majority depth.

1. Mahalanobis's depth (M_hD) [21]: Let $d(x, \mu_F)$ be the Mahalanobis distance between x and μ_F (the mean of F); that is, $d(x, \mu_F) = (x - \mu_F)'\Sigma_F^{-1}(x - \mu_F)$. We define

$$M_hD(F; x) = [1 + d(x, \mu_F)]^{-1}$$

and its sample version is

$$\widehat{M_hD}(F; x) = [1 + (x - \bar{X})'S^{-1}(x - \bar{X})]^{-1}$$

where Σ_F and S are the covariance matrix and the sample covariance matrix respectively, and $\bar{X}$ is the sample mean.

2. Tukey's depth (TD) [34]: For a point x, the Tukey's depth at x is defined as

$$TD(F;x) = \inf_H \{F(H) : \text{H is a closed half space containing x}\}.$$

The sample version of $TD(F;x)$ is defined by replacing F by F_m, where F_m is the empirical distribution of the sample $X_1, \ldots, X_m$. When $p = 1$, $TD(F;x) = \min \{F(x), 1 - F(x-)\}$.

3. Majority depth (M_jD) of x w.r.t. F is defined to be [32]

$$M_jD(F;x) = P_F\{x \text{ is in a major side determined by } (X_1, \ldots, X_d)\}.$$

A major side is the half-space bounded by the hyperplane containing $(X_1, \ldots, X_d)$ which has probability greater than or equal to 0.5. The sample version is $M_jD(F_m;x)$. Note that when $d = 1$, $M_jD(F;x) = 1/2 + \min\{F(x), 1 - F(x-)\}$.

Simplicial depth was introduced by Liu [16], and its basic properties were investigated.

Definition 2.4.1 Simplicial Data Depth

Given a multivariate distribution G in R^d, and $Y_1, \ldots, Y_{d+1}$ are independent observations from G. The simplicial data depth of y w.r.t to G is defined as

$$SD_G(y) = P_G\{y \in S[Y_1, \ldots, Y_{d+1}]\}$$

where $S[Y_1, \ldots, Y_{d+1}]$ is the simplex whose vertices are $Y_1, \ldots, Y_{d+1}$.

The sample simplicial depth of y is defined as follows:

$$SD_{G_m}(y) = \binom{m}{d+1}^{-1} \sum_{(*)} I(y \in S[Y_{i_1}, \ldots, Y_{i_{d+1}}]),$$

which measures how deep y within the data cloud $\{Y_1, \ldots, Y_m\}$. Here I is the indicator function and $(*)$ runs over all possible subsets of $Y_1, \ldots, Y_{d+1}$ of size $d + 1$.

Lemma 2.4.2 (Liu [18])

As $m \to \infty$, $SD_{G_m}(\cdot)$ converges to $SD_G(\cdot)$ uniformly.

We illustrate the simplicial data depth for $d = 2$. In the bivariate case, a simplex is a triangle. The sample simplicial data depth of y w.r.t $\{Y_1, \ldots, Y_m\}$is defined as

$$SD_{G_m}(y) = \binom{m}{3}^{-1} \sum_{(*)} I(y \in \Delta\ (Y_{i_1}, Y_{i_2}, Y_{i_3})).$$

where Δ represent the triangle with the vertices $Y_{i_1}, Y_{i_2}, Y_{i_3}$ and $(*)$ runs over all possible triangles of the data cloud of $\{Y_1, \ldots, Y_m\}$.

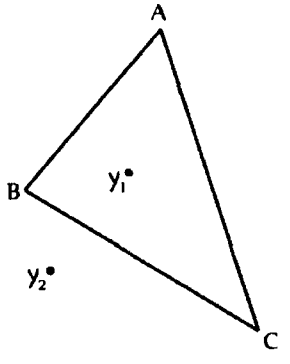

Figure 2.3: Computation of bivariate simplicial data depth.

Figure 2.3 shows the simplicial data depth of y_1 w.r.t $S = A, B, C$ is 1 and the simplicial data depth of y_2 is 0. Let

$$r_G(y) = P\{SD_G(Y) < SD_G(y)\}.$$

The empirical estimate of $r_G(y)$ is

$$r_{G_m}(y) = \sharp\{Y_i | SD_{G_m}(Y_i) < SD_{G_m}(y), i = 1, ..., m\}/m$$

Proposition 2.4.3 (Liu, Singh and Teng[19], Proposition 1)
Assume that $F = G$, $X \sim F$ and $Y_1, \ldots, Y_m \sim G$. Let $U[0, 1]$ denote the uniform

distribution on the unit interval $[0,1]$, *and let the notation* $\rightarrow^L$ *stand for the convergence in law. If* $SD_G(X)$ *has a continuous distribution, then*
(1) $r_G(X) \sim U[0,1]$, *and*
(2) as $m \rightarrow \infty$, $r_{G_m}(X) \rightarrow^L U[0,1]$ *along almost all* $Y_1, \ldots, Y_m$ *sequences.*

Proof: (1) By definition,

$$\begin{aligned} r_G(x) &= P\{SD_G(Y) < SD_G(x)\} \\ &= F_{SD_G(Y)}(SD_G(x)) \end{aligned}$$

Therefore, we conclude that

$$\begin{aligned} r_G(X) &= F_{SD_G(Y)}(SD_G(X)) \\ &\sim U[0,1] \end{aligned}$$

(2) First,

$$\begin{aligned} r_{G_m}(x) &= \sharp\{Y_i | SD_{G_m}(Y_i) < SD_{G_m}(x), i = 1, \ldots, m\}/m \\ &= \frac{1}{m}\sum_{i=1}^{m} I_{\{SD_{G_m}(Y_i) - SD_{G_m}(x) < 0\}} \end{aligned}$$

Now define

$$\begin{aligned} r_G^*(x) &= \sharp\{Y_i | SD_G(Y_i) < SD_G(x), i = 1, \ldots, m\}/m \\ &= \frac{1}{m}\sum_{i=1}^{m} I_{\{SD_G(Y_i) - SD_G(x) < 0\}} \end{aligned}$$

where I is the indictor function. By Glivenko-Cantelli, $r_G^*(x) \rightarrow r_G(x)$ uniformly in x when $m \rightarrow \infty$.

Since,

$$\begin{aligned} & SD_{G_m}(Y_i) - SD_{G_m}(x) < 0 \\ \Leftrightarrow\ & SD_{G_m}(Y_i) - SD_{G_m}(x) + SD_G(x) - SD_G(x) < 0 \\ \Leftrightarrow\ & -SD_G(x) < SD_{G_m}(x) - SD_G(x) - SD_{G_m}(Y_i) \\ \Leftrightarrow\ & SD_G(Y_i) - SD_G(x) < SD_{G_m}(x) - SD_G(x) + SD_G(Y_i) - SD_{G_m}(Y_i) \end{aligned}$$

On the other hand, as $m \to \infty$, $SD_{G_m}(\cdot)$ converges to $SD_G(\cdot)$ uniformly by lemma 2.4.2. Therefore, $I_{\{SD_{G_m}(Y_i)-SD_{G_m}(x)<0\}} = I_{\{SD_G(Y_i)-SD_G(x)<0\}}$ as $m \to \infty$. Hence $r_{G_m}(x) \to r_G^*(x) \to r_G(x) \sim U[0,1]$.

Q.E.D.

Burr, Rafalin and Souvaine proposed a modification to compute the revised simplicial data depth [3].

