

A GLR Chart for Monitoring a Zero-Inflated Poisson Process

Mi Lim Choi^a · Jaeheon Lee^{a,1}

^aDepartment of Applied Statistics, Chung-Ang University

(Received January 9, 2014; Revised February 13, 2014; Accepted March 6, 2014)

Abstract

The number of nonconformities in a unit is commonly modeled by a Poisson distribution. As an extension of a Poisson distribution, a zero-inflated Poisson(ZIP) process can be used to fit count data with an excessive number of zeroes. In this paper, we propose a generalized likelihood ratio(GLR) chart to monitor shifts in the two parameters of the ZIP process. We also compare the proposed GLR chart with the combined cumulative sum(CUSUM) chart and the single CUSUM chart. It is shown that the overall performance of the GLR chart is comparable with CUSUM charts and is significantly better in some cases where the actual directions of the shifts are different from the pre-specified directions in CUSUM charts.

Keywords: Attribute data, CUSUM chart, GLR chart, Poisson distribution, ZIP process.

1. 서론

통계적 공정관리에서 공정변수 X 가 단위 영역의 결점수를 나타내는 계수형 자료(attribute data)일 경우, 일반적으로 X 는 모수가 λ 인 Poisson 분포를 따른다고 가정한다. 이와 같은 공정을 관리하기 위해 사용하는 관리도로서 Shewhart의 c 관리도와 u 관리도가 있으며, 그 외 Lucas (1985)는 Poisson 분포에 기초한 CUSUM(cumulative sum) 관리도를 제안하였고, Borror 등 (1998) 역시 Poisson 분포에 기초한 EWMA(exponentially weighted moving average) 관리도를 제안하였다.

그러나 현대의 생산공정에서 불량률은 일반적으로 매우 낮기 때문에 만일 결점수가 거의 발생하지 않는 경우, 즉 데이터에 0이 많은 경우에도 Poisson 분포를 가정한다면 제1국면(phase I)에서 X 의 평균과 분산을 과소 추정하게 된다. 그리고 과소 추정된 평균과 분산을 토대로 관리한계를 설정할 경우 관리한계가 좁게 설정되어 오경보율(false alarm rate)이 커지게 된다는 문제점이 발생한다 (Sim과 Lim, 2008). 이와 같은 문제점을 보완하기 위해 Xie와 Goh (1993)는 ZIP(zero-inflated Poisson) 모형을 제안하였다. 이 모형은 랜덤쇼크(random shock)가 확률 p 를 가지고 독립적으로 발생하고, 랜덤쇼크가 발생한 경우에는 결점수가 모수 λ 인 Poisson 분포를 따르고 그렇지 않은 경우에는 결점수가 0임을 나타낸 모형이다. 즉, 랜덤쇼크의 발생 여부는 Bernoulli 시행이라 할 수 있고, 랜덤쇼크가 발생했을 때 결

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0009571).

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Korea. E-mail: jaeheon@cau.ac.kr

점수는 기존의 가정과 같이 Poisson 분포를 따르는 것이다. Xie와 Goh (1993), Xie 등 (2001), 그리고 He 등 (2003)은 ZIP 공정을 관리하는 Shewhart 관리도와 이와 관련된 연구를 진행하였고, He 등 (2012)은 ZIP 공정을 관리하는 CUSUM 관리도를 제안하였다. 특히 He 등 (2012)은 ZIP 공정에서 모수가 p 와 λ 두개이기 때문에 이를 각각 관리하는 두개의 CUSUM 관리도를 병행하는 절차와 두 모수의 변화를 동시에 탐지할 수 있는 CUSUM 관리도를 제안하였다. Chen 등 (2008)은 ZIP 공정을 일반화시킨, 즉 다수의 랜덤쇼크가 발생할 수 있다고 가정하는 GZIP(generalized zero-inflated Poisson) 모형을 제안하고, 이 경우에 사용할 수 있는 관리도들을 서로 비교하였다.

이 논문에서는 ZIP 공정을 관리하는 GLR(generalized likelihood ratio) 관리도 절차를 제안하고자 한다. GLR 관리도는 우도비검정을 기반으로 제안된 관리도로서, 최근 그 연구가 활발하게 진행되고 있다. GLR 관리도의 가장 큰 특징은 CUSUM 관리도와는 달리 사전에 공정모수의 변화량을 지정할 필요가 없다는데 있다 (Reynolds 등, 2013; Choi와 Lee, 2014). 공정모수의 변화량을 매시점마다 추정하여 사용하기 때문에 어떠한 공정모수의 변화에도 효율이 크게 나빠지지 않는다는 장점을 가지고 있다. 또한 이 논문에서는 제안된 GLR 관리도와 He 등 (2012)이 제안한 CUSUM 관리도들을 비교하여 그 효율을 평가하고자 한다.

2장에서는 ZIP 공정을 소개하고, 3장에서는 ZIP 공정을 관리하는 GLR 관리도 절차를 제안한다. 4장에서는 제안된 GLR 관리도의 효율을 살펴보기 위해 모의실험을 통하여 He 등 (2012)이 제안한 CUSUM 관리도들과 비교한다. 마지막 5장에서는 본 연구에 대한 결론을 제시하고 있다.

