國立交通大學

統計學研究所

碩 士 論 文

多維貝氏切割法在半導體製程改善之應用

Multiple Bayesian Segmentation for Process Improvement in Semiconductor Manufacturing

研 究 生：黃銘勳

指導教授：盧鴻興　教授

中 華 民 國 九 十 八 年 六 月

Multiple Bayesian Segmentation for Process Improvement in Semiconductor Manufacturing

研究生：黃銘勳　　Student: Ming-Syun Huang

指導教授：盧鴻興 博士　　Advisor: Dr. Horng-Shing Lu

國立交通大學

統計學研究所

碩士論文

A Thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

In

Statistics

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

多維貝氏切割法在半導體製程改善之應用

研究生：黃銘勳　　指導教授：盧鴻興 教授

國立交通大學統計所碩士班

摘　要

因爲半導體產業是一種高投資產業，每間公司在製程上所投資的成本都很高，因此，對每間公司而言，如何使製程穩定便是一個很重要的目標。因爲想要使製程穩定，所以當製程發生問題時，工程師便需要去找到製程發生問題的地方，但由於半導體產業的製程相當複雜，與製程相關的參數非常的多，因此當製程發生問題時，如何利用這麼多的製程參數去找到製程發生變動的所在位置，對工程師而言實爲一大挑戰。

在現有文獻中，對於檢驗製程是否發生變異並找出其正確位置，往往都著重在針對單一參數去找尋製程變動的位置，因此開發多變量偵測製程變異之系統極爲重要。在本篇文章中，將利用數學方式建立模型，提供多變量偵測製程之方式，幫助工程師更有效率解決製程中有問題的地方。

本篇論文所提供的方式，主要想法來自於賴政言於 2008 年所探討的 SBS (Single variable Bayesian Segmentation)，並對其做改良，進而將此方法推廣至多個維度上。由於半導體產業的製造過程中，常常會出現離群值，而對數學上建立模型產生困擾，因此，本文中對於離群值之出現，也提供一個方式來解決離群值對建立模型的影響。

Multiple Bayesian Segmentation for Process Improvement in Semiconductor Manufacturing

Student: Ming-Syun Huang Advisor: Horng-Shing Lu

ABSTRACT

Because the semiconductor industry is a high-invest industry, all companies invest a lot of money in the process. So, it is an important target to make the process stable for every company. Because the target is to make the process stable, when some changes occur to the process, engineers need to find the position where a change occurs. Because the semiconductor industry's process is quite complex, there are many different parameters that are relating to the process. When there are some changes in the process, how do engineers use these parameters' data to find the position where the change occurs is an enormous challenge for engineers.

In current literature, the most methods are using one parameter's data to find a position where changes occur to the process. Therefore, it is very important to develop a system that can use all parameters' data to find a position where the change of the process occurs. In this paper, we will establish a model using mathematics, provide a multivariate method to detect the process, and provide engineers a more effective solution to find the position where change occurs.

The main ideal of this paper comes from SBS (Single variable Bayesian Segmentation) that was investigated in Zheng-Yan Lai (2008). This paper improves on it and provides a new method that can use many parameters' data to find the position where change occurs. In the manufacturing process in the semiconductor industry, outliers often occur and affect the model that is established in mathematics. Therefore, this paper also provides a method to solve it.

誌　　謝

這篇論文的完成最主要要感謝我的指導教授盧鴻興教授與已畢業的博班學長涂凱文學長，在盧鴻興教授的指導下，除了讓我學到許多專業的統計知識之外，還讓我學到更多有關從事研究應有的態度與方法，而透過涂凱文學長的幫助，讓我清楚的知道，業界中與學界中做事的差異，同時也讓我了解，將統計應用在生活中並不需要很高深的理論，只要靠著一些簡單的統計方法與正確的統計概念，就可以創造出統計的價值；除此之外，我還要感謝跟我一起度過研究生生涯的這些第 16 屆的統計所同學們，包括了在同在盧教授指導下的同學柏成(丁丁)、庭瑋(leap)、孟樵(香菸)，在同一個教授下接受指導，讓我們常常聚在一起討論、一起打球、一起抱怨，除此之外，還有同在一間研究室(410)的同學新樺(阿貝)、婉慈(小豬)、欽友(下巴)、宇卿、澄竹、依賢，在壓力最重的碩二生活裡，我們常常互相得到一些新資訊，不管是學業的還是找工作上，都幫了我不少忙，還有要感謝隔壁研究室(411)的唱歌掛同好育珊、雅芳、智飛、芝賢、411 研究室室草少源與其他的同學，你們讓我的碩士生活過的更好，最後要感謝的就是我的家人，包括我的父母、姐弟與我最愛的老婆盈君，多虧你們的支持，我才能專心的在學業上發展，謝謝你們，同時也要感謝所上的其他教授對我的指導，相信在完成學業踏出交大之後，我一定能夠成爲一個不辜負大家期望的人。

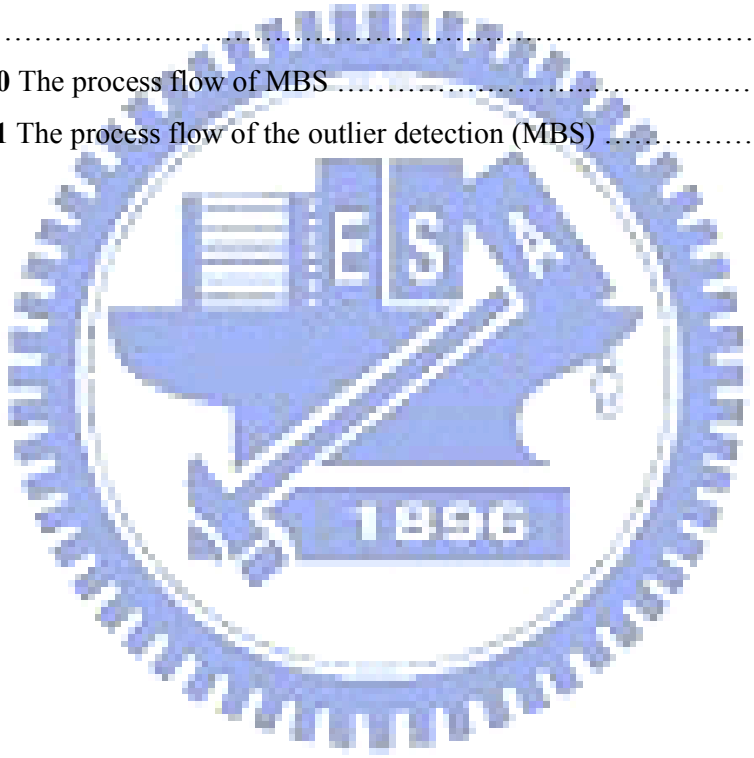
黃銘勳 謹 誌于

國立交通大學統計學研究所

中華民國九十八年六月

Table of Contents

List of Figures



List of Tables

Chapter 1. Introduction

In the semiconductor industry, the cost of a 12-inch silicon wafer fab is around 90 billion dollars, and the cost of a modern wafer fabrication facility is around 1 billion dollars. Thus, the semiconductor industry is a high-investment industry. Semiconductor engineers always try to make the process stable and improve the yield of the process. They will use any possible methods to help them study and monitor the process and avoid changes in the process. Possible methods can be divided into on-line methods and off-line methods.

The purpose of on-line methods is to use process parameters to monitor the process. Semiconductor engineers also can use on-line methods to find the position where the process change occurs in a minute. This is the advantage of on-line methods, and many studies are about on-line methods. Popular on-line methods that are used by semiconductor engineers are univariate statistical process control (SPC) and multivariate statistical process control (MVSPC).

Important reasons that off-line methods exist are: (1) some data can't be retrieved in a minute, (2) sometimes on-line SPC can't be used because the control limit is unreasonable, (3) according to all parameters' historical data, the semiconductor engineers can use off-line methods to study and understand the process. In this paper, I will introduce the Single Variable Bayesian Segmentation method (SBS) first. SBS is an off-line Bayesian method. It was investigated in Zheng-Yan Lai (2008). The purpose of SBS is to segment one parameter's historical data at a time and help semiconductor engineers study the process. Then, I will provide a new method, called Multivariate Bayesian Segmentation (MBS). MBS is similar to SBS. MBS is also an off-line Bayesian method, and the purpose is to help semiconductor engineers understand the process. The difference between MBS and SBS is that MBS can be used to segment more than one parameter's historical data at a time.

The motive of my study is that SBS can only be used to segment one parameter's historical data at a time, but there are many different parameters that are related to the process in the semiconductor industry. For example, a wafer's data from start to finish is around 10~100 MB. There are many kinds of the parameters, and the amount of the parameter's historical data is huge. It is assumed that semiconductor engineers can only segment a parameter's historical data at a time. If there are S different parameters now, they need to segment S times. When there are more kinds of parameters, semiconductor engineers need to segment more times. That is not efficient. If there is a method that can be used to segment all different parameters' historical data at a time, it is more efficient when semiconductor engineers study the process. Besides, interactions between every parameter are hard to be quantified and recorded. If semiconductor engineers segment one parameter's historical data at a time, they can't consider the effect that comes from interactions between every parameter. For example, there are two different parameters' historical data as shown in Figure 1. The correlation is at the level of 0.8. Because the correlation is very large, these two parameters' historical data have similar trends most of time. That is, when one parameter's historical data tend to increase significantly, another parameter's historical data also tend to increase significantly. However, parameter 1's data increase but parameter 2's data decrease significantly. So position B could be the position where there is a change occurring to the process. But if we only use parameter 1's data to find the position where the change occurs to the process, we may consider that position A is the position where the change occurs to the process because the value at position A is more similar to the values after position.

