The Korean Journal of Applied Statistics (2015) DOI: http://dx.doi.org/10.5351/KJAS.2015.28.5.907
28(5), 907–920

Comparisons of the Performance with Bayes Estimator and MLE for Control Charts Based on Geometric Distribution

Hwiju Hong[a] · Jaeheon Lee[a,1]

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

(Received June 5, 2015; Revised July 27, 2015; Accepted September 21, 2015)

Abstract

Charts based on geometric distribution are effective to monitor the proportion of nonconforming items in high-quality processes where the in-control proportion nonconforming is low. The implementation of this chart is often based on the assumption that in-control proportion nonconforming is known or accurately estimated. However, accurate parameter estimation is very difficult and may require a larger sample size than that available in practice for high-quality process where the proportion of nonconforming items is very small. An inaccurate estimate of the parameter can result in estimated control limits that cause unreliability in the monitoring process. The maximum likelihood estimator (MLE) is often used to estimate in-control proportion nonconforming. In this paper, we recommend a Bayes estimator for the in-control proportion nonconforming to incorporate practitioner knowledge and avoid estimation issues when no nonconforming items are observed in the Phase I sample. The effects of parameter estimation on the geometric chart and the geometric CUSUM chart are considered when the MLE and the Bayes estimator are used. The results show that chart performance with estimated control limits based on the Bayes estimator is generally better than that based on the MLE.

Keywords: ARL, Bayes estimator, geometric chart, MLE, Phase I sample

1. 서론

산업혁명 이후 제조업을 비롯한 대부분의 산업은 생산자 입장에서의 생산성과 효율성을 강조하는 방향으로 발전해왔다. 하지만 소비자의 요구가 다양해지고 기업 간의 경쟁이 격화되면서 생산성과 효율성뿐만 아니라 생산품의 품질을 체계적으로 관리할 필요가 생겼다. 생산품의 품질을 균일하게 유지하기 위해 일반적으로 통계적 공정관리(statistical process control; SPC) 기법을 사용한다. 여기서 통계적 공정관리는 품질특성치(quality characteristic)를 이용하여 공정의 상태를 파악하고 관리상태에서의 모수값을 유지하도록 공정을 관리하는 과정을 말한다. 통계적 공정관리는 주로 관리도(control chart)를 사용하여 공정의 변화를 탐지한다. 관리도는 공정의 상태를 파악하고, 공정의 변화를 최대한 빠르게 탐지

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]Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 156-756, Korea. E-mail: jaeheon@cau.ac.kr

하여 공정을 효율적으로 관리할 수 있게 하는 통계적 도구이다. 관리도의 종류에는 Shewhart 관리도를 비롯해 누적합(cumulative sum; CUSUM) 관리도, 지수가중이동평균(exponentially weighted moving average; EWMA) 관리도 등이 있다.

생산자의 품질 개선에 대한 노력이 계속되어, 산업 기술이 발달하고 공정이 자동화되면서 생산품의 품질이 높아지고 있다. 이러한 공정을 고품질공정(high-quality process)이라 한다. 고품질공정의 가장 큰 특징은 공정 내에 불량품이 매우 드물게 발생하여 공정의 불량률이 매우 낮다는 것이다. 따라서 공정의 불량률의 변화를 탐지하기 위해 전통적인 Shewhart p 관리도나 np 관리도를 사용하는 것은 바람직하지 않다. 고품질공정에서 p 관리도나 np 관리도를 사용할 경우, 표본크기를 설정하는데 많은 어려움이 있기 때문이다.

이에 대한 대안으로 기하분포(geometric distribution)를 따르는 확률변수를 관리통계량으로 사용하는 관리도를 고려할 수 있는데, Shewhart 형태의 관리도를 기하 관리도(geometric chart)라고 하며 누적합 형태의 관리도를 기하누적합 관리도(geometric CUSUM chart)라고 한다. 기하 관리도는 Calvin (1983)에 의해 처음으로 제안되었으며, Kaminsky 등 (1992)과 Woodall (1997) 등에 의해 고품질공정에서 불량률 변화의 탐지에 효율적이라고 알려지게 되었다. 기하누적합 관리도의 절차는 Bourke (1991)에 의해 처음 소개되었다.

일반적으로 관리도를 사용하는 과정은 관리상태에서의 모수를 추정하여 관리한계(control limit)를 설정하는 제1국면(Phase I)과 설정된 관리한계를 이용하여 이상상태 여부를 판단하는 제2국면(Phase II)으로 나누어 생각할 수 있다. 기존의 관리도에 관한 연구는 주로 제2국면에서의 관리도의 성능에 대해 진행되어져 왔다. 하지만 제1국면에서 추정한 관리상태의 모수가 실제 값과 같을 때 제2국면에서 연구자가 원하는 오경보율을 맞출 수 있기 때문에, 제1국면에서의 추정오차가 관리도의 성능에 미치는 영향을 연구하는 것은 의미가 있는 일이다. 추정오차는 표본의 크기에 영향을 받기 때문에, 관리도의 성능은 표본의 크기에 영향을 받는다고 할 수 있다. Qusenberry (1993)는 $\bar{X}$ 관리도와 X 관리도에서 표본의 크기에 따른 관리도의 성능을 연구하였고, 그 외 Yang 등 (2002), Zhang 등 (2013), Lee 등 (2013)을 비롯한 많은 연구자들이 표본의 크기에 따른 관리도의 성능을 연구하였다.

이 논문은 Zhang 등 (2013)과 Lee 등 (2013)과 같이 제1국면에서 관리상태의 모수를 추정하는 방법으로 베이즈추정량(Bayes estimator)의 사용을 제안하고 있다. Zhang 등 (2013)은 기하 관리도에서, 그리고 Lee 등 (2013)은 Bernoulli 누적합 관리도에서 베이즈추정량의 사용을 제안했는데, 베이즈추정량의 사용이 그들의 논문의 주된 내용이 아니기 때문에 다양한 경우에 대해 성능을 비교하지는 않았다. 이 논문에서는 베이즈추정량을 사용하는 경우의 성능을 알아보기 위해 기하 관리도와 기하누적합 관리도에서 최대우도추정량(maximum likelihood estimator; MLE)을 사용하는 경우와 비교했으며, 사전분포의 모수를 다양하게 설정하여 사전분포가 기하 관리도와 기하누적합 관리도의 성능에 어떠한 영향을 미치는지에 대하여 연구하였다.

2. 기하분포에 기초한 관리도

2.1. 기하 관리도

공정의 불량률은 p, 표본군의 표본크기는 m, 그리고 표본에서 발견되는 불량품의 수는 N이라 할 때, 전통적으로 사용하는 p 관리도와 np 관리도는 N/m과 N을 관리통계량으로 사용한다. 이때 불량품의 수 N은 모수가 m과 p인 이항분포(binomial distribution)를 따르는 확률변수이다. 그런데 고품질공정에서는 불량률이 매우 작기 때문에 적절한 표본크기를 설정하기가 쉽지 않다는 어려움이 있다. 이에 대한 대안으로 사용할 수 있는 기하 관리도는 불량품 사이에 관측된 양품의 수를 관리통계량으로 사용한다.

