

## *DD*-plot for Detecting the Out-of-Control State in Multivariate Process

Dae-Heung Jang<sup>a</sup> · Seongbaek Yi<sup>a</sup> · Youngil Kim<sup>b,1</sup>

<sup>a</sup>Department of Statistics, Pukyong National University

<sup>b</sup>Division of Management, Chung-Ang University

(Received October 29, 2012; Revised December 17, 2012; Accepted March 20, 2013)

---

### Abstract

It is well known that the *DD*-plot is a useful graphical tool for non-parametric classification. In this paper, we propose another use of *DD*-plot for detecting the out-of-control state in multivariate process. We suggested a dynamic version of *DD*-plot and its accompanying a quality index plot in such case.

Keywords: Multivariate process, data depth, dynamic *DD*-plot.

---

### 1. 서론

자료깊이(data depth)는 지난 20여년 동안 다변량통계의 다양한 영역에서 강력한 도구로 사용되어 오고 있다. 그 이유는 이 방법이 분포의 중심성(centrality)을 알기 위한 비모수적이고 강건한 통계량을 제시하기 때문이다. 자료깊이에는 simplicial depth, Oja depth, Mahalanobis depth, half-space depth, projection depth, convex hull peeling depth, majority depth, likelihood depth 등과 같은 다양한 종류가 있으며 자료깊이의 개념에 대해서는 Liu 등 (1999)와 Li 등 (2012)에 잘 제시되어 있다. *DD*-plot이란 두 개의 다변량분포를 비교하기 위하여 이 두 개의 다변량분포에서 추출된 두 개의 표본을 대상으로 서로의 자료깊이를 계산하여 산점도 형태로 그린 그림이다. Li 등 (2012)은 시뮬레이션을 통하여 자료깊이를 기반으로 하는 이 *DD*-plot이 비모수적 분류자로서 탁월한 성능이 있음을 밝혔다. 또한 자료의 일부가 오염되어 있거나 특이값이 존재해도 강건함을 보였다.

다변량공정에서 이상상태를 탐지하기 위한 도구로서 우리는 주로 다변량  $T^2$  관리도를 사용한다. 그러나 이러한 다변량  $T^2$  관리도를 사용하기 위해서는 다변량공정이 다변량정규분포를 따른다는 가정이 필요하다. 그러므로 다변량공정이 다변량정규분포를 따르지 않는 경우 우리는 다변량  $T^2$  관리도를 사용하기 곤란하다. 다변량공정에서 이상상태를 탐지하기 위한 또 다른 접근방법으로서 Liu (1995)는 자료깊이에 기반을 둔 관리도( $r$  관리도,  $Q$  관리도,  $S$  관리도)를 제안하였다. 또한 Liu와 Singh (1993)은 두 개의 다변량분포를 비교하기 위하여 이 두 개의 다변량분포에서 추출한 두 개의 표본을 대상으로 서로의 자료깊이를 계산하여 얻어지는 통계량인 품질지수(quality index)를 제시하고 이 통계량을 이용한 비모수검정을 제안하였다. 본 논문을 통하여 우리는 *DD*-plot이 분류문제 뿐만이 아니라 다변량공정의 이상

---

<sup>1</sup>Corresponding author: Professor, Division of Management, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 153-756, Korea. E-mail: yik01@cau.ac.kr

상태를 탐지하기 위한 그래픽 도구로서 유용하게 쓰일 수 있음을 보이고자 한다. 이러한  $DD$ -plot은 다변량공정의 분포를 특정분포(예로 다변량정규분포)로 가정할 필요가 없는 분포무관방법이므로 모든 다변량공정의 이상상태를 탐지하는 데 사용할 수 있다. 다변량  $T^2$  관리도의 목적이 평균 변화를 탐지하는 데 있는 반면  $DD$ -plot은 특정 모수의 비교가 아닌 다변량분포의 이상 상태를 전체적으로 탐지하는 데 있다. 또한 본 논문을 통하여 우리는 다변량공정의 변화에 따른  $DD$ -plot의 변화를 동적으로 보기 위한 그래픽 수단으로서 동적(dynamic)  $DD$ -plot과 동적 품질지수그림을 제안하고자 한다. 우리는 이러한 동적  $DD$ -plot과 동적 품질지수그림을 사용하여 시간의 경과에 따른 다변량공정의 변화를 동적으로 확인하여 다변량공정의 이상상태를 동적으로 탐지할 수 있게 된다.

2절과 3절에서 각각 모의실험과 실제 자료에 대하여  $DD$ -plot이 다변량공정의 이상상태를 탐지하기 위한 그래픽 도구로서 유용하게 쓰일 수 있음을 보였고 4절에서 결론을 내렸다.