2. Zero-inflated Poisson 공정

1장에서 언급한 바와 같이 ZIP 모형은 결점수가 거의 발생하지 않는 공정에서 결점수가 Poisson 분포를 따른다고 가정하였을 때 생기는 문제점을 보완하기 위해 제안된 모형이다 (Xie와 Goh, 1993). 이 모형은 랜덤쇼크가 발생하지 않는 경우 결점수가 0이고 발생하는 경우 모수가 λ 인 Poisson 분포를 따르며, 랜덤쇼크는 확률 p 를 가지고 독립적으로 발생함을 가정하고 있다. ZIP 모형을 가정한 공정변수 X 의 확률분포는 다음과 같이 정의할 수 있다.

$$\Pr(X = x; p, \lambda) = \begin{cases} 1 - p + p e^{-\lambda}, & x = 0, \\ p \frac{\lambda^x e^{-\lambda}}{x!}, & x > 0. \end{cases}$$

즉, $x = 0$ 인 경우에는 랜덤쇼크가 발생하지 않을 확률 $(1 - p)$ 과 랜덤쇼크가 발생하지만 결점수가 0이 될 확률 $(p e^{-\lambda})$ 의 합으로 나타낼 수 있고, $x > 0$ 인 경우에는 랜덤쇼크가 발생하여 Poisson 분포를 따르는 것을 나타내고 있다. ZIP 분포의 평균과 분산은 다음과 같이 계산할 수 있다.

$$\begin{aligned} E(X) &= p\lambda, \\ \text{Var}(X) &= p\lambda + p(1 - p)\lambda^2. \end{aligned}$$

3. ZIP 공정을 관리하는 GLR 관리도

시점 n 까지 관측한, 서로 독립이고 ZIP 분포를 따르는 결점수를 X_1, X_2, \dots, X_n 이라고 하자. 이와 같은 ZIP 공정을 관리하는 목적은 관리상태에서 공정모수가 p_0 와 λ_0 인 ZIP 분포를 따르는데, 이상원인으로 인하여 두개의 공정모수가 변화하는 지를 탐지하는데 있다. 이때 관리상태일 때 공정모수 값인 p_0 와 λ_0 는 알려져 있거나, 제1국면에서 정확하게 추정할 수 있음을 가정한다. 또한 일반적으로 생산공정의 상황이 나빠지는 경우에 관심이 있기 때문에, 이 논문에서는 이상원인으로 인하여 랜덤쇼크가 발생할 확

률이 p_0 에서 $p_1 (> p_0)$ 으로 커지는 경우와 랜덤쇼크가 발생한 후 λ_0 가 $\lambda_1 (> \lambda_0)$ 으로 평균 결점수가 많아지는 경우만 고려하기로 한다.

ZIP 공정을 관리하는 GLR 관리도를 GLR-ZIP 관리도라고 할 때, 이 관리도 절차는 다음과 같이 유도할 수 있다. 우선, 시점 n 에서 공정이 관리상태라고 가정하는 귀무가설 하에서의 우도함수(likelihood function)는 다음과 같다.

$$L(p_0, \lambda_0 | X_1, X_2, \dots, X_n) = \left(1 - p_0 + p_0 e^{-\lambda_0}\right)^{n_0} p_0^{n-n_0} \frac{\lambda_0^{\sum_{t=1}^n X_t} e^{-\lambda_0(n-n_0)}}{\prod_{t=1}^n X_t!},$$

여기서 n_0 는 n 개의 관측치 중에서 $X_t = 0$ ($t = 1, 2, \dots, n$)의 개수이다.

시점 τ 와 $\tau + 1$ 에서 이상원인이 발생했다고 가정하자. 즉, 시점 τ 까지는 관리상태이고 시점 $\tau + 1$ 부터 이상상태가 되는 것이다. 이때 τ 를 공정의 변화시점(process change point)이라고 한다. 이상원인으로부터 인하여 공정모수가 p_1 과 λ_1 으로 변화했음을 나타내는 대립가설 하에서의 우도함수는 다음과 같다.

$$\begin{aligned} L(p_1, \lambda_1 | X_1, X_2, \dots, X_n) &= \left(1 - p_0 + p_0 e^{-\lambda_0}\right)^{n_0^u(\tau)} p_0^{\tau-n_0^u(\tau)} \frac{\lambda_0^{\sum_{t=1}^{\tau} X_t} e^{-\lambda_0(\tau-n_0^u(\tau))}}{\prod_{t=1}^{\tau} X_t!} \\ &\quad \times \left(1 - p_1 + p_1 e^{-\lambda_1}\right)^{n_0^a(\tau)} p_1^{n-\tau-n_0^a(\tau)} \frac{\lambda_1^{\sum_{t=\tau+1}^n X_t} e^{-\lambda_1(n-\tau-n_0^a(\tau))}}{\prod_{t=\tau+1}^n X_t!}, \end{aligned}$$

여기서 $n_0^u(\tau)$ 는 시점 τ 까지 $X_t = 0$ ($t = 1, 2, \dots, \tau$)의 개수이고, $n_0^a(\tau)$ 는 시점 τ 이후 시점 n 까지 $X_t = 0$ ($t = \tau + 1, \tau + 2, \dots, n$)의 개수이다. 따라서 $n_0 = n_0^u(\tau) + n_0^a(\tau)$ 를 만족하게 된다. 단순한 계산을 통하여 대립가설 하에서 p_1 과 λ_1 의 최대우도추정량(maximum likelihood estimator; MLE)은 다음과 같이 유도할 수 있다.

$$\begin{aligned} \hat{p}_1 &= \frac{n - \tau - n_0^a(\tau)}{(n - \tau)(1 - e^{-\lambda_1})}, \\ \hat{\lambda}_1 &= \frac{\sum_{t=\tau+1}^n X_t}{(n - \tau) \hat{p}_1}, \end{aligned}$$

여기서 \hat{p}_1 와 $\hat{\lambda}_1$ 은 닫혀진 형태(closed form)로 표현되지 않기 때문에 비선형방정식의 해를 구하는 절차를 통해 계산할 수 있다.