기하 관리도는 기하분포를 따르는 확률변수에 기반한 Shewhart 관리도이다. 확률변수 Y_i는 $(i-1)$번째 불량품과 i번째 불량품 사이에 발생한 양품의 수로 정의할 경우, Y_i는 모수가 p인 기하분포를 따른다는 사실이 잘 알려져 있다. 공정이 관리상태일 때의 불량률을 p_0라 할 때, Y_i의 확률질량함수는 다음과 같다.

$$f(y_i) = (1-p_0)^{y_i} \cdot p_0, \quad y_i = 0, 1, 2, \ldots.$$

관리도를 적용하기 위해서는 먼저 관리한계를 설정해야 하는데, 관리한계는 일반적으로 3σ 한계(3σ limit)를 사용한다. 3σ 한계는 관리통계량의 분포가 대칭형일 경우 유용하게 사용할 수 있지만, 기하분포와 같이 한쪽으로 치우쳐진 분포를 가정하는 경우 3σ 한계 보다는 확률한계(probability limit)를 사용하는 것이 더 타당하다. 확률한계는 양쪽 꼬리 확률이 오경보율과 같아지는 분위수를 사용하여 설정하는데, 오른쪽 꼬리 확률이 $\alpha/2$가 되는 관리한계를 관리상한(upper control limit; UCL)으로, 그리고 왼쪽 꼬리 확률이 $\alpha/2$가 되는 관리한계를 관리하한(lower control limit; LCL)으로 설정하는 것이다. 이때 오경보율 α는 공정이 관리상태일 때 관리통계량이 관리한계 밖으로 나갈 확률을 의미한다. 확률한계를 이용한 기하 관리도의 관리상한과 관리하한은 각각 식 (2.1)과 (2.2)를 만족하게 된다.

$$\sum_{y_i=0}^{\mathrm{LCL}-1} (1-p_0)^{y_i} \cdot p_0 = \frac{\alpha}{2}, \tag{2.1}$$

$$\sum_{y_i=\mathrm{UCL}+1}^{\infty} (1-p_0)^{y_i} \cdot p_0 = \frac{\alpha}{2}. \tag{2.2}$$

식 (2.1)과 (2.2)를 통해 식 (2.3)과 (2.4)와 같은 관리한계를 얻을 수 있다. 이때 Y_i는 이산형 확률변수이기 때문에 α를 정확하게 만족하는 관리한계는 얻을 수 없으며, 가장 가까운 정수값을 사용하고 있다.

$$\mathrm{LCL} = \frac{\ln(1-\alpha/2)}{\ln(1-p_0)}, \tag{2.3}$$

$$\mathrm{UCL} = \frac{\ln(\alpha/2)}{\ln(1-p_0)} - 1. \tag{2.4}$$

관리상태일 때의 불량률인 p_0가 알려져 있는 경우 식 (2.3)과 (2.4)를 사용할 수 있지만, 일반적으로 제1국면에서 표본을 추출하여 이를 추정하게 된다. 추정된 값을 $\hat{p}_0$이라 하면, 추정된 관리한계는 식 (2.5)와 (2.6)과 같이 표현할 수 있다. 표본크기 m은 고정된 값이라 할 때, 추정값 $\hat{p}_0$은 불량품의 수 N에 의존하기 때문에 $\widehat{\mathrm{LCL}}(N)$과 $\widehat{\mathrm{UCL}}(N)$이란 기호를 사용하였다.

$$\widehat{\mathrm{LCL}}(N) = \frac{\ln(1-\alpha/2)}{\ln(1-\hat{p}_0)}, \tag{2.5}$$

$$\widehat{\mathrm{UCL}}(N) = \frac{\ln(\alpha/2)}{\ln(1-\hat{p}_0)} - 1. \tag{2.6}$$

이 논문에서는 추정된 관리한계를 사용하는 기하 관리도의 성능을 평균런길이인 ARL(average run length)의 기댓값과 표준편차를 이용하여 살펴보고자 한다. 여기서 런길이(run length)는 주어진 공정의 상태하에서 이상신호를 줄 때까지 추출한 표본의 수를 나타내는데, 공정의 상태에 따라 관리상태일 때의 런길이와 이상상태일 때의 런길이로 구분할 수 있다. 관리상태일 때의 ARL은 오경보율과 관련이 있으며, 이상상태일 때의 ARL은 이상원인의 탐지 능력과 관련이 있다. 이 논문에서는 추정된 관리한계를 사용할 경우 주어진 오경보율, 즉 주어진 관리상태에서의 ARL을 평균적으로 잘 달성할 수 있는지와 ARL의 산포를 이용하여 관리도의 성능을 평가하고자 한다.

먼저 공정의 불량률이 p이고 식 (2.5)와 (2.6)의 관리한계를 사용하는 기하 관리도에서 i번째 관리통계량 Y_i가 관리한계를 벗어날 확률을 $\alpha(p,N)$이라 하면, $\alpha(p,N)$은

$$\begin{aligned}\alpha(p,N) &= \Pr\left(Y_i < \widehat{\mathrm{LCL}}(N)\right) + \Pr\left(Y_i > \widehat{\mathrm{UCL}}(N)\right)\\ &= \left\{1-(1-p)^{\widehat{\mathrm{LCL}}(N)}\right\} + (1-p)^{\widehat{\mathrm{UCL}}(N)+1}\end{aligned} \tag{2.7}$$

로 계산할 수 있다.

런길이를 R이라 할 때, N이 주어진 경우 런길이 R, 즉 $R\,|\,N$의 분포는 모수가 $\alpha(p,N)$인 기하분포를 따르며, N은 관리상태인 제1국면에서 발견된 불량품의 수이기 때문에 모수가 m과 p_0인 이항분포를 따르는 확률변수이다. ARL은 실제 공정의 불량률 p와 추정량 $\hat{p}_0$ (또는 N)에 의존하기 때문에 $\mathrm{ARL}(p,N)$으로 표현할 경우, 이에 대한 기댓값 $E_N[\mathrm{ARL}(p,N)]$과 표준편차 $\mathrm{SD}_N[\mathrm{ARL}(p,N)]$은

$$E_N[\mathrm{ARL}(p,N)] = E_N[E(R\,|\,N)] = E_N\left[\frac{1}{\alpha(p,N)}\right], \tag{2.8}$$

$$\begin{aligned}\mathrm{SD}_N[\mathrm{ARL}(p,N)] &= \mathrm{SD}_N[E(R\,|\,N)]\\ &= \sqrt{E_N[E^2(R\,|\,N)] - E_N^2[E(R\,|\,N)]}\\ &= \sqrt{E_N\left[\frac{1}{\alpha^2(p,N)}\right] - E_N^2\left[\frac{1}{\alpha(p,N)}\right]}\end{aligned} \tag{2.9}$$

로 나타낼 수 있으며,

$$\begin{aligned}E_N\left[\frac{1}{\alpha(p,N)}\right] &= \sum_{n=0}^{m}\frac{1}{\alpha(p,n)}\binom{m}{n}{p_0}^n(1-p_0)^{m-n},\\ E_N\left[\frac{1}{\alpha^2(p,N)}\right] &= \sum_{n=0}^{m}\frac{1}{\alpha^2(p,n)}\binom{m}{n}{p_0}^n(1-p_0)^{m-n}\end{aligned}$$

을 이용하여 계산할 수 있다.

