Journal of the Korean Data & Information Science Society
2017, **28**(2), 339–348
http://dx.doi.org/10.7465/jkdi.2017.28.2.339
한국데이터정보과학회지

비모수적 Shewhart-Lepage 관리도의 최적 설계†

이성민[1] · 이재헌[2]

[12]중앙대학교 응용통계학과

접수 2017년 2월 20일, 수정 2017년 3월 11일, 게재확정 2017년 3월 13일

요 약

전통적인 통계적 공정관리에서 품질특성치의 위치모수와 척도모수의 변화를 탐지하는 것은 주된 관심사였고, 이를 위해 일반적으로 두 개의 관리도를 병행하여 사용한다. 그러나 하나의 관리도를 사용하여 두 모수의 변화를 동시에 탐지하는 절차에 대한 연구도 많이 진행되어 왔다. 하나 또는 두 개의 관리도를 사용할 때, 제1국면 (phase I)을 통하여 모수를 추정하여 관리한계를 설정하여 제2국면 (phase II)의 관리도를 운영하는데 이때 정규성 가정의 만족 여부는 아주 중요한 점검 사항이다. 실제 공정에서는 종종 분포에 대한 가정을 하기 어렵거나 정규분포를 따른다고 가정하기 어려운 경우가 있는데, 이러한 경우에는 비모수적 관리도를 사용할 수 있다. 이 논문에서는 비모수적 관리도이면서, 하나의 관리도를 사용하여 위치모수와 척도모수의 변화를 탐지하는 Shewhart-Lepage 관리도를 소개하고, 위치모수와 척도모수가 동시에 변화할 때 진단 단계에서 변화의 원인을 가장 정확하게 진단할 수 있는 최적의 진단한계를 모의실험을 통해 제시하고 그 효율에 대해 연구하였다.

주요용어: 비모수적 관리도, 위치모수, 척도모수, Ansari-Bradley 검정, Shewhart-Lepage 관리도, Wilcoxon 순위합 검정.

1. 서론

1924년 Shewhart가 공정을 효율적으로 관리하기 위해서 제안한 관리도 (control chart)는 현재에도 지속적으로 다양한 역할을 하고 있다. 공정이 관리상태일 때 목표로 하는 제1종 오류 α를 만족하도록 관리한계 (control limit)를 설정하고, 공정이 이상상태일 때 관리한계를 벗어날 때까지 관측한 표본의 수인 런길이의 평균, 즉 ARL (average run length)을 이용해서 관리도의 성능을 평가한다.

공정을 효율적으로 관리하기 위해 모니터링하는 대상은 연속형 측도의 품질특성치 (quality characteristic)의 경우 대부분 위치모수와 척도모수, 예를 들면 평균과 분산이며, 전통적인 접근 방법으로 위치모수의 변화를 탐지하는 관리도와 척도모수의 변화를 탐지하는 관리도 두 개를 병행해서 사용한다. Jones와 Case (1981), Saniga (1989), 그리고 Rahim과 Costa (2000)와 같은 연구자는 두 개의 Shewhart 형태의 관리도를 이용하여 정규분포의 평균과 분산을 동시에 탐지하는 절차를 연구하였으며, 이 절차들은 일반적으로 큰 변화를 탐지하는데 좋은 성능을 나타냈다. Reynolds와 Stoumbos (2001, 2006)는 평균과 분산의 변화를 동시에 탐지하기 위해 EWMA (exponentially weighted moving average) 관리도를 병행하는 절차에 대해 연구하였다.

† 이 논문은 2014년도 정부 (교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2014R1A1A2054200).
[1] (06974) 서울시 동작구 흑석로 84, 중앙대학교 응용통계학과, 석사과정.
[2] 교신저자: (06974) 서울시 동작구 흑석로 84, 중앙대학교 응용통계학과, 교수. E-mail: jaeheon@cau.ac.kr

그러나 두 개의 관리도를 병행할 경우, 전체적으로 주어진 관리상태에서의 특성을 만족하도록 각각의 관리한계를 설정해야 하고, 또한 CUSUM (cumulative sum)과 EWMA 관리도의 경우에는 관리한계 외에 설정하는 튜닝 모수 (tuning parameter)를 각 관리도에 대해 설정해야 하는 어려움이 있다. 이에 대한 대안으로써, 하나의 관리도를 사용하여 위치모수와 척도모수를 동시에 탐지하는 절차에 대한 연구들이 있는데, 이 연구들은 다음과 같다. Chao와 Cheng (1996)이 제안한 Semicircle 관리도와 Chen과 Cheng (1998)이 제안한 Max 관리도는 두 개의 Shewhart 관리도를 기반으로 한 단일 관리도이다. Chen 등 (2001)은 두 개의 EWMA 관리도를 하나로 결합한 MaxEWMA 관리도를 제안하였고, Memar와 Niaki (2010)는 개별 관측값에 대한 일반화된 MaxEWMA 관리도를 제안하였다. Costa와 Rahim (2004)은 평균과 분산을 동시에 탐지하는 비중심 카이제곱 (noncentral chi-square) 관리도라는 단일 관리도를 제안하였고, Zhang 등 (2010)은 평균과 분산의 변화에 대해 우도비 검정에 기초한 관리도를 제안하였다. Cheng과 Thaga (2006)와 Jung과 Lee (2008)는 평균과 분산을 동시에 탐지하는 여러 가지 단일 관리도에 대해 그 효율을 비교하였다. 하나의 관리도를 사용하여 여러 모수의 변화를 탐지하는 절차는 하나의 관리도로 여러 변수의 변화를 탐지하고 이상신호 발생 후 이를 진단하는 단계가 있다는 측면에서 다변량 관리도 (multivariate control chart)와도 유사한 점이 있다 (Choi와 Cho, 2016; Hwang, 2016).