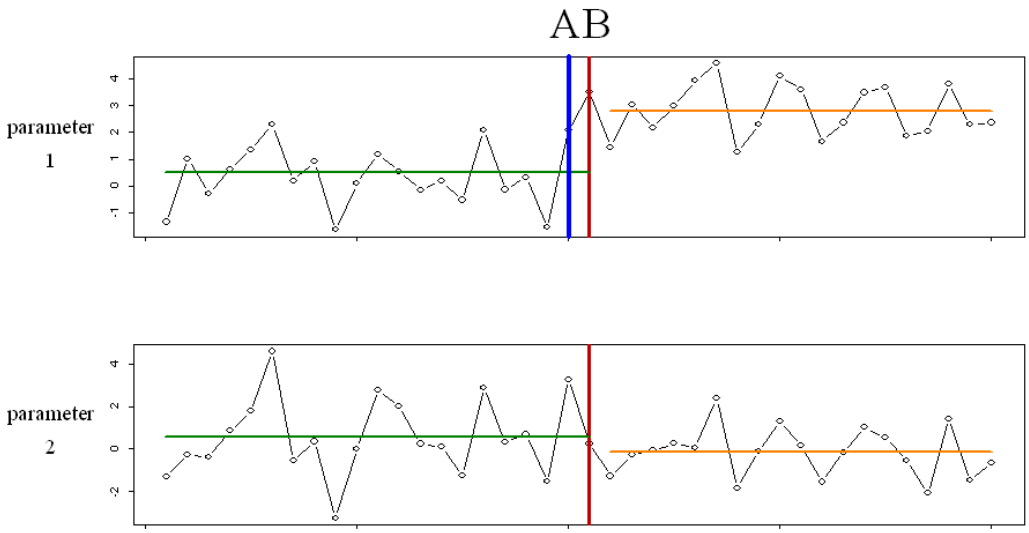


Figure 1. Mean-shift occurs in position B

In other words, if we segment all parameters' historical data at a time, we can get better information for considering the parameters' interaction. Besides, because the semiconductor process is quite complex, sometimes data must be taken down by people, and some unpredictable measurement deviations occur to the measurement apparatus. Therefore, outliers could sometimes occur, and they may affect the result that semiconductor engineers use MBS to get. In this paper, I will provide a method to detect outliers. This method can be used to avoid the outlier influencing the result, and the method can make MBS more robust.

Chapter 2. Literature Review

2.1 Introduction of Single variable Bayesian Segmentation (SBS)

The purpose of SBS is to segment one parameter's historical data, and the main idea is to find the position where a mean-shift occurs from one parameter's historical data. A mean-shift means that the data is composed of different distributions, and their means are different. For example, a parameter's historical data are shown in Figure 2. From Figure 2, we can find that the data before position A and the data after position A come from different distributions, and their means are different. So position A is a position where the mean-shift occurs. Because the mean-shift only occurs in position A, the data is split into two groups. The first group is the part of the data in front of position A, and second group is the part of the data after position A.

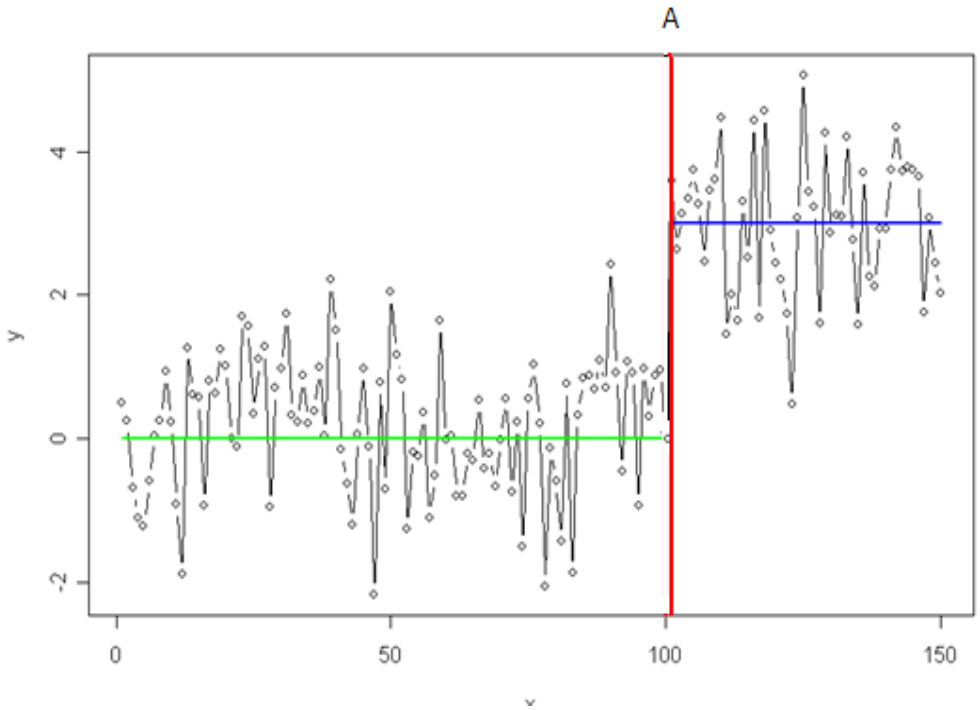


Figure 2. mean-shift occurs in position A.

Definition 1. Segmentation point: a position where mean-shift occurs.

For example, position A is a segmentation point in Figure 2.

Definition 2. Y is a sequence of observation with length N, $Y=(Y_1,\cdots,Y_N)$,

and Y_i can be described as the following formula for $1\le i\le N$:

$$Y_i=\mu_i+\varepsilon_i,\ 1\le i\le N,\ \varepsilon_i\sim \mathrm{N}(0,\ \sigma_\varepsilon^2), \qquad (2.1)$$

where μ_i is the mean of i-th component,

σ_ε^2 is the variance of i-th component,

N is the sample size.

Definition 3. R is a binary sequence with length N, $R=(R_1,\cdots,R_N)$, R_i is assumed to be i.i.d, where $R_i=\begin{cases}1 \text{ if a mean-shift occurs to the } i\text{-th position,}\\ 0 \text{ if a mean-shift doesn't occur to the } i\text{-th position,}\end{cases}$ $1\le i\le N$; (2.2)

and $\begin{cases}\text{the probability of } R_i=1 \text{ is } \lambda,\\ \text{the probability of } R_i=0 \text{ is } 1\text{-}\lambda,\end{cases}$ $0<\lambda<1$.

Assume there are $K-1$ segmentation points in the data, and their locations are $t_1,t_2,\cdots,t_{K-1}$ with $1=t_0<t_1<\cdots<t_{K-1}<t_K=N$. Because there are $K-1$ segmentation points now, the observation sequence Y is split into K groups. We call them $G_1,G_2,\cdots,G_K$ in order, and let their means are $\theta_1,\theta_2,\cdots,\theta_K$ in order. That is, if Y_i belongs to G_j, μ_i is equal to θ_j.

Because we don't know the real number of groups and the right locations of segmentation points, the number of groups (K) and the locations of segmentation points are unsure. Here, I will apply Bayesian method and SBS to find the most probable segmentation points and their locations. In the following, SBS is discussed:

We know that $$\mathrm{P}(R \mid Y) = \frac{\mathrm{P}(Y \mid R)\mathrm{P}(R)}{\mathrm{P}(Y)} \propto \mathrm{P}(Y \mid R)\mathrm{P}(R), \tag{2.3}$$

$$\mathrm{P}(Y \mid R) = \prod_{k=1}^{K} \prod_{y_i \in \text{Group } k} \frac{e^{\frac{(y_i-\theta_k)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi\sigma_\varepsilon^2}}, \tag{2.4}$$

$$\text{and } \mathrm{P}(R) = \lambda^{K-1}(1-\lambda)^{N-K+1}. \tag{2.5}$$

From $(2.3),(2.4),(2.5)$, we can get

$$\mathrm{P}(R \mid Y) = \left(\prod_{k=1}^{K} \prod_{y_i \in \text{Group } k} \frac{e^{\frac{(y_i-\theta_k)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi\sigma_\varepsilon^2}} \right) \times \left(\lambda^{K-1}(1-\lambda)^{N-K+1} \right). \tag{2.6}$$

And the maximum value of the posterior probability is

$$\underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Max} \{\mathrm{P}(R \mid Y)\} = \underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Max} \left\{ \left[\prod_{k=1}^{K} \prod_{y_i \in \text{Group } k} \frac{e^{\frac{(y_i-\theta_k)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi\sigma_\varepsilon^2}} \right] \times \lambda^{K-1}(1-\lambda)^{N-K+1} \right\} \tag{2.7}$$

$$\propto \underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Min} \left\{ -\log \left(\left[\prod_{k=1}^{K} \prod_{y_i \in \text{Group } k} \frac{e^{\frac{(y_i-\theta_k)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi\sigma_\varepsilon^2}} \right] \times \lambda^{K-1}(1-\lambda)^{N-K+1} \right) \right\}$$

$$= \underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Min} \left\{ -\log \left(\prod_{k=1}^{K} \prod_{y_i \in \text{Group } k} e^{\frac{(y_i-\theta_k)^2}{2\sigma_\varepsilon^2}} \times \left(\frac{\lambda}{1-\lambda} \right)^{K-1} \times \left(\frac{1-\lambda}{\sqrt{2\pi\sigma_\varepsilon^2}} \right)^{N} \right) \right\}$$

$$\propto \underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Min} \left\{ \sum_{k=1}^{K} \sum_{y_i \in \text{Group } k} \frac{(x_i-\theta_k)^2}{2\sigma_\varepsilon^2} + K \log\left(\frac{1-\lambda}{\lambda}\right) \right\} \quad \text{(because } \lambda \text{ and } \sigma_\varepsilon \text{ are constant)}$$

$$\propto \underset{K,\theta_1,\cdots,\theta_K,t_1,\cdots,t_{k-1}}{Min} \left\{ \sum_{k=1}^{K} \sum_{y_i \in \text{Group } k} (x_i-\theta_k)^2 + 2\sigma_\varepsilon^2 \times K \log\left(\frac{1-\lambda}{\lambda}\right) \right\}, \tag{2.8}$$

where K is the group number.

Here, we let

$$\hat{s}^2 = \frac{\sum_{k=1}^{K} \sum_{y_i \in \text{Group } k} \left(y_i - \hat{\theta}_k \right)}{N - K}, \qquad (2.9)$$

where N is the number of data points,

K is the group number,

$\hat{\theta}_k$ is the sample mean of group k.

Because $\hat{s}^2$ is the unbiased estimator of σ_ε^2, we use $\hat{s}^2$ to estimate σ_ε^2.

Because λ is the probability of mean-shift, λ means the probability of unusual conditions occurring to the data. Suppose that a semiconductor engineer is interested in a certain parameter X and X follows Normal$\left(\mu_x, \sigma_x^2\right)$. There is a 99.73% chance that X will be in $[\mu_x - 3\sigma_x, \mu_x + 3\sigma_x]$, and there is a 0.27% chance that X will be outside of $[\mu_x - 3\sigma_x, \mu_x + 3\sigma_x]$. So if X is bigger than $\mu_x + 3\sigma_x$ or smaller than $\mu_x - 3\sigma_x$, the semiconductor engineer may consider that a unusual condition occurs to the process. That is, the semiconductor engineer may consider that there is a segmentation point in the process and the capability of the process has changed. Here, one possible value of λ could be

$$\lambda = \mathrm{P}\left(X > \mu_x + 3\sigma_x\right) + \mathrm{P}\left(X < \mu_x - 3\sigma_x\right) = 0.0027. \qquad (2.10)$$

The other choices of λ are possible to describe the probability of process change.