2.2. 기하누적합 관리도

일반적으로 공정모수의 큰 변화를 탐지하고자 하는 경우 Shewhart 관리도를 사용하고, 작은 변화를 탐지하고자 하는 경우 누적합 관리도를 사용한다. 누적합 관리도는 현재 시점의 관측값뿐만 아니라 과거의 관측값을 함께 이용하기 때문에 작은 변화에도 민감하게 반응한다는 장점이 있다. 그러나 Shewhart 관리도는 공정이 관리상태일 때의 모수값만 사전에 설정하면 되지만, 누적합 관리도는 공정이 관리상태일 때의 모수값뿐만 아니라 탐지하고자 하는 모수값까지 설정해야 관리도를 적용할 수 있다. 또한 누적합 관리도는 공정모수의 증가를 탐지하는 절차와 감소를 탐지하는 절차를 따로 적용해야 한다는 특징이 있다. 일반적으로 불량률이 증가하는 것에 더 많은 관심이 있기 때문에, 이 논문에서는 불량률이 증가하는지, 즉 불량품 사이의 양품의 평균 개수가 감소하는지를 탐지하는 음의 누적합 관리도 절차를 고려하고자 한다.

시점 i에서 음의 기하누적합 관리도의 관리통계량 C_i는

$$C_i = \max\{\,0, C_{i-1} - Y_i + k\}$$

로 정의할 수 있으며, 관리한계 H에 대해서 $C_i \geq H$인 경우 이상상태의 신호를 주게 된다. 여기서 k는

참고값(reference value)이라 하며, 로그확률비에 근거하여

$$k = \frac{\ln\left(\dfrac{p_1(1-p_0)}{p_0(1-p_1)}\right)}{\ln\left(\dfrac{1-p_0}{1-p_1}\right)} \tag{2.10}$$

를 사용한다.

기하누적합 관리도의 성능은 ANOS(average number of observations to signal)의 기댓값과 표준편차를 이용하여 살펴보고자 한다. 여기서 ANOS는 ARL과 유사한 개념으로 이상신호까지 관측한 관측값의 평균 개수를 나타낸다. 일반적으로 누적합 관리도에서는 ARL과 ANOS가 닫혀진 형태로 표현되지 않기 때문에 모의실험 또는 마코프체인(Markov chain)을 이용하여 계산할 수 있다. 그런데 Chang과 Gan (2001)은 관리한계의 특정한 범위 내에서 기하누적합 관리도의 ANOS를 계산하는 정확한 표현식을 다음과 같이 제안하였고, 이 논문에서는 이 식을 사용하였다.

$$\mathrm{ANOS}(p) = \begin{cases} \dfrac{2\left\{1 - q^{c-1}\left((H-1)\,p+1\right)\right\} + q^{c-H}}{p\,(1-q^{c-1}-(H-1)p\,q^{c-1})}, & 0 \le H \le c, \\ \dfrac{a}{b}, & c+1 \le \ H \le 2c, \end{cases} \tag{2.11}$$

p는 실제 공정의 불량률, $q = 1-p$, c는 식 (2.10)의 참고값 k보다 같거나 작은 정수값을 나타낸다. 또한 a와 b는

$$\begin{aligned} a &= q^{c-1} + 3\,q^{H-c} - 3\,q^{H-1} + p\,q^{c-1}\left\{\left(1+q^{c-1}\right)(c-H) + (c-1)\right\} \\ &\quad + \binom{H-c+1}{2} p^2\,q^{c-2}\left(q^c - 3\,q^H\right) + 3\,p\,q^{H-1}\left\{q^{c-1}(H-c)(p(H-c)+1) - H + 1\right\}, \\ b &= p^3\,q^{H+c-2}\left\{(c-H)^2 - \binom{H-c+1}{2}\right\} + (c-H)\,p^2\,q^{H-1}\left(1-q^{c-1}\right) \\ &\quad + p\,q^{H-1}\left(q^{1-c} - p\,(c-1) - 1\right) \end{aligned}$$

을 나타낸다.

제1국면을 통하여 관리상태의 불량률 p_0를 추정할 경우 추정량 $\hat{p}_0$과 $\hat{p}_1$을 식 (2.10)에 대입한다. 이때 탐지하기를 원하는 불량률은 일반적으로 $\hat{p}_1 = \delta\hat{p}_0$으로 설정한다. 따라서 k, c, 그리고 주어진 관리상태에서의 특성을 만족하는 관리한계 H는 추정량 $\hat{p}_0$, 즉 불량품의 수 N에 의존하게 되고, ANOS 또한 N에 의존하는 값이 된다. 추정량 $\hat{p}_0$을 사용한 ANOS를 ANOS(p,N)으로 표현할 경우, 이에 대한 기댓값 $E_N[\mathrm{ANOS}(p,N)]$과 표준편차 $\mathrm{SD}_N[\mathrm{ANOS}(p,N)]$은

$$E_N[\mathrm{ANOS}(p,N)] = \sum_{n=0}^{m} \mathrm{ANOS}(p,n)\binom{m}{n} {p_0}^n (1-p_0)^{m-n}, \tag{2.12}$$

$$\mathrm{SD}_N[\mathrm{ANOS}(p,N)] = \sqrt{E_N[\mathrm{ANOS}^2(p,N)] - E_N^2[\mathrm{ANOS}(p,N)]} \tag{2.13}$$

으로 표현되며,

$$E_N\left[\mathrm{ANOS}^2(p,N)\right] = \sum_{n=0}^{m} \mathrm{ANOS}^2(p,n)\binom{m}{n} {p_0}^n (1-p_0)^{m-n}$$

을 이용하여 계산할 수 있다.

Chang과 Gan (2001)이 제안한 식 (2.11)에는 ANOS를 계산할 수 있는 관리한계에 제약이 있다. 식 (2.11)에 의하면 $0 \leq H \leq 2c$인 경우에만 ANOS를 구할 수 있기 때문에, 본 연구의 모의실험에서는 이 범위에 해당하는 경우를 고려하였다.

