응용통계연구 제 **21**권 3호, *2008*년, *pp. 509 521*

하나의 관리도로 공정 평균과 분산의 변화를 탐지하는 절차*

정상현[1)] 이재헌[2)]

요 약

평균과 분산이 동시에 변화할 수 있는 공정을 관리할 경우, 평균의 변화를 탐지하는 관리도와 분산의 변화를 탐지하는 관리도를 병행하여 사용하는 것이 일반적이다. 여러 연구자들이 하나의 관리도를 사용하여 공정 평균과 분산의 변화를 동시에 탐지할 수 있는 절차를 제안했는데, 이 논문에서는 이와 같은 관리도들을 소개하고 그 효율을 비교하였다. 그 결과 GLR 관리도, Omnibus EWMA 관리도 그리고 Interval 관리도는 충분히 좋은 효율을 가짐을 알 수 있었다.

주요용어: 공정관리, GLR 관리도, EWMA 관리도, Interval 관리도, 평균런길이.

1. 서론

통계적 공정관리(statistical process control: SPC)에서 관리도(control chart)는 생산 공정에서 변동의 원인이 되는 공정 모수의 변화를 탐지하는 도구로서 널리 사용되어 왔다. 대표적인 관리도로는 Shewhart 관리도, CUSUM(cumulative sum) 관리도 그리고 EWMA(exponentially weighted moving average) 관리도 등이 있다. 관리도는 관리통계량이 미리 설정된 관리한계(control limit)를 벗어날 경우 이상원인(special cause)이 발생했다는 신호를 주게 된다.

실제 공정에서 품질특성치(quality characteristic)는 연속형 변수로 측정되며, 정규분포를 가정하는 것이 일반적이다. 따라서 공정의 평균과 분산을 동시에 관리하며, 이 경우 평균의 변화를 탐지하는 관리도와 분산의 변화를 탐지하는 관리도를 병행하여 사용하는 것이 일반적이다. 만일 하나의 관리도를 사용하여 공정 평균과 분산의 변화를 동시에 탐지할 수 있고 두 개의 관리도를 병행하는 절차에 비하여 효율이 크게 떨어지지 않는다면, 관리자 입장에서는 하나의 관리도를 사용하는 것을 더 선호할 것이다.

하나의 관리도를 사용하여 공정 평균과 분산의 변화를 동시에 탐지하는 절차에 대해서는 꾸준하게 연구되어져 왔다. 이 논문의 목적은 개별적으로 연구되어 왔고 상대적으로 잘 알려져 있지 않았던 하나의 관리도를 사용하여 평균과 분산의 변화를 탐지하는 절차를 소개하고 그 효율을 비교함으로써, 각 관리도의 강점과 약점을 살펴보는 것에 있다. 소개할 관리도는 GLR 관리도, Omnibus EWMA 관리도, MaxMin EWMA 관리도, Max EWMA 관리도 그리

* 이 논문은 2007년도 중앙대학교 학술연구비(일반연구비) 지원에 의한 것임.
1) (156-756) 서울특별시 동작구 흑석동 221, 중앙대학교 대학원 통계학과, 석사과정.
E-mail: kami1001@nate.com
2) (156-756) 교신저자. 서울특별시 동작구 흑석동 221, 중앙대학교 수학통계학부, 교수.
E-mail: jaeheon@cau.ac.kr

고 Interval 관리도 등이며, 이들의 효율을 EWMA-$\bar{X}$ 관리도와 EWMA- $\ln S^2$ 관리도를 병행하는 절차와 비교함으로써 하나의 관리도를 사용하는 절차에 대한 효율을 알아보았다.

2. 공정 평균과 분산의 변화를 동시에 탐지하는 관리도

X는 공정에서 관측되는 품질특성치를 나타내는 공정 변수이고, 평균이 μ이고 표준편차가 σ인 정규분포(normal distribution)를 따른다고 가정하자. 또한 평균이 μ_0이고 표준편차가 σ_0인 경우 공정이 관리상태라고 정의한다. 이 논문에서 고려하는 공정 관리의 목적은 평균을 μ_0에서 μ_a로 변화시키며 동시에 표준편차를 σ_0에서 σ_a로 변화시키는 이상원인을 탐지하는 것이다.