2.2 The influence of outliers

In real cases, because the semiconductor process is quite complex, sometimes data must be taken down by people, and some unpredictable measurement deviations occur to the measurement apparatus. Outliers may occur in the process. Here, there is an outlier detection method that can avoid an outlier's effect.

The outlier detection method includes two steps:

(1) Choosing the possible outliers first.

(2) Checking the possible outliers in order. If the possible outlier will affect the result, we will consider that the possible outlier is an outlier and delete it.

The usual method of outlier detection is to check any observation lower than $Q_1 - S \times IQR$ or higher than $Q_3 + S \times IQR$. If there is an observation lower than $Q_1 - S \times IQR$ or higher than $Q_3 + S \times IQR$, we consider that the observation is an outlier. In step (1), SBS lets $S=1.5$ and uses the same method to find a possible outlier. That is, for each group, if any observation that belongs to the same group is lower than $Q_1 - 1.5 \times IQR$ or higher than $Q_3 + 1.5 \times IQR$, we will consider that the observation is a possible outlier.

If a possible outlier exists, in step (2), we will delete the possible outlier from the data and segment the data again. If we can get the same result, we consider that the possible outlier is not an outlier. If we can't get the same result, we consider that the possible outlier is an outlier and delete it.

2.3 The algorithm

(1) Let input data = initial data, and initial group number(K_0) = 1.

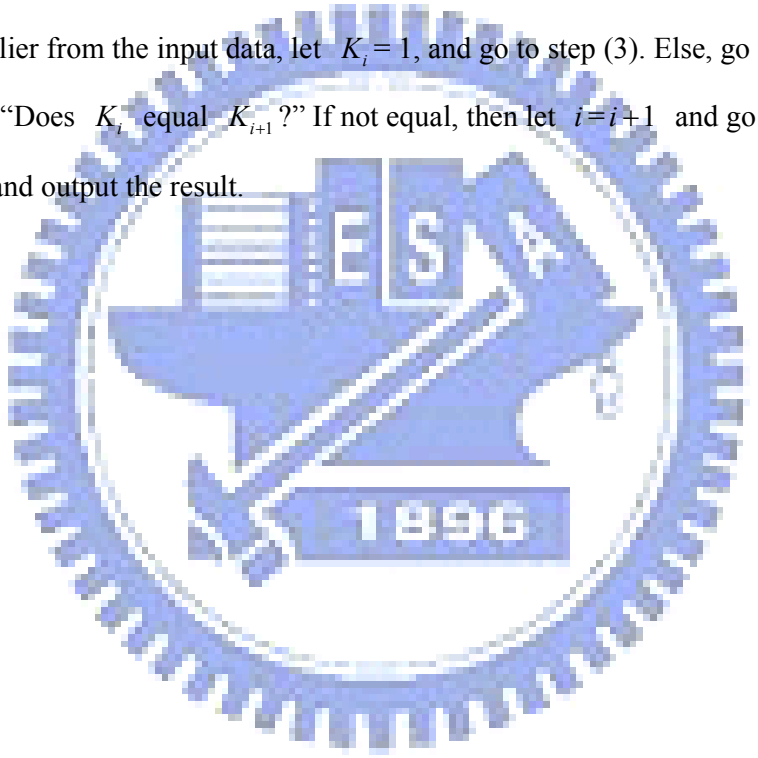
(2) Using input data to estimate σ_ε^2 under K_i group.

(3) Using $\hat{s}^2$ to segment each group, and get new group number K_{i+1}.

(4) Find a possible outlier from each new group. If a possible outlier exists, go to step (5). Else, go to step (6).

(5) Check "Is the possible outlier an outlier?" If the possible outlier is an outlier, then delete the outlier from the input data, let $K_i = 1$, and go to step (3). Else, go to step (6).

(6) Check "Does K_i equal K_{i+1}?" If not equal, then let $i = i + 1$ and go to step (2).Else, break and output the result.



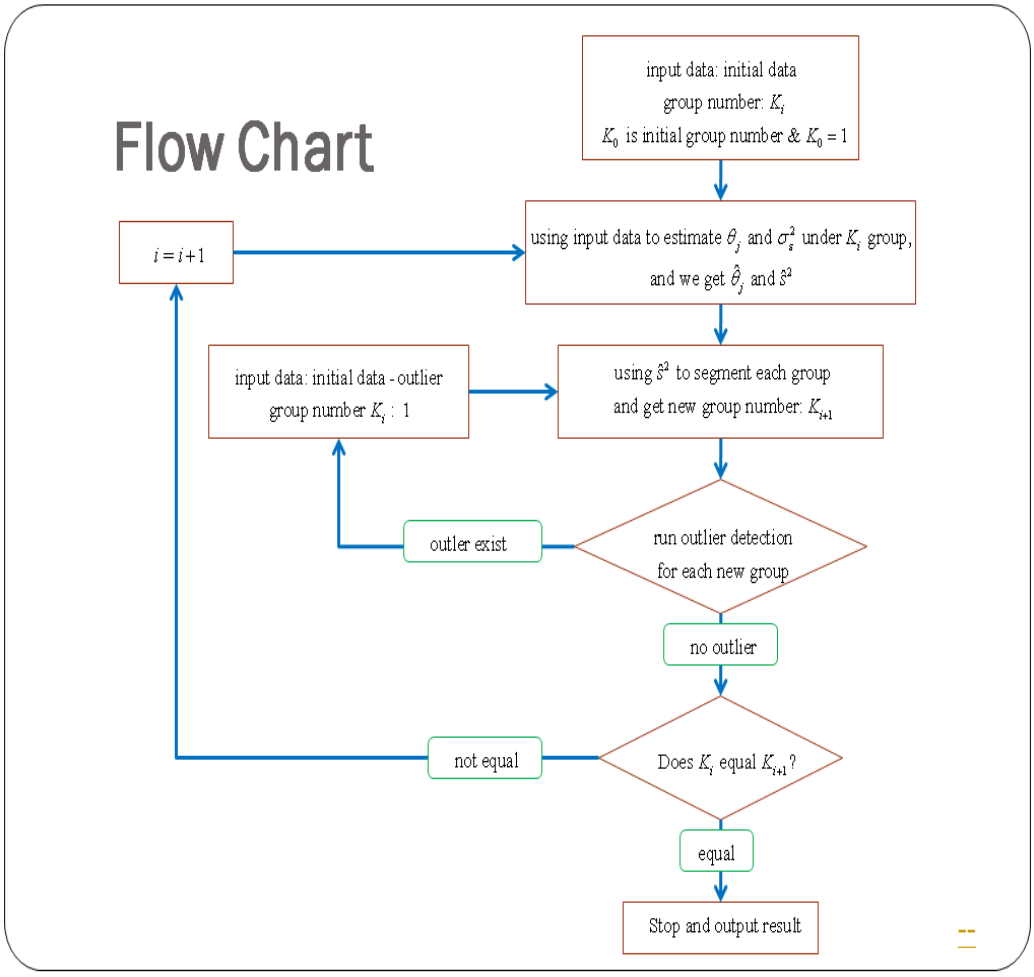


Figure 3. The process flow of SBS

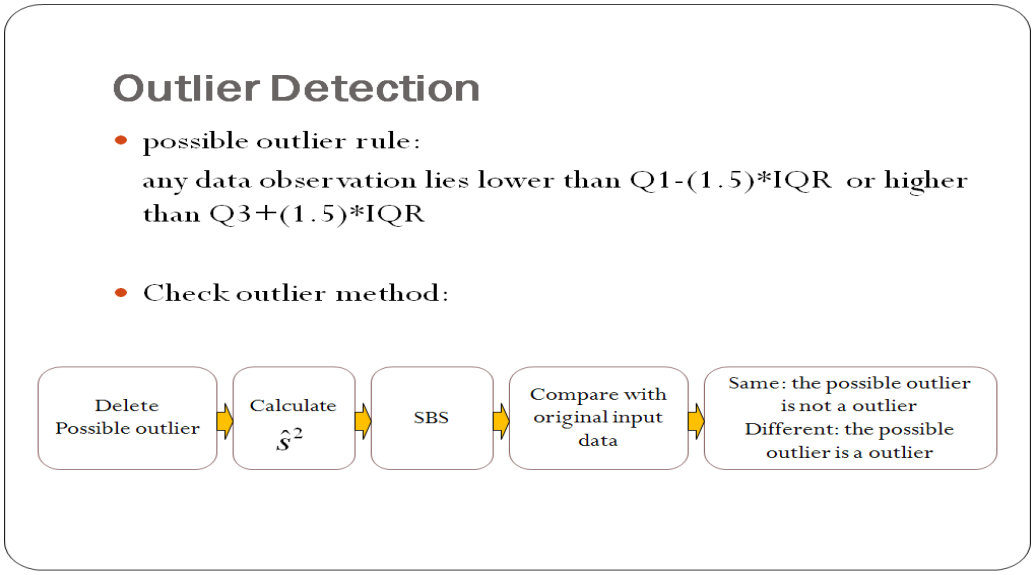


Figure 4. The process flow of the outlier detection (SBS)

Chapter 3. New Method

3.1 Introduction of Multivariate Bayesian Segmentation (MBS)

Multivariate Bayesian Segmentation extends from Single variable Bayesian Segmentation. The purpose of MBS is to segment many parameters' historical data at a time. In other words, the purpose of MBS is to find segmentation points from many parameters' historical data.

Here, a segmentation point is a position where a mean-shift occurs to more than one parameter's data. For example, there are 3 different parameters' data as shown in Figure 5. In Figure 5, we can find:

(1) In position A, a mean-shift occurs to parameter 1's data.

(2) In position B, a mean-shift occurs to parameter 1's data and parameter 2's data.

(3) In position C, a mean-shift occurs to all parameters' data.

So there are three segmentation points in the data, and their locations are A, B, and C. Because there are three segmentation points in the data, the data is split into four groups. We call them G_1, G_2, G_3, G_4 in order.