3. p_0의 추정량

제1국면을 통해 관리상태일 때의 불량률 p_0를 추정할 때, 전통적으로 사용하는 추정량은 최대우도추정량이다. 표본크기가 m이고 불량품의 수가 N인 경우 최대우도추정량 $\hat{p}_{0,M}$은

$$\hat{p}_{0,M} = \frac{N}{m} \tag{3.1}$$

이 된다. 그러나 최대우도추정량을 사용할 때 가장 큰 문제점은 표본에서 불량품이 발견되지 않은 경우, 즉 $N = 0$인 경우 $\hat{p}_{0,M} = 0$이 되어 식 (2.5)와 (2.6)의 관리한계가 정의되지 않는다는 것이다. 특히 이 논문에서 촛점을 맞추고 있는 고품질공정에서는 표본크기 m에 따라서 이런 경우가 발생할 가능성을 무시할 수 없다. 이러한 문제점에 대해 Zhang 등 (2013)과 Lee 등 (2013)에서 언급한 바 있다.

최대우도추정량의 대안으로서 베이즈추정량을 고려할 수 있다. 베이즈추정량은 실무자의 사전지식을 반영할 수 있다는 장점이 있지만, 이를 수량화한 사전분포(prior distribution)를 미리 정의해야 한다. 이 논문에서는 p_0에 대한 사전분포로서 베타분포(beta distribution)를 사용하고자 한다. p_0의 사전분포로 Beta(a, b)를 가정하고, 표본크기가 m이고 불량품의 수가 N인 경우 p_0의 사후분포(posterior distribution)는 Beta$(a + N, b + m - N)$이 됨을 쉽게 알 수 있다. 따라서 p_0에 대한 베이즈추정량 $\hat{p}_{0,B}$은

$$\hat{p}_{0,B} = \frac{a+N}{a+b+m} \tag{3.2}$$

이 된다. 식 (3.2)의 베이즈추정량은

$$\hat{p}_{0,B} = \left(\frac{a+b}{a+b+m}\right)\left(\frac{a}{a+b}\right) + \left(\frac{m}{a+b+m}\right)\left(\frac{N}{m}\right)$$

으로 표현할 수 있는데, 이것은 사전분포의 평균과 표본에서 얻어지는 최대우도추정량의 가중평균의 형태가 된다. 따라서 표본크기 m이 큰 경우 최대우도추정량에 상대적으로 더 큰 가중이 부여되며, 그렇지 않은 경우 사전분포의 평균에 더 큰 가중이 부여됨을 알 수 있다.

사전분포로 사용하는 베타분포는 모수 a와 b에 따라 다양한 형태를 가지는데, 이 논문에서는 p_0가 작은 경우에 관심이 있기 때문에 지수분포(exponential distribution) 형태인 $a = 1$, $b > 1$인 경우와 사전분포의 평균이 작은 값을 가지는 단봉(unimodal)형의 $a > 1$, $b > 1$인 경우를 고려하였다.

식 (3.1)의 $\hat{p}_{0,M}$과 식 (3.2)의 $\hat{p}_{0,B}$를 식 (2.5)와 (2.6)에 $\hat{p}_0$ 대신 대입하면 기하 관리도의 추정된 관리한계를 얻을 수 있으며, 식 (2.7), (2.8), (2.9)를 사용하면 최대우도추정량과 베이즈추정량을 이용한 기하 관리도의 ARL의 기댓값과 표준편차를 구할 수 있다. 또한 식 (2.10)에 대입하고 식 (2.11), (2.12), (2.13)을 사용하면 기하누적합 관리도의 ANOS의 기댓값과 표준편차를 구할 수 있다.

4. 추정량의 성능 비교

4.1. 기하 관리도에서 베이즈추정량의 성능

기하 관리도에서 베이즈추정량의 성능을 알아보기 위해 최대우도추정량을 사용한 경우와 비교하고

Table 4.1. ARL and SDARL values of the g-chart using Bayes estimator when $\text{ARL}_0 = 370.4$ and $p_0 = 0.0001$

	m	Beta $(1, 19999)$	Beta $(2, 39998)$	Beta $(1, 9999)$	Beta $(2, 19998)$	Beta $(1, 4999)$	Beta $(2, 9998)$	MLE
Mean		0.00005		0.0001		0.0002		
Variance		2.50E-09	1.25E-09	1.00E-08	5.00E-09	4.00E-08	2.00E-08	
ARL	50,000	361.4	403.6	313.2	327.6	280.5	258.3	291.8
	100,000	371.9	405.8	334.2	340.4	311.7	294.3	326.3
	200,000	377.5	402.1	350.7	352.8	335.9	323.0	348.2
	500,000	378.2	392.3	363.4	363.8	355.7	348.4	363.0
	1,000,000	376.2	384.5	367.6	367.7	363.2	358.9	367.5
	2,000,000	374.0	378.6	369.3	369.4	367.0	364.6	369.3
	5,000,000	372.1	374.0	370.1	370.1	369.1	368.1	370.1
	∞	370.4	370.4	370.4	370.4	370.4	370.4	370.4
SDARL	50,000	139.1	112.0	161.1	156.7	169.1	169.2	166.0
	100,000	134.0	113.2	149.9	146.3	156.1	156.5	153.6
	200,000	125.9	113.1	135.3	132.8	139.1	139.5	137.8
	500,000	105.9	100.8	109.4	108.2	110.9	111.0	110.7
	1,000,000	86.2	84.0	87.6	86.9	88.2	88.1	88.2
	2,000,000	66.4	65.6	66.9	66.6	67.1	67.1	67.2
	5,000,000	44.5	44.3	44.6	44.6	44.7	44.7	44.7
	∞	0.0	0.0	0.0	0.0	0.0	0.0	0.0

MLE = maximum likelihood estimator; ARL = average run length; SDARL = standard deviation of ARL.

자 한다. 제1국면에서 더 좋은 추정량이라고 할 수 있는 기준은 연구자에 의해 설정된 관리상태의 ARL(ARL_0) 또는 ANOS(ANOS_0)에 평균적으로 더 가깝고 변동이 작은 것이 될 것이다.

Table 4.1부터 Table 4.3까지는 기하 관리도(g-chart)에서 베이즈추정량과 최대우도추정량을 사용하는 경우 표본크기 m에 따른 ARL_0의 기댓값(ARL)과 표준편차(SDARL)를 계산한 결과이다. 여기서 $m = \infty$인 경우는 p_0의 참값을 알고 있는 경우를 나타낸다. 고품질공정을 가정하기 때문에 p_0는 0.0001, 0.0005, 0.001을 고려하였고, ARL_0는 Shewhart 관리도에서 가장 널리 사용되는 3σ 한계의 ARL_0인 370.4를 고려하였다.