따라서 시점 n 에서 GLR-ZIP 관리도의 관리통계량 R_n 은 다음과 같이 정의된다.

$$\begin{aligned} R_n &= \ln \frac{\max_{0 \leq \tau \leq n-m} L(p_1, \lambda_1 | X_1, X_2, \dots, X_n)}{L(p_0, \lambda_0 | X_1, X_2, \dots, X_n)} \\ &= \max_{0 \leq \tau \leq n-m} \left\{ n_0^a(\tau) \ln \frac{1 - \hat{p}_1 + \hat{p}_1 e^{-\hat{\lambda}_1}}{1 - p_0 + p_0 e^{-\lambda_0}} \right. \\ &\quad \left. + (n - \tau - n_0^a(\tau)) \left(\ln \frac{\hat{p}_1}{p_0} - \hat{\lambda}_1 + \lambda_0 \right) + \ln \frac{\hat{\lambda}_1}{\lambda_0} \sum_{t=\tau+1}^n X_t \right\}. \end{aligned} \tag{3.1}$$

τ 는 관리상태의 제일 마지막 시점을 나타내기 때문에 $\tau = 0$ 인 경우는 공정이 시작부터 이상상태인 것을 의미한다. GLR-ZIP 관리도 절차는 시점 n 에서 통계량 R_n 이 관리한계 h_{GLR} 보다 큰 경우 이상신호를 주는 것이다. 여기서 h_{GLR} 는 관리상태일 때 주어진 특성을 만족하도록 설정할 수 있다. 일반적으로 관리상태에서의 ARL(average run length), ATS(average time to signal), 또는 ANOS(average

number of observations to signal)가 주어진 값을 만족하도록 설정하는데, 이 논문에서는 ANOS라는 측도를 사용하였다. 이 측도에 대해서는 다음 장에서 설명하고자 한다. 또한 GLR 관리도에서 이와 같은 측도들은 Markov chain 등을 이용하여 이론적으로 계산할 수 없기 때문에 모의실험을 통한 시행착오방법(trial and error method)을 사용하여 관리한계 h_{GLR} 를 찾아낼 수 있다. 식 (3.1)에서 m 은 모수를 추정하기 위해 사용하는 관측값의 최소 개수를 나타낸다. Lai (1995)는 m 은 일반적으로 추정해야 할 모수의 개수와 같게 한다고 언급했기 때문에, 이 논문에서 $m = 2$ 를 사용하였다.

4. GLR-ZIP 관리도의 효율

4.1. 관리도 효율의 측도

이 논문에서는 관리도의 효율에 대한 측도로 ANOS를 사용하는데, ANOS는 공정의 시작부터 신호까지 관측한 관측치 수의 기댓값으로 정의된다. 일반적으로 관리상태에서 오경보률은 작고 이상상태에서 이상신호를 빨리 주는 관리도가 효율이 좋다고 평가하고 있다. 따라서 관리상태에서의 ANOS($ANOS_0$)를 주어진 값으로 고정한 후, 이상상태의 ANOS($ANOS_1$)가 얼마나 작아지는 지를 통하여 관리도의 효율을 판단하고자 한다.

이상상태의 ANOS는 공정이 이상상태에서 시작했다고 가정하고 계산하는 zero-state ANOS와 공정이 관리상태에서 시작하여 안정적인 상태가 된 후 이상상태로 변화하는 것을 가정하고 계산하는 steady-state ANOS로 구분할 수 있다. steady-state ANOS에서의 가정이 좀 더 현실적일 수 있지만, 두 가지의 ANOS 모두 유사한 경향을 가지며 He 등 (2012)의 연구 결과와 비교하기 위해 이 논문에서는 zero-state ANOS를 측도로 사용하기로 한다.

4.2. ZIP 모형을 관리하는 CUSUM 관리도

이 논문에서는 GLR-ZIP 관리도의 효율을 알아보기 위해 He 등 (2012)이 제안한 $p - \lambda$ CUSUM 관리도 및 t -CUSUM 관리도와 비교하고 있다. 따라서 He 등 (2012)이 제안한 CUSUM 관리도의 절차를 소개하고자 한다.

4.2.1. $p - \lambda$ CUSUM 관리도 $p - \lambda$ CUSUM 관리도는 모수 p 의 변화를 탐지하는 p -CUSUM 관리도와 모수 λ 의 변화를 탐지하는 λ -CUSUM 관리도를 병행하는 절차를 일컫는다. 먼저 모수 p 가 p_0 에서 p_1 으로 변하는 지를 탐지하는 p -CUSUM 관리도의 절차는 다음과 같다. 시점 n 에서 p -CUSUM 관리도의 관리통계량 B_n 은

$$B_n = \max(0, B_{n-1} + K_n), \quad n = 1, 2, \dots \quad (4.1)$$

으로 정의된다. 여기서 $B_0 = 0$ 이고, K_n 은 다음과 같다.

$$K_n = \begin{cases} \ln \frac{1 - p_1 + p_1 e^{-\lambda_0}}{1 - p_0 + p_0 e^{-\lambda_0}}, & X_n = 0, \\ \ln \frac{p_1}{p_0}, & X_n > 0, \end{cases}$$

p -CUSUM 관리도 절차는 시점 n 에서 통계량 B_n 이 관리한계 h_p 보다 큰 경우 이상신호를 주는 것이다. 여기서 h_p 는 $ANOS_0$ 가 주어진 값을 만족하도록 설정하게 된다.