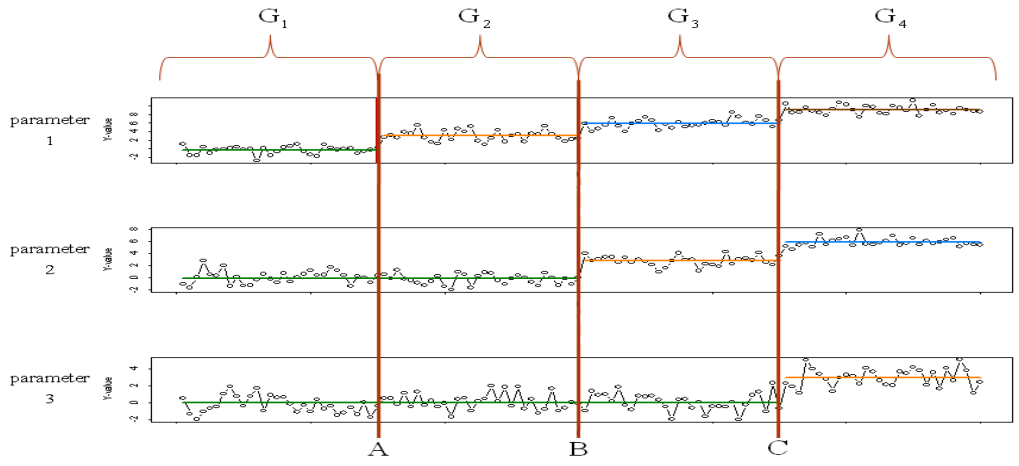


Figure 5. There are 3 segmentation points in the data.

Next, I will introduce MBS:

Assume there are S different parameters and N observation vectors ($\tilde{Y}_1, \cdots, \tilde{Y}_N$).

For $i = 1, \cdots, N$, $\tilde{Y}_i$ can be shown in the following formula:

$$\tilde{Y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iS} \end{bmatrix} = \begin{bmatrix} \mu_{i1} \\ \vdots \\ \mu_{iS} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iS} \end{bmatrix} = \tilde{\mu}_i + \tilde{\varepsilon}_i;\ i = 1, \cdots, N;\ \tilde{\varepsilon}_i \sim \mathrm{N}_S\left(\tilde{0},\ \Sigma\right), \qquad (3.1)$$

where $\tilde{\mu}_i$ is the mean vector of the i-th observation vector $\tilde{Y}_i$

Σ is the covariance matrix

Assume there are $K-1$ segmentation points in the data, and their locations are $t_1, \cdots, t_{K-1}$ with $0 < t_1 < t_2 < \cdots < t_{K-1} < N$. Because there are $K-1$ segmentation points in the data, the data ($\tilde{Y}_1, \cdots, \tilde{Y}_N$) can be split into K groups. We call them $\mathrm{G}_1, \cdots, \mathrm{G}_K$ in order. Because there are K different groups now, there are K different mean vectors. We call them $\tilde{\theta}_1, \cdots, \tilde{\theta}_K$ in order. That is, if $\tilde{Y}_i \in \mathrm{G}_j$, $\tilde{\mu}_i = \tilde{\theta}_j$.

Let E_i mean the change situation of the i-th position, $1 \le i \le N$.

For example, if there are 3 different parameters, there are 8 different change situations as shown in Figure 6.

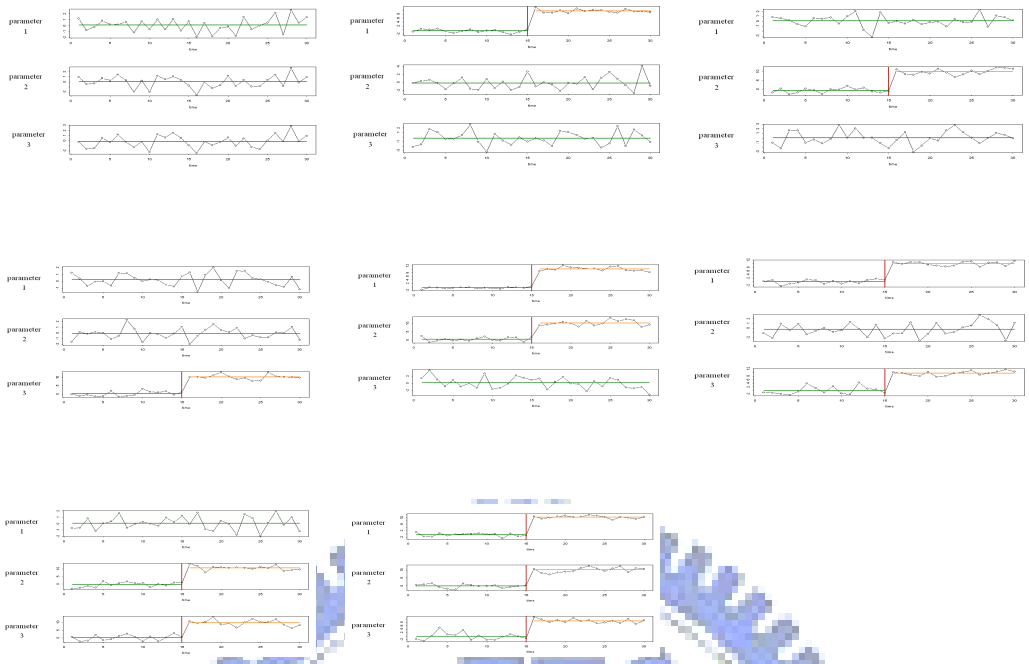


Figure 6. All probability change situations with 3 parameters

Let p_i mean the number of the parameter where the mean-shift occurs.

Assume that $\mathrm{E}_1, \mathrm{E}_2, \cdots, \mathrm{E}_N$ are independent.

$$\Rightarrow \mathrm{P}(\mathrm{E}_1, \cdots, \mathrm{E}_N) = \mathrm{P}(\mathrm{E}_1) \times \mathrm{P}(\mathrm{E}_2) \times \cdots \times \mathrm{P}(\mathrm{E}_N).$$

Assume that the probability of mean-shift for all parameters is λ.

$$\Rightarrow \mathrm{P}(\mathrm{E}_i) = \lambda^{p_i} \cdot (1-\lambda)^{S-p_i}.$$

Because we don't know the real number of groups and the right locations of the segmentation points and the change situations of the segmentation points, $K, \tilde{\theta}_1, \cdots, \tilde{\theta}_K, p_1, \cdots, p_N, \mathrm{E}_1, \cdots, \mathrm{E}_N, t_1, \cdots, t_{k-1}$ are unknown. Here, I will apply Bayesian method and MBS to find the most probable segmentation points and their locations.

$$P(E_1,\cdots,E_N \mid \tilde{y}_1,\cdots,\tilde{y}_N) = \frac{P(\tilde{y}_1,\cdots,\tilde{y}_N \mid E_1,\cdots,E_N)\cdot P(E_1,\cdots,E_N)}{P(\tilde{y}_1,\cdots,\tilde{y}_N)}$$

$$\propto P(\tilde{y}_1,\cdots,\tilde{y}_N \mid E_1,\cdots,E_N)\cdot P(E_1,\cdots,E_N)$$

$$\propto \left\{\prod_{j=1}^{K}\prod_{y_i\in G_j}\left[(2\pi)^{-\frac{S}{2}}\cdot(|\Sigma|)^{-\frac{1}{2}}\cdot\exp\left\{-\frac{1}{2}(\tilde{y}_i-\bar{\theta}_j)^{T}\Sigma^{-1}(\tilde{y}_i-\bar{\theta}_j)\right\}\right]\right\}\times\left\{\prod_{i=1}^{N}\lambda^{p_i}\cdot(1-\lambda)^{S-p_i}\right\}.$$

And the maximum of the posterior probability is

$$\max_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,E_1,\cdots,E_N,t_1,\cdots,t_{k-1}}\left\{P(E_1,\cdots,E_N \mid \tilde{y}_1,\cdots,\tilde{y}_N)\right\} \tag{3.2}$$

$$= \max_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,E_1,\cdots,E_N,t_1,\cdots,t_{k-1}}\left\{P(\tilde{y}_1,\cdots,\tilde{y}_N \mid E_1,\cdots,E_N)\cdot P(E_1,\cdots,E_N)\right\}$$

$$= \min_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,E_1,\cdots,E_N,t_1,\cdots,t_{k-1}}\left\{-\log\left[P(\tilde{y}_1,\cdots,\tilde{y}_N \mid E_1,\cdots,E_N)\cdot P(E_1,\cdots,E_N)\right]\right\}$$

$$= \min_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,E_1,\cdots,E_N,t_1,\cdots,t_{k-1}}\left\{\frac{1}{2}\sum_{j=1}^{K}\sum_{\tilde{y}_i\in G_j}(\tilde{y}_i-\tilde{\theta}_j)^{T}\Sigma^{-1}(\tilde{y}_i-\tilde{\theta}_j)+\sum_{i=1}^{N}p_i\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\right\}$$

$$= \min_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,E_1,\cdots,E_N,t_1,\cdots,t_{k-1}}\left\{\sum_{j=1}^{K}\sum_{\tilde{y}_i\in G_j}(\tilde{y}_i-\tilde{\theta}_j)^{T}\Sigma^{-1}(\tilde{y}_i-\tilde{\theta}_j)+2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\cdot\sum_{i=1}^{N}p_i\right\}, \tag{3.3}$$

where K is the number of groups.

Here, we still need to solve some questions:

1. How do we estimate the mean vectors?
2. How do we estimate the covariance matrix?
3. How do we choose the parameter λ?

3.1.1 How do we estimate the mean vectors?

We can use the change situation of the process to estimate the mean vector ($\tilde{\theta}_j$). For example, there are 3 different parameters and N observation vectors now. And there are 3

segmentation points in the data. Their locations are t_1, t_2, t_3 with $1 = t_0 < t_1 < t_2 < t_3 < t_4 = N$.

(1) In position t_1, a mean-shift occurs to parameter 1's data.

(2) In position t_2, a mean-shift occurs to parameter 1's data and parameter 2's data.

(3) In position t_3, a mean-shift occurs to all parameters' data.

And we can show them in Figure 7.