μ_0와 σ_0는 알려져 있는 경우도 있지만, 일반적으로 예비표본 등을 이용하여 추정하게 된다(이를 Phase I 단계라고 부른다). 이때 추정된 값은 사용한 예비표본의 표본수가 어느 정도 큰 값이고 예비표본으로부터 공정이 관리상태라고 판정될 때 의미있는 값이 된다. 이 논문에서 이후에는 μ_0와 σ_0는 알려져 있는 값이라고 가정해도 충분할 정도로 Phase I 단계가 잘 수행되어 참값과 차이가 없음을 가정한다.

$\boldsymbol{X_t} = (X_{t1}, X_{t2}, \ldots, X_{tn})$은 시점 t에서 추출한 크기가 n인 표본이고, $\bar{X}_t$와 S_t^2은 이 표본으로부터 계산된 표본평균과 표본분산이라 하자. 또한 $Z_t = \sqrt{n}(\bar{X}_t - \mu_0)/\sigma_0$라고 하자. 이때 표본내에서 개별 관측값 사이의 추출간격은 무시할 정도로 작으며, 표본내와 표본들간의 관측값들은 모두 독립임을 가정한다. 또한 이상원인으로 인하여 변화하는 평균과 표준편차의 양은 각각 $\delta = (\mu - \mu_0)/\sigma_0$와 $\gamma = \sigma/\sigma_0$를 이용하여 표기하기로 하고, δ와 γ는 모르는 값이라고 가정한다.

2.1. GLR 관리도

여기서 소개하는 GLR(generalized likelihood ratio) 관리도는 관리통계량으로 GLR을 사용하는 것으로 SPC 절차에 적용된 것은 그리 오래되지 않는다 (Vander Wiel, 1996; Apley와 Shi, 1999). 주로 서로 상관된 공정(autocorrelated process)에서 모수 변화를 탐지할 때 적용하던 GLR 관리도 절차를 서로 독립인 공정의 평균과 분산의 변화를 탐지하는 경우에 적용하여 소개하고자 한다.

이상원인이 알려지지 않은 시점 τ와 $\tau + 1$사이에서 발생하고, τ를 공정의 변화시점(change point)이라고 정의한다. 즉, $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_\tau}$는 관리상태인 $N(\mu_0, \sigma_0^2)$에서 추출한 표본이고, $\boldsymbol{X_{\tau+1}}, \boldsymbol{X_{\tau+2}}, \ldots$는 이상상태인 $N(\mu_a, \sigma_a^2)$에서 추출한 표본이 된다.

$\boldsymbol{X_k}$가 $N(\mu_k, \sigma_k^2)$에서 추출한 표본이라 할 때, 시점 t에서 관리상태와 이상상태를 각각 가설검정의 귀무가설과 대립가설로 표현하면

$$
\begin{aligned}
H_0 &: \mu_1 = \mu_2 = \cdots = \mu_t = \mu_0, \quad \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_t^2 = \sigma_0^2, \\
H_a &: \mu_1 = \mu_2 = \cdots = \mu_\tau = \mu_0, \quad \mu_{\tau+1} = \mu_{\tau+2} = \cdots = \mu_t = \mu_a\ , \\
&\quad \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_\tau^2 = \sigma_0^2, \quad \sigma_{\tau+1}^2 = \sigma_{\tau+2}^2 = \cdots = \sigma_t^2 = \sigma_a^2
\end{aligned}
$$

가 된다. 위의 가설에 대한 로그우도비(log likelihood ratio)는

$$\ln LR_t = \frac{1}{2}\left[\sum_{k=\tau+1}^{t} Z_k^2 + \sum_{k=\tau+1}^{t} V_k - n(t-\tau)\ln\gamma^2 - \frac{1}{\gamma^2}\left\{\sum_{k=\tau+1}^{t}(Z_k-\sqrt{n}\,\delta)^2 + \sum_{k=\tau+1}^{t} V_k\right\}\right] \tag{2.1}$$