Definition 2.4.4 Revised Simplicial Data Depth [3]

Given a data set $S = \{X_1, \ldots, X_m\}$, the simplicial depth of x is the average of the fraction of closed simplices containing x and the fraction of open simplices containing x, that is

$$SD_{BRS}(x) = \frac{1}{2}\binom{m}{d+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{d+1} \le m} \{\mathrm{I}(x \in S[X_{i_1}, X_{i_2}, X_{i_3}]) + \mathrm{I}(x \in \mathrm{int}(S[X_{i_1}, X_{i_2}, X_{i_3}]))\}.$$

where $\mathrm{int}(S[X_{i_1}, X_{i_2}, X_{i_3}])$ *refers to the open relative interior of* $S[X_{i_1}, X_{i_2}, X_{i_3}]$. *Equivalently, the formula can be changed as,* $SD_R(x) = \rho(x) + 1/(2\sigma(x))$, *where* $\rho(x)$ *is the number of simplices with data points as vertices which contain* x *in their open interior, and* $\sigma(x)$ *is the number of simplices with data points as vertices which contain* x *in their boundary.*

2.4.2 Control Charts Based on Data Depth

Data depth helps to reduce multivariate measurement to a univariate index. Many researchers have developed nonparametric statistical control charts based on it. Liu proposed three control charts: the r chart, Q chart and S chart based on simplicial depth which can be viewed as analogues of the X chart, $\bar{X}$ chart, and CUSUM chart [17].

Assume that the observations $Y_1, \ldots, Y_m$ and $X_1, \ldots, X_n$, our main concern is a possible shift in the mean X_i's. Compute $r_G(X_1), \ldots, r_G(X_n)$ if $Y_1, \ldots, Y_m$ are from

the distribution G (or $r_{G_m}(X_1), \ldots, r_{G_m}(X_n)$, if $Y_1, \ldots, Y_m$ are available but not G). The r chart is a plot of $r_G(X_i)$'s (or $r_{G_m}(X_i)$'s) against observation i $(i = 1, \ldots, n)$ with control limit h. When $r_G(X_i)$ (or $r_{G_m}(X_i)$) falls below h, the process is out-of-control.

Proposition 2.4.3 indicates that $r_G(X) \sim U[0,1]$ if $F = G$ and the expected value of $r_G(X)$ is 0.5. If the number of observations is large, the expected value of $r_{G_m}(X)$ is 0.5 almost surely for all sequences $Y_1, \ldots, Y_m$. Let $h = \alpha$, where α is the significance level for the following hypotheses

$$H_0 : F = G$$

$$H_\alpha : \text{there is a location shift from G to F}$$

Note that F is the off-target distribution.

The Q chart is similar to the $\bar{X}$ chart and plots the average of $r_G(X_i)$'s (or $r_{G_m}(X_i)$'s). Let the subgroup size is q, and define the average of $r_G(X_i)$'s and $r_{G_m}(X_i)$'s the following

$$Q(G, F_q) = \frac{1}{q} \sum_{i=1}^{q} r_G(X_i)$$

$$Q(G_m, F_q) = \frac{1}{q} \sum_{i=1}^{q} r_{G_m}(X_i)$$

The chart plots $Q(G, F_q^1), \ldots, Q(G, F_q^n)$ or $(Q(G_m, F_q^1), \ldots, Q(G_m, F_q^n))$, and the process is out of control when $Q(G, F_q)$ (or $Q(G_m, F_q)$) is less than h. The choice of h depends on the size of q. When q is large, h should be $0.5 - z_\alpha(12q)^{-1/2}$ for plotting $Q(G, F_q)$'s and $0.5 - z_\alpha\sqrt{\frac{1}{12}[(\frac{1}{m}) + (\frac{1}{q})]}$ for plotting $Q(G_m, F_q)$'s. If q is quite small, h is exactly $(q!\alpha)^{1/q}/q$. The details about the design of h, see [17].

The third chart proposed by Liu [17] is the S chart, which is similar to the idea of the CUSUM chart. It plots the values $S_n(G)$ and $S_n(G_m)$ defined by

$$S_n(G) = \sum_{i=1}^{n} \left[r_G(X_i) - \frac{1}{2} \right]$$

$$S_n(G_m) = \sum_{i=1}^{n} \left[r_{G_m}(X_i) - \frac{1}{2} \right]$$

Proposition 2.4.5 (Liu[17], Proposition 3.4)

(1) $\frac{\sqrt{12}S_n(G)}{\sqrt{n}} \rightarrow^L \mathrm{N}(0,1)$ *as* $n \rightarrow \infty$, *and*

(2) $\frac{\sqrt{12}S_n(G_m)}{n\sqrt{(1/m+1/n)}} \rightarrow^L \mathrm{N}(0,1)$, *as* $min(m,n) \rightarrow \infty$.

Proposition 2.4.5 implies that the control limit h for $S_n(G)$ is $-(z_\alpha(n/12)^{1/2})$ and for $S_n(G_m)$ is $-(z_\alpha\sqrt{n^2[(1/m)+(1/n)]/12})$.

Messaoud [25] introduced the $rMEWMA$ chart which is based on the sequential rank of simplicial data depth. Let $X_1, \ldots, X_t$ be independent random observation from the control process, $RS = (X_{t-m+1}, \ldots, X_t)$ denotes a reference sample which includes the m most recent observations from the process at time t. The simplicial data depth $SD(RS, X_i)$, $i = t-m+1, ..., t$, is calculated w.r.t. RS.

Let R_t^* represent the sequential rank of $SD(RS, X_t)$ among $SD(RS, X_{t-m+1}), \ldots,$ $SD(RS, X_t)$. It is defined by

$$R_t^* = 1 + \sum_{i=t-m+1}^{t} I(SD(RS, X_t) > SD(RS, X_i))$$

where I is the indicator function. When in control, R_t^* is uniformly distributed on the $1, 2, \ldots, m$. Hackl and Ledolter [11] proposed the standardized sequential rank R_t^m

$$R_t^m = \frac{2}{m}\left(R_t^* - \frac{m+1}{2}\right)$$

which is uniformly distributed on $\{\frac{1}{m}-1, \frac{3}{m}-1, \ldots, 1-\frac{1}{m}\}$ with $\mu_{R_t^m} = 0$ and variance $\sigma_{R_t^m} = \frac{m^2-1}{3m^2}$, when the process is in control.

The control statistic T_i of rMEWMA chart is computed in the following way:

$$T_i = min\{B, (1-\lambda)T_{t-1} + \lambda R_t^m\}$$

where $0 < \lambda \leq 1$ is a smoothing parameter and $B > 0$ is a reflecting boundary. The process generates an out-of-control signal when $T_i < h$, where $h < 0$ is a lower control limit.

The choices of the combinations of λ, h, B and m ensures the performance meets statistical criteria. Crowder [8] proposed the integral equation to approximate the in-control ARL. Let $L(u)$ be the ARL of rMEWMA chart. Then the integral equation is given by [25]

$$L(u) = 1 + L(B)Pr\left(q \geq \frac{B-(1-\lambda)u}{\lambda}\right) + \int_h^B L((1-\lambda)u + \lambda q)dF(q)$$

where $T_0 = u$ and $F(q)$ is the cumulative distribution of $q = R_t^m$, and R_t^m are uniformly distributed on the m points $\{1/m - 1, 3/m - 1, \ldots, 1 - 1/m\}$.

With a sufficiently large reference sample size, the rMEWMA is able to detect the change quickly. The relationship between the smoothing parameter λ and the reference sample size m also affects the performance of rMEWMA chart.

A numerical illustrative example of rMEWMA is given in [25]. Table 2.7 shows 20 observations from the bivariate normal distribution with mean vector $\mu = (0, 0)'$ and covariance matrix equal to identity matrix. The smoothing parameters are $\lambda = 0.2$, $h = -0.435$ and $B = h$. The first ten observations are used as the reference samples.