다양한 연구를 통해 단일 관리도 절차에 대한 진보가 있었지만, 이를 사용하는 경우 다음과 같은 두 가지 문제점이 발생할 수 있다. (물론 두 관리도를 병행하는 경우에도 발생할 수 있다.) 첫 번째는 실제 공정에서 계량형 품질특성치의 분포를 모르거나 또는 정규분포를 가정하기가 어려운 경우가 있다는 것이다. 이러한 경우가 문제가 되는 이유는 비정규분포에서 평균과 분산을 탐지하는 관리도는 제1종 오류인 오경보율을 계산하기 어렵기 때문에 관리한계를 설정하는 것이 간단하지 않다. 두 번째는 품질특성치의 분포뿐만 아니라 관리상태 (in-control)에서의 위치모수와 척도모수를 모르는 경우가 일반적이라는 것이다. 이 경우 관리상태의 모수는 제1국면 (phase I)에서 추정해야 하는데, 어느 정도 안정된 추정량을 얻기 위해서는 많은 수의 제1국면 표본이 필요하며 때로는 많은 수의 표본을 추출하기가 어려운 공정이 있을 수 있다. 제1국면의 추정에 대해서는 Montgomery (2013)를 참고할 수 있다.

이와 같은 두 가지 문제점은 실제 공정에서 종종 발생할 수 있으며, 이러한 경우 사용할 수 있는 비모수적 형태의 관리도에 대한 연구가 많이 진행되었다. 비모수적 관리도 (nonparametric contol chart)에 대한 자세한 내용은 Chakraborti 등 (2010)을 참고할 수 있다. 이 논문에서 소개할 관리도는 Shewhart-Lepage 관리도인데, Mukherjee와 Chakraborti (2012)에 의해 처음 제안되었다. 이 관리도는 일반적으로 모수의 큰 변화를 잘 탐지하는 Shewhart 형태의 비모수적 관리도로서, 품질특성치의 분포를 가정하지 않으며 알려지지 않은 위치모수와 척도모수를 동시에 탐지하는 것이 관리도 사용의 목적이다.

이 논문의 목적은 Mukherjee와 Chakraborti (2012)가 제안한 Shewhart-Lepage 관리도를 소개하고, 이 관리도에서 이상신호가 발생했을 때 공정의 위치모수와 척도모수 변화의 진단을 위해 사용하는 진단한계의 최적값을 모의실험 결과를 통해 제안하고 기존의 방법과 그 효율을 비교하였다. 기존의 방법인 Mukherjee와 Chakraborti (2012)의 방법은 진단한계를 설정할 때 위치모수와 척도모수의 변화 확률이 동일하다는 가정을 하는데, 이 가정은 진단한계를 비교적 간단하게 구할 수 있게 하는 장점은 있지만 진단의 효율은 떨어지게 할 것으로 예상하고 있다.

이 논문의 나머지 구성을 살펴보면, 2절에서는 Shewhart-Lepage 관리도에 대한 기본적인 내용, 3절에서는 Shewhart-Lepage 관리도의 설계 및 운영, 4절에서는 모의실험의 결과 제시, 그리고 5절에서 논문의 요약과 결론을 내리면서 마무리한다.

2. Shewhart-Lepage 관리도의 기본적 내용

먼저 제2국면 (phase II)의 모니터링을 시작할 때, 크기가 m인 참고표본 (reference sample)을 제1국면에서 사용가능한 표본으로 가정하고 제2국면의 각 시점마다 추출하는 크기가 n인 표본을 검사표본 (test sample)이라고 가정하자. 이 표본들로 관리도에 타점하는 Lepage 통계량을 계산할 수 있고, 이 통계량은 위치모수의 변화 탐지를 위한 Wilcoxon 순위합 검정 (rank sum test)의 통계량과 척도모수의 변화 탐지를 위한 Ansari-Bradley 검정의 통계량이 결합된 형태로 이루어져 있다. 비모수적 검정에 대한 상세한 내용은 Song 등 (2016)의 비모수통계학 교재를 참고하길 바란다.

만약 계산된 통계량이 관리한계를 초과하면 이상상태 (out-of-control)라는 신호를 주고, 이 신호의 원인을 찾는 진단 (diagnosis)을 시작해야 한다. 두 모수의 변화를 하나의 관리도를 사용하여 탐지하기 때문에 신호후 진단하는 절차는 중요한 과정이다. 이와 같은 진단 단계의 목적은 이상신호의 원인이 위치모수 또는 척도모수만의 문제인지, 아니면 두 모수가 동시에 변화한 것인지를 판단하는 것이다. 이와 반대로, 계산된 통계량이 관리한계 내에 있으면 공정은 관리상태라고 판단하고 다음의 검사표본을 추출해서 모니터링을 계속 수행하게 된다.