가 됨을 알 수 있다. 여기서 $V_k = (n-1)S_k^2/\sigma_0^2$이며, 자세한 유도 과정은 부록 A에 수록하였다. 이때 $\ln LR_t$의 값이 클 경우 귀무가설을 기각하게 된다. 식 (2.1)의 통계량에서 모르는 값은 δ, γ^2 그리고 τ인데, GLR 관리도에서 사용하는 시점 t에서의 관리통계량은 이들 모르는 모수에 대하여 가장 큰 값이 되는

$$\max_{\delta,\gamma^2,\tau} \ln \mathrm{LR}_t$$

을 사용하며, 이 값이 관리한계 보다 클 경우 이상상태의 신호를 주는 것이다. 따라서 GLR 관리도의 관리통계량은 $\hat{\delta}$, $\hat{\gamma}^2$ 그리고 $\hat{\tau}$을 각각 δ, γ^2 그리고 τ의 MLE(maximum likelihood estimator)라 할 때,

$$\max_{\delta,\gamma^2,\tau} \ln \mathrm{LR}_t = \frac{1}{2}\left[\sum_{k=\hat{\tau}+1}^{t} Z_k^2 + \sum_{k=\hat{\tau}+1}^{t} V_k - n(t-\hat{\tau})\ln\hat{\gamma}^2 - \frac{1}{\hat{\gamma}^2}\left\{\sum_{k=\hat{\tau}+1}^{t}(Z_k-\sqrt{n}\,\hat{\delta})^2 + \sum_{k=\hat{\tau}+1}^{t} V_k\right\}\right] \tag{2.2}$$

가 된다.

위의 식에서 δ, γ^2 그리고 τ의 MLE는

$$\begin{aligned}
\hat{\delta} &= \frac{1}{\sqrt{n}}\bar{Z}_{\hat{\tau}+1,t},\\
\hat{\gamma}^2 &= \frac{\sum_{k=\hat{\tau}+1}^{t}(Z_k-\bar{Z}_{\hat{\tau}+1,t})^2 + \sum_{k=\hat{\tau}+1}^{t} V_k}{n(t-\hat{\tau})},\\
\hat{\tau} &= \arg\min_{0\le\tau<t}\left\{\sum_{k=1}^{\tau} Z_k^2 + \sum_{k=1}^{\tau} V_k + n(t-\tau)\left[\ln\left(\frac{\sum_{k=\tau+1}^{t}(Z_k-\bar{Z}_{\tau+1,t})^2 + \sum_{k=\tau+1}^{t} V_k}{n(t-\tau)}\right)+1\right]\right\}
\end{aligned} \tag{2.3}$$

로 유도된다. 여기서 $0 < a \le b$인 정수 a와 b에 대하여 $\bar{Z}_{a,b} = \sum_{k=a}^{b} Z_k/(b-a+1)$이며, 자세한 유도 과정은 부록 B에 수록하였다.

따라서 식 (2.2)에 (2.3)의 MLE를 대입하여 정리하면

$$\max_{\delta,\gamma^2,\tau} \ln \mathrm{LR}_t = \frac{1}{2}\left[\sum_{k=\hat{\tau}+1}^{t} Z_k^2 + \sum_{k=\hat{\tau}+1}^{t} V_k - n(t-\hat{\tau})\left\{\ln\hat{\gamma}^2+1\right\}\right] \tag{2.4}$$

이 된다. 따라서 공정의 평균과 분산의 변화를 동시에 탐지하는 GLR 관리도의 절차는 시점 t에서 식 (2.4)의 관리통계량이 관리한계 h_G보다 클 경우 이상신호를 주는 것이다. 이때, 관리한계 h_G는 주어진 관리상태에서의 평균런길이(average run length: ARL) 값을 만족하도록 설정할 수 있다.

2.2. Omnibus EWMA 관리도

Omnibus EWMA 관리도는 Domangue와 Patch (1991)가 제안하였고, 관리통계량은

$$O_t = \lambda_O\,|Z_t|^\alpha + (1-\lambda_O)\,O_{t-1} \tag{2.5}$$

을 사용한다. 여기서 λ_O는 평활모수(smoothing parameter)로서 $0 < \lambda_O \leq 1$인 값을 갖는다(이후에 모든 평활모수는 이와 같은 범위를 가짐을 가정한다). 또한 O_0는 초기치로서 관리상태에서 $|Z_t|^\alpha$의 평균을 사용한다.