All the nonparametric control charts in the above are based on the simplicial depth, therefore the computation of the simplicial depth is the requirement for these charts. For the simplicial computation in R^2, Rousseeuw and Ruts [30] pointed out an effective algorithm for calculating simplicial depth which has a time complexity O $(n \log n)$, where n is the number of observations. When the dimension increases to 3, Cheng and Ouyang [6] showed an $O(n^3)$ time algorithm. If the dimension is larger than 3, there are no known algorithms faster than the definition of simplicial depth because computing the simplicial depth for R^d $(d \geqslant 4)$ is a time-consuming job. Because of the difficulty of computing the simplicial depth, the nonparametric control charts based on simplicial depth are restricted to the bivariate case.

2.4.3 A Rank-Based Multivariate CUSUM

Qiu and Hawkins proposed a multivariate distribution free CUSUM procedure which is based on the cross-sectional antiranks of the measurements [29]. Let $X(i) =$

$(X_1(i), X_2(i), \ldots X_p(i))'$ denote a p-dimensional observation at the ith time point in the process. When the process is in-control and its mean vector is $\mu = (\mu_1, \ldots, \mu_p)' = 0$. When the process is out of control, the mean vector μ has a shift from 0. The null hypothesis involved in multivariate process control is,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = 0.$$

which is equivalent to the combination of the following two hypotheses:

$$H_0^{(1)} : \mu_1 = \mu_2 = \cdots = \mu_p$$

$$H_0^{(2)} : \sum_{j=1}^{p} \mu_j = 0.$$

Define the antirank vector $A(i) = (A_1(i), A_2(i), \ldots A_p(i))'$, where $A(i)$ is a permutation of $(1, 2, \ldots, p)'$ such that $X_{A_1(i)}(i) \leq X_{A_2(i)}(i) \leq \cdots \leq X_{A_p(i)}(i)$ are the order statistics of $\{X_j(i), j = 1, 2, \ldots, p\}$. For $1 \leq j \leq p$, define the indicator function $\xi_{1,j}(i)$

$$\xi_{1,j}(i) = \begin{cases} 1 & \text{if } A_1(i) = j \\ 0 & \text{otherwise,} \end{cases}$$

and $\xi_1(i) = (\xi_{1,1}(i), \xi_{1,2}(i), \ldots, \xi_{1,p})'$. Under $H_0^{(1)}$, for $j = 1, 2, \ldots, p$, $E(\xi_{1,j}(i)) = g_{1,j}$. The probability distribution of $A_1(i)$ is $\{g_{1,j}, j = 1, 2, \ldots, p\}$. When the mean vector is shifted, the corresponding distribution of $A(i)$ is denoted by $\{g^*_{1,j}, j = 1, 2, \ldots, p\}$. The CUSUM procedure is the following

$$\begin{cases} S_n^{(1)} = 0 & \\ S_n^{(2)} = 0 & \text{if } C_n \leq k_1 \\ S_n^{(1)} = (S_{n-1}^{(1)} + \xi_1(n))(C_n - k_1)/C_n & \\ S_n^{(2)} = (S_{n-1}^{(2)} + g_1)(C_n - k_1)/C_n & \text{if } C_n > k_1 \end{cases}$$

and

$$\begin{aligned} C_n &= [(S_{n-1}^{(1)} - S_{n-1}^{(2)}) + (\xi_1(n) - g_1)]' \\ &\times \operatorname{diag}((S_{n-1,1}^{(2)} + g_{1,1})^{-1}, \ldots, (S_{n-1,p}^{(2)} + g_{1,p})^{-1}) \\ &\times [(S_{n-1}^{(1)} - S_{n-1}^{(2)}) + (\xi_1(n) - g_1)], \end{aligned}$$

where $g_1 = (g_{1,1}, g_{1,2}, \ldots, g_{1,p})'$, $S_0^{(1)} = S_0^{(2)} = 0$, and $k_1 \geq 0$. Define

$$y_n = (S_n^{(1)} - S_n^{(2)})' \operatorname{diag}(1/S_{n,1}^{(2)}, \ldots, 1/S_{n,p}^{(2)})(S_n^{(1)} - S_n^{(2)})$$

Then

$$y_n > h_1$$

signals a shift, where h_1 is a control limit.

The parameter k_1 is chosen from the interval

$$\left[0, \max_{\iota=1}^{p} \frac{\sum_{j \neq \iota} g_{1,j}}{g_{1,\iota}}\right).$$

If k_1 is chosen outside the interval, the in-control ARL could not be achieved. The control limit h_1 is determined by simulation. For explanatory details about k_1 and h_1, see [29].

Table 2.7: Illustrative example of rMEWMA chart

i	$x_{i,1}$	$x_{i,2}$	$SD(RS, X_t)$	Q_t^*	Q_t^m	T_t
1	0.13	-0.09				
2	1.67	0.73				
3	1.00	-1.28				
4	-2.40	-0.68				
5	-0.04	0.89				
6	-0.02	-1.30				
7	-0.67	0.18				
8	0.83	-0.55				
9	-0.64	0.01				
10	-0.67	-0.83	0.250	8	0.5	0.100
11	0.61	-0.37	0.317	10	0.9	0.260
12	-0.29	-0.92	0.317	10	0.9	0.388
13	-0.58	0.06	0.342	10	0.9	0.435
14	0.05	-0.75	0.292	9	0.7	0.435
15	-0.14	1.48	0.150	3	-0.5	0.248
16	-0.21	-0.26	0.375	10	0.9	0.378
17	-0.14	-2.54	0.150	3	-0.5	0.203
18	0.58	-0.04	0.150	3.5	-0.4	0.082
19	-0.23	0.72	0.250	8	0.5	0.166
20	1.58	-0.39	0.150	2.5	-0.6	0.013

Chapter 3

Voronoi Rank Cusum

3.1 Voronoi Diagram

Descartes was the first scholar to introduce the idea of the Voronoi diagram. Back in the 17th century, he claimed that the solar system consists of vertices. He showed a decomposition of space into convex regions; however, he did not give a mathematical definition of these regions. The Russian mathematician Voronoi was the first who formally introduced the concept. Voronoi [39] defined the Voronoi diagram by specifying that any two point sites in the two-dimensional plane be connected if their Voronoi regions have a boundary in common. He extended the study of the structure to the n-dimensional case in 1908. The structure has been called a Voronoi diagram or Voronoi tesselation. Delaunay [9] conceived of the same structure and used it for the study of quadratic forms.

Definition 3.1.1 *Let a given set S consist of $n \geqslant 3$ points, such as $a, b, c, \ldots$ in d-dimensional Euclidean space E^d. The points of S are called sites. The Voronoi diagram of S splits E^d into regions with one region for each site. Each of the points in a region corresponding to a given site $s \in S$ are closer to the site s than any other site in $s' \in S$.*

Let $d(a,b)$ denote the Euclidean distance between two points $a = (a_1, a_2)$ and $b = (b_1, b_2)$,

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

The line segment from a to b is represented by $\overline{ab}$ and the closure of a set A will be denoted by $\overline{A}$. The mathematical definition of Voronoi diagram is the following:

Definition 3.1.2 Voronoi Diagram (Aurenhammer and Klein [2])

For a, $b \in S$ let

$$B(a,b) = \{x | d(a,x) = d(b,x)\}$$

be the bisector of a and b. $B(a,b)$ is the perpendicular line through the center of the line segment $\overline{ab}$. It separates the halfplane

$$D(a,b) = \{x | d(a,x) < d(b,x)\}$$

containing a from the halfplane $D(b,a)$ containing b. We call

$$VR(a,S) = \bigcap_{b \in S, b \neq a} D(a,b)$$

the Voronoi region of a with respect to S. Finally, the Voronoi diagram of S is defined by

$$V(S) = \bigcup_{a,b \in S, a \neq b} \overline{VR(a,S)} \cap \overline{VR(b,S)}.$$

Each Voronoi region $VR(a,S)$ is the intersection of $n-1$ open halfplanes containing the site a. Therefore, different Voronoi regions are disjoint. The common boundary of two Voronoi regions is called a Voronoi edge.