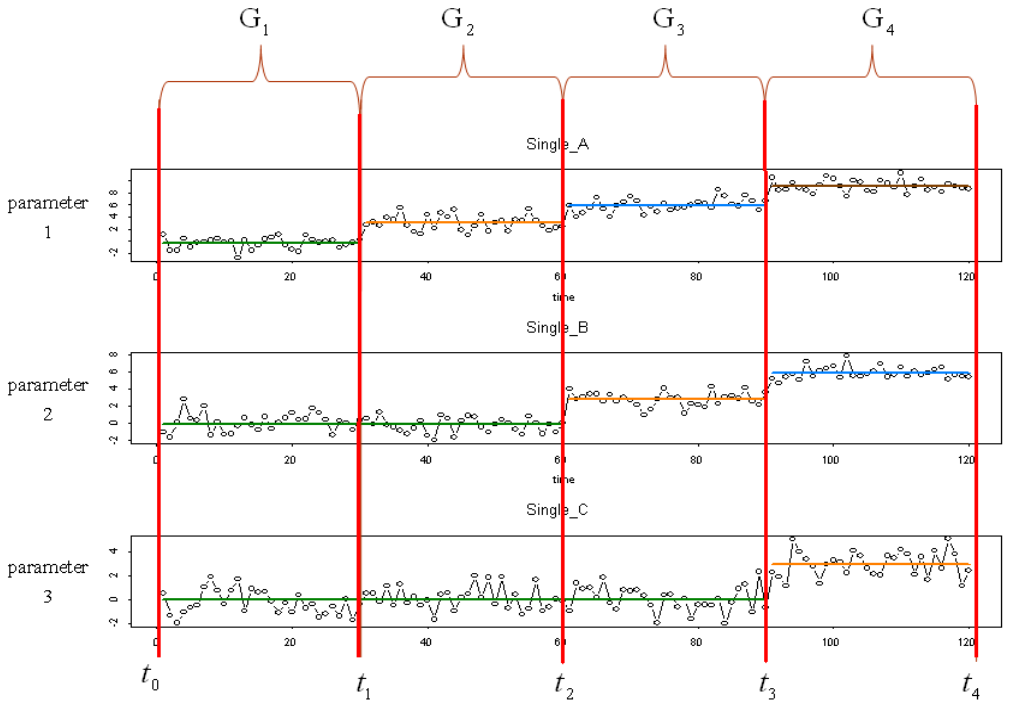


Figure 7. There are 3 segmentation points in the data

Because there are 3 segmentation points in the data, the data can be split into 4 groups. We call them G_1, G_2, G_3, G_4 in order. That is, If $\tilde{Y}_i \in G_j$, then $\tilde{Y}_i \sim N_3(\tilde{\theta}_j, \Sigma)$; $\tilde{\theta}_j = \begin{bmatrix} \theta_{j(1)} \\ \theta_{j(2)} \\ \theta_{j(3)} \end{bmatrix}$;

$j = 1,2,3,4$.

For each $\tilde{\theta}_j$, the usual method is using the data that belong to G_j to calculate the

sample mean ($\bar{Y}_j = \begin{bmatrix} \bar{Y}_{j(1)} \\ \bar{Y}_{j(2)} \\ \bar{Y}_{j(3)} \end{bmatrix}$), and let $\tilde{\theta}_j$ be equal to $\bar{Y}_j$; j=1,2,3,4. For example, if we want to estimate $\theta_{1(2)}$ and $\theta_{2(2)}$, we can use parameter 2's partial data that is in $[t_0, t_1]$ to calculate its sample mean ($\bar{Y}_{1(2)}$), and let $\bar{Y}_{1(2)}$ be the estimator of $\theta_{1(2)}$. We can use parameter 2's partial data that is in $[t_1, t_2]$ to calculate its sample mean ($\bar{Y}_{2(2)}$), and let $\bar{Y}_{2(2)}$ be the estimator of $\theta_{2(2)}$. That is, let $\hat{\theta}_{1(2)}$ be equal to $\bar{Y}_{1(2)}$ and $\hat{\theta}_{2(2)}$ be equal to $\bar{Y}_{2(2)}$.

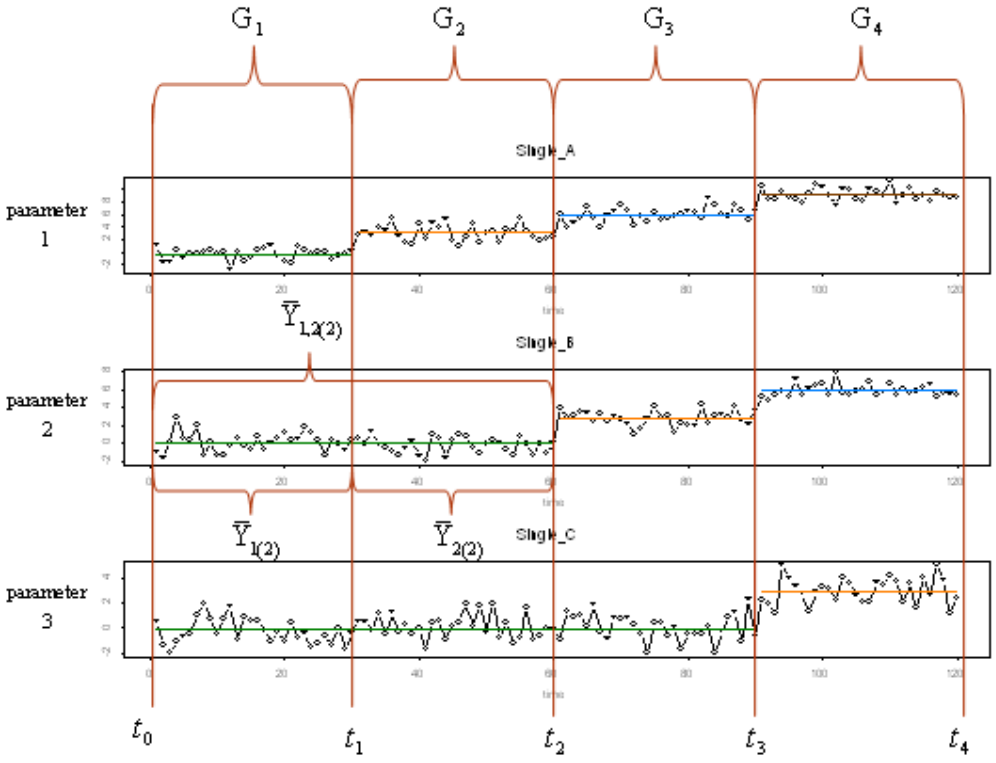


Figure 8. There are 4 groups in the data

Here, I will consider the change situation of the process to get the better estimator of $\tilde{\theta}_j$. Following the previous example, we want to estimate $\theta_{1(2)}$ and $\theta_{2(2)}$. But in position t_1, we

can find that a mean-shift only occurs to parameter 1. A mean-shift doesn't occur to parameter 2 in position t_1. $\theta_{1(2)}$ and $\theta_{2(2)}$ should be equal, and we can estimate them together. So we can use parameter 2's partial data that is in $[t_0, t_2]$ to calculate its sample mean ($\bar{Y}_{1,2(2)}$), and let $\bar{Y}_{1,2(2)}$ be the estimator of $\theta_{1(2)}$ and $\theta_{2(2)}$. That is, let $\hat{\theta}_{1(2)} = \hat{\theta}_{2(2)} = \bar{Y}_{1,2(2)}$.

3.1.2 How do we estimate the covariance matrix?

Definition 4.

$$\hat{\Sigma} = \frac{1}{N-K} \sum_{j=1}^{K} \sum_{\tilde{y}_i \in \text{Group } j} \left(\tilde{Y}_i - \bar{Y}_j\right)\left(\tilde{Y}_i - \bar{Y}_j\right)^{T}, \qquad (3.4)$$

where N is the number of observations,

K is the group number.

It is assumed that data is sufficiently large to estimate covariance matrix. Because $\hat{\Sigma}$ is an unbiased estimator of Σ, I will use $\hat{\Sigma}$ to estimate Σ. Following the previous example, there are 3 different parameters and N observation vectors now. And there are 3 segmentation points in the data. The data is shown as Figure 7. Because there are 4 groups (G_1, G_2, G_3, G_4), K is equal to 4. And we can use the method that is introduced in the section 3.1.1 to estimate the mean vectors ($\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4$). Let the estimator of $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4$ be $\tilde{\theta}_1^*, \tilde{\theta}_2^*, \tilde{\theta}_3^*, \tilde{\theta}_4^*$. Finally, we put the information into the formula (3.4), and we can get

$$\hat{\Sigma} = \frac{1}{N-4} \sum_{j=1}^{4} \sum_{\tilde{Y}_i \in G_j} \left(\tilde{Y}_i - \tilde{\theta}_j^*\right)\left(\tilde{Y}_i - \tilde{\theta}_j^*\right)^{T}.$$

3.1.3 How do we find the parameter lambda?

Because λ is the probability of a mean-shift, λ means the probability of unusual conditions occur to the data. If a semiconductor engineer is interested in a certain parameter

X , X follows Normal$\left(\mu_x, \sigma_x^2\right)$. There is a 99.73% chance that X will be in [$\mu_x - 3\sigma_x$, $\mu_x + 3\sigma_x$], and there is a 0.27% chance that X will be outside of [$\mu_x - 3\sigma_x$, $\mu_x + 3\sigma_x$]. So if X is bigger than $\mu_x + 3\sigma_x$ or smaller than $\mu_x - 3\sigma_x$, the semiconductor engineer will consider that an unusual condition occurs to the process. That is, the semiconductor engineer will consider that there is a segmentation point in the process and that the capability of the process has changed. Here, we let

$$\lambda = \mathrm{P}\left(X > \mu_x + 3\sigma_x\right) + \mathrm{P}\left(X < \mu_x - 3\sigma_x\right) = 0.0027. \qquad (3.5)$$

But the semiconductor engineer can use the property of the product process to determine λ .

3.2 How to find the optimal solution quickly

From the formula (3.2) and the formula (3.3), we know that

$$\max_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,\mathrm{E}_1,\cdots,\mathrm{E}_N,t_1,\cdots,t_{k-1}} \left\{\mathrm{P}\left(\mathrm{E}_1,\cdots,\mathrm{E}_N \mid \tilde{y}_1,\cdots,\tilde{y}_N\right)\right\}$$

$$\propto \min_{K,\tilde{\theta}_1,\cdots,\tilde{\theta}_K,p_1,\cdots,p_N,\mathrm{E}_1,\cdots,\mathrm{E}_N,t_1,\cdots,t_{k-1}} \left\{\sum_{j=1}^{K}\sum_{\tilde{y}_i\in \mathrm{G}_j}\left(\tilde{y}_i-\tilde{\theta}_j\right)^{\mathrm{T}}\Sigma^{-1}\left(\tilde{y}_i-\tilde{\theta}_j\right)+2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\cdot\sum_{i=1}^{N}p_i\right\}.$$

So our purpose is to find a change situation that can minimize

$$\sum_{j=1}^{K}\sum_{\tilde{y}_i\in \mathrm{G}_j}\left(\tilde{y}_i-\tilde{\theta}_j\right)^{\mathrm{T}}\Sigma^{-1}\left(\tilde{y}_i-\tilde{\theta}_j\right)+2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\cdot\sum_{i=1}^{N}p_i.$$

But if the number of the parameters or the number of observation vectors increase, the possible change situation will increase. For example, assume there are 2 different parameters and N observation vectors. For each observation vector, there are 4 different change situations. They are

(1) No mean-shift

(2) A mean-shift occurs to the first parameter.

(3) A mean-shift occurs to the second parameter.

(4) A mean-shift occurs to the first and the second parameter.

Then, the number of the possible change situation is 4^N.