O_t를 통계량 $|Z_t|^\alpha$ $(t=1,2,\ldots)$로 표현하면

$$O_t = \lambda_O \sum_{j=0}^{t-1}(1-\lambda_O)^j\,|Z_{t-j}|^\alpha + (1-\lambda_O)^t\,O_0 \tag{2.6}$$

가 되고, $|Z|^\alpha$의 k차 적률(moment)은

$$E\left[|Z|^{\alpha k}\right] = \frac{2^{\frac{\alpha k}{2}}\,\Gamma\left[\dfrac{\alpha k+1}{2}\right]}{\sqrt{\pi}} \tag{2.7}$$

이 됨이 알려져 있다. 식 (2.5), (2.6) 그리고 (2.7)을 이용하면 관리상태에서 관리통계량 O_t의 평균과 분산을 다음과 같이 구할 수 있다 (Domangue와 Patch, 1991).

$$\begin{aligned}
E[O_t] &= \left(\frac{2^\alpha}{\pi}\right)^{\frac{1}{2}}\Gamma\left(\frac{\alpha+1}{2}\right)\left\{1-(1-\lambda_O)^t\right\} + O_0\,(1-\lambda_O)^t,\\
\mathrm{Var}[O_t] &= \frac{2^\alpha\,\lambda_O}{(2-\lambda_O)\,\pi}\left[\sqrt{\pi}\,\Gamma(\alpha+0.5) - \left\{\Gamma\left(\frac{\alpha+1}{2}\right)\right\}^2\right]\left\{1-(1-\lambda_O)^{2t}\right\}.
\end{aligned}$$

$E^*[O_t]$와 $\mathrm{Var}^*[O_t]$를 $t\to\infty$일 때 $E[O_t]$와 $\mathrm{Var}[O_t]$의 극한값이라 하면

$$\begin{aligned}
E^*[O_t] &= \left(\frac{2^\alpha}{\pi}\right)^{\frac{1}{2}}\Gamma\left(\frac{\alpha+1}{2}\right),\\
\mathrm{Var}^*[O_t] &= \frac{2^\alpha\lambda_O}{(2-\lambda_O)\,\pi}\left[\sqrt{\pi}\,\Gamma(\alpha+0.5) - \left\{\Gamma\left(\frac{\alpha+1}{2}\right)\right\}^2\right]
\end{aligned}$$

가 된다.

일반적으로 α는 0.5와 2를 많이 사용한다. 만일 $\alpha = 0.5$인 경우 위의 극한값은

$$E^*[O_t] \approx 0.8222, \qquad \mathrm{Var}^*[O_t] \approx \frac{0.1219\,\lambda_O}{2-\lambda_O}$$

이고, $\alpha = 2$인 경우에는

$$E^*[O_t] = 1, \qquad \mathrm{Var}^*[O_t] = \frac{2\,\lambda_O}{2-\lambda_O}$$

이 된다. Omnibus EWMA 관리도는 관리한계 h_O에 대하여 $O_t \geq h_O$인 경우 공정에 이상이 있다는 신호를 주며, 관리한계는 $h_O = E^*[O_t] + K_O\sqrt{\mathrm{Var}^*[O_t]}$를 사용한다. 여기서 K_O는 주어진 관리상태에서의 ARL 값을 만족하도록 설정할 수 있다.

2.3. MaxMin EWMA 관리도

어떤 시점에서 추출한 표본의 최대값과 최소값을 이용하여 공정을 관리하는 절차는 예전부터 연구되어 왔는데, 그 시초로 Howell (1949)은 표본의 관측값들 중에서 가장 큰 관측값이 관리상한을 벗어나거나 혹은 가장 작은 관측값이 관리하한을 벗어날 때 신호를 주는 Shewhart 형태의 관리도를 제안하였다. 그 후 Sarkadi와 Vincze (1974)는 순서통계량(order statistics)을 이용한 관리도를 제안하였고, Amin과 Wolff (1995)는 표본의 관측값들의 절대값 중 가장 큰 값을 사용하는 EXTV(extreme value) EWMA 관리도를 제안하였다. EXTV EWMA 관리도의 관리통계량은 다음과 같다.

