

## Notes on identifying source of out-of-control signals in phase II multivariate process monitoring

Sungim Lee<sup>a,1</sup>

<sup>a</sup>Department of Applied Statistics, Dankook University

(Received January 3, 2018; Revised January 17, 2018; Accepted January 17, 2018)

---

### Abstract

Multivariate process control has become important in various applied fields. For instance, there are many situations in which the simultaneous monitoring of multivariate quality characteristics is necessary for the manufacturing industry. Despite its importance, its practical usage is not as convenient because it is difficult to identify the source of the out-of-control signal in a multivariate control chart. In this paper, we will introduce how to detect the source of the out-of-control by using confidence intervals for new observations, and will discuss the identification and interpretation of the out-of-control variable through simulation studies.

Keywords: Hotelling's  $T^2$ , multivariate statistical process control, identification of the out-of-control variable

---

### 1. 서론

일변량 품질특성치  $x$ 가 공정평균이  $\mu$ 이고, 공정분산이  $\sigma^2$ 인 분포를 따른다고 가정할 때, Shewhart (1926)가 제안한  $\bar{X}$  관리도는 공정평균이 정상상태인지 이상상태인지를 모니터링하기 위한 것으로, 새로운 관측치가 관리한계선을 넘는다면 그 해석은 자연스럽게 품질특성치의 공정평균이 변화했다는 것을 의미한다. 공정평균이  $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$ 이고 공정공분산행렬이  $\Sigma = (\sigma_{ij})_{p \times p}$ 인 다변량 품질특성치  $\mathbf{x}^T = (X_1, X_2, \dots, X_p)$ 의 경우에는 공정평균의 변화를 모니터링하기 위해 품질특성치별로 각각의  $\bar{X}$  관리도를 작성하기보다, 다변량 관측치들은 품질특성치 간에 상관성이 높기 때문에 일반적으로 호텔링의  $T^2$  관리도 같은 다변량 관리도를 이용하여 모니터링 하는 것이 공정의 변화를 좀 더 민감하게 감지할 수 있다고 알려져 있다 (Montgomery, 2005). 실제로 Kenett와 Halevy (1984)는 미국 국방부에서 발행한 군수품의 규격을 설정할 때, 제품 특성치들 간에 서로 상관관계가 있기 때문에 다변량 접근이 필요함을 보여주었고, Kang과 Kim (2016) 등도 실제 산업 분야에서 다변량 품질특성치에 대한 다변량 관리도의 필요성을 강조하였다. 그런데, 다변량 관리도로부터 이상신호가 발생한 경우에는 일변량 관리도와는 달리 이에 대한 해석이 그리 간단하지가 않다. 공정평균벡터를 모니터링할 때 이상신호가 발생했다면 그 이유는 그 시점에  $p$ 개의 변수 중 어떤 한 변수의 공정평균이 변화했거나, 둘 혹은 그 이상의 변수들 간의 직선관계가 관리상태하에서의 관계와 다름에 기인할 수 있다. 또는 그 둘의 복합적 이유로 어떤 변수는 공정평균이 변화했고 동시에 다른 변수들 간의 직선관계가 반대로 나타났을 수도 있을 것이

---

This Research was supported by the Dankook University Research Grants 2016.

<sup>1</sup>Department of Applied Statistics, Dankook University, 152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do 16890, Korea. E-mail: [silee@dankook.ac.kr](mailto:silee@dankook.ac.kr)

다. Jackson (1991)은 다변량 품질관리절차에서 이상신호가 발생할 경우 이에 대한 해석을 제시해야 한다고 강조하고 있는데, 다변량 관리도를 이용한 모니터링은 공정의 변화를 감지한다고 해도 무엇이 문제인지 그 의미를 해석하고 이와 관련한 조치를 취하기가 간단하지 않아 실제 문제에서는 그 필요성 만큼 활용되지 못하는 것이 현실이다.

물론 호텔링의  $T^2$  관리도에서 이상신호의 원인을 식별하기 위한 다양한 연구가 있어 왔다 (Alt, 1985; Doganaksoy 등, 1991; Hayter과 Tsui, 1994; Mason 등, 1995, 1997; Timm, 1996 등). Alt (1985)는 가짜알람률을 Bonferroni 방식으로 조정하여 변수별로  $\bar{X}$  관리도의 관리한계선을 제시하였다. Mason 등 (1995)은 호텔링의  $T^2$  통계량을 서로 독립적인  $p$ 개의 항으로 직교분할하였는데, 각 항은 개별적인 변수의 기여도를 반영하도록 했다. 이를 Mason-Young-Tracy (MYT) 분해라고도 한다.  $T^2$  통계량에 대한 직교분해를 생각하면 주성분을 이용한 직교분해를 생각해볼 수 있는데, 주성분은 모든 변수들의 선형 결합으로 나타나기 때문에 이상신호의 원인을 개별변수의 관점에서 해석하기 쉽지 않다. 이로 인해 다변량관리도에서는 이상신호에 대해 해석이 가능한 MYT 분해가 널리 알려져 있다. 그런데, MYT 분해는  $p$ 개의 품질특성치에 대해 가능한 분할방식의 수가  $p \times p!$ 개 존재하며 실제 계산해야할 서로 다른 항은  $p \times 2^{p-1}$ 개로  $p$ 가 큰 경우에는 비효율적인 뿐 아니라 각 항을 해석하기가 간단하지 않다. Mason 등 (1997)은 좀 더 합리적인 계산량으로 이상상태의 변수를 선택하는 알고리즘을 제시하였고, 이것은 R의 MSQC 패키지에서도 이용 가능하다. 이 절차는 이상신호가 발생하면 우선 각 변수에 대한 개별적인 검정 통계량으로부터 관리상태하에서의 공정평균과 다르다고 할 수 있는지 검정을 하게 된다. 이들 연구에서 알 수 있듯이 다변량관리도에서 이상신호가 발생한 후에 다시 개별적인 변수의 관리상태를 평가하는 것은 매우 자연스러운 단계이다. 본 논문에서는 호텔링의  $T^2$  관리도를 통해 이상신호가 발생했을 경우 변수별로 공정평균 변화유무를 탐지하기 위해 새로운 관측치에 대해 변수별 신뢰영역을 구하는 서로 다른 세 가지 방법을 알아보고, 이들 방법 간에 어떠한 차이점이 있는지 알아보기로 한다. 또한, 이에 대한 성능을 비교 평가함으로써 실제 다변량 품질특성치의 모니터링에서 이상신호 발생 시 그 원인을 파악하고 해석할 때 유의할 사항을 알아보도록 한다. 각 진단방식을 시각적으로 이해하기 위해 본 논문에서는  $p = 2$ 인 이변량 데이터에 대한 이단계 모니터링(phase II monitoring)으로 문제를 한정하고, 관리상태 하에서의 평균벡터나 공분산행렬이 알려지지 않은 경우를 가정하였다. 이단계 모니터링이란 과거 데이터 또는 참조표본으로부터 관리상태 하에서의 모수를 추정한 후, 미래 관측치에 대해 관리상태 여부를 판단하는 모니터링을 지칭하는 것으로서 실질적인 공정 모니터링은 대부분 여기에 해당된다고 할 수 있다. 이에 반해 일단계 모니터링(phase I monitoring)이란 수집된 과거 데이터의 관리상태 여부를 판정하는 것이며, 이를 통해 관리상태 하에서의 모수를 추정하게 된다.