제1국면에서 베이즈추정량을 사용하는 것에 대한 효율을 평가할 때 사전분포의 영향을 함께 살펴볼 필요가 있다. 사전분포의 평균이 p_0와 일치하는 경우뿐만 아니라 p_0보다 작은 경우와 p_0보다 큰 경우의 사전분포도 함께 고려하였다. 4장에서 언급한 바와 같이, 평균이 동일한 사전분포에 대해 분포가 지수적으로 감소하는 형태($a = 1, b > 1$)와 단봉형인 형태($a > 1, b > 1$)인 두 가지 경우를 고려하였다. 제1국면에서 최대우도추정량을 사용할 때 표본에서 불량품이 발견되지 않은 경우($N = 0$) 제2국면에서 불량품이 발견되는 즉시 이상상태의 신호를 주는 것, 즉 식 (2.7)에서 $\alpha(p, N = 0) = 1$로 처리하였다.

Table 4.1은 $p_0 = 0.0001$인 경우 기하 관리도의 ARL_0의 기댓값과 표준편차를 나타낸 표이다. Table 4.1의 3열부터 8열은 베이즈추정량을 사용한 경우이고, 9열은 최대우도추정량을 사용한 경우이다. Table 4.1은 실제 p_0값이 0.0001인 경우인데, 사전분포의 평균이 p_0의 1/2배인 0.00005, p_0와 동일한 0.0001, p_0의 2배인 0.0002를 갖는 베타분포를 고려하였다. 편의상 단봉형인 $a > 1$인 경우로 $a = 2$인 경우만 고려하였다. 사전분포의 평균이 동일한 경우 $a = 1$인 베타분포의 분산은 $a = 2$인 경우의 분산에 비해 거의 2배 정도 더 크다. 물론 여기서 고려하는 베타분포의 분산은 아주 작은 값이기 때문에 그 차이는 미미하다고 할 수 있다.

Table 4.2. ARL and SDARL values of the g-chart using Bayes estimator when $ARL_0 = 370.4$ and $p_0 = 0.0005$

	m	Beta $(1, 3999)$	Beta $(2, 7998)$	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	MLE
Mean		0.00025		0.0005		0.001		
Variance		6.25E-08	3.12E-08	2.50E-07	1.25E-07	9.99E-07	5.00E-07	
ARL	50,000	378.3	400.1	354.7	356.2	342.0	330.5	353.1
	100,000	378.2	392.3	363.4	363.8	355.7	348.4	363.0
	200,000	376.2	384.5	367.6	367.7	363.2	358.9	367.5
	500,000	373.4	377.2	369.6	369.6	367.7	365.8	369.6
	1,000,000	372.1	374.0	370.1	370.1	369.1	368.1	370.1
	2,000,000	371.3	372.3	370.3	370.3	369.7	369.2	370.3
	5,000,000	370.7	371.2	370.3	370.3	370.1	369.9	370.3
	∞	370.4	370.4	370.4	370.4	370.4	370.4	370.4
SDARL	50,000	122.0	111.4	129.6	127.5	132.8	133.0	131.8
	100,000	105.9	100.8	109.4	108.2	110.9	111.0	110.7
	200,000	86.2	84.0	87.6	86.9	88.2	88.1	88.2
	500,000	60.5	59.9	60.9	60.7	61.0	61.0	61.1
	1,000,000	44.5	44.3	44.6	44.6	44.7	44.7	44.7
	2,000,000	32.1	32.1	32.2	32.2	32.2	32.2	32.2
	5,000,000	20.6	20.6	20.6	20.6	20.6	20.6	20.6
	∞	0.0	0.0	0.0	0.0	0.0	0.0	0.0

MLE = maximum likelihood estimator; ARL = average run length; SDARL = standard deviation of ARL.

Table 4.1의 결과를 살펴보면, p_0가 아주 작은 공정에서 제1국면의 표본크기로 아주 큰 값을 사용하지 않을 경우(m < 1,000,000) 사전에 설정한 ARL_0를 평균적으로 잘 달성하지 못하며, 그 산포도 매우 크다는 것을 알 수 있다. 사전분포의 평균이 실제 p_0값보다 큰 베타분포로 설정한 경우를 제외하고, 작거나 같은 베타분포로 설정한 경우 베이즈추정량을 사용한 기하 관리도의 효율이 최대우도추정량을 사용한 경우에 비해 좋은 것으로 나타났다. 표본크기 m이 크지 않은 경우 사전분포로 Beta$(1, 19999)$를 사용한 경우의 효율이 가장 좋음을 알 수 있다. 또한 m이 점점 커질수록 사전분포의 평균이 실제값과 같은 Beta$(1, 9999)$와 Beta$(2, 19998)$로 설정한 경우의 효율이 좋아짐을 알 수 있다. m이 아주 큰 경우 베이즈추정량을 사용하는 경우와 최대우도추정량을 사용하는 경우의 효율에는 큰 차이가 나지 않는다.

Table 4.2는 $p_0 = 0.0005$인 경우이고, Table 4.3은 $p_0 = 0.001$인 경우의 결과이다. Table 4.2와 Table 4.3의 결과는 Table 4.1의 결과와 유사한 경향을 나타냈으며, p_0가 커질수록 표본크기 m에 대해 사전분포 및 추정량에 따른 차이는 점점 작아짐을 알 수 있다. 사전분포의 특징을 살펴보면, 평균이 실제 p_0보다 작거나 큰 베타분포로 설정된 경우에는 전반적으로 $a = 1$일 때 효율이 더 좋았으며, 평균이 p_0와 동일한 베타분포로 설정된 경우에는 $a = 2$일 때 효율이 더 좋음을 알 수 있다.

4.2. 기하누적합 관리도에서 베이즈추정량의 성능

이번 절에서는 기하누적합 관리도에서 베이즈추정량의 성능에 대해 살펴보고자 한다. 기하 관리도에서 추정량의 성능을 알아보기 위해 ARL을 사용한 것과 달리, 기하누적합 관리도에서는 Chang과 Gan (2001)의 결과를 이용하기 위해 ANOS를 사용하였다.