## 2. 모의실험자료에 대한 $DD$ -plot

본 절에서는 모의실험을 통하여  $DD$ -plot이 다변량공정의 이상상태를 탐지하기 위한 그래픽 도구로서 유용하게 쓰일 수 있음을 보이고자 한다. 모의실험에서는 비교를 위하여 Park (2012)이 사용한 예를 이용하였다. Park (2012)은 마할라노비스거리(Mahalanobis distance)를 이용하여 공정모수의 재조정 절차에 대하여 연구하였다.

관리상태일 때 평균이  $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (0, 0, 0, 0)$ 이고, 분산-공분산행렬이

$$\Sigma = \begin{pmatrix} 1.000 & -0.424 & 0.957 & 0.922 \\ -0.424 & 1.000 & -0.540 & -0.063 \\ 0.957 & -0.540 & 1.000 & 0.822 \\ 0.922 & -0.063 & 0.822 & 1.000 \end{pmatrix}$$

인 다변량정규분포를 따르는 다변량공정에서 다음과 같은 세 가지 시나리오를 고려한다.

시나리오 1 : 첫 번째 변수의 평균  $\mu_1$ 이 0에서 2로 서서히 변한다.

시나리오 2 :  $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (2, 0, 0, 0)$ 에서 네 번째 변수의 평균  $\mu_4$ 가 0에서 3으로 서서히 변한다.

시나리오 3 : 세 번째 변수의 분산  $\sigma_3^2$ 이 1에서 9로 서서히 변한다.

각 시나리오에서 다변량공정의 변화에 따른  $DD$ -plot의 변화를 동적으로 보기 위한 그래픽 수단으로 우리는 동적  $DD$ -plot를 제안할 수 있다. 이 동적  $DD$ -plot를 이용하면 평균이나 분산의 변화에 대하여  $DD$ -plot의 동적변화를 확인할 수 있다. 컴퓨터 화면에서 슬라이드바를 이용하여 다변량공정에 변화를 주었을 때 동적으로 변하는  $DD$ -plot의 패턴을 통하여 우리는 다변량공정 하에서의 이상상태 탐지를 확인할 수 있게 된다.

시나리오 1: 첫 번째 변수의 평균  $\mu_1$ 이 0에서 2로 서서히 변할 때 동적  $DD$ -plot을 통하여  $DD$ -plot의 변화를 동적으로 볼 수 있다. Figure 2.1은 정상상태하의 다변량공정에서 추출한 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시,  $\mu_1 = 0$ )를 표시한 동적  $DD$ -plot과 동적 품질지수그림이다. 자료깊이는 마할라노비스깊이(Mahalanobis depth)를 이용하였다.  $DD$ -plot의 경우 200개의 점들이 45° 각도로 하나의 직선을 중심으로 흩어져 있음을 알 수 있으며 이를 통하여 검사용 데이터 100개는 정상 데이터임을 확인할 수 있다. 즉 공정이 정상상태임을 알 수 있다. 품질지수그림의 경우

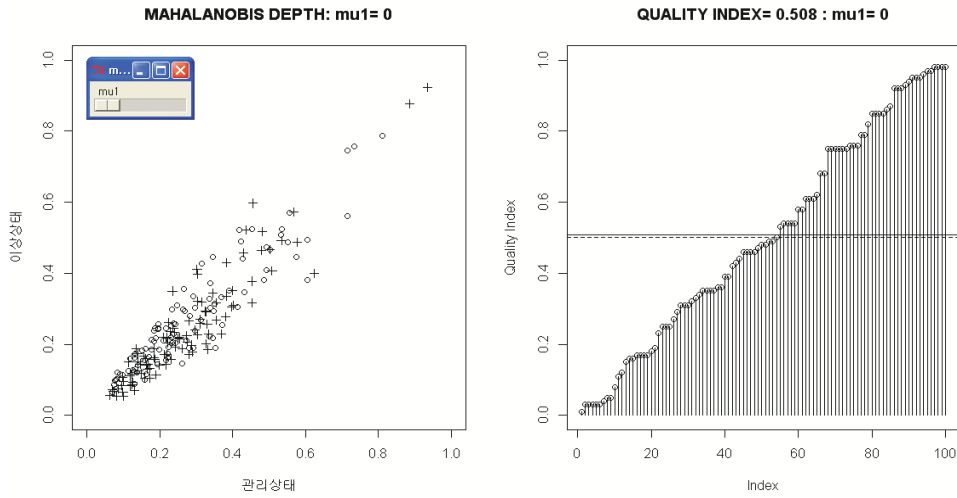


Figure 2.1. Dynamic DD-plot and dynamic quality index plot for the in-control state( $\mu_1 = 0$ ) in a multivariate process using Mahalanobis depth