Figure 3.1 shows a two-dimensional Voronoi diagram. There is another way of defining the Voronoi diagram $V(S)$. It illustrates that the Voronoi regions form a decomposition of the plane.

Let x represent an arbitrary point in the plane. At x, we make a circle C centered at x and its radius grows starting from 0. At some stage, the expanding radius will hit other points from the set S.

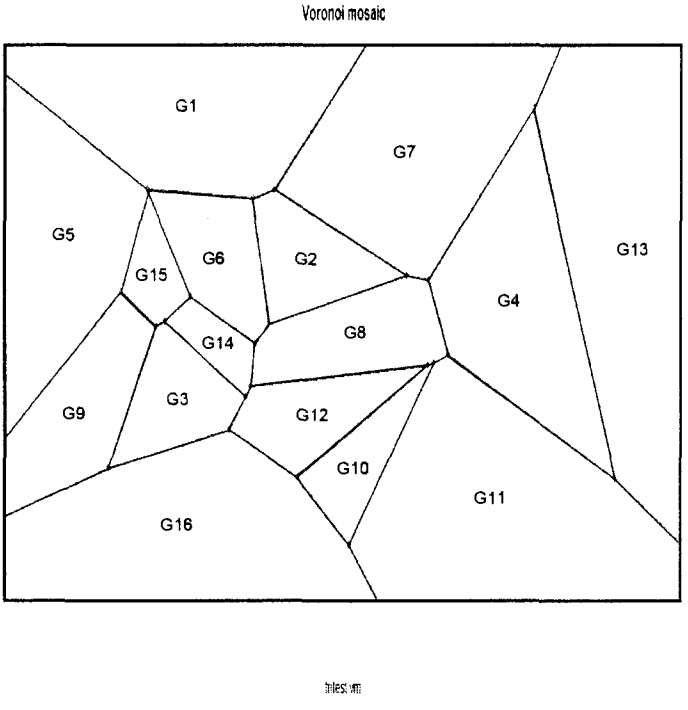

Figure 3.1: A sample Voronoi diagram.

Lemma 3.1.3 (Aurenhammer and Klein [2])

If the circle C expanding from x hits exactly one point a, then x belongs to $VR(a, S)$. If C hits exactly two points, a and b, then x is an interior point of a Voronoi edge separating the regions of a and b. If C hits three or more points simultaneously, then x is a Voronoi vertex adjacent to those regions whose sites have been hit.

Given a new point p, which point in a points set S is the closest to p? This problem can be solved by forming the Voronoi diagram for the set S. The point p will fall in the Voronoi region corresponding to the site in S nearest to p, i.e. the point $s \in S$ satisfying $d(p, s) = \min\{d(p, s'); s' \in S\}$ where $d(p, s)$ is the Euclidean distance.

The Voronoi diagram and its geometric dual have proved to be one of the most important data structures in Computational Geometry. the Voronoi diagram and its

applications has been used in a variety of fields, including biology, crystallography, geology, metallurgy, mesh generation, curve and surface reconstruction, meteorology, mathematics, robotics, geography, and marketing. We use the properties of the Voronoi diagram to develop the Voronoi Cusum in the next section.

3.2 Voronoi Cusum

The Voronoi tesselation of n points in $\mathbf{R}^d$ denoted by $X_1, X_2, \ldots, X_n$ is a set of n regions $G_1, G_2, \ldots, G_n$ where region G_i denotes the set points closest to arrival i; i.e. to X_i. If a new point X_{n+1} falls in G_i, let $R_{n+1}^{(1)} = i$. More generally if the point X_k in region G_k is the r^{th} closest point to X_{n+1} let $R_{n+1}^{(r)} = k$. In other words the ranks $\{R_{n+1}^{(r)}, r = 1, \ldots, c\}$ are the indices of the c points of $X_1, \ldots, X_n$ closest to X_{n+1}. Also define $U_{n+1}^{(r)} = R_{n+1}^{(r)}/(n+1)$.

Conditioned on the $\{X_i : i \in \{1, \ldots, n\}\}$ the distribution of $U_{n+1}^{(r)}; r = 1, \ldots, c$ of course depends on the volume of the $\{G_i : i \in \{1, \ldots, n\}\}$ and on the distribution of the X's but, via simulation, we observe the variables are asymptotically independent and uniform. We prove the following Lemma for $r = 1$. We conjecture it is true for $r > 1$.

Lemma 3.2.1 (Mao, McDonald and Zarepour [23])

Let X_i be a sequence of independent identically distributed random variables with continuous bounded density for the continuous components.

$$\lim_{n\to\infty} P(U_{n+1}^{(r)} \leq t_r; r = 1, \ldots, c | \{X_i : i \in \{1, \ldots, n\}) = \prod_{r=1}^{c} t_r \text{ in distribution.}$$

Proof:

Let f denote the density of i.i.d. observations $X_1, X_2, \cdots, X_n$.

$$P(U_{n+1}^{(1)} \leq t | \{X_i : i \in \{1, \ldots, n\}) = \sum_{i \leq [nt]} \int_{G_i} f(x)dx.$$

Clearly $E\int_{G_i} f(x)dx = 1/n$ by symmetry. It suffices to prove that

$$E\left(\sum_{i\leq[nt]}\left(\int_{G_i} f(x)dx - \frac{1}{n}\right)\right)^2 \to 0.$$

as $n \to \infty$.

First of all, as $n \to \infty$, $E\left(\max_{1\leq i\leq n}\int_{G_i} f(x)dx\right) \to 0$. For any ϵ there exists an L such that $\int_{[-L,L]^d} f(z)dz > 1-\epsilon$. Now divide $[-L,L]^d$ into a grid with grid lengths δ so $L = k\delta$ giving a total of k^d blocks $B_1, \ldots B_{k^d}$. Pick δ sufficiently small that $\int_{B_i} f(z)dz < \epsilon$ for each i. Almost surely there exists an N such that for $n \geq N$ at least one point of $X_1, X_2, \cdots, X_n$ lies in each block. Now consider the Voronoi cell G_i around the point X_i. Suppose X_i has $(2d+1)^d$ contiguous blocks surrounding the point X_i (including the block B_i containing X_i). View this as a cube made of d blocks above B_i and d blocks below B_i in each dimension. G_i is contained in the union of these $(2d+1)^d$ contiguous blocks because the boundary of the associated cube cannot be in G_i. This is so because the distance of X_i from this boundary is at least $d\delta$. However any point on this boundary is no more than $\sqrt{d}\delta$ from some point X_j in the block touching this boundary so the boundary is closer to X_j than X_i. Consequently

$$\int_{G_i} f(z)dz \leq \sum_{i\in(2d+1)^d \text{ contiguous blocks}} \int_{B_i} f(z)dz < \epsilon(2d+1)^d$$

and this is arbitrarily small. If X_i is near the boundary of this cube or even outside the cube then G_i is contained in at most $(2d+1)^d$ blocks plus the complement of the cube so at worst $\int_{G_i} f(z)dz < \epsilon((2d+1)^d + 1)$. Either way $\int_{G_i} f(z)dz$ is arbitrarily small for n large so we conclude the $\max_{1\leq i\leq n}\int_{G_i} f(x)dx \downarrow 0$ almost surely and $E\left(\max_{1\leq i\leq n}\int_{G_i} f(x)dx\right) = o(n)$.