다음으로 λ 가 λ_0 에서 λ_1 으로 변하는 지를 탐지하는 λ -CUSUM 관리도의 절차를 살펴보자. 시점 n 에서 λ -CUSUM 관리도의 관리통계량 L_n 은 다음과 같이 정의된다.

$$L_n = \max(0, L_{n-1} + M_n), \quad n = 1, 2, \dots, \quad (4.2)$$

여기서 $L_0 = 0$ 이고, M_n 는 다음과 같다.

$$M_n = \begin{cases} \log \frac{1 - p_0 + p_0 e^{-\lambda_1}}{1 - p_0 + p_0 e^{-\lambda_0}}, & X_n = 0, \\ X_n \ln \frac{\lambda_1}{\lambda_0} + (\lambda_0 - \lambda_1), & X_n > 0. \end{cases}$$

λ -CUSUM 관리도 절차는 시점 n 에서 통계량 L_n 이 관리한계 h_λ 보다 큰 경우 이상신호를 주는 것이다. 여기서 h_λ 는 ANOS₀가 주어진 값을 만족하도록 설정하게 된다.

위에서 언급한 바와 같이 $p - \lambda$ CUSUM 관리도는 p -CUSUM 관리도와 λ -CUSUM 관리도를 병행하는 절차로서, 식 (4.1)에 정의된 통계량 B_n 이 h_p 보다 크거나 식 (4.2)에 정의된 통계량 L_n 이 h_λ 보다 큰 경우 이상신호를 주게 된다. 이때 관리한계 h_p 와 h_λ 는 각 관리도를 개별적으로 사용하는 경우 ANOS₀가 유사한 값을 가지면서, 병행하여 사용하는 경우 ANOS₀가 주어진 값을 만족하도록 설정한다. $p - \lambda$ CUSUM 관리도는 두 가지 모수의 변화를 동시에 탐지할 수 있다는 장점이 있지만, 이상상태에서의 모수값인 p_1 과 λ_1 , 그리고 관리한계인 h_p 와 h_λ 등 모두 4가지 값을 사전에 결정해야 한다는 단점을 가지고 있다.

4.2.2. t -CUSUM 관리도 t -CUSUM 관리도는 $p - \lambda$ CUSUM 관리도와는 달리, 하나의 관리도로 p 와 λ 의 변화를 동시에 탐지하는 CUSUM 관리도이다. 시점 n 에서 t -CUSUM 관리도의 관리통계량 T_n 은 다음과 같이 정의된다.

$$T_n = \max(0, T_{n-1} + N_n), \quad n = 1, 2, \dots,$$

여기서 $T_0 = 0$ 이고, N_n 은 다음과 같다.

$$N_n = \begin{cases} \log \frac{1 - p_1 + p_1 e^{-\lambda_1}}{1 - p_0 + p_0 e^{-\lambda_0}}, & X_n = 0, \\ X_n \ln \frac{\lambda_1}{\lambda_0} + (\lambda_0 - \lambda_1) + \ln \frac{p_1}{p_0}, & X_n > 0. \end{cases}$$

t -CUSUM 관리도 절차는 시점 n 에서 통계량 T_n 이 h_t 보다 큰 경우 이상신호를 주는 것이다. 여기서 h_t 는 ANOS₀가 주어진 값을 만족하도록 설정하게 된다. 따라서 t -CUSUM 관리도에서는 이상상태에서의 모수값인 p_1 과 λ_1 , 그리고 관리한계인 h_t 등 3가지 값을 사전에 결정해야 한다.

4.3. 모의실험

이 절에서는 이 논문에서 제안한 GLR-ZIP 관리도와 CUSUM 관리도들의 효율을 ANOS를 이용하여 비교하였다. GLR-ZIP 관리도의 ANOS는 모의실험을 통하여 계산하였고, $p - \lambda$ CUSUM 관리도와 t -CUSUM 관리도의 ANOS는 He 등 (2012)에 나와있는 값을 그대로 사용하였다. 모든 모의실험은 100,000번 반복하였다. He 등 (2012)은 $p - \lambda$ CUSUM 관리도와 t -CUSUM 관리도에서 미리 설정해야 하는 p_1 과 λ_1 으로 $p_1 = 1.25p_0$, $p_1 = 1.5p_0$, $\lambda_1 = \lambda_0 + 1$, 그리고 $\lambda_1 = \lambda_0 + 2$ 를 고려하였다. 또한

Table 4.1. ANOS values of GLR-ZIP and CUSUM charts with only one parameter shift when $p_0 = 0.2$ and $p_1 = 1.25p_0$.