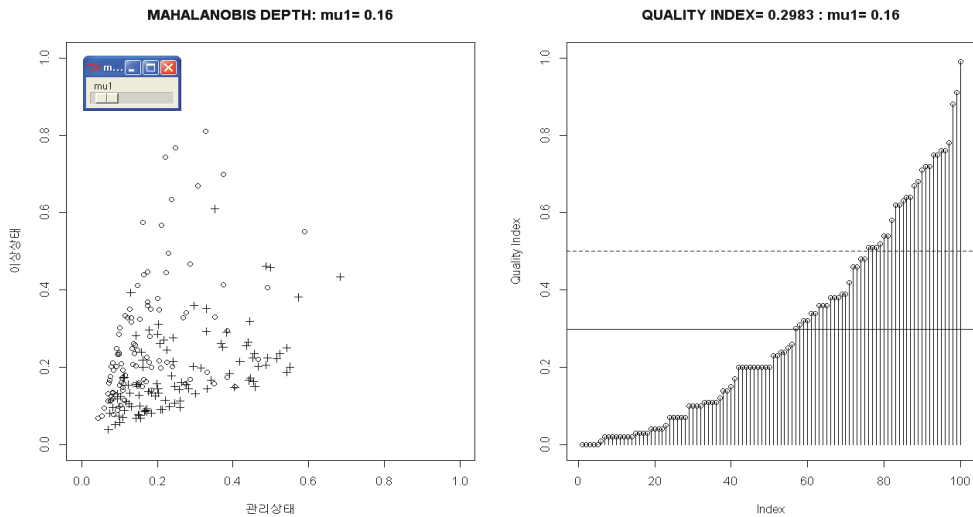
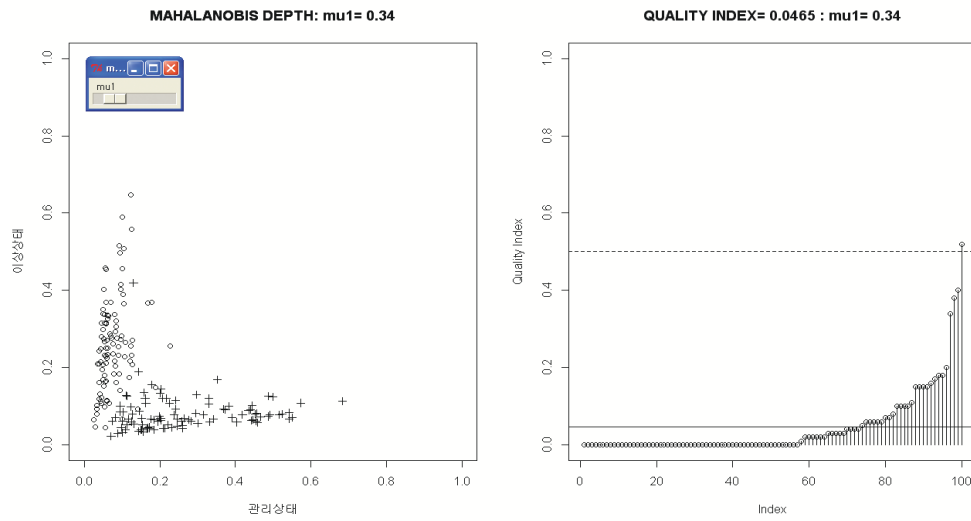


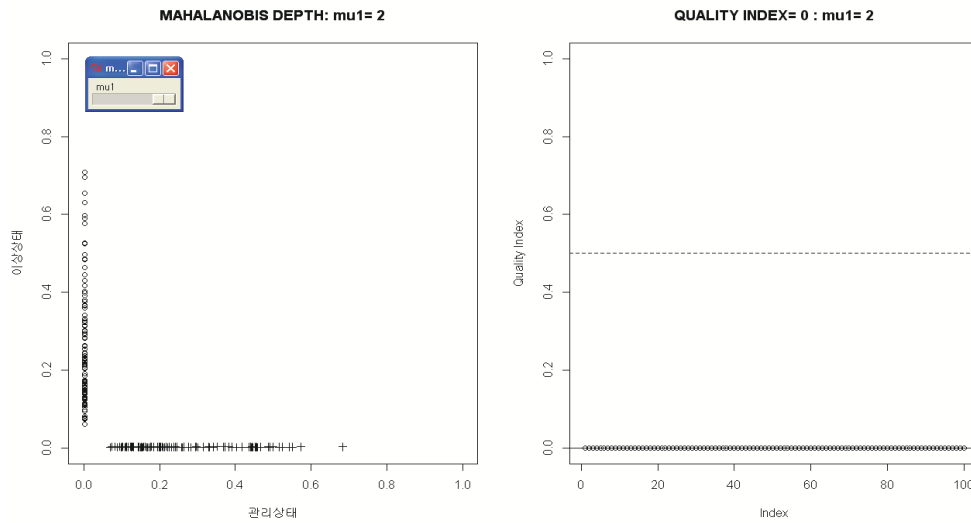
Figure 2.2. Dynamic DD-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 0.16$ ) in a multivariate process using Mahalanobis depth

검사용 데이터 100개에 대한 품질지수 값이 0.5 근방인 0.508에 있으며 이를 통해서도 공정이 정상상태임을 알 수 있다.