Now consider the sum

$$\begin{aligned}\sum_{i=1}^n\left(\int_{G_i} f(x)dx\right)^2 &\leq \max_i\left(\int_{G_i} f(x)dx\right)\left(\sum_{i=1}^n\left(\int_{G_i} f(x)dx\right)\right)\\ &= \max_i\int_{G_i} f(x)dx\end{aligned}$$

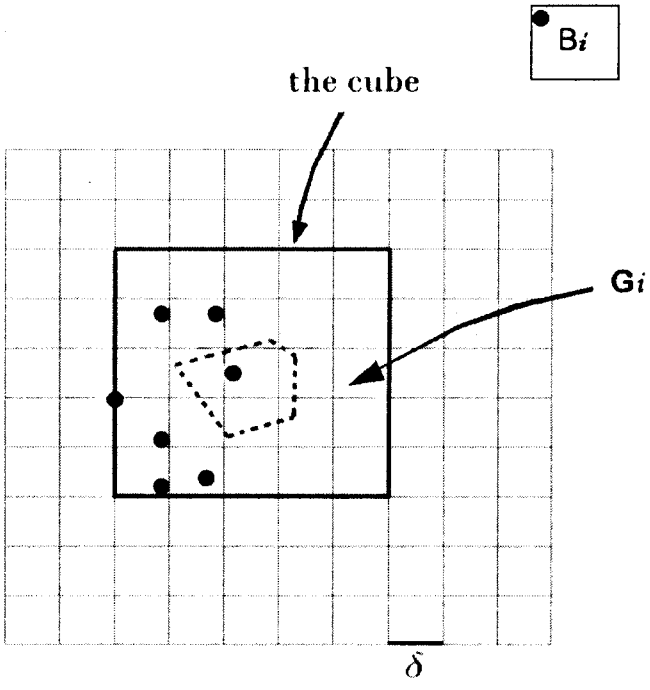

Figure 3.2: Showing the Voronoi region G_i is in the cube for $d = 2$.

Hence, by symmetry, $E\left(\int_{G_i} f(x)dx\right)^2 = E\left(\max_i \int_{G_i} f(x)dx\right)/n$ so $E\left(\int_{G_i} f(x)dx\right)^2 = o(n)/n$.

Again, by symmetry,

$$\begin{aligned} 0 &= E\left(\sum_{1\le i\le n}\left(\int_{G_i} f(x)dx - \frac{1}{n}\right)\right)^2 \\ &= nE\left(\int_{G_1} f(x)dx - \frac{1}{n}\right)^2 + 2n(n-1)E\left(\int_{G_1} f(x)dx - \frac{1}{n}\right)\left(\int_{G_2} f(x)dx - \frac{1}{n}\right). \end{aligned}$$

Hence,

$$E\left(\int_{G_1} f(x)dx - \frac{1}{n}\right)\left(\int_{G_2} f(x)dx - \frac{1}{n}\right) = \frac{1}{2(n-1)}E\left(\int_{G_1} f(x)dx - \frac{1}{n}\right)^2.$$

Finally,

$$
\begin{aligned}
& E\left(\sum_{i\le[nt]}\left(\int_{G_i} f(x)dx-\frac{1}{n}\right)\right)^2 \\
&= \sum_{i\le[nt]} E\left(\int_{G_i} f(x)dx-\frac{1}{n}\right)^2+\sum_{i,j\le[nt],i\neq j} E\left(\int_{G_i} f(x)dx-\frac{1}{n}\right)\left(\int_{G_j} f(x)dx-\frac{1}{n}\right) \\
&= [nt]E\left(\int_{G_1} f(x)dx-\frac{1}{n}\right)^2+2[nt]([nt]-1)E\left(\int_{G_1} f(x)dx-\frac{1}{n}\right)\left(\int_{G_2} f(x)dx-\frac{1}{n}\right) \\
&= \left([nt]+\frac{[nt]([nt]-1)}{n(n-1)}\right)E\left(\int_{G_1} f(x)dx-\frac{1}{n}\right)^2 \\
&\le 2\left([nt]+\frac{[nt]([nt]-1)}{(n-1)}\right)\left(E\left(\int_{G_1} f(x)dx\right)^2+\left(\frac{1}{n}\right)^2\right) \\
&= 2\left([nt]+\frac{[nt]([nt]-1)}{(n-1)}\right)\frac{o(n)}{n}=o(n).
\end{aligned}
$$

This gives our result for $r=1$.

Q.E.D

To extend the above proof to $r=2$ for instance we would first have to look at the region T_i of points inside region G_i where X_i is the second closest point among $\{X_1,\ldots,X_n\}$. By symmetry $ET_i = 1/(N(N-1))$ where N is defined in the proof of lemma. We can then proceed as above. We won't pursue this because there is a bigger issue; the sequence $U_n^{(r)}; r=1,\ldots,c$ indexed by n is *not* asymptotically independent. To see why notice that if $R_n^{(1)}=i$ and $R_n^{(2)}=j$ for some n then if for some $m>n$, $R_m^{(1)}=i$ again then there is a good chance that $R_m^{(2)}=j$ again.

Score the approximate uniforms $U_{n+1}^{(r)}$ into approximate normals $N_{n+1}^{(r)}$ by $N_{n+1}^{(r)} = \Phi^{-1}(U_{n+1}^{(r)}), r=1,\ldots,c)$ where Φ is the cumulative distribution of a standard normal. Then define

$$M_{n+1}^c = (N_{n+1}^{(1)}+\cdots+N_{n+1}^{(c)})/c$$

By experiment, $c=\min\{9,\lfloor\sqrt{n}\rfloor\}$ is a good choice. Then add the new point to the tessillation at the center of the new region G_{n+1}. We don't actually calculate the

Voronoi region since we only compare distances from the new point to the $X_1, \ldots, X_n$. If the sequence of points $X_1, X_2, \ldots, X_n, X_{n+1}$ is i.i.d. then the M^c_{n+1} is approximately a standard normal. We can use the M's to construct a standard normal Cusum:

$$C_{n+1} = \max\{0, C_n + M^c_{n+1} - k\}$$

The procedure will signal an alarm if $C_{n+1} \geq h$.

Taking the average M^c_{n+1} over c neighbours instead of just one has an advantage. We found we get much better power because at first the off-target points are rather sparse among the on-target points. Averaging makes sure the off-target points nearby are properly weighted. There is however an issue when n is fairly small. In the on-target situation the Cusum is actually smaller than that formed by real i.i.d. standard normals and the average on-target run length is longer (not a bad thing). However the procedure has longer off-target run lengths for change points less than 30 or so. To see why notice that C_{10} is necessarily zero even in an off-target situation! The reason is that the nine closest points are necessarily in the regions $\{1, 2, \ldots, 9\}$ so $M^9_{10} = 0$. For small values of n the $\{U^{(r)}_{n+1} : r = 1, \ldots, c\}$ are not approximately i.i.d. uniforms. Moreover the sequence of M^c_n will not be independent since they tend to incorporate the same U's. The problem occurs when we want the on-target run length to be 200 or less.

We give a numerical example to illustrate how the Voronoi rank works. In Figure 3.1, $G_1, G_2, \ldots, G_{16}$ represents the Voronoi regions of points $X_1, X_2, \ldots, X_{16}$ in R^2. A new point X_{17} arrives and falls into region G_{15}, so $R^{(1)}_{17} = 15$. From the above algorithm, $c = 4$ and $R^{(2)}_{17} = 6$, $R^{(3)}_{17} = 5$, $R^{(4)}_{17} = 14$. Define $U^{(1)}_{17} = 15/17$, $U^{(2)}_{17} = 6/17$, $U^{(3)}_{17} = 5/17$ and $U^{(4)}_{17} = 14/17$; and transform the approximate uniforms into approximate normal $N^{(1)}_{17} = 1.1868$, $N^{(2)}_{17} = -0.3774$, $N^{(3)}_{17} = -0.5414$ and $N^{(4)}_{17} = 0.9289$. Finally, we construct $M^4_{17} = (1.1868 - 0.3774 - 0.5414 + 0.9289)/4 = 0.3001$ to apply in the normal Cusum procedure.