Lepage 통계량을 구성하는 2개의 개별 통계량에 대한 기본적인 사항은 다음과 같다. 알려지지 않은 연속분포함수 $F(u)$와 $G(v) = F(\delta v + \theta)$, $(\delta > 0, -\infty < \theta < \infty)$를 따르는 두개의 모집단으로부터 추출한 각각 크기 m과 n인 독립인 확률표본을 $U_1, U_2, \cdots, U_m$과 $V_1, V_2, \cdots, V_n$이라 하자. 또한 N은 $m+n$이라 할 때, $U_i\,(i = 1, 2, \cdots, m)$와 $V_j\,(j = 1, 2, \cdots, n)$의 혼합표본에서 $Z_k\,(k = 1, 2, \cdots, N)$라는 지시변수 (indicator variable)를 다음과 같이 정의하자.

$$Z_k = \begin{cases} 1, & k\text{번째 순서통계량이 } V\text{일 때} \\ 0, & k\text{번째 순서통계량이 } U\text{일 때} \end{cases}$$

두 위치모수의 동일성을 검정하기 위해 Wilcoxon의 순위합 검정을 이용할 수 있으며, 검정통계량을 T_1이라 할 때 T_1의 정의와 공정이 관리상태 (in-control: IC)일 때 평균과 분산은 다음과 같다.

$$T_1 = \sum_{k=1}^{N} k\, Z_k \tag{2.1}$$

$$E(T_1\,|\,IC) = \mu_1 = \frac{1}{2}n(N+1) \tag{2.2}$$

$$Var(T_1\,|\,IC) = \sigma_1^2 = \frac{1}{12}mn(N+1) \tag{2.3}$$

다음으로, 두 척도모수의 동일성을 검정하기 위해 Ansari-Bradley 검정을 이용할 수 있으며, 검정통계량 T_2의 정의와 공정이 관리상태일 때 평균과 분산은 다음과 같다.

$$T_2 = \sum_{k=1}^{N} \left| k - \frac{1}{2}(N+1) \right| Z_k \tag{2.4}$$

$$E(T_2\,|\,IC) = \mu_2 = \begin{cases} \dfrac{nN}{4}, & N\text{이 짝수일 때} \\ \dfrac{n(N^2-1)}{4N}, & N\text{이 홀수일 때} \end{cases} \tag{2.5}$$

$$Var(T_2\,|\,IC) = \sigma_2^2 = \begin{cases} \dfrac{mn(N^2-4)}{48(N-1)}, & N\text{이 짝수일 때} \\ \dfrac{mn(N+1)(N^2+3)}{48N^2}, & N\text{이 홀수일 때} \end{cases} \tag{2.6}$$

3. Shewhart-Lepage 관리도의 설계 및 운영

공정이 관리상태일 때 수집하는 제1국면의 m개 표본을 $\mathbf{X}_m = (X_1, X_2, \cdots, X_m)$이라 하고, 제2국면일 때 시점 i일 때 추출하는 n개의 표본을 $\mathbf{Y}_{i,n} = (Y_{i1}, Y_{i2}, \cdots, Y_{in})$이라고 하자. $\mathbf{X}_m$은 2절에서 언급한 확률표본 U에 대응되고 $\mathbf{Y}_{i,n}$은 확률표본 V에 대응된다. 표본 $\mathbf{X}_m$과 $\mathbf{Y}_{i,n}$에 대해 식 (2.1)부터 (2.6)까지를 이용하면, i번째 검정통계량 T_{1i}, T_{2i}, 그리고 평균과 분산을 계산할 수 있다. 또한 다음과 같이 표준화된 Wilcoxon 순위합 검정통계량 S_{1i}와 표준화된 Ansari-Bradley 검정통계량 S_{2i}을 구성할 수 있다.

$$S_{1i} = \frac{T_{1i} - \mu_1}{\sigma_1}\ ,\ S_{2i} = \frac{T_{2i} - \mu_2}{\sigma_2}$$

다음으로 관리도에 타점을 하기 위한 Shewhart-Lepage 관리통계량은 다음과 같이 정의한다. 여기서 Shewhart 형태의 관리도를 사용하기 때문에 Shewhart-Lepage 관리통계량이라 부르기로 한다.

$$S_i^2 = S_{1i}^2 + S_{2i}^2\ ,\ i = 1, 2, 3, \cdots$$

Mukherjee와 Chakraborti (2012)는 관리상한으로 H를 사용하였고, $S_i^2 \geq 0$이기 때문에 관리하한은 사용하지 않았다. 이때 H는 주어진 관리상태의 특성, 예를 들면 관리상태에서의 ARL인 ARL_0가 미리 정해진 값을 만족하도록 설정할 수 있다. 이 경우 i번째 표본의 검정통계량 S_i^2이 H보다 작으면 관리상태라고 판단하여 $i+1$번째 표본을 조사하고, 만일 S_i^2이 H보다 커지면 이상상태의 신호를 주게 된다.