Assume there are 3 different parameters and N observation vectors. For each observation vector, there are 8 different change situations. They are

(1) No mean-shift.

(2) A mean-shift occurs to the first parameter.

(3) A mean-shift occurs to the second parameter.

(4) A mean-shift occurs to the third parameter.

(5) A mean-shift occurs to the first and the second parameters.

(6) A mean-shift occurs to the first and the third parameters.

(7) A mean-shift occurs to the second and the third parameters.

(8) A mean-shift occurs to all parameters.

Then, the number of the possible change situation is 8^N.

So I will provide a method that can help us to find a change situation that can minimize

$$\sum_{j=1}^{K}\sum_{\tilde{y}_i\in \mathrm{G}_j}\left(\tilde{y}_i-\tilde{\theta}_j\right)^{\mathrm{T}}\Sigma^{-1}\left(\tilde{y}_i-\tilde{\theta}_j\right)+2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\cdot\sum_{i=1}^{N}p_i.$$

The method includes 3 steps:

(1) Decide the order of the segmentation points.

(2) Decide the change situation of the segmentation points.

(3) Decide the best solution from all possible change situations.

3.2.1 Decide the order of the segmentation points

My idea is to find a position from all the positions first. Then I want to find a position from the positions that remain. By the same way, I can decide the order of segmentation points.

So we need to know how to choose a position from the possible position. Next, I will explain how to choose a position from the possible position with an example. Assume there are 3 different parameters and N observation vectors ($\tilde{y}_1,\cdots,\tilde{y}_N$), and $\tilde{y}_i=\begin{bmatrix} y_{i(1)} \\ y_{i(2)} \\ y_{i(3)} \end{bmatrix}$; $i=1,\cdots,N$. Assume we have known that there are S segmentation points in the data, and their locations are $t_1,t_2,\cdots,t_S$.

Let

$$\mathrm{F}_S(q)=\sum_{i=1}^{u_1-1}\left\{\begin{bmatrix} y_{i(1)}-\overline{Y}_{1,1} & y_{i(2)}-\overline{Y}_{1,2} & y_{i(3)}-\overline{Y}_{1,3} \end{bmatrix}\Sigma^{-1}\begin{bmatrix} y_{i(1)}-\overline{Y}_{1,1} \\ y_{i(2)}-\overline{Y}_{1,2} \\ y_{i(3)}-\overline{Y}_{1,3} \end{bmatrix}\right\}$$

$$+\sum_{i=u_1}^{u_2-1}\left\{\begin{bmatrix} y_{i(1)}-\overline{Y}_{2,1} & y_{i(2)}-\overline{Y}_{2,2} & y_{i(3)}-\overline{Y}_{2,3} \end{bmatrix}\Sigma^{-1}\begin{bmatrix} y_{i(1)}-\overline{Y}_{2,1} \\ y_{i(2)}-\overline{Y}_{2,2} \\ y_{i(3)}-\overline{Y}_{2,3} \end{bmatrix}\right\}$$

$$+\cdots$$

$$+\sum_{i=u_{S+1}}^{N}\left\{\begin{bmatrix} y_{i(1)}-\overline{Y}_{S+2,1} & y_{i(2)}-\overline{Y}_{S+2,2} & y_{i(3)}-\overline{Y}_{S+2,3} \end{bmatrix}\Sigma^{-1}\begin{bmatrix} y_{i(1)}-\overline{Y}_{S+2,1} \\ y_{i(2)}-\overline{Y}_{S+2,2} \\ y_{i(3)}-\overline{Y}_{S+2,3} \end{bmatrix}\right\}$$

$$+\left[2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\right]\cdot\left[3\cdot(S+1)\right],$$

where $q\in\{2,\cdots,N\}\setminus\{t_1,\cdots,t_S\}$,

and $(u_1,u_2,\cdots,u_{S+1})$: the order of $(t_1,t_2,\cdots,t_S,q)$, where $u_1<\cdots<u_{S+1}$.

and $\overline{Y}_{1,1}, \overline{Y}_{1,2}, \overline{Y}_{1,3}, \overline{Y}_{2,1}, \overline{Y}_{2,2}, \overline{Y}_{2,3}, \cdots, \overline{Y}_{S+2,1}, \overline{Y}_{S+2,2}, \overline{Y}_{S+2,3}$: the sample mean of some observation. (show as Figure 9)

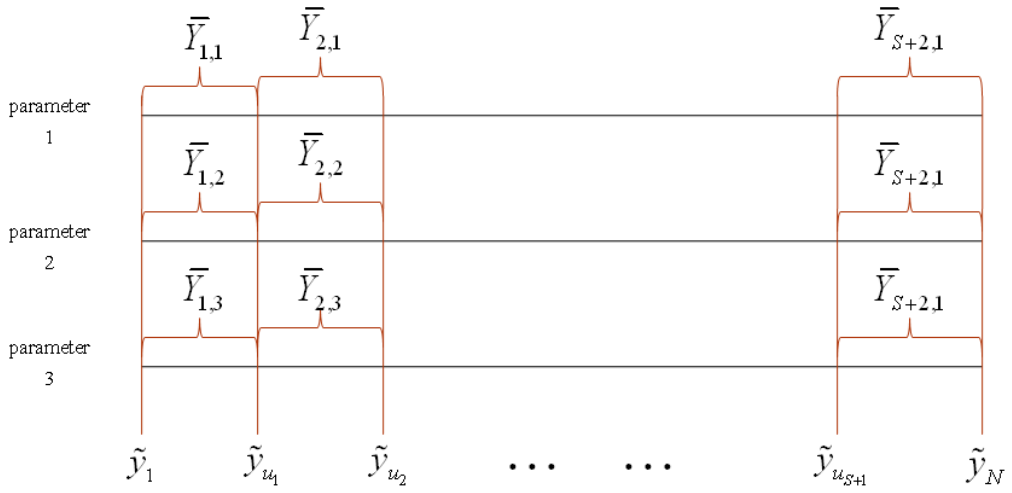


Figure 9.

We plug in values from small to large in $\{2,\cdots,N\}\setminus\{t_1,\cdots,t_S\}$ as q, and we calculate the value of $\mathrm{F}_S(q)$. Then we can choose a position whose value of $\mathrm{F}_S(q)$ is the minimum, and let the position be the $S+1$ -th segmentation point. That is, if $h \in \{2,\cdots,N\}\setminus\{t_1,\cdots,t_S\}$ and $\mathrm{F}_S(h) = \min\{\mathrm{F}_S(q) \mid \forall q \in \{2,\cdots,N\}\setminus\{t_1,\cdots,t_S\}\}$, $\tilde{y}_h$ is the $S+1$-th segmentation point. By the same way, we can decide the order of the segmentation point.

3.2.2 Decide the change situation of the segmentation points

In section 3.2.1, we learned how to decide the order of the segmentation points. In this section, I will decide the change situation of the segmentation points. First, we used the method that was introduced in section 3.2.1 to find the order of the segmentation points.

Second, according to different situations, we used the method that was introduced in the section 3.1.1 to estimate $\tilde{\theta}_j$. Finally, we need to choose a situation that can minimize

$$\sum_{j=1}^{K}\sum_{\tilde{y}_i\in G_j}\left(\tilde{y}_i-\tilde{\theta}_j\right)^{T}\Sigma^{-1}\left(\tilde{y}_i-\tilde{\theta}_j\right)+2\cdot\log\left(\frac{1-\lambda}{\lambda}\right)\cdot\sum_{i=1}^{N}p_i. \qquad (3.6)$$

For example, there are 3 different parameters and 5 observation vectors ($\tilde{y}_1,\cdots,\tilde{y}_5$). And we use the method that was introduced in the section 3.2.1 to get the order of the segmentation points: $\tilde{y}_3\to\tilde{y}_4\to\tilde{y}_2\to\tilde{y}_5$. Because there are only 3 parameters, there are 7 possible change situations of a segmentation point.

(1) A mean-shift occurs to the first parameter.

(2) A mean-shift occurs to the second parameter.

(3) A mean-shift occurs to the third parameter.

(4) A mean-shift occurs to the first and the second parameters.

(5) A mean-shift occurs to the first and the third parameters.

(6) A mean-shift occurs to the second and the third parameters.

(7) A mean-shift occurs to all parameters.

First, we assume only $\tilde{y}_3$ is a segmentation point, and others are not segmentation points.

1. Let a mean-shift occur to the first parameter in $\tilde{y}_3$, and use the method that was introduced in section 3.1.1 to estimate the mean vector ($\tilde{\theta}_j$). Then we input them into the formula (3.6) and get a value (called value_1).
2. Let a mean-shift occur to the second parameter in $\tilde{y}_3$. In the same way, we can estimate the mean vector ($\tilde{\theta}_j$) and get a value (called value_2).
3. Let a mean-shift occur to the third parameter in $\tilde{y}_3$. In the same way, we can estimate

the mean vector ($\tilde{\theta}_j$) and get a value (called value_3).

4. Let a mean-shift occur to the first parameter and the second parameter in $\tilde{y}_3$. In the same way, we can estimate the mean vector ($\tilde{\theta}_j$) and get a value (called value_4).
5. Let a mean-shift occur to the first parameter and the third parameter in $\tilde{y}_3$. In the same way, we can estimate the mean vector ($\tilde{\theta}_j$) and get a value (called value_5).
6. Let a mean-shift occur to the second parameter and the third parameter in $\tilde{y}_3$. In the same way, we can estimate the mean vector ($\tilde{\theta}_j$) and get a value (called value_6).
7. Let a mean-shift occur to all the parameters in $\tilde{y}_3$. In the same way, we can estimate the mean vector ($\tilde{\theta}_j$) and get a value (called value_7).

Then we choose a situation whose value is the minimum of (value_1, value_2, value_3, value_4, value_5, value_6, value_7), and let the change situation of $\tilde{y}_3$ be that situation. Here, we assume that the change situation of $\tilde{y}_3$ is (m), $1 \le m \le 7$.

Second, we assume both $\tilde{y}_3$ and $\tilde{y}_4$ are segmentation points and the change situation of $\tilde{y}_3$ is (m). By the same way, we can let

1. A mean-shift occurs to the first parameter in $\tilde{y}_4$.
2. A mean-shift occurs to the second parameter in $\tilde{y}_4$.
3. A mean-shift occurs to the third parameter in $\tilde{y}_4$.
4. A mean-shift occurs to the first and the second parameters in $\tilde{y}_4$.
5. A mean-shift occurs to the first and the third parameters in $\tilde{y}_4$.
6. A mean-shift occurs to the second and the third parameters in $\tilde{y}_4$.
7. A mean-shift occurs to all the parameters in $\tilde{y}_4$.