3.3 Simulation

In this section, simulations are conducted in order to assess the application of the proposed Voronoi Cusum Chart. All the multivariate control charts compared in this section are assumed to have a group size of 1.

Table 3.1 gives the reference value k and the signal level h to fix the on-target run length of our algorithm to 200, 500 and 1000. The h value is determined by simulation based on 100000 replications. The simulation data are generated from the bivariate normal distribution with mean $\mu = (0,0)'$ and covariance matrix I.

Table 3.1: k and h required to set the run length of our procedure

ARL	k	h	h^*
200	0.5	3.15	3.5
500	0.5	4.276	4.375
1000	0.5	5.11	5.071

One can compare with h^* which is the corresponding signal level for a normal Cusum obtained from Table 3.2 in [12]. We see that as the run length increases, h increases beyond h^*. This is due to the sequential dependence of the M_n^c as mentioned above. Hence, while M_n^c is approximately normal, the sequence of M_n^c are not asymptotically independent. Hence we can't expect our Cusum to be close to a Cusum of i.i.d. normals. Nevertheless the reference value k and signal level h given in Table 3.1 do fix the on-target run length as desired in *any* dimension.

One might wonder why this procedure should have good off-target behaviour. First of all the above lemma shows that when on target the distribution of the U_i's is close to uniform even conditionally so an off-target condition will cause a departure from uniform. Intuitively one sees that off-target observations will tend to fall together and will have larger $R_i^{(r)}, r = 1, \ldots, c$ values since they are more recent. Hence the M_{n+1}^c will tend to be larger after the change point. Our procedure is therefore

rather like a sign test. It follows that our procedure is not as powerful as the non-parametric Cusum given in [24] which is essentially based on likelihoods. In [24] one sequentially ranks the one dimensional observations. Let $\hat{F}^n$ denote the empirical distribution of $X_1, X_2, \ldots, X_n$. Then the sequential rank of observation X_{n+1} is given by $R_{n+1} = n\hat{F}^n(X_{n+1})$. The procedure in [24] rescales these sequential ranks into $U_{n+1} = R_{n+1}/(n+1)$. Notice that for $t \in [0,1]$,

$$\begin{aligned} P(U_{n+1} \leq t | \{X_i : i \in \{1,\ldots,n\}) &= P(\frac{n}{n+1}\hat{F}^n(X_{n+1}) \leq t | \{X_i : i \in \{1,\ldots,n\}) \\ &\rightarrow P(F(X_1) \leq t) = t \end{aligned}$$

since $\hat{F}^n$ converges almost surely to F by the Glivenko-Cantelli theorem. The fact that U_{n+1} is asymptotically uniform even conditioned on the $X_1, \ldots, X_n$ is what makes the procedure so robust. The same is true for the multivariate procedure proposed here.

Table 3.2 gives the performance results for our algorithm for observations from a bivariate normal distribution with on-target mean $\mu_0 = (0,0)'$ and covariance matrix I and off-target mean $\mu_\delta = (\delta, \delta)$ and covariance matrix I. Here, λ is given by $\lambda = \sqrt{(\mu_1 - \mu_\delta)'\Sigma_X^{-1}(\mu_1 - \mu_\delta)}$. The estimate is based on 3000 replications. Simulations are done with reference values k and the signal level h for the on-target run length to 200, 500 and 1000. The lines $\delta = 0$ in Table 3.2 give the average on-target run length. In all off-target cases we first simulated 30 on-target observations before switching to the off-target distribution. Switching after 100 on-target observations yields about the same average off-target run length.

The on-target run lengths are very close to their targets set at 200, 500 and 1000. We also notice that the off-target run lengths are a lot larger than those of parametric procedures when the off-target shift δ is small like 0.5. This is to be expected because our procedure is essentially like a sign test which is certainly much less powerful than a parametric t test.

In Table 3.3 we redo the simulation in Table 3.2 the average on-target run lengths with reference values k and signal level h for the on-target run length are 200, 500 and

Table 3.2: Average On-target and Off-target Run Lengths: Observations are bivariate normals. $d = 2$.

ARL_{On}	k	h	δ	λ	$\widehat{ARL}$
200	0.5	3.15	0	0	199.85
	0.5	3.15	0.5	0.7071	50.27
	0.5	3.15	0.7071	1	28.81
	0.5	3.15	0.8660	1.2247	19.99
	0.5	3.15	1	1.4142	16.23
500	0.5	4.276	0	0	499.54
	0.5	4.276	0.5	0.7071	98.27
	0.5	4.276	0.7071	1	38.68
	0.5	4.276	0.8660	1.2247	25.82
	0.5	4.276	1	1.4142	20.15
1000	0.5	5.11	0	0	995.35
	0.5	5.11	0.5	0.7071	195.45
	0.5	5.11	0.7071	1	52.75
	0.5	5.11	0.8660	1.2247	30.23
	0.5	5.11	1	1.4142	22.28

1000. We tabulate the standard deviations as well and find they are essentially equal to the average run length as they should be because the run lengths are approximately geometrically distributed. Again estimates are based on 3000 replications.

In Table 3.4 the on-target distribution is a 10-dimensional multinormal distribution with a mean $\mu_0 = (0,0,0,0,0,0,0,0,0,0)'$ and covariance matrix I while the off-target mean $\mu_\delta = (\delta, \delta, \ldots, \delta)$ and covariance matrix I. Again the reference values k and h are chosen from Table 3.1 to fix the on-target average run length as 200, 500 and 1000. The off-target average run length of the Voronoi Cusum is determined by simulation. ARL_{On} is the on-target average run length. The simulation is based on

Table 3.3: On-Target Average Run Lengths and Standard Deviations: Observations are bivariate normals. $d = 2$.

k	h	$\widehat{ARL}$	$SD(\widehat{ARL})$
0.5	3.15	199.85	197.91
0.5	4.276	499.54	513.23
0.5	5.11	995.35	987.55

3000 replications. We see the on-target run length of our procedure is close to our predicted run length and the off-target run length is short (although not of course as short as the parametric cusum gives!).

Table 3.4: Multinormal to Multinormal Off-Target Average Run Length: Observations are normals. $d = 10$.

ARL_{On}	k	h	δ	λ	$\widehat{ARL}$
200	0.5	3.15	0	0	200.79
	0.5	3.15	1	3.16	8.44
500	0.5	4.276	0	0	489.92
	0.5	4.276	1	3.16	9.77
1000	0.5	5.11	0	0	990.88
	0.5	5.11	1	3.16	10.66

In Table 3.5 we consider a 3-dimensional distribution where one component is discrete. The the first two components of the on-target observations are standard normals while the third one is Binomial $(10, .1)$. The first two components of the off-target vector are are standard normals while the third one is Binomial $(10, .3)$. For reference values k and h chosen so the target average run length is 500 and off target average run length are determined numerically. The estimation is based on 3000 replications. Again we see the on-target run length of our procedure is as predicted and the off-target run length is fairly short. On the other hand a parametric

procedure doesn't work at all in this case.

Table 3.5: Two components are normal but the third is binomial. $d = 3$.

k	h	$\widehat{ARL}_{On}$	$\widehat{ARL}_{Off}$
0.5	4.276	499.56	14.96

In Table 3.6 we compare the off-target run length for our procedure with the multivariate EWMA Control Chart [33]. In the table, the d-variate t-distribution with v degrees of freedom is represent by $t_d(v)$. We denote the independent d-variate gamma distribution with shape parameter α and scale parameter β by $Gamma_d(\alpha, \beta)$. The values $k = 0.5$ and $h = 3.15$ are from the Table 3.1 so the the on-target run length is set to 200. From Table 3.6 we see the on-target run length is predicted fairly accurately and certainly better than for the multivariate EWMA Control Chart. On the other hand the off-target run lengths are certainly longer.