논문의 구성은 다음과 같다. 본 연구에서는 이변량 품질특성치의 평균벡터에 대한 이단계 모니터링을 가정하고, 2절에서는 호텔링의  $T^2$  관리도로부터 이상신호가 발생할 경우 이에 대한 원인이 되는 변수를 식별하기 위한 방법으로, 관리상태하에서 기대되는 새로운 관측값에 대해 개별적인 신뢰구간을 찾는 세 가지 방법을 소개한다. 3절에서는 2절에서 소개한 세 가지 신뢰구간에 어떠한 차이점이 있는지 모의실험을 통해 비교 분석해 보고자 한다. 4절에서는 연구결과를 요약하고 앞으로의 연구방향에 대해 고찰해 보기로 한다.

## 2. 개별 변수에 대한 신뢰구간으로부터 이상신호의 원인 식별

이변량 데이터  $\mathbf{x}^T = (X_1, X_2)$ 는 관리상태하에서 평균벡터가  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ 이고, 공분산 행렬이  $\boldsymbol{\Sigma} = (\sigma_{ij})_{2 \times 2}$ 인 이변량 정규분포를 따른다고 가정하자. 실제 문제에서 관리상태하에서의 이들 모수 값은 알려지지 않은 경우가 대부분이므로 과거의 데이터로부터 모수를 추정하여 이를 통해 관리한계선을 추정하고, 미래 데이터의 이상상태 발생 여부를 모니터링하게 된다. 일반적으로  $m \times 2$  행렬로 구성된 이변

량 데이터  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)^T$ 가 관리상태하에서의 과거 데이터로 주어졌다고 가정한다. 이로부터 관리상태하에서의 평균벡터와 공분산행렬을 다음과 같이 추정한다

$$\begin{aligned}\bar{\mathbf{x}} &= \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i, \\ S_{\mathbf{x}} &= \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.\end{aligned}$$

새로운 이변량 데이터  $\mathbf{y}^T = (Y_1, Y_2)$ 로부터  $\mu$ 의 변화를 모니터링하기 위한 호텔링  $T^2$  통계량은

$$T^2 = (\mathbf{y} - \bar{\mathbf{x}})^T S_{\mathbf{x}}^{-1} (\mathbf{y} - \bar{\mathbf{x}}) \sim \frac{2(m+1)(m-1)}{m(m-2)} F_{(2, m-2)} \quad (2.1)$$

로 주어진다 (Montgomery, 2005). 이것은 공정평균이 관리상태라면 새로운 관측치  $\mathbf{y}$ 가 참조표본의 평균벡터와 공분산행렬을 기반으로 타원형의 경계 안에서 관측된다는 것을 의미한다. 서론에서 언급했듯이, 새로운 관측치가 관리한계선 밖에 존재할 때 이러한 이상신호는 어떤 변수에 기인한 것인지 해석하는 절차가 매우 중요한데, 식 (2.1)로부터 구하게 되는 새로운 관측치에 대한 신뢰영역으로는 이러한 해석이 쉽지 않다. 이에 개별적인 품질특성치에 대한 신뢰구간을 통해 원인 변수를 식별하려는 노력은 매우 자연스러운 절차이다. 즉, 다변량 품질특성치의 변화를 모니터링 할 때 다변량 검정통계량이 효과적이지만, 이상신호가 발생한 후 그 원인을 식별하는데 있어서는 개별적인 검정통계량을 확인할 필요가 있다.

이변량 데이터에 대하여 각 변수별로  $100(1-\alpha)\%$  신뢰영역을 찾는 방법은 크게 세 가지 방법이 존재한다. 첫 번째 방법은 식 (2.1)을 이용하여 새로운 데이터  $\mathbf{y}$ 의 모든 가능한 선형함수들에 대한 동시신뢰영역(simultaneous confidence region)으로부터 개별변수에 대한 신뢰구간을 찾는 방법으로, 신뢰구간을 구할 때  $F$ 분위수를 사용하게 되어 이 방법을 본 논문에서는 S-신뢰구간 추정법이라고 부르기로 하고 정리 2.1에 소개한다.