Then we use the method that was introduced in section 3.1.1 to estimate the mean vector ($\tilde{\theta}_j$) in order. And we input them into the formula (3.6). We will get a value for each

situation, and we will have 7 values. Finally, we choose a situation whose value is the minimum of the 7 values, and let the change situation of $\tilde{y}_4$ be that situation. In the same way, we can decide all the change of the segmentation points.

3.2.3 Decide the best solution from all the possible change situations

In this section, I will use the formula (3.3) to decide the number of the segmentation points. And we can find the best solution of the formula (3.3). If there are N observation vectors now, the possible number of the segmentation points is from 0 to $N-1$. Assume that the number of the segmentation points is r and $0 \le r \le N-1$. Then we can use the method that was introduced in the section 3.2.1 to find the first r segmentation points. And we can use the method that was introduced in section 3.2.2 to decide the change situations of the first r segmentation points. Let r be from 0 to $N-1$. We use the method that was introduced in section 3.1.1 to estimate $\tilde{\theta}_j$ and put them into the formula (3.6) for each r. Finally, we only want to find a change situation whose value is the minimum.

3.3 The influence of outliers

In real case, because the semiconductor process is quite complex, sometimes data must be taken down by people, or some unpredictable measurement deviations occur to the measurement apparatus. Outliers might occur in the process. In the section 2.2, SBS provided an outlier detection method to avoid the influence of outliers. In this section, I will use the outlier detection method that was introduced in the section 2.2 to avoid the influence of outliers.

The outlier detection method includes two steps:

(1) Choosing the possible outlier first.

(2) Checking the possible outliers in order. If the possible outlier will affect the result, we

will consider that the possible outlier is a outlier and delete it.

The usual method of outlier detection is to check that is any observation lower than $Q_1 - S \times IQR$ or higher than $Q_3 + S \times IQR$. If there is an observation lower than $Q_1 - S \times IQR$ or higher than $Q_3 + S \times IQR$, we consider that the observation is an outlier. In step (1), SBS lets $S=1.5$, so I also let $S=1.5$ here. And we use the same method to find a possible outlier. That is, for each group, If any observation that belongs to the same group is lower than $Q_1 - 1.5 \times IQR$ or higher than $Q_3 + 1.5 \times IQR$, we will consider that the observation is a possible outlier.

If a possible outlier exists, in step (2), we will delete the possible outlier from the data and segment the data again. If we can get the same result, we consider that the possible outlier is not an outlier. If we can't get the same result, we consider that the possible outlier is an outlier and delete it. When the remaining data is too few, we may not consider outlier detection.

3.4 The algorithm

(1) Let input data = initial data, and initial group number(K_0) = 1.

(2) Use the input data to estimate Σ under K_i group.

(3) Use $\hat{\Sigma}$ to segment each group, and get new group number K_{i+1}.

(4) Find a possible outlier from each new group. If a possible outlier exists, go to step (5). Else, go to step (6).

(5) Check "Is the possible outlier an outlier?" If the possible outlier is an outlier, then delete the outlier from the input data, let $K_i = 1$, and go to step (3). Else, go to step (6).

(6) Check "Does K_i equal K_{i+1}?" If not equal, then let $i = i + 1$ and go to step (2). Else, break and output the result.

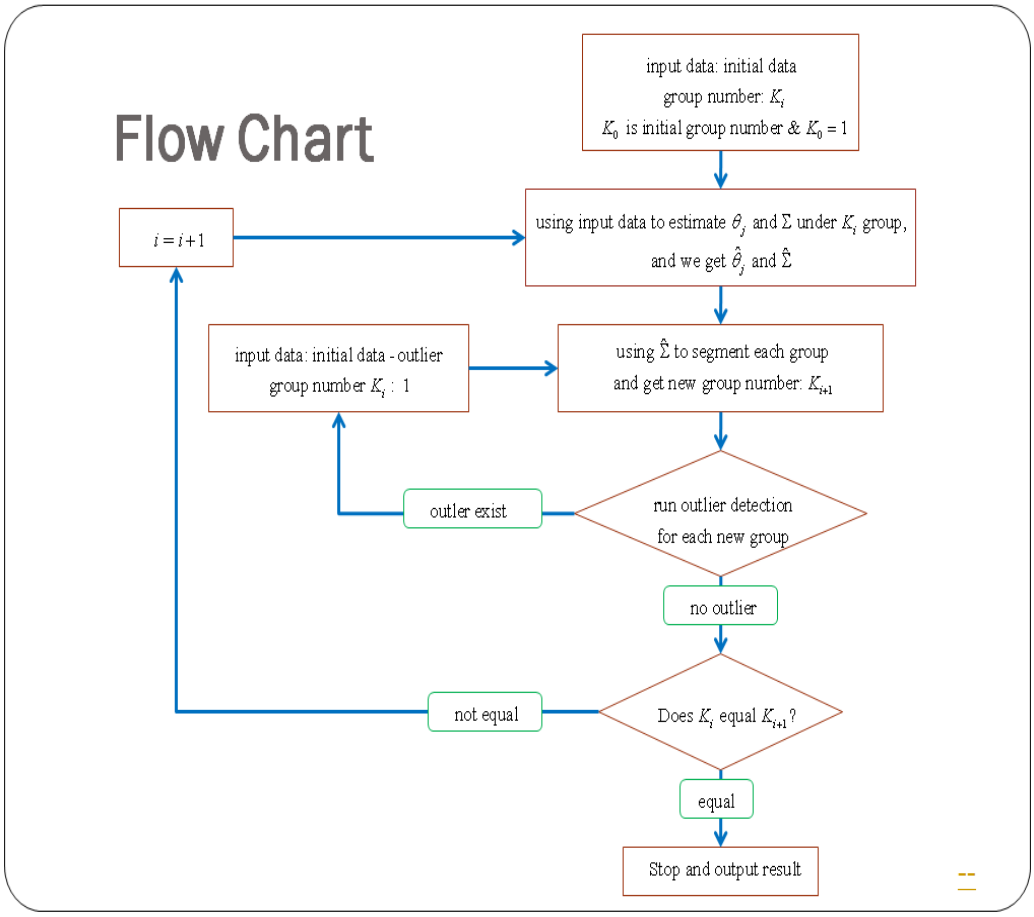


Figure 10. The process flow of MBS

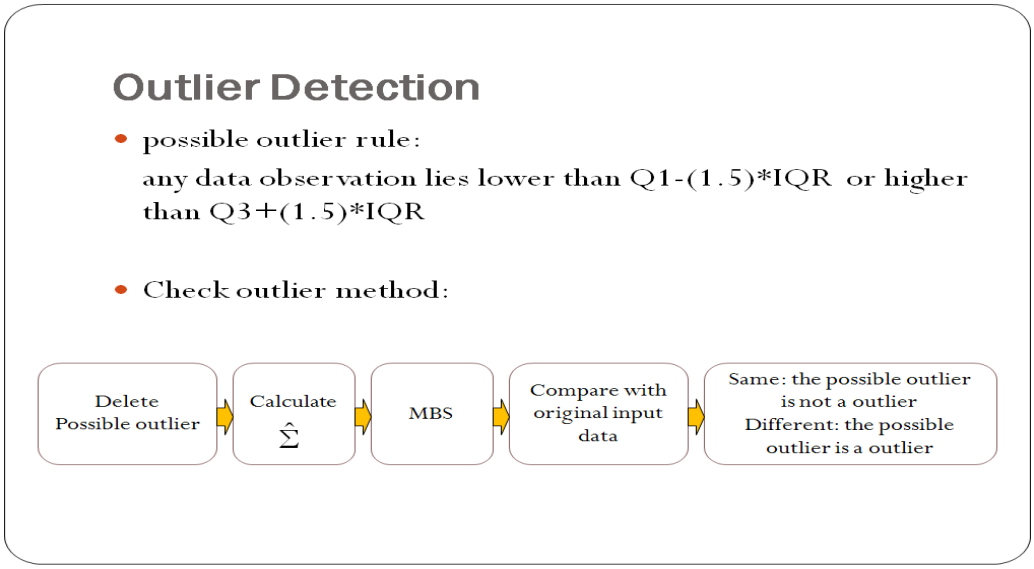


Figure 11. The process flow of the outlier detection (MBS)

Chapter 4. Experiment

In this chapter, we will discuss many cases of three dimensional simulation results. We can use these cases to compare Multivariate Bayesian Segmentation with Single Variable Bayesian Segmentation. In each case, the data are simulated by joint distribution of several distributions ($E_1, E_2, \cdots$), where E_k is a multivariate normal distribution. $E_1, E_2, \cdots$ with the same covariance matrix(Σ) but different mean vectors. That is,

$$E_k \sim \text{Multinormal}(\tilde{\mu}_k, \Sigma);\ k=1,2,\cdots,$$

where $\tilde{\mu}_k = \begin{bmatrix} \mu_{k(1)} \\ \mu_{k(2)} \\ \mu_{k(3)} \end{bmatrix}$ is the mean vector of E_k,

$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ is the covariance matrix of E_k.

Here, I assume that the correlation matrix is a AR1-matrix, and σ_m is the standard deviation of the m-th parameter, $m = 1, 2, 3$.

$$\Rightarrow \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_1 \times \sigma_1 & \rho \times \sigma_1 \times \sigma_2 & \rho^2 \times \sigma_1 \times \sigma_3 \\ \rho \times \sigma_1 \times \sigma_2 & \sigma_2 \times \sigma_2 & \rho \times \sigma_2 \times \sigma_3 \\ \rho^2 \times \sigma_1 \times \sigma_3 & \rho \times \sigma_2 \times \sigma_3 & \sigma_3 \times \sigma_3 \end{bmatrix}.$$

The main cases we want to discuss following below:

(1) No shift.

(2) Shift three times and balanced.

(3) Shift three times and unbalanced.

Here, no shift means that there are no segmentation points in the data, and shift three times means there are three segmentation points in the data. Balanced means that every group's number is the same, and the levels of the shifts are the same. Unbalanced means that every group's number is different, and the levels of the shifts are different.

For each main case, we will discuss the levels of the correlation coefficient (ρ), so we will let ρ equal 0, 0.4, 0.8. We will discuss the effect of the standard deviations ($\sigma_1, \sigma_2, \sigma_3$), so we will let them be the same or be different. We will discuss the effect of an outlier, so we will replace two outliers from the data whose values are 5 and 6 times IQR higher than the third quartile. In each case, we will simulate 500 times to the simulated classification rate.