$$T_t = \lambda \max_{1\leq j\leq n} |\boldsymbol{Y_t}| + (1-\lambda)\,T_{t-1},$$

여기서 λ는 평활모수이고, $\boldsymbol{X_t} = (X_{t1}, \ldots, X_{tn})$에 대하여 $|\boldsymbol{Y_t}| = (|Y_{t1}|, \ldots, |Y_{tn}|)$이며 $Y_{tj} = (X_{tj} - \mu_0)/\sigma_0$이다. 초기치 T_0는 관리상태에서 $\max|\boldsymbol{Y_t}|$의 평균을 사용한다. 공정의 이상신호는 평균이 변하거나 분산이 증가해서 통계량 T_t가 관리상한을 벗어나면 발생한다. 이 관리도는 이상신호 이후 공정 평균과 분산 중 어떤 모수가 변화했는지 그리고 공정 평균이 어떤 방향(증가 또는 감소)으로 변했는지 판단하기 어렵다는 단점이 있다.

Amin 등 (1999)은 이러한 단점을 보완하기 위하여, 가장 작은 관측값에 대한 EWMA와 가장 큰 관측값에 대한 EWMA를 하나의 관리도에 작성하는 절차를 제안하였고, 이것을 MaxMin EWMA 관리도라고 칭하였다. EXTV EWMA 관리도는 관측값들 중에서 목표값으로부터 가장 멀리 떨어져 있는 값을 이용하는 것이고, MaxMin EWMA 관리도는 관측값들 중에서 목표값으로부터 각각 상한과 하한쪽으로 가장 멀리 떨어져 있는 값을 이용하는 것이다.

시점 t에서 Min EWMA는

$$L_t = \lambda_L \min_{1\leq j\leq n} (\boldsymbol{Y_t}) + (1-\lambda_L)\,L_{t-1}$$

이고, Max EWMA는

$$H_t = \lambda_H \max_{1\leq j\leq n} (\boldsymbol{Y_t}) + (1-\lambda_H)\,H_{t-1}$$

로 정의한다. 여기서 λ_L과 λ_H는 평활모수이며, 초기치 L_0와 H_0는 각각 관리상태에서 $\min(\boldsymbol{Y_t})$와 $\max(\boldsymbol{Y_t})$의 평균을 사용한다. MaxMin EWMA 관리도는 관리한계 h_{MM}에 대하여 $H_t \geq h_{MM}$ 또는 $L_t \leq -h_{MM}$인 경우 공정에 이상이 있다는 신호를 주며, 여기서 h_{MM}은 주어진 관리상태에서의 ARL 값을 만족하도록 설정할 수 있다.

공정 평균과 분산의 변화에 대한 MaxMin EWMA 관리도의 관리통계량 L_t와 H_t의 변화는 다음과 같이 요약할 수 있다. 먼저 분산에는 변화가 없지만, 평균이 증가하는 경우 두 통계량은 모두 증가하고 평균이 감소하는 경우 두 통계량은 모두 감소하는 경향을 나타낸다. 또한 평균에는 변화가 없지만, 분산이 증가하는 경우 두 통계량의 폭은 넓어지고 분산이 감소하는 경우 그 폭은 좁아지는 경향을 나타낸다. 만일 평균과 분산이 모두 변할 경우 위의 두가지 경향이 병행하여 나타나게 된다.

MaxMin EWMA 관리도는 표본 내에서 오직 두 개의 관측값만을 이용하여 평균과 분산의 변화를 동시에 탐지할 수 있다는 장점이 있다. 공정 분산이 감소하는 경우 이 관리도로는 탐지할 수 없지만, 이 경우 H_t와 L_t 각각에 관리상한과 관리하한을 설정하는 수정된 관리도를 이용하면 이를 해결할 수 있다.