**정리 2.1** (S-신뢰구간) 호텔링  $T^2$  통계량을 이용하여 공정평균벡터의 변화에 대한 이단계 모니터링을 실시할 때, 새로운 이변량 관측치  $\mathbf{y}$ 가 관리상태하에 있다면, 각 변수에 대한  $100(1-\alpha)\%$  신뢰구간은 다음과 같다.

$$\bar{X}_i \pm \sqrt{\frac{2(m+1)(m-1)}{m(m-2)} F_{(\alpha, 2, m-2)} \sqrt{S_{ii}}}. \quad (2.2)$$

이 때,  $F_{(\alpha, 2, m-2)}$ 는 자유도  $(2, m-2)$ 인  $F$ -분포의 제  $100(1-\alpha)$ 백분위수를 나타낸다.  $\bar{X}_i$ 와  $S_{ii}$  ( $i = 1, 2$ )는 참조표본에서 구한 각 변수별 표본평균과 표본분산을 나타내고,  $m$ 은 참조표본의 크기를 나타낸다.

증명: 새로운 데이터  $\mathbf{y}$ 가 관리상태하에서의 관측값이라면, 식 (2.1)로부터 다음이 성립함을 알 수 있다.

$$\begin{aligned}P \left\{ \max_a \frac{(a^T \mathbf{y} - a^T \bar{\mathbf{x}})^2}{a^T S_{\mathbf{x}} a} \leq \frac{2(m+1)(m-1)}{m(m-2)} F_{(\alpha, 2, m-2)} \right\} &= 1 - \alpha \\ \Leftrightarrow P \left\{ (a^T \mathbf{y} - a^T \bar{\mathbf{x}})^2 \leq \frac{2(m+1)(m-1)}{m(m-2)} F_{(\alpha, 2, m-2)} \cdot a^T S_{\mathbf{x}} a \text{ for any } a \right\} &= 1 - \alpha \\ \Leftrightarrow P \left\{ a^T \mathbf{y} \in a^T \bar{\mathbf{x}} \pm \sqrt{\frac{2(m+1)(m-1)}{m(m-2)} F_{(\alpha, 2, m-2)} a^T S_{\mathbf{x}} a} \text{ for any } a \right\} &= 1 - \alpha.\end{aligned} \quad (2.3)$$

이상신호에 대한 원인을 식별할 때 일차적 관심은 각 변수별 변화에 있으므로,  $\mathbf{a} = (1, 0)^T$  또는  $\mathbf{a} = (0, 1)^T$ 로 하여 신뢰영역을 찾아보면 식 (2.2)와 같다.  $\square$

식 (2.2)에서 구한 변수별 S-신뢰구간은 직사각형으로 나타나는데, 만약 관리상태하에서의  $(\mu, \Sigma)$ 가 알려져 있다면 타원형의 신뢰영역을 각각의 변수축에서 둘러싼 영역으로 나타낸다. 따라서 새로운 관측치가 타원안에서 관측되어 관리상태하에 있다면 S-신뢰구간에서도 이상신호로 탐지되지 않아 가짜알람률은 영이 된다. 두번째 방법은 Mason 등 (1997)에서 소개한 절차로 새로운 관측치의 각 변수로부터 이상상태 유무를 판단하기 위해 다음의 검정통계량을 사용한다.

$$T_i^2 = \frac{(Y_i - \bar{X}_i)^2}{S_{ii}} \sim \frac{(m+1)}{m} F_{(1, m-1)}. \quad (2.4)$$

이로부터 새로운 변수별 신뢰구간을 구할 수 있는데 이를 본 논문에서는 M-신뢰구간 추정법이라 부르기로 하고 정리 2.2에 소개한다.

**정리 2.2** (M-신뢰구간) 새로운 이변량 관측치  $\mathbf{y}$ 가 관리상태하에 있다면, 각 변수에 대한  $100(1 - \alpha)\%$  신뢰구간은 다음과 같다.

$$\bar{X}_i \pm \sqrt{F_{(\alpha, 1, m-1)}} \sqrt{\left(1 + \frac{1}{m}\right) S_{ii}}. \quad (2.5)$$

R의 MSQC 패키지의 `mult.chart` 함수는 식 (2.1)에 대해 이상신호가 나타날 경우, 개별적인 변수에 대해 식 (2.4)를 이용하여 검정한다. 세 번째 방법은 새로운 데이터  $\mathbf{y}$ 의 변수별 신뢰구간을 구하면서 전체 유의수준을 Bonferonni 방법으로 수정한 것으로 이를 본 논문에서는 B-신뢰구간 추정법이라 부르기로 하고 정리 2.3에 소개한다.

**정리 2.3** (B-신뢰구간) 새로운 이변량 관측치  $\mathbf{y}$ 가 관리상태하에 있다면, 각 변수에 대한  $100(1 - \alpha)\%$  신뢰구간은 다음과 같다.

$$\bar{X}_i \pm t_{\left(\frac{\alpha}{2p}, m-1\right)} \sqrt{\left(1 + \frac{1}{m}\right) S_{ii}} \quad (2.6)$$

이 때,  $t_{(\alpha/2p, m-1)}$ 는 자유도  $(m-1)$ 인  $t$ -분포의 제  $100(1 - \alpha/2p)$ 백분위수를 나타낸다.

증명: 새로운 데이터  $\mathbf{y}$ 가 관리상태하에서의 관측치 일 때, 개별 관측치  $Y_i$ 에 대하여  $E(Y_i - \bar{X}_i) = 0$ ,  $\text{Var}(Y_i - \bar{X}_i) = (1 + 1/m)\sigma_{ii}^2$ 이 성립하여 다음을 만족한다.

$$\begin{aligned} & P \left\{ \frac{|Y_i - \bar{X}_i|}{\sqrt{(1 + 1/m)S_{ii}}} \leq t_{\left(\frac{\alpha}{2p}, m-1\right)} \right\} \geq 1 - \alpha \\ \Leftrightarrow & P \left\{ |Y_i - \bar{X}_i| \leq t_{\left(\frac{\alpha}{2p}, m-1\right)} \cdot \sqrt{\left(1 + \frac{1}{m}\right) S_{ii}} \right\} \geq 1 - \alpha \\ \Leftrightarrow & P \left\{ Y_i \in \bar{X}_i \pm t_{\left(\frac{\alpha}{2p}, m-1\right)} \sqrt{\left(1 + \frac{1}{m}\right) S_{ii}} \right\} \geq 1 - \alpha. \end{aligned} \quad (2.7)$$

□

두 번째와 세 번째에서 구한 신뢰구간은 결국 각 변수에 대한 주변분포에서 구한 신뢰구간이 되는데, M-신뢰구간은 각 변수에 대해  $\alpha$ 검정을 한 것이고, B-신뢰구간은 동시신뢰영역에 대해  $\alpha$ 검정을 하기 위해 개별적인 신뢰구간에 대해 유의수준을 조정한 것이다. 따라서 B-신뢰구간이 조금 더 넓게 나타나는 경향이 있으나 가짜 알람률은 M-신뢰구간에서 크게 나타나게 된다.