이상신호를 준 경우, 위치모수와 척도모수 중 어떠한 모수에 변화가 발생했는가를 판단하기 위해 진단을 시작한다. Mukherjee와 Chakraborti (2012)는 진단한계 H_1과 H_2 (H_1, $H_2 < H$)를 사용하여 다음과 같은 세 가지로 경우로 구분하여 진단하였다.

(1) $S_{1i}^2 > H_1$이지만 $S_{2i}^2 \leq H_2$인 경우 위치모수만 변화했다고 판단한다.

(2) $S_{1i}^2 \leq H_1$이지만 $S_{2i}^2 > H_2$인 경우 척도모수만 변화했다고 판단한다.

(3) $S_{1i}^2 > H_1$이고 $S_{2i}^2 > H_2$인 경우 위치모수와 척도모수가 모두 변화했다고 판단한다.

기존의 연구인 Mukherjee와 Chakraborti (2012)는 진단한계 H_1과 H_2, 그리고 관리한계 H에 대해서 다음과 같은 특징이 있음을 밝혔다. 여기서 R은 런길이를 나타내는 확률변수이다.

(1) 진단한계 H_1과 H_2는 모두 양수이며, $H_1 + H_2 = H$인 관계가 성립한다.

(2) r번째 표본에서 이상신호가 발생, 즉 $R = r$인 조건 하에서, $[S_r^2 > H]$라는 사상 (event)은 다음의 세 가지 서로 배반인 사상, $[S_{1r}^2 > H_1, S_{2r}^2 \leq H_2 \,|\, R = r]$, $[S_{1r}^2 \leq H_1, S_{2r}^2 > H_2 \,|\, R = r]$, $[S_{1r}^2 > H_1, S_{2r}^2 > H_2 \,|\, R = r]$로 분할된다.

또한 Mukherjee와 Chakraborti (2012)는 다음과 같은 관리상태 하에서의 확률을 각각 $P_0[S_{1r}^2 > H_1 \,|\, R = r] = \gamma_1$과 $P_0[S_{2r}^2 > H_2 \,|\, R = r] = \gamma_2$라 할 때, 모의실험을 통하여 $\gamma_1 = \gamma_2$가 되도록 진단한계 H_1과 H_2을 결정하는 방법을 제안하였다. 그들이 제안한 방법은 간편하다는 장점은 있지만, H_1과 H_2가 위치모수와 척도모수의 변화량과는 관계없이 참고표본의 크기 m과 검사표본의 크기 n에만 의존하게 된다.

이 논문에서는 공정이 관리상태일 때 위치모수와 척도모수의 변화가 같은 확률로 발생한다는 가정 대신, 우리가 가장 관심이 있는 상황인 위치모수와 척도모수가 동시에 변화하는 경우 이를 가장 잘 진단할 수 있는 최적의 H_1과 H_2를 찾고자 한다. 이때 변화를 진단하는 성능의 판단 기준은 다음과 같이 정의되는 $P(C)$인데, 이는 r번째 검사표본에서 위치모수와 척도모수가 동시에 변화하여 공정에 이상신호가 발생한 경우 이를 정확하게 진단할 확률을 나타낸다. 즉, 주어진 위치모수와 척도모수의 변화량에 대해 $P(C)$를 최대로 하도록 진단한계 H_1과 H_2을 설정하는 것이다.

$$P(C) = Pr(S_{1r}^2 > H_1\ ,\ S_{2r}^2 > H_2 \,|\, R = r) \tag{3.1}$$

4. 모의실험

이 절에서는 모의실험을 실시하여 앞에서 제안한 최적의 진단한계 설정 방법에 대한 효율을 기존의 방법과 비교하고자 한다. 모의실험에서 사용한 관리모수들의 값은 기본적으로 Mukherjee와 Chakraborti (2012)의 경우와 동일하게 설정하였다. 즉, 참고표본의 크기는 $m = 30$과 50, 검사표본의 크기는 $n = 5$를 사용하였다. 또한 관리상태일 때 평균런길이는 $\text{ARL}_0 = 500$으로 설정하였고, 이때 이를 만족하는 관리상한은 $m = 30$, $n = 5$인 경우 $H = 9.40$이고, $m = 50$, $n = 5$인 경우 $H = 10.32$가 된다. 모든 모의실험의 결과는 50,000번의 반복을 수행하여 산출하였다.

앞 절에서 설명한 바와 같이 이상신호가 발생한 경우 진단을 위한 진단한계 H_1과 H_2를 결정할 때, Mukherjee와 Chakraborti (2012)는 공정이 관리상태일 때 위치모수와 척도모수의 변화가 같은 확률로 발생한다는 가정을 하고 이를 만족하는 H_1과 H_2를 계산하였다. 그 결과, 위치모수와 척도모수의 변화량과는 관계없이 $m = 30$, $n = 5$인 경우 $H_1 = 5.75$, $H_2 = 3.65$이고 $m = 50$, $n = 5$인 경우 $H_1 = 6.52$, $H_2 = 3.80$을 얻을 수 있다.