Figure 2.2는 마할라노비스깊이를 이용하여 공정상태가 정상인 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시,  $\mu_1 = 0.16$ )를 표시한 동적 DD-plot과 동적 품질지수그림이다. 정상 데이터와 이상 데이터가 한 직선상에 놓이지 않고 서로 혼재되어 있으면서도 분리되기 시작함을 알 수 있다. 검사용 데이터 100개에 대한 품질지수 값이 0.2983임을 통해서도 공정이 이상상태로 이동하고 있음을



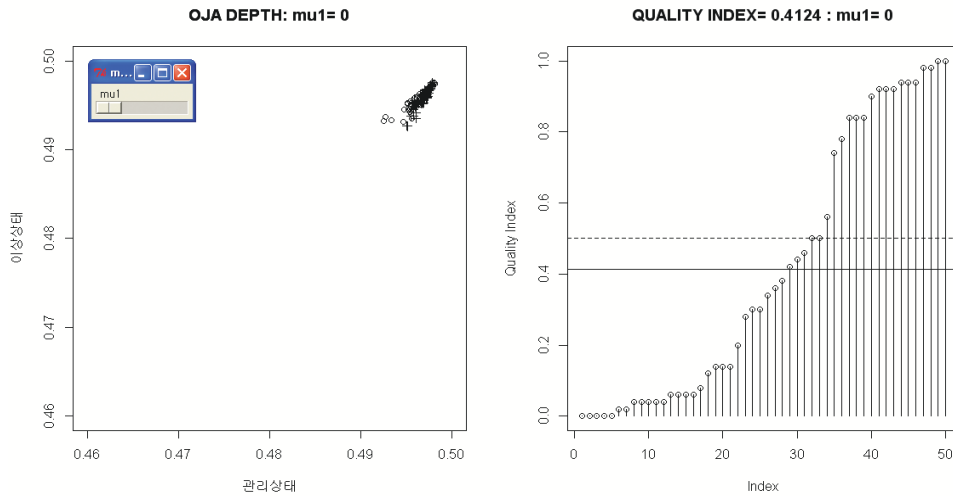
**Figure 2.3.** Dynamic *DD*-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 0.34$ ) in a multivariate process using Mahalanobis depth



**Figure 2.4.** Dynamic *DD*-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 2$ ) in a multivariate process using Mahalanobis depth

알 수 있다.

Figure 2.3은 마할라노비스깊이를 이용하여 공정상태가 정상인 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시,  $\mu_1 = 0.34$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. *DD*-plot에서 정상 데이터와 이상 데이터가 거의 대부분 분리되어 있음을 알 수 있다. 품질지수그림에서 검사용 데이터 100개에 대한 품질지수 값은 0.0465이므로 유의수준 0.05에서 귀무가설( $H_0$  : 검사용 데이터는 정상그룹이다)를 기각할 수 있고 정상상태와 이상상태를 거의 확실하게 구분할 수 있다.



**Figure 2.5.** Dynamic *DD*-plot and dynamic quality index plot for the in-control state ( $\mu_1 = 0$ ) in a multivariate process using Oja depth

Figure 2.4는 마할라노비스깊이를 이용하여 공정상태가 정상인 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시,  $\mu_1 = 2$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. *DD*-plot의 경우 정상 데이터와 이상 데이터가 완전 분리되어 있음을 알 수 있다. 품질지수그림의 경우 검사용 데이터 100개에 대한 품질지수 값이 0임을 통해서도 정상상태와 이상상태를 확실하게 구분할 수 있다. 이상과 같은 절차를  $\mu_1$ 의 값을 서서히 변화시키면서 동적 *DD*-plot과 동적 품질지수그림을 살펴보면  $\mu_1 \geq 0.68$ 일 때부터 정상상태와 이상상태를 확실하게 구분할 수 있음을 알 수 있다.

우리는 다른 자료깊이를 이용하여 동적 *DD*-plot을 그려 볼 수 있으며 다른 깊이를 사용하여도 우리는 비슷한 결과를 얻는다. 예로 Oja depth를 이용하여 첫 번째 변수의 평균  $\mu_1$ 을 0에서 2로 서서히 변화시키면서 동적 *DD*-plot을 통하여 *DD*-plot의 변화를 동적으로 확인해 볼 수 있다. Figure 2.5는 Oja depth를 이용하여 공정상태가 정상인 정상 데이터 50개(십자표시)와 검사용 데이터 50개(동그라미표시,  $\mu_1 = 0$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. *DD*-plot의 경우 100개의 점들이 각도가  $45^\circ$ 인 하나의 직선을 중심으로 주위에 집중되어 늘어서 있음을 알 수 있다. 이를 통하여 검사용 데이터 50개는 정상 데이터임을 알 수 있다. 즉 공정이 정상상태임을 확인할 수 있다. 검사용 데이터 50개에 대한 품질지수그림의 경우 품질지수 값이 0.4124로서 0.5에서 크게 벗어나지 않음을 통해서도 역시 공정이 정상상태임을 알 수 있다.