### 3. 모의실험

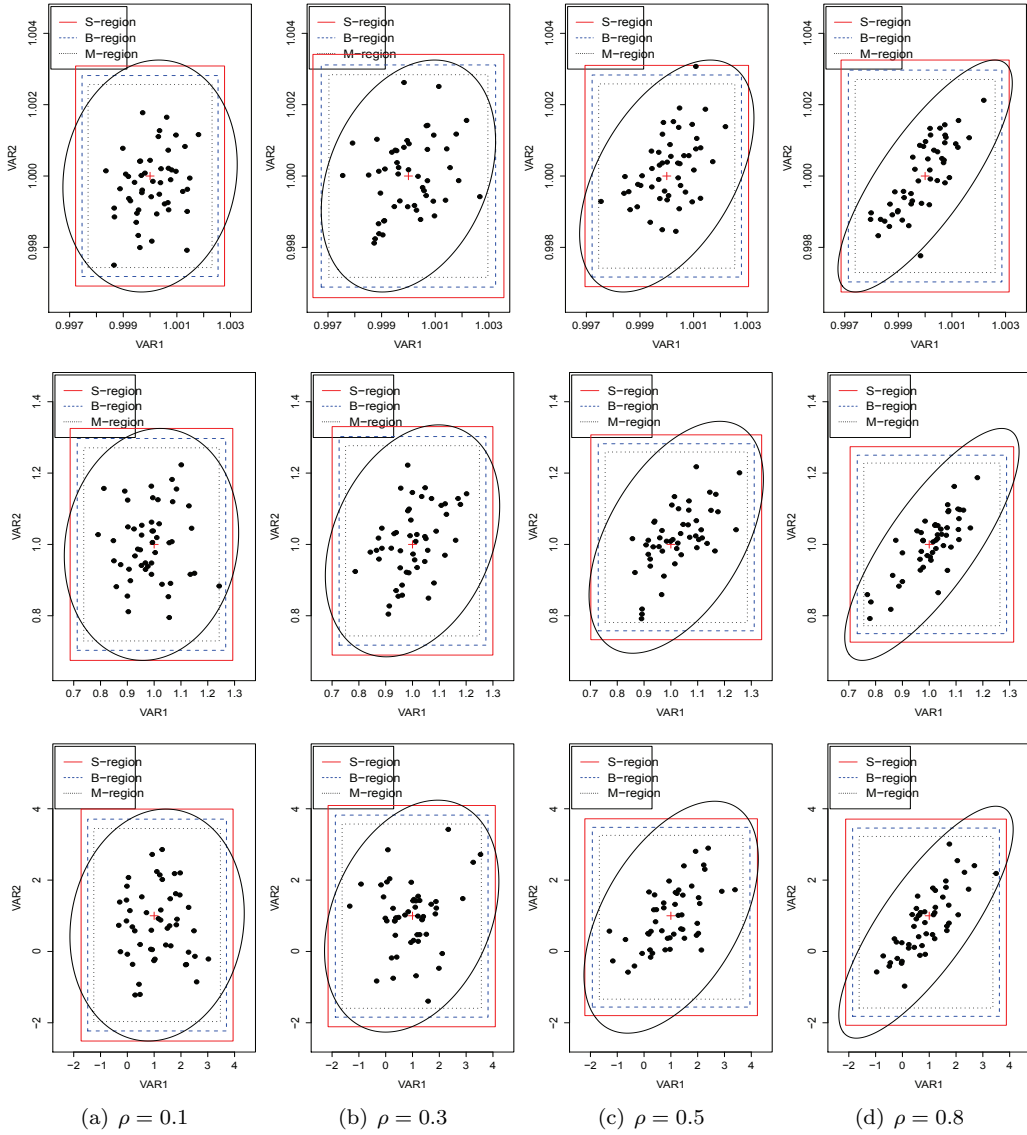
이 절에서는 이변량 데이터에 대해 호텔링  $T^2$  통계량을 통한 이단계 모니터링을 가정하고, 이상신호 발생 후 그 원인파악을 위한 조치로써 품질특성치별로 이상상태 여부를 판단하기 위해 2절에서 소개한 3가지 신뢰구간을 사용할 때 어떠한 차이가 있는지 모의실험을 통해 알아보려고 한다. 즉, (1) 호텔링의  $T^2$  통계량의 동시신뢰영역으로부터 개별 신뢰구간을 이용하는 방법 (식 (2.2)), (2) 품질특성치별로 새로운 관측치에 대한 신뢰구간을 구한 방법 (식 (2.5)), 그리고 (3) 개별 신뢰구간을 구할 때 Bonferroni 방법으로 신뢰수준을 조정된 신뢰구간을 이용하는 방법 (식 (2.6))으로 각각 S-신뢰구간, M-신뢰구간, 그리고 B-신뢰구간을 비교해 보고자 한다. 모의실험을 위해 이변량 데이터  $\mathbf{x}^T = (X_1, X_2)$ 의 분포는 관리상태하에서 이변량 정규확률분포  $(\mu_0, \Sigma)$ 를 따른다고 가정한다. 이 때,  $\mu_0 = (1, 1)^T$ 이고,  $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ 이다.