본 논문에서 제안한 방법은 주어진 위치모수와 척도모수의 변화량에 대해 식 (3.1)에 정의된 $P(C)$를 가장 최대로 하도록 H_1과 H_2을 결정하는 것이다. 이를 위하여 모의실험에서 H_1을 0과 H 사이에서 0.1 단위로 변화시키면서 (이때 $H_2 = H - H_1$) 각 경우의 $P(C)$를 계산하여 이를 최대로 하는 H_1과 H_2를 산출하였다.

모의실험에서 품질특성치의 분포는 정규분포 (standard normal distribution)와 이와 모양은 유사하지만 꼬리 부분이 더 두터운 라플라스분포 (Laplace distiribution)를 가정하였고, 위치모수의 변화량 θ는 0.25부터 0.25 단위로 최대 1.5까지 설정하였고, 척도모수의 변화량 δ는 1.25부터 0.25 단위로 최대 2.0까지 설정하였다. Mukherjee와 Chakraborti (2012)의 진단한계와 이 논문에서 제안한 방법의 진단한계, 그리고 이때의 $P(C)$ 값이 Table 4.1에서 Table 4.4까지 주어져 있다. 표에서 'M & C'는 Mukherjee와 Chakraborti (2012)의 방법이고, 'Proposed'는 본 논문에서 제안한 방법을 나타낸다.

Table 4.1 Vaues of H_1, H_2, and $P(C)$ for $N(\theta, \delta)$ ($m = 30$, $n = 5$, $H = 9.40$)

		M & C				Proposed		
θ	δ	H_1	H_2	$P(C)$		H_1	H_2	$P(C)$
0.25	1.25	5.75	3.65	0.17130		7.4	2.0	0.50716
0.5	1.25	5.75	3.65	0.22548		7.4	2.0	0.59944
0.75	1.25	5.75	3.65	0.28786		7.4	2.0	0.67504
1.0	1.25	5.75	3.65	0.34938		7.4	2.0	0.73058
1.25	1.25	5.75	3.65	0.42238		7.4	2.0	0.78506
1.5	1.25	5.75	3.65	0.49394		7.4	2.0	0.82272
0.25	1.5	5.75	3.65	0.17540		7.4	2.0	0.39678
0.5	1.5	5.75	3.65	0.22634		7.4	2.0	0.47076
0.75	1.5	5.75	3.65	0.28948		7.4	2.0	0.55240
1.0	1.5	5.75	3.65	0.34932		7.4	2.0	0.61960
1.25	1.5	5.75	3.65	0.41306		7.4	2.0	0.67424
1.5	1.5	5.75	3.65	0.47968		7.4	2.0	0.72940
0.25	1.75	5.75	3.65	0.16902	*	3.0	6.4	0.33092
						7.4	2.0	0.30686
0.5	1.75	5.75	3.65	0.21656		7.4	2.0	0.37698
0.75	1.75	5.75	3.65	0.27308		7.4	2.0	0.45306
1.0	1.75	5.75	3.65	0.33498		7.4	2.0	0.52328
1.25	1.75	5.75	3.65	0.39688		7.4	2.0	0.58544
1.5	1.75	5.75	3.65	0.45346		7.4	2.0	0.63516
0.25	2.0	5.75	3.65	0.15932	*	2.0	7.4	0.36660
						7.4	2.0	0.26190
0.5	2.0	5.75	3.65	0.19662	*	2.9	6.5	0.35790
						7.4	2.0	0.31056
0.75	2.0	5.75	3.65	0.24900		7.4	2.0	0.37082
1.0	2.0	5.75	3.65	0.30634		7.4	2.0	0.43970
1.25	2.0	5.75	3.65	0.36198		7.4	2.0	0.50142
1.5	2.0	5.75	3.65	0.41888		7.4	2.0	0.55774

Table 4.1은 $m = 30$이고 $n = 5$일 때 정규분포 $N(\theta, \delta)$에 대한 결과이다. (θ, δ)가 (0.25, 1.75), (0.25, 2.0), (0.5, 2.0)인 경우 (표에서 '∗' 표시된 행)를 제외하고는 최적의 H_1과 H_2가 각각 7.4와 2.0로 결정되었다. 또한 Table 4.2는 $m = 50$이고 $n = 5$일 때 정규분포 $N(\theta, \delta)$에 대한 결과인데, (θ, δ)가 (0.25, 2.0), (0.5, 2.0)인 경우를 제외하고는 최적의 H_1과 H_2가 각각 7.8와 2.5로 결정되었다. 이는 모의실험을 진행하기 전에 전혀 예상하지 못한 결과인데, 본래 이 연구는 위치모수 θ와 척도모수 δ의 변화에 따라 최적의 H_1과 H_2이 어떻게 변화하는가에 대한 것이었다. 그러나 주어진 (θ, δ)에 거의 관계없이 동일한 값으로 결정된 것이다. 동일한 값으로 결정되지 않은 경우에는 표에서 바로 아래 행에 동일한 값을 사용했을 때의 $P(C)$를 제시하였다. 이 논문에서 제안된 방법으로 H_1과 H_2을 결정할 경우, 기본의 방법에 비해 위치모수와 척도모수가 동시에 변화한 경우 이를 제대로 진단할 확률을 많이 향상시키는 것으로 나타났다. 또한 몇 가지 경우에서 최적의 진단한계가 동일한 값 (7.4와 2.0, 또는 7.8와 2.5)으로 나오지 않았지만, 이 경우에 동일한 값을 사용해도 $P(C)$에 큰 차이가 나지 않는다고 판단된다.