2.4. Max EWMA 관리도

Max EWMA 관리도는 Chen 등 (2001)에 의하여 제안되었다. Max EWMA 관리도는 앞에서 정의한 Z_t와

$$W_t = \Phi^{-1}\left\{H\left(\frac{(n-1)\,S_t^2}{\sigma_0^2}\ ;\ n-1\right)\right\} \tag{2.8}$$

을 이용하는 것으로, $\Phi(\cdot)$는 표준정규분포의 누적분포함수, $\Phi^{-1}(\cdot)$는 $\Phi(\cdot)$의 역함수 그리고 $W \sim \chi_v^2$라 할 때 $H(w;v)$는 W의 누적분포함수를 나타낸다. 식 (2.8)에 정의된 W_t를 이용하는 것의 장점은 Z_t와 서로 독립이고 관리상태에서 Z_t와 마찬가지로 표준정규분포를 따르기 때문에, 하나의 관리도에 표현하기가 용이하다는 것이다.

이제 Z_t와 W_t를 이용하여 다음과 같은 EWMA를 정의해 보자.

$$\begin{aligned} C_t &= \lambda_M\, Z_t + (1-\lambda_M)\, C_{t-1}, \\ D_t &= \lambda_M\, W_t + (1-\lambda_M)\, D_{t-1}, \end{aligned}$$

여기서 λ_M은 평활상수이고, C_0와 D_0는 초기치로서 각각 관리상태에서 Z_t와 W_t의 평균인 0를 사용한다. Max EWMA 관리도의 관리통계량은 C_t와 D_t 중 최대값인

$$M_t = \max\{|C_t|, |D_t|\}$$

를 사용하며, 이러한 이유에서 Max EWMA 관리도라 부르게 되었다. 관리통계량 M_t는 공정평균이 변하거나 분산이 변하면 큰 값을 가지게 되며, 공정 평균과 분산이 목표값에 가깝게 되면 M_t는 작은 값을 가지게 된다. 그러나 M_t는 음수값을 가질 수 없기 때문에, Max EWMA 관리도는 관리한계 h_M에 대하여 $M_t \geq h_M$인 경우 공정에 이상이 있다는 신호를 주게 된다. h_M은 관리상태에서 M_t의 평균과 분산을 이용하여 $h_M = E(M_t) + K_M\sqrt{\mathrm{Var}(M_t)}$를 사용하며, 여기서 K_M은 주어진 관리상태에서의 ARL 값을 만족하도록 설정할 수 있다.

시점 t에서 관리통계량 M_t가 관리한계를 벗어났을 경우, $|C_t|$와 $|D_t|$ 값을 관리한계와 비교함으로써 공정 평균, 분산, 또는 평균과 분산 모두에 변화가 발생했는지 파악할 수 있다. 만일 두 통계량 값 중 $|C_t|$가 관리한계를 벗어난 경우, $Z_t > 0$이면 공정 평균이 증가한 것이고, $Z_t < 0$이면 평균이 감소한 것이라고 판단할 수 있다. 또한 $|D_t|$가 관리한계를 벗어난 경우, $W_t > 0$이면 공정 분산이 증가한 것이고, $W_t < 0$이면 분산이 감소한 것이라고 판단할 수 있다. $|C_t|$와 $|D_t|$가 모두 관리한계를 벗어난 경우에도 각각 Z_t와 W_t의 변화를 통하여 공정 평균과 분산의 변화를 파악할 수 있다.

2.5. Interval 관리도

Interval 관리도는 공정 평균과 분산의 변화를 하나의 관리도로 탐지하기 위하여 구간(interval)을 이용하는 것으로, Gan 등 (2004)이 제안하였다. 이 관리도는 관리통계량으로 $\bar{X}_t - r_I\, S_t$와 $\bar{X}_t + r_I\, S_t$를 사용하는데, 매 시점에서 $\bar{X}_t - r_I\, S_t$와 $\bar{X}_t + r_I\, S_t$ 값을 세로로 연결하여 선분을 만든 후, 관리도에 표시해 나가는 것이다. 여기서 r_I는 상수인데, Gan 등 (2004)은 0.25를 사용하는 것을 권장하고 있다.

Interval 관리도는 관리한계 h_I^+와 h_I^-에 대하여 $\bar{X}_t + r_I\, S_t \geq h_I^+$ 또는 $\bar{X}_t - r_I\, S_t \leq h_I^-$인 경우 공정에 이상이 있다는 신호를 주며, 관리한계는 $h_I^+ = \mu_0 + K_I\, \sigma_0$와 $h_I^- = \mu_0 - K_I\, \sigma_0$를 사용한다. 여기서 K_I는 주어진 관리상태에서의 ARL 값을 만족하도록 설정할 수 있다.