- (1) 이단계 모니터링을 가정하고 있으므로, 관리상태하의 모수  $(\mu_0, \Sigma)$ 를 추정하기 위한 참조표본의 크기를  $m = 50$ 인 경우와  $m = 1000$ 인 경우를 가정한다.
- (2) 평균벡터의 변화유무에 대해 통계적 유의성을 검정할 때, 분산행렬에 의해 검정력에 차이가 있으므로, CV를 기준으로 하여 세 가지 분산형태를 가정한다. 즉, (i)  $\sigma_1 = \sigma_2 = 0.001$ , (ii)  $\sigma_1 = \sigma_2 = 0.01$ , (iii)  $\sigma_1 = \sigma_2 = 1$ 인 경우를 가정한다.
- (3) 두 변수의 상관정도에 따른 차이를 알아보기 위해  $\rho = 0.1, 0.3, 0.5, 0.8$ 인 경우를 가정한다.
- (4) 평균벡터가 다음과 같이 변화했을 때, 검정력을 비교해 본다. (i)  $\mu_1 = (1 + \sigma_1, 1)^T$ , (ii)  $\mu_2 = (1 - \sigma_1, 1 + \sigma_2)^T$ .

#### 3.1. 참조표본의 크기가 $m = 50$ 인 경우

먼저 관리상태하에서의  $(\mu_0, \Sigma)$ 를 추정하기 위한 참조표본의 크기가  $m = 50$ 인 경우를 가정한다. 참조표본을 추출하여 세 가지의 신뢰구간을 그리면 분산가정에 따라 행렬로 Figure 3.1과 같이 나타난다. 각 행의 그림들은 다시 변수 간 상관정도에 따라 4가지 경우로 나타난다. 이들 그림에서 타원은 공정이 관리상태일 때 새로운 이변량 관측치가 속할 동시신뢰영역을 나타낸다 (식 (2.1)). 동시신뢰영역의 ‘VAR 1’ 방향과 ‘VAR 2’ 방향을 둘러싸는 각 신뢰구간을 S-신뢰구간(실선으로 표시됨)이라고 표현하였는데, Figure 3.1에서 알 수 있듯이 동시신뢰영역이 실선(solid line)을 벗어나는 경우가 생기게 된다. 이것은 Lee (2015)에서도 언급했듯이 이단계 모니터링의 경우 관리상태의 공정평균과 분산행렬을 추정함으로써 발생하는 문제가 된다. 참조표본의 크기가 커짐에 따라 이러한 문제는 보정이 가능한 것으로 나타난다 (3.2절 참조).

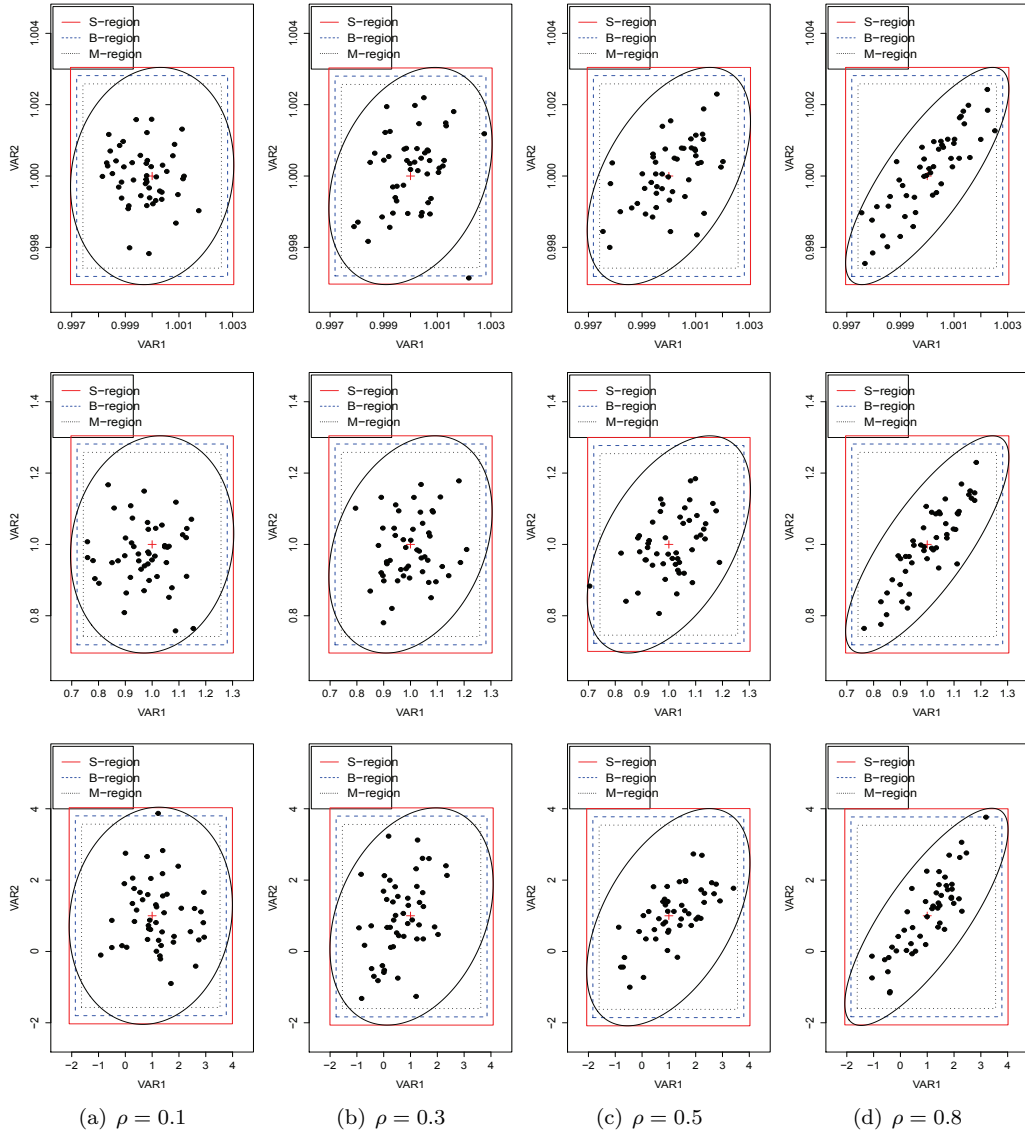
B-신뢰구간은 파선(dashed line)으로 S-신뢰구간보다 조금 더 작은 영역을 구성하는데, 이것은 새로운 데이터가 타원의 경계에 존재할 때 B-신뢰구간에 의해 평균벡터 변화를 모니터링 하는 경우 S-신뢰구간보다 가짜 알람률이 커질 수 있음을 나타낸다. 물론 변수들 간의 상관성이 커짐에 따라 B-신뢰구간에 의해 새로운 관측치의 변화를 확인하는 것이 S-신뢰구간보다 효율적일 수 있다. M-신뢰구간은 점선(dotted line)으로 나타났는데, 가장 작은 신뢰영역을 구성하게 된다. 이것은 가짜 알람률은 커질 수 있지만 문제가 되는 변수를 식별하는데 있어 다른 방법들보다 장점이 있음을 알 수 있다. 그러나 두 변수가 강한 상관관계를 나타낼수록 새로운 관측치에 대하여 개별 신뢰구간에 의해 생성된 신뢰영역과 동시신뢰영역간의 차이가 두드러져 이상신호의 원인을 개별적인 변수로 이해하기 어렵다는 것을 알 수 있다.