관리상태일 때의 불량률 p_0는 기하 관리도의 경우와 동일하게 0.0001, 0.0005, 0.001로 설정했는데, 기하누적합 관리도를 사용하기 위해서는 3장에서 언급한 바와 같이 탐지하고자 하는 기각품질수준의 값

Table 4.3. ARL and SDARL values of the g-chart using Bayes estimator when $\mathrm{ARL}_0 = 370.4$ and $p_0 = 0.001$

	m	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	Beta $(1, 499)$	Beta $(2, 998)$	MLE
Mean		0.0005		0.001		0.002		
variance		2.50E-07	1.25E-07	9.99E-07	5.00E-07	3.99E-06	2.00E-06	
ARL	50,000	378.2	392.3	363.4	363.8	355.7	348.4	363.0
	100,000	376.2	384.5	367.6	367.7	363.2	358.9	367.5
	200,000	374.0	378.6	369.3	369.4	367.0	364.6	369.3
	500,000	372.1	374.0	370.1	370.1	369.1	368.1	370.1
	1,000,000	371.3	372.3	370.3	370.3	369.7	369.2	370.2
	2,000,000	370.8	371.3	370.3	370.3	370.1	369.8	370.3
	5,000,000	370.6	370.8	370.4	370.4	370.2	370.1	370.4
	∞	370.4	370.4	370.4	370.4	370.4	370.4	370.4
SDANOS	50,000	105.9	100.8	109.4	108.2	110.9	111.0	110.7
	100,000	86.2	84.0	87.6	86.9	88.2	88.1	88.2
	200,000	66.4	65.6	66.9	66.6	67.1	67.1	67.2
	500,000	44.5	44.3	44.6	44.6	44.7	44.7	44.7
	1,000,000	32.1	32.1	32.2	32.2	32.2	32.2	32.2
	2,000,000	23.0	22.9	23.0	23.0	23.0	23.0	23.0
	5,000,000	14.6	14.6	14.6	14.6	14.6	14.6	14.6
	∞	0.0	0.0	0.0	0.0	0.0	0.0	0.0

MLE = maximum likelihood estimator; ARL = average run length; SDARL = standard deviation of ARL.

Table 4.4. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when $\mathrm{ANOS}_0 = 300{,}000$, $p_0 = 0.0001$, and $\delta = 3$

	m	Beta $(1, 19999)$	Beta $(2, 39998)$	Beta $(1, 9999)$	Beta $(2, 19998)$	Beta $(1, 4999)$	Beta $(2, 9998)$	MLE
Mean		0.00005		0.0001		0.0002		
Variance		2.50E-09	1.25E-09	1.00E-08	5.00E-09	4.00E-08	2.00E-08	
ANOS	200,000	248,988	231,058	266,419	265,031	258,490	278,544	238,983
	500,000	286,378	277,620	298,964	298,329	300,646	310,149	299,646
	1,000,000	302,066	292,996	310,028	311,035	315,332	318,792	310,305
	2,000,000	302,840	297,898	307,905	307,828	310,513	312,987	307,991
	5,000,000	301,230	299,227	303,257	303,239	304,275	305,284	303,271
	∞	300,009	300,009	300,009	300,009	300,009	300,009	300,009
SDANOS	200,000	153,693	134,706	169,922	163,698	170,254	181,157	159,456
	500,000	127,811	120,807	135,721	133,347	138,558	141,859	138,225
	1,000,000	100,303	95,722	103,593	103,025	105,814	106,463	104,671
	2,000,000	72,030	70,216	73,548	73,156	74,334	74,700	73,939
	5,000,000	45,163	44,690	45,552	45,455	45,744	45,846	45,644
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

을 미리 설정해야한다. 이 논문에서 고려하는 불량률은 아주 작은 경우이기 때문에, 그 차이가 의미가 있도록 δ값은 3과 5를 고려하였다. 베이즈추정량을 위해 가정하는 사전분포는 기하 관리도의 경우와 동일한 베타분포를 사용하였고, 주어진 ANOS_0와 p_0에 대해 추정량을 사용할 경우 ANOS_0의 기댓

Table 4.5. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when ANOS_0 = 300,000, $p_0 = 0.0001$, and $\delta = 5$

	m	Beta $(1, 19999)$	Beta $(2, 39998)$	Beta $(1, 9999)$	Beta $(2, 19998)$	Beta $(1, 4999)$	Beta $(2, 9998)$	MLE
Mean		0.00005		0.0001		0.0002		
Variance		2.50E-09	1.25E-09	1.00E-08	5.00E-09	4.00E-08	2.00E-08	
ANOS	200,000	306,760	267,963	354,326	349,173	382,355	404,627	360,279
	500,000	303,690	286,826	321,985	321,126	331,777	340,671	322,903
	1,000,000	302,027	293,381	311,044	310,810	315,695	320,139	311,264
	2,000,000	301,070	296,688	305,537	305,478	307,803	310,023	305,587
	5,000,000	300,442	298,679	302,223	302,209	303,107	304,002	302,226
	∞	300,016	300,016	300,016	300,016	300,016	300,016	300,016
SDANOS	200,000	203,049	164,213	241,771	226,311	265,039	270,247	259,527
	500,000	126,986	116,251	136,184	133,113	141,149	142,828	139,428
	1,000,000	89,179	85,290	92,365	91,370	94,018	94,647	93,372
	2,000,000	62,812	61,412	63,927	63,592	64,487	64,722	64,252
	5,000,000	39,620	39,267	39,903	39,818	40,042	40,102	39,982
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

Table 4.6. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when ANOS_0 = 70,000, $p_0 = 0.0005$, and $\delta = 3$

	m	Beta $(1, 3999)$	Beta $(2, 7998)$	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	MLE
Mean		0.00025		0.0005		0.001		
Variance		6.25E-08	3.12E-08	2.50E-07	1.25E-07	9.99E-07	5.00E-07	
ANOS	100,000	42,766	43,768	41,544	41,661	42,911	40,050	41,426
	200,000	50,741	51,323	50,015	50,043	48,247	49,101	47,411
	500,000	59,110	59,323	58,827	58,834	58,061	58,469	58,821
	1,000,000	65,152	65,490	65,164	65,168	64,922	65,149	65,164
	2,000,000	68,973	68,823	69,019	69,121	69,249	69,154	69,020
	5,000,000	70,047	69,948	70,145	70,144	70,189	70,241	70,143
	∞	70,010	70,010	70,010	70,010	70,010	70,010	70,010
SDANOS	100,000	26,513	25,074	27,837	27,806	28,900	29,002	27,875
	200,000	26,191	24,910	27,436	27,398	28,109	28,619	27,609
	500,000	22,810	21,900	23,719	23,706	24,457	24,630	23,732
	1,000,000	17,375	16,384	17,961	17,953	18,495	18,559	17,968
	2,000,000	10,468	10,238	10,908	10,699	10,729	11,162	10,912
	5,000,000	5,001	4,992	5,016	5,015	5,037	5,031	5,025
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

값(ANOS)과 표준편차(SDANOS)를 Table 4.4에서 Table 4.9에 제시하였다.