Interval 관리도는 이상신호가 발생할 때 공정 평균 또는 분산이 변화했는지, 아니면 둘 다 변화했는지 쉽게 확인할 수 있지만, 분산이 감소하는 경우는 탐지할 수 없다. 이 경우 구간의 폭인 $2\, r_I\, S_t$에 대하여 관리상한과 관리하한을 설정하여 비교하는 절차를 추가함으로 이를 해결할 수 있다.

2.6. EWMA-$\bar{X}$와 EWMA-$\ln S^2$ 관리도

앞에서는 하나의 관리도를 사용하여 공정 평균과 분산에 대한 변화를 동시에 탐지하는 관리도에 대해서 알아보았다. 이 절에서는 공정 평균과 분산의 변화를 탐지하기 위하여 전통적으로 많이 사용하는 EWMA-$\bar{X}$ 관리도와 EWMA-$\ln S^2$ 관리도를 병행하는 절차를 소개하고자 한다. 이 절차는 EWMA-$\bar{X}$ 관리도와 EWMA-$\ln S^2$ 중 어느 한쪽에서 관리한계를 벗어나면 공정에 이상이 있다는 신호를 주게 된다. 공정 분산에 대한 변화를 탐지하기 위하여 사용하는 EWMA-$\ln S^2$ 관리도는 Crowder와 Hamilton (1992)이 제안하였으며, S^2 대신 $\ln S^2$을 사용하는 것에 여러가지 장점이 있다고 알려져 있다.

EWMA-$\bar{X}$ 관리도의 관리통계량은

$$E_t^{\mu} = \lambda_\mu\, Z_t + (1 - \lambda_\mu)\, E_{t-1}^{\mu}$$

이고 EWMA-$\ln S^2$ 관리도의 관리통계량은

$$E_t^{\sigma^2} = \max\left\{\lambda_{\sigma^2}\, \ln(S_t^2) + (1 - \lambda_{\sigma^2})\, E_{t-1}^{\sigma^2},\, \ln(\sigma_0^2)\right\}$$

으로 정의하며, λ_μ와 λ_{σ^2}은 평활모수이고, 초기치는 $E_0^{\mu} = 0$와 $E_0^{\sigma^2} = \ln(\sigma_0^2)$를 사용한다. 여기서 EWMA-$\ln S^2$ 관리도는 공정 분산이 증가하는 경우만을 탐지하는 절차를 나타낸다.

EWMA-$\bar{X}$ 관리도와 EWMA-$\ln S^2$ 관리도는 관리한계 h_μ와 h_{σ^2}에 대하여 $|E_t^\mu| \geq h_\mu$ 또는 $E_t^{\sigma^2} \geq h_{\sigma^2}$인 경우 공정에 이상이 있다는 신호를 주며, 관리한계는

$$\begin{aligned} h_\mu &\approx K_\mu \sqrt{\frac{\lambda_\mu}{2-\lambda_\mu}}, \\ h_{\sigma^2} &= K_{\sigma^2} \sqrt{\mathrm{Var}(E_t^{\sigma^2})} \\ &\approx K_{\sigma^2} \sqrt{\left(\frac{\lambda_{\sigma^2}}{2-\lambda_{\sigma^2}}\right) \cdot \left\{\frac{2}{n-1} + \frac{2}{(n-1)^2} + \frac{4}{3\,(n-1)^3} - \frac{16}{15\,(n-1)^5}\right\}} \end{aligned}$$

를 사용한다 (Crowder와 Hamilton, 1992). 여기서 K_μ와 K_{σ^2}은 주어진 관리상태에서의 ARL 값을 만족하도록 설정할 수 있다.