Table 3.7 gives the off-target run length comparison between our procedure and the multivariate EWMA Control Chart [33]. In the table, a p-variate t-distribution with v degrees of freedom is represent by $t_p(v)$. We denote the p-variate gamma distribution with shape parameter α and scale parameter β by $Gamma_p(\alpha, \beta)$. $k = 0.5$ and $h = 4.276$ are from the Table 3.1, the on-target run length is 500. The EWMA Control chart parameters $r = 0.10$ and $h_4 = 10.75$ are found in [20]. Again we see the on-target run length is set properly by our procedure but the EWMA control chart is so far off it is pointless to compare off-target run lengths.

Table 3.8 gives the off-target run length comparison between our procedure and the nonparametric multivariate rMEWMA Control Chart described in [25]. In the table, a bivariate normal distribution N_2 has an on-target mean $\mu_0 = (0, 0)'$ and covariance matrix I. We denote the p-variate t-distribution with v degrees of freedom by $t_p(v)$ and the p-variate gamma distribution with shape parameter α and scale parameter β by $Gamma_p(\alpha, \beta)$. The on-target run length is 200, $k = 0.5$ and $h = 3.15$ are from Table 3.1. The rWEWMA Control chart parameters $r = 0.30$ and $m = 200$

are found in [25]. From table 3.8 we see our procedure detects a small shift in the mean vector faster than the rMEWMA Control Chart.

Table 3.6: Comparing Run Length to those of EMWA

ARL_{On}		λ	$\widehat{ARL}$	ARL_{EMWA}
200	$t_2(3)$	0	199.97	207.3
	$t_2(3)$	0.25	149.96	73.0
	$t_2(3)$	0.5	95.47	27.9
	$t_2(3)$	1	35.58	11.4
	$t_2(3)$	3	7.52	3.6
200	$t_5(3)$	0	196.09	173.51
	$t_5(3)$	0.25	168.63	70.32
	$t_5(3)$	0.5	128.13	28.24
	$t_5(3)$	1	58.17	31.92
	$t_5(3)$	3	10.41	3.63
200	$t_{10}(3)$	0	194.31	155.53
	$t_{10}(3)$	0.25	183.09	76.15
	$t_{10}(3)$	0.5	152.28	28.3
	$t_{10}(3)$	1	84.88	11.54
	$t_{10}(3)$	3	14.87	3.74
200	$Gamma_2(1,1)$	0	200.06	198.7
	$Gamma_2(1,1)$	0.25	171.85	66.4
	$Gamma_2(1,1)$	0.5	138.38	26.9
	$Gamma_2(1,1)$	1	31.33	11.4
	$Gamma_2(1,1)$	3	5.01	3.6
200	$Gamma_5(1,1)$	0	204.26	173.48
	$Gamma_5(1,1)$	0.25	179.98	64.45
	$Gamma_5(1,1)$	0.5	150.87	26.73
	$Gamma_5(1,1)$	1	40.01	11.36
	$Gamma_5(1,1)$	3	5.01	3.6
200	$Gamma_{10}(1,1)$	0	201.75	161.5
	$Gamma_{10}(1,1)$	0.25	183.61	61.38
	$Gamma_{10}(1,1)$	0.5	166.63	26.54
	$Gamma_{10}(1,1)$	1	53.48	11.33
	$Gamma_{10}(1,1)$	3	4.98	3.6

Table 3.7: Comparing run lengths for Bivariate-t and Gamma distributions with EMWA

ARL_{On}		λ	$\widehat{ARL}$	ARL_{EMWA}
500	$t_2(3)$	0	495.62	245.31
	$t_2(3)$	0.25	402.58	232.29
	$t_2(3)$	0.5	265.35	185.52
	$t_2(3)$	1	58.93	33.73
	$t_2(3)$	3	8.77	1.93
500	$Gamma_2(1,1)$	0	498.85	253.57
	$Gamma_2(1,1)$	0.25	470.69	184.94
	$Gamma_2(1,1)$	0.5	351.35	78.54
	$Gamma_2(1,1)$	1	46.21	11.68
	$Gamma_2(1,1)$	3	5.56	1.35

[1]The EWMA chart does not work when the on-target run length is 500.

Table 3.8: Comparing Run Length for Bivariate normal, t and Gamma distributions with rMEMWA

ARL_{On}		λ	$\widehat{ARL}$	ARL_{rMEMWA}
200	N_2	0	199.85	200.48
	N_2	0.5	80.17	182.31
	N_2	1	28.81	97.64
	N_2	1.5	14.55	22.27
	N_2	2	9.93	6.79
	N_2	2.5	7.32	4.41
	N_2	3	6.15	3.60
200	$t_2(3)$	0	199.97	202.48
	$t_2(3)$	0.5	95.47	156.12
	$t_2(3)$	1	35.58	56.45
	$t_2(3)$	1.5	19.14	10.16
	$t_2(3)$	2	12.46	4.63
	$t_2(3)$	2.5	9.31	3.68
	$t_2(3)$	3	7.52	3.35
200	$Gamma_2(1,1)$	0	200.06	202.85
	$Gamma_2(1,1)$	0.5	138.38	140.30
	$Gamma_2(1,1)$	1	31.33	111.38
	$Gamma_2(1,1)$	1.5	11.34	83.94
	$Gamma_2(1,1)$	2	6.74	40.79
	$Gamma_2(1,1)$	2.5	5.44	11.50
	$Gamma_2(1,1)$	3	5.01	4.59

Chapter 4

Conclusion

In the thesis, we provide a framework for statistical control charts and develop a new distribution-free control chart called the Voronoi Cusum control chart. Two important issues of the application of Voronoi Cusum control chart are discussed. First, the control statistic is the average of c neighbors' Voronoi ranks, where $c = \min\{9, \lfloor\sqrt{n}\rfloor\}$. Second, the reference value k and the decision value h of Voronoi Cusum are generated from the simulation. A simulation study is conducted in order to compare in-control and out-of-control performance of Voronoi Cusum control charts with parametric MEMA control chart and nonparametric rMEWMA control chart. Once the control chart generates a signal, it detects a shift change. The adjustment of the process is based on the interpretation of out-of-control signals. The interpretation of out-of-control signals from multivariate control charts can be quite difficult, and this question has not been addressed in this thesis. Further research is needed in this area.