**Figure 3.1.** S-confidence region (solid line), B-confidence region (dashed line), M-confidence region (dotted line) for new observations  $\mathbf{y}$  according to  $\rho$  with  $m = 50$  and data points (points) under the control. The first row is the case  $\sigma_1 = \sigma_2 = 0.001$ , the second row is the case  $\sigma_1 = \sigma_2 = 0.1$ , and the third row is the case  $\sigma_1 = \sigma_2 = 1$ .

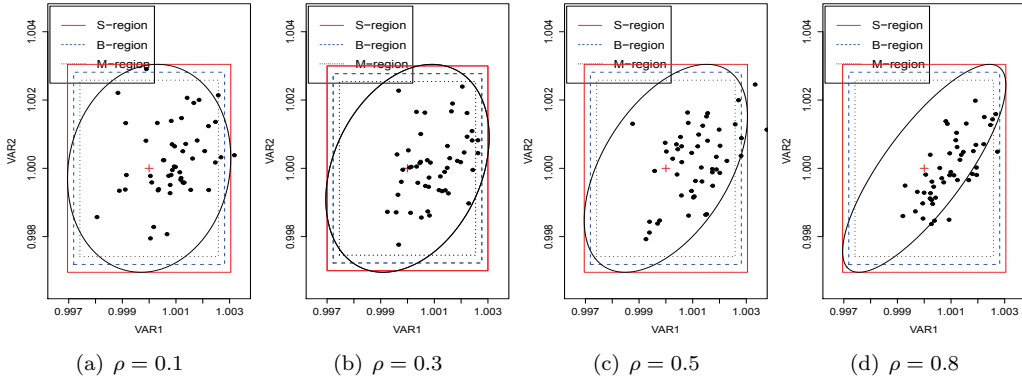
### 3.2. 참조표본의 크기가 $m = 1000$ 인 경우

관리상태하에서의  $(\mu_0, \Sigma)$ 를 추정하기 위한 참조표본의 크기가  $m = 1000$ 인 경우에는  $m = 500$ 인 경우와 비교해볼 때, Figure 3.2에서 알 수 있듯이, S-신뢰구간은 새로운 데이터  $\mathbf{y}$ 에 대한 동시신뢰영역인 타원을 둘러싸는 신뢰구간이 됨을 알 수 있다. 즉, 새로운 데이터가 신뢰영역 안에 있어 관리상태인 경우라면, S-신뢰구간을 벗어나는 경우는 없고, 가짜알람률은 영이 된다. 그러나 새로운 데이터가 이상신

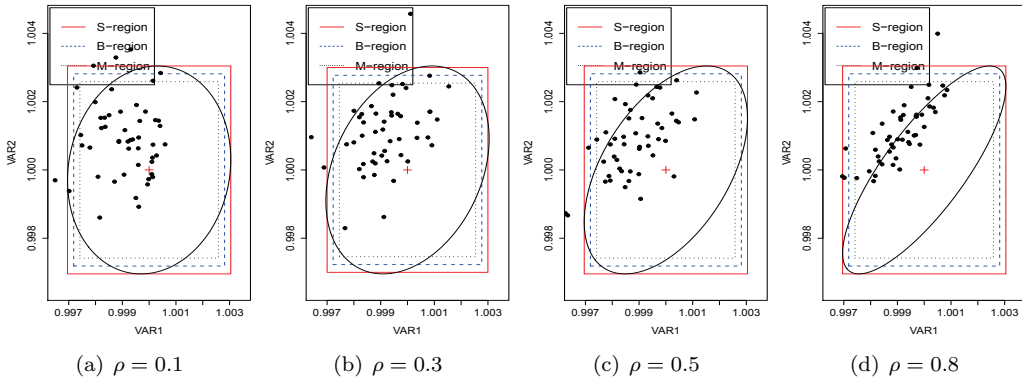


**Figure 3.2.** S-confidence region (solid line), B-confidence region (dashed line), M-confidence region (dotted line) for new observations  $\mathbf{y}$  according to  $\rho$  with  $m = 1000$  and data points (points) under the control. The first row is the case  $\sigma_1 = \sigma_2 = 0.001$ , the second row is the case  $\sigma_1 = \sigma_2 = 0.1$ , and the third row is the case  $\sigma_1 = \sigma_2 = 1$ .

호로 탐지된 경우라도 S-신뢰구간으로는 이상신호가 식별되지 않는 경우가 발생하고 변수 간의 상관성이 커질수록 이러한 제2종의 오류는 커지게 된다. 그렇기 때문에 각 변수의 이상 유무를 통해 이변량 데이터의 관리상태 유무를 모니터링하지 않게 된다. 다만, 이러한 신뢰구간은  $T^2$  통계량을 통해 이상신호 탐지 후 그 원인을 식별하기 위한 가장 일반적인 절차로써 검토되는 것이 보통이다. 이런 가정에서 살펴보면 변수별 M-신뢰구간으로 구성된 영역이 공정평균벡터에 대해 가장 민감하게 변화됨을 알 수 있다.