p	λ_0	λ	GLR-ZIP	$p - \lambda$ CUSUM		t -CUSUM	
				$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$	$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$
0.2	2	2	200.80	201.21	199.78	200.73	200.79
0.25	2	2	125.66	94.11	96.27	130.92	150.55
0.3	2	2	71.71	56.38	56.68	92.96	118.92
0.4	2	2	31.45	30.31	30.78	55.48	82.22
0.5	2	2	18.06	20.60	20.70	38.21	62.08
0.2	2	3	61.94	45.76	48.07	38.45	38.91
0.2	2	4	26.68	21.75	20.91	18.63	17.37
0.2	2	5	15.93	14.05	12.80	12.36	11.02
0.2	2	6	11.27	10.93	9.70	9.43	8.29
h			5.2923	1.4096	1.4315	1.9182	2.3480
				1.9989	2.3131		
0.2	6	6	199.87	199.47	203.49	200.23	200.51
0.25	6	6	118.65	91.62	94.04	117.66	138.89
0.3	6	6	66.74	54.16	53.97	78.64	104.06
0.4	6	6	28.41	28.16	28.59	44.49	67.51
0.5	6	6	15.90	19.04	19.10	30.17	49.21
0.2	6	7	115.86	75.09	78.99	64.91	62.50
0.2	6	8	54.97	39.13	38.16	34.53	31.07
0.2	6	9	31.98	25.77	23.98	23.15	19.97
0.2	6	10	21.67	19.73	17.78	17.49	14.75
h			5.7870	1.5122	1.5279	1.6600	2.0703
				1.6985	2.3436		

관리한계를 설정하기 위해 사전에 결정하는 $ANOS_0$ 값은 $p_0 = 0.2$ 인 경우 200, $p_0 = 0.1$ 인 경우 340을 사용하였기 때문에, 이 논문에서도 이 값들을 그대로 사용하였다.

Table 4.1과 Table 4.2는 $p_0 = 0.2$ 이고 $\lambda_0 = 2$ 와 6이며, p 또는 λ 가 증가하는 단일 변화(single shift)에 대한 ANOS 값을 나타낸다. 각 표의 상단과 하단의 제일 첫 행은 관리상태일 때 ANOS 값을 나타내고, 상단과 하단의 제일 마지막 행의 h 값은 각 관리도의 관리한계를 나타낸다. $p - \lambda$ CUSUM 관리도는 p -CUSUM 관리도와 λ -CUSUM 관리도를 병행하는 절차이므로 두개의 관리한계 값이 필요하고, 나머지 관리도들은 하나의 관리한계 값만 필요하다.

Table 4.1은 $p_0 = 0.2$ 이며 표의 상단은 $\lambda_0 = 2$ 이고 하단은 $\lambda_0 = 6$ 인 경우이다. 상단과 하단은 각각 p 가 변화하는 경우와 λ 가 변화하는 경우로 나누어져 있다. 그리고 CUSUM 관리도의 p_1 과 λ_1 는 각각 $p_1 = 1.25p_0$ 이고 $\lambda_1 = \lambda_0 + 1$ 과 $\lambda_1 = \lambda_0 + 2$ 로 설정한 경우로 구분되어 있다. Table 4.1의 결과로부터 다음의 사실을 알 수 있다. λ 는 변하지 않고 p 가 증가하는 경우, GLR-ZIP 관리도는 t -CUSUM 관리도보다 효율이 좋지만, $p - \lambda$ CUSUM 관리도보다 전반적으로 효율이 좋지 않음을 알 수 있다. 하지만 λ 는 변하지 않고 p 가 크게 증가한 경우($p = 0.5$)에는 GLR-ZIP 관리도의 효율이 가장 좋은 것으로 나타났다. 그러나 p 는 변하지 않고 λ 가 증가하는 경우, GLR-ZIP 관리도는 $p - \lambda$ CUSUM 관리도와 t -CUSUM 관리도에 비하여 효율이 나쁨을 알 수 있다. 따라서 GLR-ZIP 관리도는 랜덤쇼크의 발생 확률인 p 의 증가를 탐지하는 경우에는 효율적이지만, λ 의 증가를 탐지하는 경우에는 효율적이지 못하다고 할 수 있다. Table 4.2는 CUSUM 관리도에서 $p_1 = 1.5p_0$ 로 설정한 것만 Table 4.1의 설정과 차

Table 4.2. ANOS values of GLR-ZIP and CUSUM charts with only one parameter shift when $p_0 = 0.2$ and $p_1 = 1.5p_0$.

p	λ_0	λ	GLR-ZIP	$p - \lambda$ CUSUM		t -CUSUM	
				$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$	$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$
0.2	2	2	200.80	202.03	196.44	200.02	196.39
0.25	2	2	125.66	96.30	95.14	120.93	143.77
0.3	2	2	71.71	55.59	55.61	80.16	110.66
0.4	2	2	31.45	28.32	28.04	43.00	72.63
0.5	2	2	18.06	18.61	18.58	27.64	52.32
0.2	2	3	61.94	46.09	46.68	40.57	38.81
0.2	2	4	26.68	21.64	20.59	19.70	17.46
0.2	2	5	15.93	14.36	12.88	12.91	11.08
0.2	2	6	11.27	11.02	9.59	9.71	8.32
h			5.2923	2.1966	2.1865	2.2420	2.5656
				2.0242	2.2990		
0.2	6	6	199.87	201.22	199.46	200.72	200.61
0.25	6	6	118.65	93.87	92.87	103.55	128.61
0.3	6	6	66.74	53.61	51.59	62.82	90.10
0.4	6	6	28.41	26.26	25.72	31.65	52.71
0.5	6	6	15.90	17.31	17.12	20.32	35.69
0.2	6	7	115.86	76.73	76.84	74.01	65.15
0.2	6	8	54.97	39.93	37.13	40.79	32.87
0.2	6	9	31.98	26.58	24.16	27.53	21.10
0.2	6	10	21.67	20.10	17.67	20.76	15.47
h			5.7870	2.3270	2.3218	2.0740	2.3140
				1.7053	2.3434		