Appendix A: Cusum Code in Language R

```
# The on-traget distribution is Bivariate Normal with mu=(0,0)' and
# sigma=I. The on-target ARL sets to 200, the simulation is based on
# 100000 replications. The program is developed in language R.

library(MASS)
# Covariance Matrix
Sigma<- diag(2)
# On-target mean
mu<-matrix(c(0,0),2,1)
# off-target mean
mu1<-matrix(c(1,1),2,1)
# Reference value
kvalue<-0.5
# Total run length
total<-0
# Iteration number
counter<-0
# Decision interval
hvalue<-3.15

# Calculate the distance between two points
Edistance<-function(x1, x2)
{
  temp <- x1 - x2
  sum(temp * temp)
```

```
}

for(i in 1:100000) {

        # initial the sample with three observations
        Imatrix<-mvrnorm(n=3, mu, Sigma)

        Tvalue<-0
        # m counts the number of observations
        m<-4

        while(Tvalue<hvalue)
        {
              # Generate the sample point
              point<-(mvrnorm(n=1,mu,Sigma))
              Ilength<-nrow(Imatrix)
              Dmatrix<-matrix(0,Ilength,1)

              for(u in 1:Ilength)
              {
                    Dmatrix[u]<-Edistance(point, Imatrix[u,])
              }
              # Find the nearest regions to the new sample point
              backmatrix<-order(Dmatrix)
              # Calculate the M value
              M<-0
              if(m<81)
              {
```

```
        Cregion<-floor(sqrt(m))
        for(r in 1:Cregion)
        {
            M<-M+qnorm((backmatrix[r]/(Ilength+1)),0,1)
        }
        M<-M/Cregion
}
else
{
        for(r in 1:9)
        {
            M<-M+qnorm((backmatrix[r]/(Ilength+1)),0,1)
        }
        M<-M/9
}
Imatrix<-rbind(Imatrix, point)
# Cusum procedure
if(m<81)
{
            Cregion<-floor(sqrt(m))
            cvalue<-Tvalue+sqrt(Cregion)*M-kvalue
}
else
{
            cvalue<-Tvalue+3*M-kvalue
}
if(cvalue>0)
{
```

```
            Tvalue<-cvalue
        }
        if(cvalue<0)
        {
            Tvalue<-0
        }
        # Detect the change point and break the loop
        if(Tvalue>hvalue||Tvalue==hvalue)
        {
            run<-m
            break
        }
        m<-m+1
    }
    total=total+run
}
# Compute the ARL
ARL=total/100000
```

Bibliography

[1] ABELLANAS, M., CLAVEROL, M. AND HURTADO F. (2007). Point set stratification and Delaunay depth. *Computational Statistics & Data Analysis* **51**, 2513-2530.

[2] AURENHAMMER, F. AND KLEIN, R. (2000). Handbook of Computational Geometry. Elsevier. 201-290.

[3] BURR, M. A, RAFALIN, E. AND SOUVAINE, D. L. (2004). Simplicial Depth: An Improved Definition, Anaysis and Efficiency for the Finite Sample Case. *In: Proceedings of the 16th Canadaian Conference on Computational Geometry*, 136-139.

[4] BAKIR, S. T. AND REYNOLDS, M. R. (1979). A Nonparametric for Process Control Based on Within-Group Ranking. *Technometrics*, **21**, 175-183.

[5] CHAKRABORTI, S., VAN DER LAAN, P. AND BAKIR, S. T. (2001). Nonparmaetric Control Charts: An Overview and Some Results. *Journal of Quality Technology*, **33**, 304-315.

[6] CHENG, A. Y, OUYANG, M. (2001). On algorithm for Simplicial Depth. *In: Proceedings of the 16th Canadaian Conference on Computational Geometry*, 53-56.

[7] CROSIER, R.B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* **30**, 291-303.

[8] CROWDER, S.V. (1987). A Simple Method for Studying Run-length Distribution of Expontially Weighted Moving Average Charts. *Technometrics* **29**, 401-407.

[9] DELAUNAY, B. (1934). Sur la *sphère* vide. A la memoire de Georges Voronoi. *Izv. Akad. Nauk SSSR, Otdelenie Mathematicheskih i Estesvennyh Nauk* **7**, 793-800.

[10] JACKSON J. (1991) A user guide to principal components. John Wiley: New York.

[11] HACKL, P. AND LEDOLTER, J., (1992). A New Nonparametric Quality Control Technique. *Communications in Statistics-Simulation and Computation.* **21**, 423-443.

[12] HAWKINS, D.M. AND OLWELL, D.H., (1997). Cumulative Sum Charts and Charting for Quality Improvement. *Statistics for Engineering and Physical Science.* New York: Springer-Verlag.

[13] HEALY, J. D. (1987). A Note on Multivariate CUSUM Procedures. *Technometrics.* **29**, 409-412.

[14] HOTELLING, H. (1947). Multivariate quality control - illustrated by the air testing of sample bomb sights. *Techniques of Statistical Analysis* Ed. C. Eisenhart, M. W. Hastay and W. A. Wallis. New York: McGraw Hill.

[15] LIU, R.Y., ROBERT J.S., DIANE L.S. (2004). Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications. AMS Bookstore.

[16] LIU, R.Y. (1990). On a notation of Data Depth Based on Random Simplices. *The Annals of Statistics.* **18**, 405-414.

[17] LIU, R.Y. (1995). Control Charts for Multivariate Processes. *Jornal of the American Statistical Association.* **90**, 1380-1388.

[18] LIU, R.Y. AND SINGH, K. (2000). A Quality Index Based on Data Depth and Multivariate Rank Tests. *J. AMS.* **88**, 252-260.

[19] LIU, R.Y, SINGH, K. AND TENG, J.H. (2004). DDMA-charts: Nonparametric multivariate moving average control charts based on data depth.*Allgemeines Statistisches Archiv.* **88**, 235-258.

[20] LOWRY, C., WOODALL, W., CHAMP, W. AND RIGDOM, S. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics.*, **34**, 46-53.

[21] MAHALANOBIS, P. C. (1936). On the Generalized Distance in Statistics. *in Proceedings of the National Academy of India.* **12**, 49-55.

[22] MASON, R. L. AND YOUNG, J. C. (2001). Implementing multivariate process control using Hotelling's T^2 statistic. *Quality Process.* **34**, 71-73.

[23] MAO, J., MCDONALD, D. AND ZAREPOUR, M. (2009). On-line Quality Control Using Voronoi Depth. (Submitted)

[24] MCDONALD, D. (1990). A Cusum procedure based on sequential ranks. *Naval Res. Logist.*, **39**, 627-646.

[25] MESSAOUD, A. (2006). Monitoring Strategies for Chatter Detecion In a Drilling Process. *Doctoral Thesis*, University of Dortmund.

[26] MONTGOMERY, D.C. AND RUNGER, G.C. (2006) Applied Statistics and Probability for Engineers. John Wiley.

[27] PAGE, E.S. (1954) Continuous inepection schemes. *Biometrika.*, Vol. 41. 100-115.

[28] PIGNATIELLO, J. JR. AND RUNGER, G. C. (1990). Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology*, **22**, 173-186.

[29] QIU, P. H. AND HAWKINS, D. (2001). A Rank-Based Multivariate CUSUM Procedure. *Technometrics.* **43**, 120-132.

[30] ROUSSEUW, P. J. AND RUTS, I. (1996). Bivariate Location Depth. *Applied Statistics*, **45**, 516-526.

[31] SHEWHART, M. (1931). Economic Control of Quality of Manufactured Product. New York: Van Nostrand.

[32] SINGH, K. (1991). A Notion of Majority Depth. *technical report*, Rutgers Univerity, Dept. of Statistics.

[33] STOUMBOS, Z AND SULLIVAN (2002). Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology.*, **34**, 260-276.

[34] TUKEY, J. W. (1975). Mathematics and Picturing Data. *in Proceedings of the 1974 International Congress of Mathematicians*, Vancouver, **2**, 523-531.

[35] WILCOXON, F., KATTI, S. K. AND WILCOX, R. A. (1972). Critical Values and Probability Levels for the Wilcoson Rank Sum Test and the Wilcoxon Signed Rank Test. *Selected Tables in Mathematical Statistics*, Vol. 1 American Mathematical Society, Providence, R. I.

[36] WOODALL, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology.*, **32**, 341-350.

[37] WOODALL, W. H. AND MONTGOMERY, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology.*, **31**, 376-386.

[38] WOODALL, W. H. AND NCUBE, M. M. (1985). Multivariate CUSUM quality-control procedures. *Technomatrics*, **27**, 285-292.

[39] VORONOI, G. M. (1908). Nouvelles applicaitons des *paramètres* continus *à* la *théorie* des formes quadratiques. *deuxième Mémoire*: Recherches sur les *parallélloèdres* primitifs. *J. Reine Angew. Math.* **134**, 198-287.