**Figure 3.3.** S-confidence region (solid line), B-confidence region (dashed line), M-confidence region (dotted line) for new observations  $\mathbf{y}$  when  $m = 1000$ ,  $\sigma_1 = \sigma_2 = 0.001$  at  $\alpha = 0.01$  and data points out of control with  $\mu = \mu_1$ .



**Figure 3.4.** S-confidence region (solid line), B-confidence region (dashed line), M-confidence region (dotted line) for new observations  $\mathbf{y}$  when  $m = 1000$ ,  $\sigma_1 = \sigma_2 = 0.001$  at  $\alpha = 0.01$  and data points out of control with  $\mu = \mu_2$ .

### 3.3. 검정력 비교

3.2절의 결과에서도 알 수 있듯이 새로운 이변량 관측치의 관리상태 유무를 판단하기 위해서는 각각의 변수에 대한 관리상태를 모니터링하는 것보다 호텔링의  $T^2$  검정 통계량을 이용한 모니터링이 검정력이 좋은 것으로 기대할 수 있다. 두 변수의 상관성이 커짐에 따라 새로운 관측치에 대한 동시신뢰영역이 타원형으로 좁게 나타나는 경향이 있어 이러한 사실을 쉽게 짐작할 수 있다. 본 절에서는  $\mu$ 의 변화에 따라  $T^2$  검정 통계량의 검정력과, 각 변수별 3가지 형태의 신뢰구간으로 구성되는 신뢰영역에 의해 나타나는 검정력을 비교하기 위해  $m = 1000$ ,  $\sigma_1 = \sigma_2 = 0.001$ 인 경우를 가정하였으며 모의실험을 10,000번 실시하였다. 각 변수별로 새로운 관측치에 대한 신뢰구간을 추정하고 이를 통해 모평균벡터의 변화유무를 검정할 때, 변수별로 유의수준을  $\alpha = 0.01$ 로 했기 때문에 두 변수를 동시에 검정하는 경우 가짜 알람률이 정해진 수준  $\alpha = 0.01$ 보다 커질 수 있다. 또한, 이러한 가짜 알람률을 조정하기 위해 B-신뢰구간을 사용한 것에 유의할 필요가 있다. 이 때,  $\mu = \mu_1$ 인 랜덤포본과 이 때의 신뢰구간들은 Figure 3.3에서,  $\mu = \mu_2$ 인 랜덤포본과 신뢰구간들은 Figure 3.4를 참조하기 바란다.

Table 3.1에서  $\mu = \mu_0$ 인 경우는 공정이 관리상태인 경우를 나타내고, 이 경우  $T^2$  검정통계량은 가짜 알람률의 수준을 잘 유지하는 것으로 나타난다.  $\mu = \mu_1$ 인 경우는 한 변수(VAR1)의 평균만 변화한 경우



**Table 3.1.** A comparison of the statistical power according to  $\mu$  with  $\sigma_1 = \sigma_2 = 0.001$  at  $\alpha = 0.01$ 

	$\rho$	$T^2$ 검정	S-신뢰구간		M-신뢰구간		B-신뢰구간	
			VAR1	VAR2	VAR1	VAR2	VAR1	VAR2
$\mu = \mu_0$ 표본	0.1	0.010	0.002	0.003	0.011	0.012	0.005	0.006
	0.3	0.010	0.003	0.002	0.010	0.009	0.006	0.005
	0.5	0.011	0.003	0.002	0.012	0.010	0.004	0.004
	0.8	0.012	0.003	0.002	0.011	0.010	0.007	0.005
$\mu = \mu_1$ 표본	0.1	0.042	0.021	0.003	0.056	0.013	0.035	0.006
	0.3	0.045	0.022	0.003	0.063	0.010	0.040	0.006
	0.5	0.050	0.002	0.002	0.058	0.009	0.035	0.004
	0.8	0.146	0.026	0.004	0.068	0.010	0.043	0.005
$\mu = \mu_2$ 표본	0.1	0.114	0.021	0.026	0.063	0.068	0.039	0.043
	0.3	0.140	0.022	0.022	0.064	0.063	0.038	0.038
	0.5	0.214	0.024	0.019	0.066	0.056	0.042	0.032
	0.8	0.626	0.021	0.024	0.058	0.065	0.038	0.041

인데, 이 경우 M-신뢰구간, B-신뢰구간, S-신뢰구간을 이용한 방법 순으로 검정력이 나타남을 알 수 있다. 또한 강한상관이 아니라면 호텔링  $T^2$ 에 의한 검정력과 개별신뢰구간에 의한 검정력간에 큰 차이가 없음을 알 수 있다.  $\mu = \mu_2$ 인 경우에는 두 변수의 평균이 모두 변한 경우인데 이 경우에는  $\mu = \mu_1$ 에서와 달리  $T^2$  통계량에 의한 검정력이 개별변수의 신뢰구간으로 구성된 신뢰영역에 의한 검정력보다 크게 나타났다. 변수들의 상관성이 커질수록 이러한 차이는 더 크게 나타났다. 이것은 결국 변수들의 상관이 커질수록 이상신호의 원인이 되는 개별 변수의 식별이 어려워지는 것을 의미한다. 반대로 변수들의 상관이 크지 않은 경우라면 이러한 개별변수에 의한 확인 절차가 꽤 의미있는 절차가 될 수 있음을 알 수 있다.