이가 있다. Table 4.2의 결과는 전반적으로 Table 4.1의 결과와 유사한 경향을 나타낸다. 따라서 Table 4.1과 Table 4.2를 통해 GLR-ZIP 관리도는 CUSUM 관리도와 같이 모수의 변화량을 미리 설정할 필요가 없다는 장점이 있으며, p 의 큰 증가의 탐지에 대해서는 효율적이지만, p 의 작은 증가와 λ 의 증가의 탐지에 대해서는 CUSUM 관리도에 비해 효율적이지 못함을 알 수 있다.

Table 4.3과 Table 4.4는 $p_0 = 0.2$ 와 $\lambda_0 = 6$ 이고, p 와 λ 가 동시에 증가하는 경우에 대한 ANOS 값을 나타낸 표이다. 제일 마지막 행의 h 값은 역시 각 관리도의 관리한계를 나타낸다. Table 4.3은 CUSUM 관리도에서 $p_1 = 1.25p_0$ 로 설정하였고, Table 4.4는 CUSUM 관리도에서 $p_1 = 1.5p_0$ 로 설정했다는 차이가 있다.

Table 4.3으로부터 다음의 사실을 알 수 있다. p 와 λ 가 모두 증가하는 경우 전반적으로 t -CUSUM 관리도의 효율이 좋은 것으로 나타났다. 그러나 p 가 크게 증가한 경우, 즉 $p = 0.4$ 와 $p = 0.5$ 인 경우 GLR-ZIP 관리도는 t -CUSUM 관리도보다 효율이 떨어지지만 $p - \lambda$ CUSUM 관리도보다 전반적으로 효율이 좋음을 알 수 있다. Table 4.4에서도 Table 4.3과 유사한 결론을 얻을 수 있다.

Table 4.1에서 Table 4.4까지 나타난 결과를 종합해 보면, 두 가지 모수 중 p 만 변화하는 경우에는 $p - \lambda$ CUSUM 관리도와 GLR-ZIP 관리도가, 그리고 λ 만 변화하는 경우와 두 가지 모수 모두가 변화하는 경우 t -CUSUM 관리도가 전반적으로 효율이 좋음을 알 수 있다. 그러나 GLR-ZIP 관리도는 어떠한 경우에도 효율이 크게 떨어지지 않으며 특히 p 가 크게 증가한 경우에는 효율이 아주 좋음을 알 수 있다. 또한 GLR-ZIP 관리도는 관리한계 하나의 값만 설정하고 사용할 수 있다는 장점을 가지고 있다.

Table 4.3. ANOS values of GLR-ZIP and CUSUM charts when $p_0 = 0.2$, $\lambda_0 = 6$, and $p_1 = 1.25p_0$.

p	λ	GLR-ZIP	$p - \lambda$ CUSUM		t -CUSUM	
			$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$	$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$
0.25	7	77.78	53.49	54.48	42.83	45.01
	8	40.74	30.50	29.99	24.47	23.20
	9	24.69	20.67	19.26	17.08	15.27
	10	16.96	15.33	14.23	13.21	11.45
0.3	7	49.57	38.82	38.24	31.38	34.84
	8	29.74	24.97	24.04	18.84	18.44
	9	19.07	17.05	15.85	13.50	12.34
	10	13.43	13.00	11.73	10.61	9.35
0.4	7	23.99	23.83	23.58	20.15	23.74
	8	17.07	17.55	16.97	12.85	13.03
	9	12.13	12.64	11.85	9.50	8.91
	10	9.07	9.66	8.84	7.61	6.84
0.5	7	14.22	16.93	16.77	14.75	17.91
	8	11.09	13.38	12.86	9.73	10.06
	9	8.43	10.05	9.39	7.32	6.97
	10	6.52	7.74	7.03	5.93	5.39
h	5.7870	1.5122	1.5279	1.6600	2.0703	
			1.6985	2.3436		

Table 4.4. ANOS values of GLR-ZIP and CUSUM charts when $p_0 = 0.2$, $\lambda_0 = 6$, and $p_1 = 1.5p_0$.

p	λ	GLR-ZIP	$p - \lambda$ CUSUM		t -CUSUM	
			$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$	$\lambda_1 = \lambda_0 + 1$	$\lambda_1 = \lambda_0 + 2$
0.25	7	77.78	53.69	52.25	44.33	44.28
	8	40.74	31.14	29.24	26.82	23.55
	9	24.69	21.12	18.90	19.15	15.68
	10	16.96	15.89	14.15	14.98	11.77
0.3	7	49.57	38.27	37.37	30.33	32.74
	8	29.74	25.39	23.33	19.60	18.15
	9	19.07	17.69	15.84	14.55	12.41
	10	13.43	13.27	12.00	11.66	9.48
0.4	7	23.99	22.70	21.96	17.94	20.94
	8	17.07	17.57	16.09	12.57	12.33
	9	12.13	12.71	11.60	9.76	8.73
	10	9.07	9.98	8.83	8.06	6.82
0.5	7	14.22	15.90	15.26	12.60	15.17
	8	11.09	13.10	12.09	9.21	9.29
	9	8.43	10.17	9.24	7.34	6.73
	10	6.52	7.91	7.02	6.16	5.32
h	5.7870	2.3270	2.3218	2.0740	2.3140	
			1.7053	2.3434		