Figure 2.6은 Oja depth를 이용하여 공정상태가 정상인 정상 데이터 50개(십자표시)와 검사용 데이터 50개(동그라미표시,  $\mu_1 = 0.29$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. *DD*-plot의 경우 정상 데이터와 이상 데이터가 거의 분리되어 있음을 알 수 있다. 검사용 데이터 50개에 대한 품질지수 값 또한 0.0424이어서 유의수준 0.05에서 귀무가설( $H_0$  : 검사용 데이터는 정상그룹이다)를 기각할 수 있고 정상상태와 이상상태를 거의 확실하게 구분할 수 있다.

Figure 2.7은 Oja depth를 이용하여 공정상태가 정상인 정상 데이터 50개(십자표시)와 검사용 데이터 50개(동그라미표시,  $\mu_1 = 2$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. *DD*-plot의 경우 정상 데이터와 이상 데이터가 각도가  $45^\circ$ 인 직선을 중심으로 완전히 분리되어 있음을 알 수 있다. 또한 검사

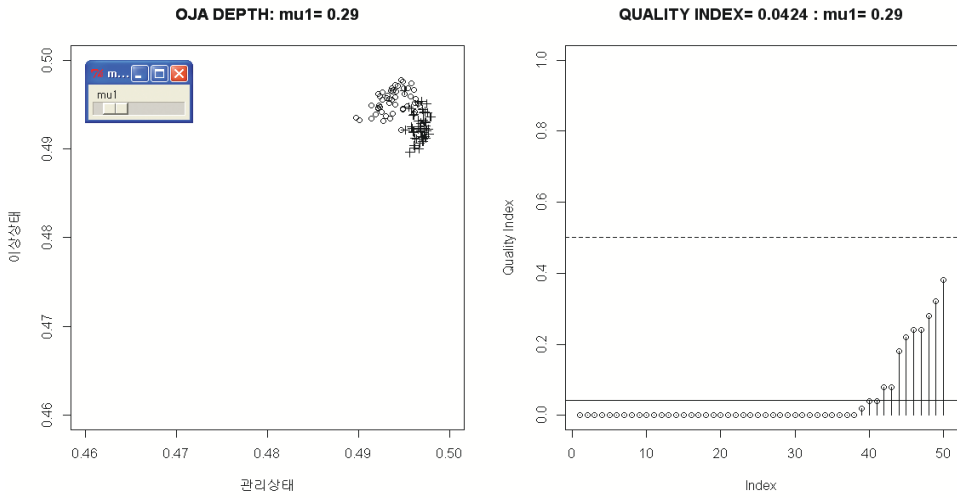


Figure 2.6. Dynamic *DD*-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 0.36$ ) in a multivariate process using Oja depth

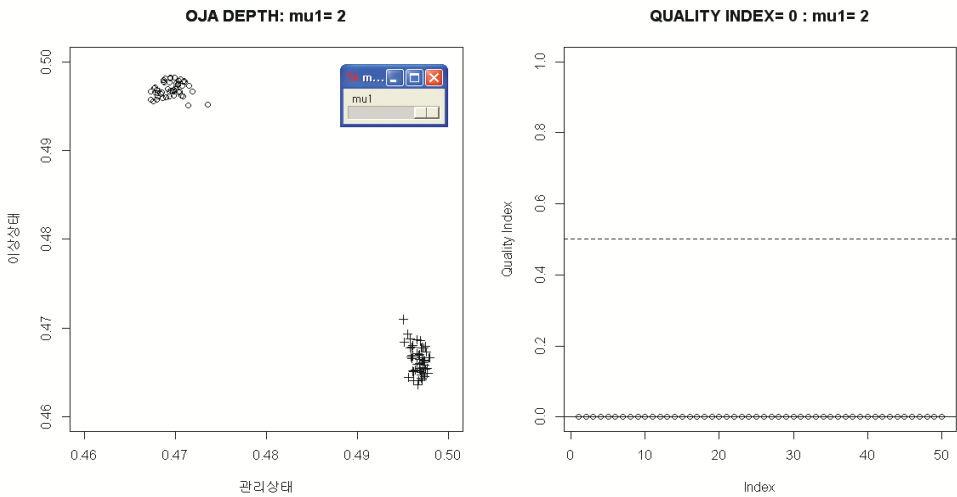


Figure 2.7. Dynamic *DD*-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 2$ ) in a multivariate process using Oja depth

용 데이터 50개에 대한 품질지수 값이 0임을 통해서도 정상상태와 이상상태를 확실하게 구분할 수 있다. 이상의 절차를  $\mu_1$ 을 서서히 변화시키면서 동적 *DD*-plot과 동적 품질지수그림을 이용하면  $\mu_1 \geq 0.47$ 일 때부터 정상상태와 이상상태를 확실하게 구분할 수 있음을 알 수 있다.