#### 4. 결론

이단계 호텔링  $T^2$  관리도에서는 새로운 데이터가 관측될 때마다 관리상태(또는 과거 상태)와 다른 평균 변화가 있는지 호텔링  $T^2$  통계량을 통해 통계적 유의성 검정을 하게 된다. 만약 평균벡터의 이상신호가 감지되면, 그 원인이 무엇인지 검토할 필요가 있고 이를 위해서 개별적인 변수에 관해 기대되는 신뢰구간을 이용하는 것은 자연스러운 절차이다. 호텔링  $T^2$  통계량으로부터 구한 변수별 S-신뢰구간은 새로운 관측값에 대한 동시신뢰영역을 둘러싸는 커다란 직사각형을 이루고 있어,  $T^2$  통계량에 의해 관리상태라고 판정한 데이터에 대해 이상신호가 발생했다고 잘못 판정할 가능성은 전혀 없다. 거꾸로 실제 이상신호가 있는데 탐지하지 못할 가능성이 커지는 문제가 존재한다. 즉, 평균벡터의 변화에 둔감한 문제가 있다. B-신뢰구간은 개별 신뢰구간으로 동시신뢰영역을 구하면서 가짜 알람률을  $\alpha$ 로 한 것으로 S-신뢰구간보다는 민감하게 이상신호를 탐지할 수 있었다. M-신뢰구간은 개별 신뢰구간을 구할 때 가짜 알람률을 각각  $\alpha$ 로 하여 관리상태일 때 이상신호라고 판정할 가능성이 가장 높지만, 다른 신뢰구간들에 비해 민감하게 이상신호를 탐지할 수 있는 방법이라고 할 수 있다. 결국, 이상신호 발생 후 변수별 영향력을 평가하기 위한 통계량으로는 가장 적절하다고 할 수 있다. 그러나 이러한 방법들도 변수 간에 강한 상관관계가 존재하면 평균벡터변화의 원인을 확인할 가능성이 매우 낮은 것을 알 수 있었다. 또한, 변수 간의 상관관계는 이상신호의 원인이 되는 변수를 판정하는데 관련이 있지만, 각 변수의 분산은 별 영향이 없음을 알 수 있다.

다변량 관리도에서는 이상신호 발생 후 그 원인이 되는 변수를 식별하기 위해 다시 개별적인 변수에 대

한 이상상태 발생여부를 검정하는 등 두 가지 절차를 시행하고 있다. 그러나 본 논문의 결과로부터 변수들의 상관성이 높은 경우 개별적인 변수의 이상유무를 통해 이상신호에 영향을 끼친 변수를 식별하기 어렵다는 것을 알 수 있었다. 이에 앞으로의 연구에서는 이 두 가지 절차를 동시에 할 수 있는 모니터링 방안에 대해 연구할 필요가 있을 것이다.

## References

- Alt, F. B. (1985). Multivariate quality control. In Kotz, S., Johnson, N. L., Read, C. R. (eds.) *Encyclopedia of Statistical Sciences*, (Vol. 6, pp. 111–122), John Wiley & Sons, New York.
- Doganaksoy, N., Faltin, F. W., and Tucker, W. T. (1991). Identification of out of control quality characteristics in a multivariate manufacturing environment, *Communications in Statistics - Theory and Methods*, **20**, 2775–2790.
- Hayter, A. J. and Tsui, K. L. (1994). Identification and quantification in multivariate quality control problems, *Journal of Quality Technology*, **26**, 197–208.
- Jackson, J. E. (1991). *A User's Guide to Principal Components*, John Wiley & Sons, New York.
- Kang, S. H. and Kim, S. B. (2016). Multivariate monitoring of the metal frame process in mobile device manufacturing, *Journal of the Korean Institute of Industrial Engineers*, **42**, 1186–1195.
- Kenett, R. S. and Halevy, A. (1984). Some statistical aspects of quality conformance inspection in military specifications documents. In *Proceedings of the 5th International Conference of the Israeli Society for Quality Assurance*, Tel Aviv, 23–35.
- Lee, S. (2015). Effects of parameter estimation in phase I on phase II control limits for monitoring autocorrelated data, *The Korean Journal of Applied Statistics*, **28**, 1025–1034.
- Mason, R. L., Tracy, N. D., and Young, J. C. (1995) Decomposition of  $T^2$  for multivariate control chart interpretation, *Journal of Quality Technology* **27**, 99–108.
- Mason, R. L., Tracy, N. D., and Young, J. C. (1997). A practical approach for interpreting Multivariate  $T^2$  control chart, *Journal of Quality Technology*, **29**, 396–406.
- Montgomery, D. C. (2005)., *Introduction to Statistical Quality Control* (5th ed), John Wiley & Sons, New York.
- Shewhart, W. A. (1926). Quality control charts, *Bell System Technical Journal*, **2**, 593–603.
- Timm, N. H. (1996). Multivariate quality control using finite intersection tests, *Journal of Quality Technology*, **28**, 233–243.

# 다변량 공정 모니터링에서 이상신호 발생시 원인 식별에 관한 연구

이성임<sup>a,1</sup>

<sup>a</sup>단국대학교 응용통계학과

(2018년 1월 3일 접수, 2018년 1월 17일 수정, 2018년 1월 17일 채택)

---

## 요약

최근 다변량 공정관리는 다양한 응용 분야에서 중요해지고 있는 추세이다. 예를 들어, 제조 산업 분야에서는 다변량 품질특성치를 동시에 모니터링할 필요가 있다. 그러나, 다변량 관리도는 이상신호가 발생한 경우 그 원인이 되는 개별적인 변수를 식별하기가 어렵기 때문에, 실제로는 기대만큼 유용하게 쓰이고 있지 않은 형편이다. 이에 본 논문에서는 새로운 관측치에 대한 개별적인 신뢰구간을 사용하여 이상신호의 원인을 탐지하는 세 가지 방법을 소개하고, 시뮬레이션 연구를 통해 이상신호의 원인이 되는 개별적인 변수를 식별하고 해석하는 데 있어 주의할 점이 무엇인지 살펴보기로 한다.

주요용어: 호텔링  $T^2$ , 다변량 SPC, 이상신호 원인 확인

---

이 연구는 2016년도 단국대학교 대학연구비의 지원으로 연구되었음.

<sup>1</sup>(16890) 경기도 용인시 수지구 죽전로 152, 단국대학교 응용통계학과. E-mail: silee@dankook.ac.kr