Table 4.2 Vaues of H_1, H_2, and $P(C)$ for $N(\theta, \delta)$ ($m = 50$, $n = 5$, $H = 10.32$)

		M & C				Proposed		
θ	δ	H_1	H_2	$P(C)$		H_1	H_2	$P(C)$
0.25	1.25	6.52	3.80	0.26712		7.8	2.5	0.49578
0.5	1.25	6.52	3.80	0.33806		7.8	2.5	0.64666
0.75	1.25	6.52	3.80	0.40794		7.8	2.5	0.63938
1.0	1.25	6.52	3.80	0.47874		7.8	2.5	0.70862
1.25	1.25	6.52	3.80	0.55144		7.8	2.5	0.74108
1.5	1.25	6.52	3.80	0.62858		7.8	2.5	0.78652
0.25	1.5	6.52	3.80	0.25020		7.8	2.5	0.40830
0.5	1.5	6.52	3.80	0.31760		7.8	2.5	0.48174
0.75	1.5	6.52	3.80	0.38498		7.8	2.5	0.54628
1.0	1.5	6.52	3.80	0.45818		7.8	2.5	0.60526
1.25	1.5	6.52	3.80	0.52150		7.8	2.5	0.66194
1.5	1.5	6.52	3.80	0.58674		7.8	2.5	0.70862
0.25	1.75	6.52	3.80	0.22712		7.8	2.5	0.34514
0.5	1.75	6.52	3.80	0.28132		7.8	2.5	0.41040
0.75	1.75	6.52	3.80	0.34952		7.8	2.5	0.47346
1.0	1.75	6.52	3.80	0.41526		7.8	2.5	0.54570
1.25	1.75	6.52	3.80	0.47528		7.8	2.5	0.60090
1.5	1.75	6.52	3.80	0.54620		7.8	2.5	0.63342
0.25	2.0	6.52	3.80	0.20332	∗	3.4	6.9	0.37554
						7.8	2.5	0.27800
0.5	2.0	6.52	3.80	0.25498	∗	3.3	7.0	0.38852
						7.8	2.5	0.33600
0.75	2.0	6.52	3.80	0.31186		7.8	2.5	0.40346
1.0	2.0	6.52	3.80	0.37114		7.8	2.5	0.46154
1.25	2.0	6.52	3.80	0.43588		7.8	2.5	0.51584
1.5	2.0	6.52	3.80	0.49476		7.8	2.5	0.57076

Table 4.3은 $m = 30$이고 $n = 5$이고 Table 4.4는 $m = 50$이고 $n = 5$일 때, 라플라스분포 Laplace(θ, δ)에 대한 결과이다. 라플라스분포의 경우 주어진 위치모수와 척도모수의 변화량에 관계없이, $m = 30, n = 5$인 경우 $H_1 = 7.4$와 $H_2 = 2.0$, 그리고 $m = 50, n = 5$인 경우 $H_1 = 7.8$와 $H_2 = 2.5$에서 $P(C)$값이 가장 큰 것으로 나타났다. 또한 모든 최적의 H_1과 H_2에서 $P(C)$값은 Mukherjee와 Chakraborti (2012)가 사용하는 것에 비해 많이 향상된 것임을 알 수 있다. 결론적으로 정규분포의 경우와 마찬가지로 이 논문에서 제안한 방법으로 진단한계를 결정할 경우, 위치모수와 척도모수가 모두 변화한 경우 기존의 연구에 비해 이를 잘 진단할 수 있을을 알 수 있다.

Table 4.3 Vaues of H_1, H_2, and $P(C)$ for Laplace(θ, δ) $(m = 30,\ n = 5,\ H = 9.40)$

		M & C			Proposed		
θ	δ	H_1	H_2	$P(C)$	H_1	H_2	$P(C)$
0.25	1.25	5.75	3.65	0.15230	7.4	2.0	0.57216
0.5	1.25	5.75	3.65	0.18006	7.4	2.0	0.70280
0.75	1.25	5.75	3.65	0.23152	7.4	2.0	0.78106
1.0	1.25	5.75	3.65	0.30504	7.4	2.0	0.81262
1.25	1.25	5.75	3.65	0.39000	7.4	2.0	0.82828
1.5	1.25	5.75	3.65	0.48932	7.4	2.0	0.83818
0.25	1.5	5.75	3.65	0.16384	7.4	2.0	0.48966
0.5	1.5	5.75	3.65	0.19702	7.4	2.0	0.60490
0.75	1.5	5.75	3.65	0.24674	7.4	2.0	0.68630
1.0	1.5	5.75	3.65	0.31244	7.4	2.0	0.72654
1.25	1.5	5.75	3.65	0.38884	7.4	2.0	0.75404
1.5	1.5	5.75	3.65	0.47496	7.4	2.0	0.76676
0.25	1.75	5.75	3.65	0.16714	7.4	2.0	0.41494
0.5	1.75	5.75	3.65	0.20236	7.4	2.0	0.51386
0.75	1.75	5.75	3.65	0.25260	7.4	2.0	0.59762
1.0	1.75	5.75	3.65	0.31256	7.4	2.0	0.64698
1.25	1.75	5.75	3.65	0.38176	7.4	2.0	0.67726
1.5	1.75	5.75	3.65	0.44792	7.4	2.0	0.70144
0.25	2.0	5.75	3.65	0.16530	7.4	2.0	0.35352
0.5	2.0	5.75	3.65	0.20148	7.4	2.0	0.43870
0.75	2.0	5.75	3.65	0.24814	7.4	2.0	0.51852
1.0	2.0	5.75	3.65	0.30492	7.4	2.0	0.57554
1.25	2.0	5.75	3.65	0.36194	7.4	2.0	0.60892
1.5	2.0	5.75	3.65	0.42668	7.4	2.0	0.64444