시나리오 2:  $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (2, 0, 0, 0)$ 에서 네 번째 변수의 평균  $\mu_4$ 가 0에서 3으로 서서히 변할 때 동적 *DD*-plot을 통하여 *DD*-plot의 변화를 동적으로 확인할 수 있다. Figure 2.8은 마할라노비스길 이를 이용하여 공정상태가 정상인 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시),

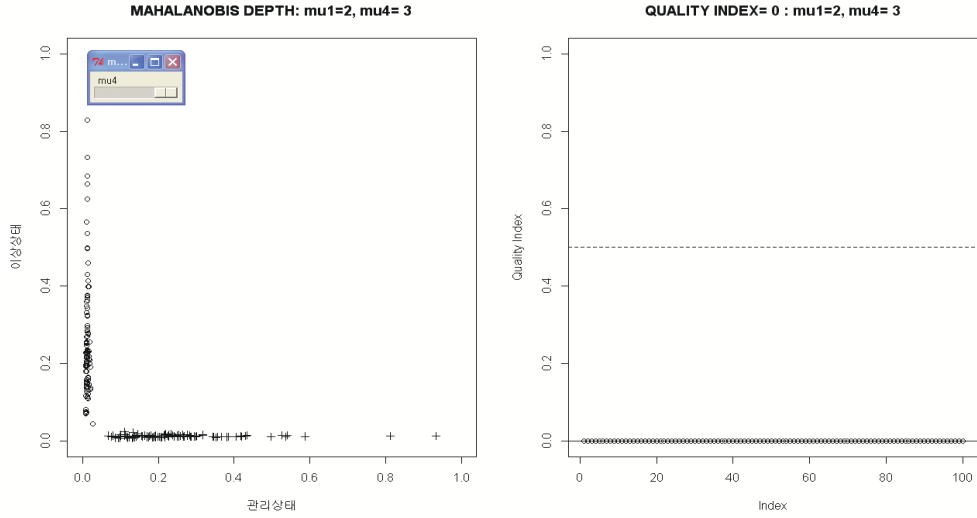


Figure 2.8. Dynamic DD-plot and dynamic quality index plot for the out-of-control state( $\mu_1 = 2, \mu_4 = 3$ ) in a multivariate process using Mahalanobis depth

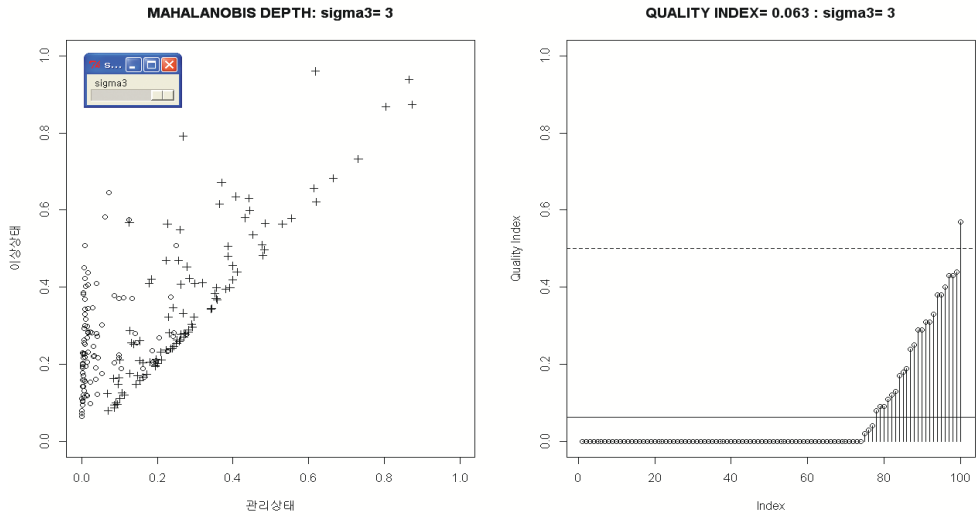


Figure 2.9. Dynamic DD-plot and dynamic quality index plot for the out-of-control state( $\sigma_3^2 = 9$ ) in a multivariate process using Mahalanobis depth

$\mu_1 = 2, \mu_4 = 3$ )를 표시한 동적 DD-plot과 동적 품질지수그림이다.  $\mu_1 = 2$ 인 상태에서 추가로  $\mu_4$ 가 0에서 3으로 서서히 변화할 때 검사용 데이터 100개에 대한 품질지수의 값은 0으로 변화가 없어 정상 데이터와 이상 데이터가 완전 분리되어 있음을 알 수 있으나  $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (2, 0, 0, 0)$ 일 때보다 완전 분리되는 정도는 오히려 더 약해졌음을 알 수 있다. 우리는 Figure 2.4와 Figure 2.8을 비교함으로써 이러한 현상을 발견할 수 있다. 이러한 현상은 다변량 분포에서의 모수간 상관관계에 기인하며 Park (2012)에서 언급한 사실과 일치한다.