Table 4.5와 Table 4.6은 $p_0 = 0.1$ 과 $\lambda_0 = 6$ 이고 CUSUM 관리도에서 $p_1 = 1.25p_0$ 와 $\lambda_1 = \lambda_0 + 2$ 로 설정했을 때, p 와 λ 가 CUSUM 관리도에서 설정한 값과 다른 방향으로 변화했을 경우의 ANOS 값을 나타낸 것이다. 앞에서 언급하 바와 같이, 이 경우 관리상태의 ANOS는 He 등 (2012)에서와 같이

Table 4.5. ANOS values of GLR-ZIP and CUSUM charts when p increases and λ decreases.

$p_0 = 0.1, \lambda_0 = 6, p_1 = 1.25p_0, \lambda_1 = \lambda_0 + 2$						
	λ	p				
		0.15	0.2	0.3	0.4	0.5
GLR-ZIP	5	95.34	49.05	21.08	12.19	8.19
$p - \lambda$ CUSUM	5	114.05	58.32	29.07	19.50	14.80
t -CUSUM	5	1055.53	701.10	405.62	281.31	214.49
GLR-ZIP	4	54.22	33.03	16.38	10.15	7.05
$p - \lambda$ CUSUM	4	122.16	60.55	29.91	19.90	15.06
t -CUSUM	4	12770.94	8807.80	5299.19	3752.12	2900.67
GLR-ZIP	3	33.33	22.37	12.46	8.19	5.92
$p - \lambda$ CUSUM	3	179.42	79.22	36.64	23.73	17.71
t -CUSUM	3	333106.14	242535.36	156049.04	114914.97	91067.25

Table 4.6. ANOS values of GLR-ZIP and CUSUM charts when p decreases and λ increases.

$p_0 = 0.1, \lambda_0 = 6, p_1 = 1.25p_0, \lambda_1 = \lambda_0 + 2$						
	p	λ				
		7	8	9	10	11
GLR-ZIP	0.09	232.67	117.89	70.15	47.93	36.03
$p - \lambda$ CUSUM	0.09	164.61	77.91	51.06	36.52	29.53
t -CUSUM	0.09	131.02	65.77	42.12	30.87	24.49
GLR-ZIP	0.075	228.33	127.07	79.04	55.30	41.63
$p - \lambda$ CUSUM	0.075	206.54	95.51	59.32	43.54	35.57
t -CUSUM	0.075	170.32	83.46	52.45	37.95	29.86
GLR-ZIP	0.05	158.37	117.72	86.07	65.47	52.20
$p - \lambda$ CUSUM	0.05	320.44	144.00	91.98	65.47	52.72
t -CUSUM	0.05	310.53	144.59	86.88	60.83	46.78

ANOS₀ = 340으로 설정하였고, 이를 만족하는 관리한계로 GLR-ZIP 관리도는 5.9554, $p - \lambda$ CUSUM 관리도는 1.3536과 2.1010, 그리고 t -CUSUM 관리도는 1.9522를 사용하였다.

Table 4.5는 CUSUM 관리도에서 p 는 설정한 방향으로 증가했지만 λ 는 설정한 방향과는 반대로 감소한 경우에 대한 것이다. 이 경우 GLR-ZIP 관리도의 효율이 $p - \lambda$ CUSUM 관리도와 t -CUSUM 관리도에 비해 월등히 좋음을 알 수 있다. 특히 t -CUSUM 관리도는 λ 가 설정한 방향과 반대로 감소한 경우 효율이 크게 떨어지고 미리 설정한 ANOS₀ 값 보다 더 커지는 경우도 다수 발생함을 알 수 있다. Table 4.6은 CUSUM 관리도에서 λ 는 설정한 방향으로 증가했지만 p 는 설정한 방향과 반대로 감소한 경우에 대한 것이다. 이 경우에는 t -CUSUM 관리도의 효율이 전반적으로 좋았으며, GLR-ZIP 관리도는 p 가 크게 감소한 경우($p = 0.05$) 효율이 좋음을 알 수 있다.

Table 4.5와 Table 4.6의 결과에서 알 수 있듯이 CUSUM 관리도는 실제 공정모수의 변화가 미리 설정한 변화량 또는 방향과 다르게 발생한 경우 효율이 크게 떨어질 수 있음을 알 수 있다. 그러나 GLR-ZIP 관리도는 공정모수의 변화량을 얻어진 관측치에 의거하여 매시점마다 추정하기 때문에 모수의 변화량 또는 변화 방향을 예측하기 어려운 경우 유용하게 사용할 수 있을 것이다.

5. 결론

결점수를 나타내는 공정변수 X 가 0값을 많이 갖는 경우 Poisson 분포를 확장한 ZIP 분포에 적합시킬

수 있다. 불량이나 결점이 극히 적게 발생하는 생산공정에서는 이 모형을 유용하게 사용할 수 있을 것이다. He 등 (2012)은 ZIP 공정을 관리하기 위해 p -CUSUM 관리도와 λ -CUSUM 관리도를 병행하는 $p - \lambda$ CUSUM 관리도와 t -CUSUM 관리도 절차를 제안하였다.