Table 4.4 Vaues of H_1, H_2, and $P(C)$ for Laplace(θ, δ) $(m = 50,\ n = 5,\ H = 10.32)$

		M & C			Proposed		
θ	δ	H_1	H_2	$P(C)$	H_1	H_2	$P(C)$
0.25	1.25	6.52	3.80	0.25330	7.8	2.5	0.65054
0.5	1.25	6.52	3.80	0.31222	7.8	2.5	0.76280
0.75	1.25	6.52	3.80	0.38212	7.8	2.5	0.82432
1.0	1.25	6.52	3.80	0.46184	7.8	2.5	0.83824
1.25	1.25	6.52	3.80	0.55218	7.8	2.5	0.84314
1.5	1.25	6.52	3.80	0.63882	7.8	2.5	0.84524
0.25	1.5	6.52	3.80	0.25446	7.8	2.5	0.54842
0.5	1.5	6.52	3.80	0.31810	7.8	2.5	0.66046
0.75	1.5	6.52	3.80	0.38250	7.8	2.5	0.72730
1.0	1.5	6.52	3.80	0.45464	7.8	2.5	0.75942
1.25	1.5	6.52	3.80	0.52662	7.8	2.5	0.76866
1.5	1.5	6.52	3.80	0.59766	7.8	2.5	0.77496
0.25	1.75	6.52	3.80	0.24482	7.8	2.5	0.46110
0.5	1.75	6.52	3.80	0.30718	7.8	2.5	0.57116
0.75	1.75	6.52	3.80	0.37238	7.8	2.5	0.64068
1.0	1.75	6.52	3.80	0.43482	7.8	2.5	0.67452
1.25	1.75	6.52	3.80	0.50400	7.8	2.5	0.69578
1.5	1.75	6.52	3.80	0.56548	7.8	2.5	0.70942
0.25	2.0	6.52	3.80	0.23044	7.8	2.5	0.39614
0.5	2.0	6.52	3.80	0.28688	7.8	2.5	0.48324
0.75	2.0	6.52	3.80	0.35048	7.8	2.5	0.55464
1.0	2.0	6.52	3.80	0.41730	7.8	2.5	0.59922
1.25	2.0	6.52	3.80	0.47084	7.8	2.5	0.62746
1.5	2.0	6.52	3.80	0.52658	7.8	2.5	0.65188

5. 결론

이 논문에서는 하나의 관리도로 위치모수와 척도모수의 변화를 동시에 탐지하는 비모수적 관리도인 Shewhart-Lepage 관리도를 소개하고, 위치모수와 척도모수가 동시에 변화한 경우 진단 단계에서 이를

가장 정확하게 판단할 수 있는 최적의 진단한계를 설정하는 방법을 제안하였다. 이 관리도의 목적이 위치모수와 척도모수의 변화를 동시에 탐지하는 것이기 때문에, 위치모수와 척도모수가 동시에 변화한 경우에 대한 진단을 우선적으로 고려하였다.

모의실험의 결과를 살펴보면, 이 논문에서 제안된 방법으로 결정한 최적의 진단한계를 사용할 경우 두 모수의 변화를 제대로 진단할 확률이 기존의 Mukherjee와 Chakraborti (2012)의 진단한계를 사용한 경우에 비해서 더 크다는 것을 확인할 수 있었다. 이러한 결과는 이 논문의 방법으로 결정한 진단한계를 사용할 경우, 좀 더 정확한 진단을 내릴 수 있다는 것을 의미한다. 즉, 진단한계 계산의 간편성을 위해 위치모수와 척도모수의 변화 확률이 동일하다는 기존 연구의 가정을 하지 않음으로 인하여, 좀 더 향상된 진단 효과가 나타나는 것으로 판단된다.

또한 모의실험으로부터 $m = 30$과 $n = 5$인 경우 최적의 진단한계는 $H_1 = 7.4$와 $H_2 = 2.0$이고, $m = 50$과 $n = 5$인 경우 $H_1 = 7.8$와 $H_2 = 2.5$로 나타났다. 위치모수와 척도모수의 변화량에 관계없이 동일한 최적의 진단한계를 얻은 결과는 예상하지 못했던 것으로, 이 수치들이 나오게 된 이론적인 배경에 대한 연구가 필요하다고 판단된다. 이러한 연구가 진행될 경우 다른 크기의 참고표본과 검사표본에 대한 최적의 진단한계도 좀 더 쉽게 계산할 수 있을 것이라 예상한다.