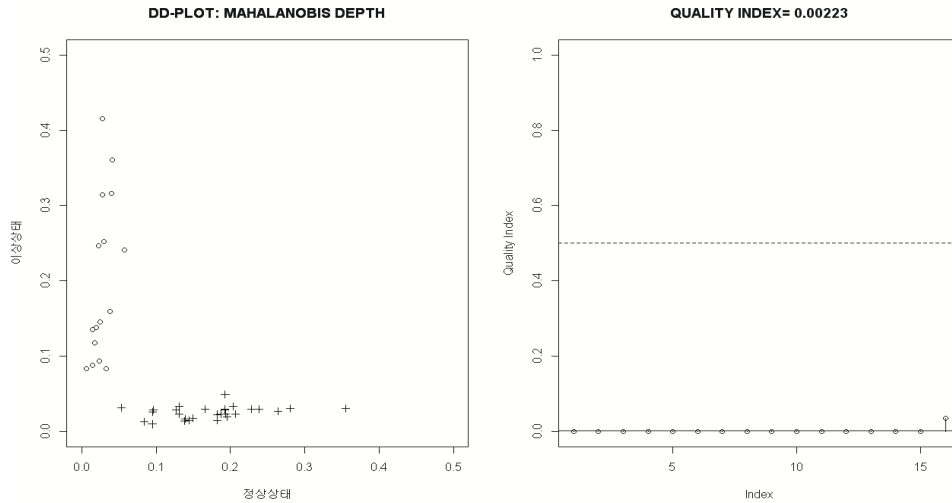


Figure 3.1. *DD*-plot and quality index plot for Steam Turbine data using Mahalanobis depth

**시나리오 3:** 세 번째 변수의 분산  $\sigma_3^2$ 이 1에서 9로 서서히 변할 때 *DD*-plot의 변화를 동적 *DD*-plot을 통하여 동적으로 확인할 수 있다. Figure 2.9는 마할라노비스깊이를 이용하여 공정상태가 정상인 정상 데이터 100개(십자표시)와 검사용 데이터 100개(동그라미표시,  $\sigma_3^2 = 9$ )를 표시한 동적 *DD*-plot과 동적 품질지수그림이다. 평균의 변화(시나리오 1, 시나리오 2)인 경우는 정상상태와 이상상태를 훨씬 용이하게 구분할 수 있는 반면 분산의 변화인 경우는 Figure 2.9처럼 품질지수 값이 0.063이어서 정상상태와 이상상태의 구분이 평균의 변화(시나리오 1, 시나리오 2)인 경우보다 상대적으로 더 어려움을 알 수 있다. 이러한 사실은 Park (2012)에서 언급한 사실과 일치한다.

### 3. 실제자료에 대한 *DD*-plot

본절에서는 두 가지 실제 자료를 이용하여 *DD*-plot이 다변량공정의 이상상태를 탐지하기 위한 그래픽 도구로서 유용하게 쓰일 수 있음을 보이고자 한다. 첫 번째 자료는 증기터빈 데이터 (Mason과 Young, 2002)로서 정상상태 하의 다변량공정 데이터(십자표시)는 변수가 6개(fuel, steam flow, steam temperature, megawatts, cooling temperature, pressure)이고 자료의 크기가 28개인 그룹으로 이루어져 있다. 검사용 데이터(동그라미표시)는 변수가 6개이고 데이터의 크기가 16개인 그룹으로 이루어져 있다. 이 검사용 데이터가 정상그룹인지 비정상그룹인지를 알아보기 위한 *DD*-plot과 품질지수그림은 Figure 3.1과 같다. *DD*-plot의 경우 정상 데이터와 검사용 데이터가 완전히 분리되어 있으며 품질지수그림의 경우 품질지수 값이 0임을 통하여 검사용 데이터는 비정상그룹임을 알 수 있다.