Table 4.4는 ANOS_0 = 300,000, $p_0 = 0.0001$, $\delta = 3$인 경우이고, Table 4.5는 $\delta = 5$인 경우 베이즈추정량을 사용한 기하누적합 관리도(g-CUSUM chart)의 ANOS와 SDANOS를 계산한 결과이다. Table

Table 4.7. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when $ANOS_0 = 100{,}000$, $p_0 = 0.0005$, and $\delta = 5$

	m	Beta $(1, 3999)$	Beta $(2, 7998)$	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	MLE
Mean		0.00025		0.0005		0.001		
Variance		6.25E-08	3.12E-08	2.50E-07	1.25E-07	9.99E-07	5.00E-07	
ANOS	100,000	102,866	96,340	109,944	109,562	113,769	117,191	110,356
	200,000	101,501	98,206	104,963	104,843	106,744	108,440	105,064
	500,000	100,614	99,283	101,977	101,948	102,658	103,333	101,997
	1,000,000	100,314	99,644	100,982	100,981	101,333	101,654	100,991
	2,000,000	100,161	99,829	100,495	100,494	100,656	100,829	100,500
	5,000,000	100,067	99,931	100,200	100,200	100,264	100,334	100,201
	∞	99,980	99,980	99,980	99,980	99,980	99,980	99,980
SDANOS	100,000	49,354	44,703	53,480	52,180	55,671	56,582	54,807
	200,000	34,144	32,477	35,523	35,142	36,234	36,565	35,925
	500,000	21,272	20,863	21,626	21,524	21,782	21,882	21,713
	1,000,000	14,974	14,822	15,088	15,063	15,140	15,177	15,119
	2,000,000	10,555	10,509	10,591	10,586	10,619	10,622	10,608
	5,000,000	6,668	6,652	6,678	6,676	6,680	6,686	6,680
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

Table 4.8. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when $ANOS_0 = 30{,}000$, $p_0 = 0.001$, and $\delta = 3$

	m	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	Beta $(1, 499)$	Beta $(2, 998)$	MLE
Mean		0.0005		0.001		0.002		
Variance		2.50E-07	1.25E-07	9.99E-07	5.00E-07	3.99E-06	2.00E-06	
ANOS	50,000	29,174	27,760	29,893	30,520	30,925	31,012	29,965
	100,000	30,205	29,363	31,131	31,102	31,536	32,039	31,160
	200,000	30,290	29,794	30,793	30,791	31,045	31,302	30,805
	500,000	30,126	29,923	30,326	30,327	30,432	30,529	30,328
	1,000,000	30,064	29,966	30,166	30,165	30,212	30,267	30,166
	2,000,000	30,033	29,981	30,083	30,083	30,105	30,133	30,082
	5,000,000	30,013	29,993	30,033	30,033	30,042	30,053	30,033
	∞	30,029	30,029	30,029	30,029	30,029	30,029	30,029
SDANOS	50,000	13,045	12,075	13,563	13,594	14,140	14,177	13,817
	100,000	10,024	9,608	10,404	10,296	10,579	10,689	10,520
	200,000	7,204	7,017	7,356	7,317	7,428	7,471	7,396
	500,000	4,514	4,468	4,554	4,543	4,573	4,583	4,563
	1,000,000	3,181	3,165	3,195	3,191	3,199	3,205	3,198
	2,000,000	2,244	2,238	2,249	2,248	2,249	2,253	2,250
	5,000,000	1,418	1,417	1,419	1,419	1,420	1,420	1,420
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

Table 4.9. ANOS and SDANOS values of the g-CUSUM chart using Bayes estimator when ANOS_0 = 50,000, $p_0 = 0.001$, and $\delta = 5$

	m	Beta $(1, 1999)$	Beta $(2, 3998)$	Beta $(1, 999)$	Beta $(2, 1998)$	Beta $(1, 499)$	Beta $(2, 998)$	MLE
Mean		0.0005		0.001		0.002		
Variance		2.50E-07	1.25E-07	9.99E-07	5.00E-07	3.99E-06	2.00E-06	
ANOS	50,000	51,424	48,166	54,959	54,768	56,869	58,576	55,198
	100,000	50,748	49,101	52,471	52,417	53,386	54,209	52,534
	200,000	50,387	49,564	51,230	51,222	51,664	52,081	51,252
	500,000	50,163	49,826	50,496	50,496	50,662	50,832	50,500
	1,000,000	50,083	49,916	50,250	50,250	50,335	50,417	50,251
	2,000,000	50,044	49,962	50,128	50,128	50,167	50,211	50,127
	5,000,000	50,015	49,981	50,048	50,048	50,069	50,081	50,049
	∞	49,947	49,947	49,947	49,947	49,947	49,947	49,947
SDANOS	50,000	24,645	22,334	26,688	26,058	27,829	28,234	27,428
	100,000	17,049	16,232	17,753	17,547	18,127	18,274	17,970
	200,000	11,920	11,619	12,166	12,095	12,290	12,346	12,226
	500,000	7,477	7,402	7,542	7,521	7,569	7,587	7,556
	1,000,000	5,278	5,252	5,298	5,294	5,305	5,314	5,300
	2,000,000	3,727	3,718	3,733	3,733	3,741	3,739	3,734
	5,000,000	2,355	2,354	2,357	2,357	2,358	2,358	2,358
	∞	0	0	0	0	0	0	0

MLE = maximum likelihood estimator;
ANOS = average number of observations to signal; SDANOS = standard deviation of ANOS.

4.4와 Table 4.5의 결과를 살펴보면, 사전분포의 평균이 실제 p_0값보다 큰 베타분포로 설정한 경우 베이즈추정량을 사용할 때의 효율은 최대우도추정량을 사용할 때의 효율에 비해 좋지 않았지만, 작은 경우와 동일한 경우에는 베이즈추정량을 사용할 때의 효율이 전반적으로 더 좋음을 알 수 있다. $\delta = 3$이고 표본크기 m이 크지 않은 경우에는 사전분포의 평균이 실제 p_0값과 동일한 베타분포인 $\text{Beta}(1, 9999)$와 $\text{Beta}(2, 19998)$로 설정한 경우의 효율이 좋았지만, m이 어느 정도 큰 경우와 $\delta = 5$인 경우에는 사전분포의 평균이 더 작은 베타분포인 $\text{Beta}(1, 19999)$와 $\text{Beta}(2, 39998)$인 경우의 효율이 전반적으로 더 좋게 나타났다.

Table 4.6과 Table 4.7은 $p_0 = 0.0005$인 경우인데, Table 4.6은 ANOS_0 = 70,000, $\delta = 3$이고 Table 4.7은 ANOS_0 = 100,000, $\delta = 5$로 설정했을 때의 결과이다. 전반적으로 $p_0 = 0.0001$의 경우와 유사한 경향을 보임을 알 수 있지만, $p_0 = 0.0005$인 경우에는 사전분포의 평균이 실제 p_0보다 더 작은 베타분포인 $\text{Beta}(1, 3999)$와 $\text{Beta}(2, 7998)$인 경우의 효율이 전반적으로 좋게 나타남을 알 수 있다.