References

Chakraborti, S., Human, S. W. and Graham, M. A. (2010). Nonparametric (distribution-free) quality control charts. In *Handbook of Methods and Applications of Statistics: Engineering, Quality Control, and Physical Sciences*, edited by N. Balakrishnan, John Wiley & Sons, Inc., New York, 298-329.

Chao, M. T. and Cheng, S. W. (1996). Semicircle control chart for variables data. *Quality Engineering*, **8**, 441-446.

Chen, G. and Cheng, S. W. (1998). Max-chart: Combining $\bar{X}$ and S chart. *Statistica Sinica*, **8**, 263-271.

Chen G., Cheng, S. and Xie, H. (2001). Monitoring process mean and variability with one EWMA chart. *Journal of Quality Technology*, **33**, 223-233.

Cheng, S. W. and Thaga, K. (2006). Single variables control charts: An overview. *Quality and Reliability Engineering International*, **22**, 811-820.

Choi, H. Y. and Cho, G.-Y. (2016). Multivariate CUSUM control charts for monitoring the covariance matrix. *Journal of the Korean Data & Information Science Society*, **27**, 539-548.

Costa, A. F. B. and Rahim, M. A. (2004). Monitoring process mean and variability with one non-central chi-square chart. *Journal of Applied Statistics*, **31**, 1171-1183.

Hwang, C. (2016). Multioutput LS-SVR based residual MCUSUM control chart for autocorrelated process. *Journal of the Korean Data & Information Science Society*, **27**, 523-530.

Jones, L. and Case, K. (1981). Economic design of a joint $\bar{X}$ and R chart. *IIE Transactions*, **13**, 182-195.

Jung, S. H. and Lee, J. (2008). Procedures for monitoring the process mean and variance with one control chart. *The Korean Journal of Applied Statistics*, **21**, 509-521.

Memar, A. O. and Niaki, S. T. A. (2010). The max EWMAMS control chart for joint monitoring of process mean and variance for individual observations. *Quality and Reliability Engineering International*, **27**, 499-514.

Montgomery, D. C. (2013). *Statistical quality control: A modern introduction*, 7th Ed., John Wiley & Sons, Inc., Hoboken, NJ.

Mukherjee, A. and Chakraborti, S. (2012). A distribution-free control chart for the joint monitoring of location and scale. *Quality and Reliability Engineering International*, **28**, 335-352.

Rahim, M. A. and Costa, A. F. B. (2000). Joint economic design of $\bar{X}$ and R charts under Weibull shock models. *International Journal of Production Research*, **38**, 2871-2889.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (2001). Individuals control schemes for monitoring the mean and variance of processes subject to drifts. *Stochastic Analysis and Applications*, **19**, 863-892.

Reynolds, M. R., Jr. and Stoumbos, Z. G. (2006). Comparisons of some exponentially weighted moving average control charts for monitoring the process mean and variance. *Technometrics*, **48**, 550-567.

Saniga, E. M. (1989). Economic statistical control-chart designs with an application to $\bar{X}$ and R charts. *Technometrics*, **31**, 313-320.

Song, M. S., Park, C. S. and Kim, H. G. (2016). *Nonparametric statistics with R*, Free Academy, Paju, Gyeonggi-Do.

Zhang, J., Zou, C. and Wang, Z. (2010). A control chart based on likelihood ratio test for monitoring process mean and variability. *Quality and Reliability Engineering International*, **26**, 63-73.

Journal of the Korean Data & Information Science Society
2017, **28**(2), 339–348

http://dx.doi.org/10.7465/jkdi.2017.28.2.339
한국데이터정보과학회지

Optimal design of a nonparametric Shewhart-Lepage control chart†

Sungmin Lee[1] · Jaeheon Lee[2]

[12]Department of Applied Statistics, Chung-Ang University

Received 20 February 2017, revised 11 March 2017, accepted 13 March 2017

Abstract

One of the major issues of statistical process control for variables data is monitoring both the mean and the standard deviation. The traditional approach to monitor these parameters is to simultaneously use two seperate control charts. However there have been some works on developing a single chart using a single plotting statistic for joint monitoring, and it is claimed that they are simpler and may be more appealing than the traditonal one from a practical point of view. When using these control charts for variables data, estimating in-control parameters and checking the normality assumption are the very important step. Nonparametric Shewhart-Lepage chart, proposed by Mukherjee and Chakraborti (2012), is an attractive option, because this chart uses only a single control statistic, and does not require the in-control parameters and the underlying continuous distribution. In this paper, we introduce the Shewhart-Lepage chart, and propose the design procedure to find the optimal diagnosis limits when the location and the scale parameters change simultaneously. We also compare the efficiency of the proposed method with that of Mukherjee and Chakraborti (2012).

Keywords: Ansari-Bradley test, location parameter, nonparametric control chart, scale paramter, Shewhart-Lepage control chart, Wilcoxon rank sum test.

---

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2054200).
[1] Graduate student, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea.
[2] Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea. E-mail: jaeheon@cau.ac.kr