두 번째 자료는 과일쥬스 데이터 (Fuchs와 Kenett, 1998)로서 정상상태 하의 다변량공정 데이터(십자표시)는 변수가 11개(11개의 아미노산, 즉 Lysine, Arginine, Aspartic Acid, Serine, Glutamine Acid, Proline, Glycine, Alanine, Valine, Phenyl Alanine, Gamma-amino Acid Butric Acid)이고 자료의 크기가 36개인 그룹으로 이루어져 있다. 검사용 데이터(동그라미표시)는 변수가 11개이고 데이터의 크기가 33개인 그룹으로 이루어져 있다. 이 검사용 데이터가 정상그룹인지 비정상그룹인지를 알아보기 위하여 그린 *DD*-plot과 품질지수그림은 Figure 3.2와 같다. *DD*-plot의 경우 정상상태의 자료와 검사용 자료가 각도가  $45^\circ$ 인 직선 주위에 혼재되어 집중되어 있고 또한 품질지수그림에서 품질지수 값이 0.5에서



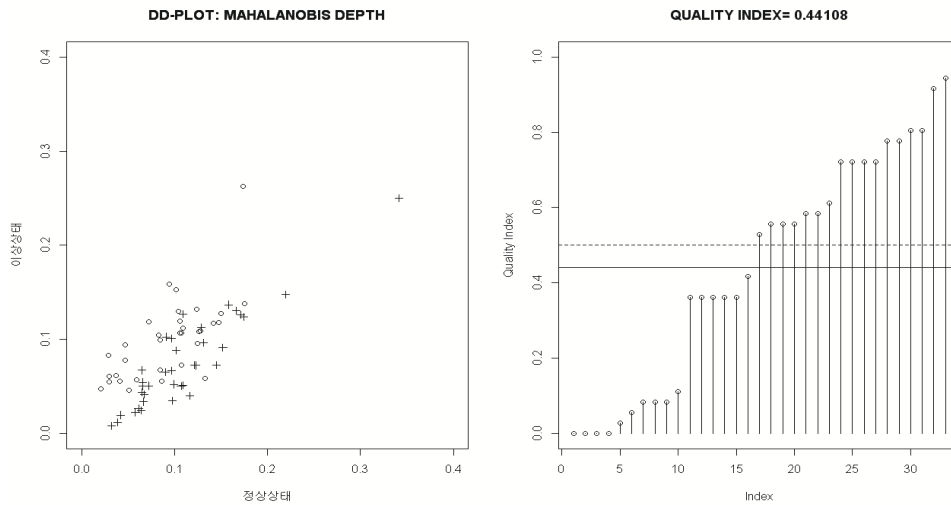


Figure 3.2. DD-plot and quality index plot for Fruit Juice data using Mahalanobis depth

크게 벗어나지 않음을 통하여 검사용 데이터는 정상그룹임을 확인할 수 있다.

#### 4. 결론

우리는 다변량공정에서 이상상태를 탐지하기 위한 그래픽 도구로서 DD-plot을 사용할 수 있다. DD-plot은 다변량공정의 분포를 특정분포로 가정할 필요가 없는 분포무관방법이므로 모든 다변량공정의 이상상태를 탐지하는 데 사용할 수 있다. DD-plot를 그릴 때 우리는 다양한 자료깊이를 사용하여 다변량공정의 이상상태 탐지결과들을 서로 비교하여 볼 수 있다. 또한 다변량공정의 변화에 따른 DD-plot의 변화를 동적으로 보기위한 그래픽 수단으로 동적 DD-plot와 동적 품질지수그림을 사용할 수 있다.

#### References

- Fuchs, C. and Kenett, R. (1998). *Multivariate Quality Control with Industrial Applications*, Marcel Dekker, New York.
- Li, J., Cuesta-Albertos, J. A. and Liu, R. (2012). DD-Classifier: Nonparametric classification procedure based on DD-plot, *Journal of the American Statistical Association*, **107**, 737–753.
- Liu, R. (1995). Control charts for multivariate processes, *Journal of the American Statistical Association*, **90**, 1380–1387.
- Liu, R., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference, *The Annals of Statistics*, **27**, 783–858.
- Liu, R. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *Journal of the American Statistical Association*, **88**, 252–260.
- Mason, R. L. and Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*, ASA-SIAM, Philadelphia.
- Park, C. (2012). A resetting scheme for process parameters using the Mahalanobis-Taguchi system, *The Korean Journal of Applied Statistics*, **25**, 589–603.

# 다변량공정에서 이상상태를 탐지하기 위한 *DD*-plot

장대흥<sup>a</sup> · 이성백<sup>a</sup> · 김영일<sup>b,1</sup>

<sup>a</sup>부경대학교 통계학과, <sup>b</sup>중앙대학교 경영학부

(2012년 10월 29일 접수, 2012년 12월 17일 수정, 2013년 3월 20일 채택)

---

## 요약

*DD*-plot은 분류문제를 풀기 위한 유용한 비모수적 방법이다. 우리는 이러한 *DD*-plot을 다변량공정에서 이상상태를 탐지하기 위한 그래픽 방법으로 사용할 수 있다. 본 논문을 통하여 이상상태를 탐지하기 위한 그래픽 방법으로서 동적 *DD*-plot과 동적 품질지수그림을 제시하고자 한다.

주요용어: 다변량공정, 자료깊이, 동적 *DD*-plot.

---

<sup>1</sup>교신저자: (153-756) 서울특별시 동작구 흑석로 84, 중앙대학교 경영학부, 교수. E-mail: yik01@cau.ac.kr