이 논문에서는 ZIP 공정을 관리하는 GLR-ZIP 관리도 절차를 제안하고, 그 효율을 He 등 (2012)이 제안한 $p - \lambda$ CUSUM 관리도 및 t -CUSUM 관리도와 비교하였다. 모의실험 결과를 종합해 보면, p 만 증가하는 경우 GLR-ZIP 관리도와 $p - \lambda$ CUSUM 관리도의 효율이 좋았으며, λ 만 증가하는 경우에는 t -CUSUM 관리도의 효율이 좋음을 알 수 있다. 만일 p 와 λ 가 함께 증가하는 경우에는 t -CUSUM 관리도의 효율이 전반적으로 좋게 나타났다. 그러나 CUSUM 관리도는 p 와 λ 가 미리 설정한 방향과 다르게 변화한 경우 효율이 크게 떨어지는 단점이 있다. 특히 p 는 설정한 방향으로 증가하지만 λ 는 설정한 방향과 반대로 감소한 경우 t -CUSUM 관리도의 효율은 크게 나빠짐을 알 수 있다. 따라서 모수의 변화량 또는 변화 방향을 미리 예측하기 힘든 경우 GLR-ZIP 관리도를 사용하는 것이 가장 좋은 대안이 될 수 있고, 예측한 방향으로 변화한 경우에도 효율이 크게 떨어지지 않기 때문에 실제 공정에서 유용하게 사용할 수 있을 것이라 기대한다. GLR-ZIP 관리도를 사용함으로써 얻을 수 있는 $\hat{\tau}$, \hat{p}_1 , 그리고 $\hat{\lambda}_1$ 를 이용하여 공정에 대한 진단에 대해서 향후 연구할 계획이다.

References

- Borror, C. M., Champ, C. W. and Rigdon, S. E. (1998). Poisson EWMA control charts, *Journal of Quality Technology*, **30**, 352–361.
- Chen, N., Zhou, S., Chang, T.-S. and Huang, H. (2008). Attribute control charts using generalized zero-inflated Poisson distribution, *Quality and Reliability Engineering International*, **24**, 793–806.
- Choi, M. L. and Lee, J. (2014). GLR charts for simultaneously monitoring a sustained shift and a linear drift in the process mean, *Communications for Statistical Applications and Methods*, to appear.
- He, B., Xie, M., Goh, T. N. and Ranjan, P. (2003). On the estimation error in zero-inflated Poisson model for process control, *International journal of Reliability, Quality, and Safety Engineering*, **10**, 159–169.
- He, S., Huang, W. and Woodall, W. H. (2012). CUSUM charts for monitoring a zero-inflated Poisson process, *Quality and Reliability Engineering International*, **28**, 181–192.
- Lai, T. L. (1995). Sequential changepoint detection in quality-control and dynamical-systems, *Journal of the Royal Statistical Society: Series B-Statistical Methodology*, **57**, 613–658.
- Lucas, J. M. (1985). Counted data CUSUM's, *Technometrics*, **27**, 129–144.
- Reynolds, M. R. Jr., Lou, J., Lee, J. and Wang, S. (2013). The design of GLR control charts for monitoring the process mean and variance, *Journal of Quality Technology*, **45**, 34–60.
- Sim, C. H. and Lim, M. H. (2008). Attribute charts for zero-inflated processes, *Communications in Statistics-Simulation and Computation*, **37**, 1440–1452.
- Xie, M. and Goh, T. N. (1993). SPC of a near zero-defect process subject to random shock, *Quality and Reliability Engineering International*, **9**, 89–93.
- Xie, M., He, B. and Goh, T. N. (2001). Zero-inflated Poisson model in statistical process control, *Computational Statistics and Data Analysis*, **38**, 191–201.

ZIP 공정을 관리하는 GLR 관리도

최미림^a · 이재현^{a,1}

^a중앙대학교 응용통계학과

(2014년 1월 9일 접수, 2014년 2월 13일 수정, 2014년 3월 6일 채택)

요약

단위 영역의 결점수는 일반적으로 Poisson 분포를 가정한다. 이 Poisson 분포의 확장된 형태로 ZIP(zero-inflated Poisson) 분포를 고려할 수 있는데, 이 모형은 데이터에 0이 많이 관측되는 경우 잘 적합된다고 알려져 있다. 이 논문에서는 ZIP 분포를 따르는 공정을 관리하는 GLR(generalized likelihood ratio) 관리도 절차를 제안하고 있다. 또한 제안된 GLR 관리도의 효율을 기존에 제안된 CUSUM 관리도들과 비교하였다. 그 결과 제안된 GLR 관리도는 모수의 다양한 변화에 대해 효율이 좋거나 또는 효율이 크게 떨어지지 않았고, 특히 CUSUM 관리도에서 모수가 미리 설정한 방향과 다르게 변화했을 때 효율이 크게 나빠지는 문제를 해결할 수 있는 대안이라는 결론을 얻을 수 있었다.

주요용어: 계수형 자료, CUSUM 관리도, GLR 관리도, Poisson 분포, ZIP 모형.

이 논문은 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2010-0009571).

¹교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 응용통계학과. Email: jaeheon@cau.ac.kr