Table 4.8과 Table 4.9는 $p_0 = 0.001$인 경우인데, Table 4.8은 ANOS_0 = 30,000, $\delta = 3$이고 Table 4.9는 ANOS_0 = 50,000, $\delta = 5$로 설정했을 때의 결과이다. 기하 관리도의 경우와 유사하게, p_0가 커질 경우 표본크기 m에 대해 사전분포 및 추정량에 따른 차이는 크지 않음을 알 수 있다. 큰 차이는 아니지만 이 경우에도 사전분포의 평균이 실제 p_0값보다 작은 베타분포인 $\text{Beta}(1, 1999)$와 $\text{Beta}(2, 3998)$인 경우의 효율이 전반적으로 제일 좋게 나타났다.

5. 결론

고품질공정에서는 불량품이 매우 드물게 관측되기 때문에 p 관리도나 np 관리도와 같은 전통적인 관리

도 대신 다른 관리도를 사용하는 것이 더 효율적일 수 있다. 이 논문에서는 고품질공정에 적합하다고 알려진 기하 관리도와 기하누적합 관리도를 고려하였다. 여기서 기하 관리도는 기하분포를 따르는 확률변수를 이용한 Shewhart 형태의 관리도이고, 기하누적합 관리도는 누적합 형태의 관리도를 말한다.

관리도를 설계할 때, 공정모수가 알려져 있지 않은 경우 제1국면에서 일반적으로 최대우도추정량을 이용하여 공정모수를 추정한다. 하지만 제1국면에서 최대우도추정량을 사용하는 경우, 표본에서 불량품이 발생하지 않았을 때에는 관리한계를 정의할 수 없다는 문제점이 발생한다. 이와 같은 문제점을 해결하기 위해 이 논문에서는 베이즈추정량을 사용할 것을 제안하고 그 성능을 살펴보았다.

베이즈추정량과 최대우도추정량을 사용했을 때 그 성능은 ARL_0(또는 ANOS_0)의 기댓값과 표준편차로 비교하였다. ARL_0(또는 ANOS_0)의 기댓값이 주어진 값을 잘 달성하며 표준편차 또한 작은 값을 가질 경우 성능이 더 좋다고 말할 수 있다. 기하 관리도와 기하누적합 관리도에서 전반적인 경향은 유사했으며, p_0값이 작을 때 표본크기가 크지 않은 경우 베이즈추정량을 사용하는 것은 효율적임을 알 수 있었다. 기하 관리도에서 표본크기가 크지 않은 경우에는 사전분포의 평균이 실제값보다 작은 베타분포로 설정한 경우의 효율이 좋았으며, 표본크기가 커질수록 평균이 실제값과 같은 베타분포로 설정한 경우의 효율이 좋게 나타났다. 기하누적합 관리도에서는 전반적으로 사전분포의 평균이 실제값보다 작은 베타분포로 설정한 경우의 효율이 좋게 나타남을 알 수 있었다. p_0값이 어느 정도 크거나 표본크기가 아주 큰 경우에는 사전분포 및 추정량에 따른 차이가 크지 않은 것으로 나타났다.

베이즈추정량을 사용하기 위해서는 사전분포를 설정하는 어려움은 있지만, 공정에 대한 실무자의 경험과 지식을 올바르게 반영할 경우 전통적으로 사용하는 최대우도추정량에 비해 더 좋은 효율을 얻을 수 있다고 판단된다. 향후 다른 관리도의 설계에서도 베이즈추정량을 사용하는 경우의 효율에 대해서 연구하는 것은 의미가 있는 일이라고 생각한다.

References

Bourke, P. D. (1991). Detecting a shift in the fraction of nonconforming items using run-length control charts with 100% inspection, *Journal of Quality Technology*, **23**, 225–238.

Calvin, T. W. (1983). Quality control techniques for zero defects, *IEEE Transactions on Components, Hybrids and Manufacturing Technology*, **CHMT-6**, 323–328.

Chang, T. C. and Gan, F. F. (2001). Cumulative sum charts for high yield processes, *Statistica Sinica*, **11**, 791–805.

Kaminsky, F. C., Benneyan, J. C., Davis, R. D. and Burke, R. J. (1992). Statistical control charts based on a geometric distribution, *Journal of Quality Technology*, **24**, 63–69.

Lee, J., Wang, N., Xu, L., Schuh, A. and Woodall, W. H. (2013). The effect of parameter estimation on upper-sided Bernoulli cumulative sum charts, *Quality and Reliability Engineering International*, **29**, 639–651.

Quesenberry, C. P. (1993). The effect of sample size on estimated control limits for $\bar{X}$ and X control charts, *Journal of Quality Technology*, **25**, 237–247.

Woodall, W. H. (1997). Control charts based on attribute data: bibliography and review, *Journal of Quality Technology*, **29**, 172–183.

Yang, Z., Xie, M., Kuralmani, V. and Tsui, K.-L. (2002). On the performance of geometric charts with estimated control limits, *Journal of Quality Technology*, **34**, 448–458.

Zhang, M., Peng, Y., Schuh, A., Megahed, F. M. and Woodall, W. H. (2013). Geometric charts with estimated control limits, *Quality and Reliability Engineering International*, **29**, 209–223.

기하분포에 기초한 관리도에서 베이즈추정량과 최대우도추정량 사용의 성능 비교

홍휘주[a] · 이재헌[a,1]

[a]중앙대학교 응용통계학과

(2015년 6월 5일 접수, 2015년 7월 27일 수정, 2015년 9월 21일 채택)

요 약

기하분포에 기초한 관리도는 불량품이 드물게 발생하는 고품질공정에서 불량률의 변화를 효율적으로 탐지할 수 있다고 알려져 있다. 이러한 관리도를 사용할 때 기본적인 가정은 관리상태일 때의 불량률이 알려져 있거나 또는 정확하게 추정되었다는 것이다. 그러나 고품질공정에서 불량률은 아주 작기 때문에 이를 정확하게 추정하기가 쉽지 않으며 또한 아주 큰 표본크기가 필요한 경우도 종종 발생한다. 일반적으로 제1국면에서 관리상태의 불량률을 추정할 때 최대우도추정량을 사용하지만, 이 논문에서는 베이즈추정량의 사용을 제안하였다. 베이즈추정량을 사용할 경우 실무자의 사전지식을 반영할 수 있으며 표본에 불량품이 발견되지 않을 경우 발생하는 최대우도추정량의 문제점을 해결할 수 있다는 장점이 있다. 기하 관리도와 기하누적합 관리도에서 베이즈추정량을 사용한 경우와 최대우도추정량을 사용한 경우를 비교한 결과, 표본의 크기가 크지 않은 경우 베이즈추정량을 사용하는 것의 효율이 더 좋음을 알 수 있었다.

주요용어: 평균런길이, 베이즈추정량, 기하 관리도, 최대우도추정량, 제1국면 표본

이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2014R1A1A2054200).

[1]교신저자: (156-756) 서울특별시 동작구 흑석동 221, 중앙대학교 응용통계학과. E-mail: jaeheon@cau.ac.kr