4.1 No Shift Case

Table 1 Result1 of no shift

					No Adding outlier		**Adding outlier**	
case	**mean**	**size**	$(\sigma_1, \sigma_2, \sigma_3)$	ρ	**SBS**	**MBS**	**SBS**	**MBS**
1	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	50	(1,1,1)	0	97.6%	98.2%	98.2%	98.2%
2		50	(1,1,1)	0.4	97.4%	97.8%	98.2%	98.8%
3		50	(1,1,1)	0.8	97.8%	98.4%	99%	99.6%
4	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	50	(1,2,3)	0	96.8%	97.2%	98.4%	98.4%
5		50	(1,2,3)	0.4	97.8%	98.6%	98.8%	99%
6		50	(1,2,3)	0.8	97.6%	98.6%	99.2%	99.8%

Table 2 Result2 of no shift

					No Adding outlier		Adding outlier	
case	**mean**	**size**	$(\sigma_1,\sigma_2,\sigma_3)$	ρ	**SBS**	**MBS**	**SBS**	**MBS**
7	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	100	(1,1,1)	0	97.2%	98%	98.4%	98.6%
8		100	(1,1,1)	0.4	95.8%	97.4%	98.6%	99.2%
9		100	(1,1,1)	0.8	98.4%	98.6%	98.8%	99.8%
10	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	100	(1,2,3)	0	96.2%	97.6%	98.6%	98.4%
11		100	(1,2,3)	0.4	97.6%	98.4%	98.8%	98.8%
12		100	(1,2,3)	0.8	97.8%	98.6%	99.4%	99.8%

From Table 1 and Table 2, we can find that the results of MBS are all better than the results of SBS for any situation.

4.2 Shift three times and Balanced Case

Table 3 Result1 of shift three times and balanced

					No Adding outlier		Adding outlier	
case	**mean**	**size**	$(\sigma_1,\sigma_2,\sigma_3)$	ρ	**SBS**	**MBS**	**SBS**	**MBS**
13	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}3\\0\\0\end{bmatrix}\begin{bmatrix}3\\3\\0\end{bmatrix}\begin{bmatrix}3\\3\\3\end{bmatrix}\right)$	(30,30,30,30)	(1,1,1)	0	57.4 %	51.8 %	60.8 %	56.3 %
14		(30,30,30,30)	(1,1,1)	0.4	60.8 %	63.2 %	61.6 %	61.2 %
15		(30,30,30,30)	(1,1,1)	0.8	61 %	88.4 %	61 %	84 %

Table 4 Result2 of shift three times and balanced

					No Adding outlier		Adding outlier	
case	**mean**	**size**	$(\sigma_1,\sigma_2,\sigma_3)$	ρ	**SBS**	**MBS**	**SBS**	**MBS**
16	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}3\\0\\0\end{bmatrix}\begin{bmatrix}3\\4.2\\0\end{bmatrix}\begin{bmatrix}3\\4.2\\5.2\end{bmatrix}\right)$	(30,30,30,30)	(1,2,3)	0	61 %	55.2 %	63.4 %	51.6 %
17		(30,30,30,30)	(1,2,3)	0.4	58 %	60.2 %	58.6 %	58.4 %
18		(30,30,30,30)	(1,2,3)	0.8	60.4 %	89.4 %	61.4 %	81.4 %
19	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}3\\0\\0\end{bmatrix}\begin{bmatrix}3\\3\\0\end{bmatrix}\begin{bmatrix}3\\3\\3\end{bmatrix}\right)$	(50,50,50,50)	(1,1,1)	0	57.8 %	53.2 %	59.6 %	55.4 %
20		(50,50,50,50)	(1,1,1)	0.4	61 %	62.8 %	61.4 %	61.4 %
21		(50,50,50,50)	(1,1,1)	0.8	61.2 %	88.8 %	60.8 %	84.8 %
22	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}3\\0\\0\end{bmatrix}\begin{bmatrix}3\\4.2\\0\end{bmatrix}\begin{bmatrix}3\\4.2\\5.2\end{bmatrix}\right)$	(50,50,50,50)	(1,2,3)	0	60.8 %	53.8 %	61.2 %	52.8 %
23		(50,50,50,50)	(1,2,3)	0.4	57.8 %	61 %	58 %	58.2 %
24		(50,50,50,50)	(1,2,3)	0.8	61.2 %	89.8 %	60.6 %	80.6 %

From Table 3 and Table 4, we can find that if the correlation coefficient (ρ) is 0 and that the results of MBS are worse than the results of SBS. If the correlation coefficient (ρ) is bigger than 0.4, the results of MBS are better than the results of SBS. And if we add outliers into the data, we still can find the same results.

4.3 Shift three times and Unbalanced Case

Table 5 Result of shift three times and unbalanced

					No Adding outlier		Adding outlier	
case	**mean**	**size**	$(\sigma_1,\sigma_2,\sigma_3)$	ρ	**SBS**	**MBS**	**SBS**	**MBS**
25	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix}\begin{bmatrix}2\\7\\0\end{bmatrix}\begin{bmatrix}2\\7\\10\end{bmatrix}\right)$	(10,25,50,35)	(1,1,1)	0	58 %	56.4 %	54 %	52.8 %
26		(10,25,50,35)	(1,1,1)	0.4	54 %	60.2 %	54.6 %	55.8 %
27		(10,25,50,35)	(1,1,1)	0.8	59.2 %	70.4 %	60 %	73.6 %
28	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix}\begin{bmatrix}2\\9.9\\0\end{bmatrix}\begin{bmatrix}2\\9.9\\17\end{bmatrix}\right)$	(10,25,50,35)	(1,2,3)	0	60.8 %	58 %	56.8 %	52 %
29		(10,25,50,35)	(1,2,3)	0.4	62.8 %	65.4 %	56 %	57 %
30		(10,25,50,35)	(1,2,3)	0.8	63.2 %	71.4 %	56.2 %	78.6 %
31	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix}\begin{bmatrix}2\\7\\0\end{bmatrix}\begin{bmatrix}2\\7\\10\end{bmatrix}\right)$	(15,35,70,30)	(1,1,1)	0	57.2 %	56.2 %	53.8 %	52.6 %
32		(15,35,70,30)	(1,1,1)	0.4	54.4 %	61.2 %	54.8 %	55.2 %
33		(15,35,70,30)	(1,1,1)	0.8	58.4 %	71.2 %	60.2 %	74.4 %
34	$\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\end{bmatrix}\begin{bmatrix}2\\9.9\\0\end{bmatrix}\begin{bmatrix}2\\9.9\\17\end{bmatrix}\right)$	(15,35,70,30)	(1,2,3)	0	61.4 %	58 %	58.2 %	54.2 %
35		(15,35,70,30)	(1,2,3)	0.4	62.2 %	64.8 %	56.8 %	57.6 %
36		(15,35,70,30)	(1,2,3)	0.8	64 %	72.8 %	54.8 %	76.4 %

From Table 5, we can find that if the correlation coefficient (ρ) is 0 and the results of MBS are worse than the results of SBS. If the correlation coefficient (ρ) is bigger than 0.4,

the results of MBS are better than the results of SBS. And if we add outliers into the data, we still can find the same results.

Chapter 5. Conclusion

In this paper, MBS is used to segment many parameters' historical data at a time and find the position where the change occurs to the process. In Chapter 4, we used many differences to compare MBS with SBS. From the comparison process and the result, we can find:

(1) If there are many different parameters, SBS must be run several times, but MBS only need to run one time. so MBS is more efficient than SBS.

(2) Under the no shift case, the results of MBS are always better than the results of SBS.

(3) Under shift three times and balanced case, if the degree of the correlation coefficient is bigger, the result of MBS is better.

(4) Under shift three times and unbalanced case, if the degree of the correlation coefficient is bigger, the result of MBS is also better.

So, if there are many parameters in the data, and the correlation coefficient is not small, it is suitable to use MBS to operate the data. MBS may be applied in the semiconductor industry, because there are many different parameters relating to the process. If an engineer uses MBS to find the segmentation point, we only need to run it one time. The engineer will work more efficiently. And we can use MBS to get a right segmentation point with consider the effect of the parameters' interaction. SBS can't be used to do that.

In future work:

(1) When correlation coefficient is small, SBS is better than MBS. When correlation coefficient is large, MBS is better than SBS. So we can investigate a method that can determine whether MBS or SBS shall be used.

(2) In this paper, our condition is laid in multivariate normal distribution with unchanged covariance matrix. Therefore, we could discuss the classification of random assignment material with the changes of covariance matrix in the future. If the covariance matrix has changed, we need to discuss how to detect the position where the covariance matrix has changed.

Reference

[1] Zheng-Yan Lai, “Improving a Regression Tree Application in Promoting Semiconductor Good Rate”, National Chiao Tung University, Master, 2008.

[2] A.R. Taylor and S.R. Duncan, “Case studies in Bayesian segmentation applied to CD control”, INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, Vol.17, pp.363-392, 2003.

[3] Marc Lavelle, “Optimal Segmentation of Random Processes”, IEEE TRANSACTIONS ON SIGNAL PROCESSING, Vol.46, No.5, MAY 1998.

[4] François Bergeret and Caroline Le Gall, “Yield Improvement Using Statistical Analysis of Process Dates”, IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING, Vol.16, No.3, AUGUST 2003.

[5] K.D. Zamba and Douglas M. Hawkins, “A Multivariate Change-Point Model for Statistical Process Control”, TECHNOMETRICS, Vol.48, No.04, pp.539-549, 2006.

[6] Chung-Hsien Wu and Chia-Hsin Hsieh, “Multiple Change-Point Audio Segmentation and Classification Using an MDL-Based Gaussian Model”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vol.14, No.4, pp.647-657, 2006.

[7] D. Aguado , A. Ferrer , J. Ferrer, A. Seco, “Multivariate SPC of a sequencing batch reactor for wastewater treatment”, Elsevier, pp.82-93, 2007.

[8] Zhen Wang and Peter Willett, “Two Algorithms to Segment White Gaussian Data with Piecewise Constant Variances”, IEEE TRANSACTIONS ON SIGNAL PROCESSING, Vol.51, No.2, pp.373-385, 2003.

[9] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., “Classification and Regression Trees”, Wadsworth, Belmont, California, 1984.

[10] Esposito, F. Malerba, D., and Semeraro, G., “A Comparative Analysis of Methods for Pruning Decision Trees”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 5, pp.476-491, 1997.

[11] Neretti, N., Remondini, D., Tatar, M., Sedivy, J.M., Pierini, M., Mazzatti, D., Powell, J., Franceschi, C., and Castellani, G.C., “Correlation analysis reveals the emergence of coherence in the gene expression dynamics following system perturbation”, BMC Bioinformatics 2007, 8(Suppl 1):S16(8 March 2007).

[12] Taylor, W.A., “Change-point analysis: A powerful new tool for detecting changes”, 2000, http://www.variation.com/cpa/tech/changepoint.html