3. 모의실험을 통한 관리도의 효율 비교

이 장에서는 앞에서 소개한 관리도들의 효율을 비교하기 위하여, 모의실험을 실시하였다. 관리도의 효율은 일반적으로 ARL_0를 같게 하고 ARL_1을 비교하는데, 앞에서 소개한 관리도들의 일부는 ARL을 이론적으로 계산하기 어렵기 때문에 모두 모의실험을 이용하여 효율을 비교하였다. 모의실험에서 일반성을 잃지 않고 $\mu_0 = 0$와 $\sigma_0^2 = 1$을 가정하였고, 표본의 크기는 $n = 4$, 공정의 변화시점은 평균이 100인 기하분포(geometric distribution)에서 추출하였으며, 반복은 100000번을 실시하였다. 또한 비교하는 관리도들의 평활모수는 모두 0.2, Omnibus 관리도에서 $\alpha = 2$ 그리고 Interval 관리도에서 $r_I = 0.25$를 사용하였다.

$\mathrm{ARL}_0 = 370.4$와 500를 만족하도록 관리한계를 설정했으며, 설정한 관리한계 및 공정 평균의 변화 δ와 표준편차의 변화 γ에 대한 이상상태에서의 ARL인 ARL_1 값을 각각 표 3.1과 3.2에 수록하였다. $\delta < 0$인 경우 유사한 결과를 얻을 수 있기 때문에 $\delta > 0$인 경우만 고려했으며, 일반적으로 분산이 증가하는 변화에 관심이 있는 경우가 대부분이기 때문에 여기서는 $\gamma > 0$인 경우만 고려하였다. 표 3.1과 3.2에서 $\delta = 0$이고 $\gamma = 1$인 경우는 관리상태에서의 ARL인 ARL_0 값을 나타내며, ARL_1 값의 위첨자는 6개 관리도 중 ARL_1 값이 작은 순서로 순위를 표시한 것으로서 3 순위까지만 나타내었다.

표 3.1과 표 3.2를 살펴보면, 다음의 결과를 알 수 있다. GLR 관리도는 공정 평균과 분산의 변화에 대하여 전반적으로 좋은 효율을 가지며, 특히 공정 분산의 변화가 어느 정도 큰 경우 효율이 좋았으며, Omnibus EWMA 관리도와 Interval 관리도는 공정 평균의 변화가 큰 경우 분산의 변화에 관계없이 효율이 좋은 것으로 나타났다. MaxMin EWMA 관리도는 이와 반대로 공정 평균의 변화는 작지만 분산의 변화는 큰 경우 효율이 좋았으며, Max EWMA 관리도는 공정 평균과 분산의 변화가 작은 일부의 경우에서 효율이 좋은 것으로 나타났다. 마지막으로 두 개의 관리도를 동시에 사용한 EWMA-$\bar{X}$와 EWMA-$\ln S^2$ 관리도는 공정 평균의 변화가 작은 경우 분산의 변화에 관계없이 효율이 좋은 것을 알 수 있었다.

이 논문에는 제시하지 않았지만 표본의 크기가 $n = 5$인 경우에도 유사한 결과를 얻었기 때문에, 위에서 언급한 특성은 표본의 크기에 크게 영향을 받지 않고 성립한다고 생각된다. 특이 사항으로 평균의 변화가 어느 정도 클 때($\delta \geq 1.5$) 동일한 δ에서 γ가 커질 경우 ARL_1 값

표 3.1: 관리도의 ARL_1 비교($\mathrm{ARL}_0 = 370.4$)

δ	γ	GLR	Omnibus EWMA	MaxMin EWMA	Max EWMA	Interval	EWMA-$\bar{X}$ & EWMA-$\ln S^2$
0.0	1.0	370.31	370.21	370.70	370.61	370.49	370.47
	1.5	9.35	15.31	7.73^2	8.33^3	15.65	7.69^1
	2.0	3.47^3	6.06	3.35^2	3.60	4.96	3.25^1
	3.0	1.59^1	2.95	1.77^2	2.03	2.06	1.80^3
0.5	1.0	13.84^3	21.91	20.71	11.26^1	46.38	11.27^2
	1.5	6.38	7.80	5.48^1	6.02^3	8.24	5.60^2
	2.0	3.06^2	4.69	3.07^3	3.33	3.93	3.04^1
	3.0	1.58^1	2.75	1.74^2	2.00	1.97	1.78^3
1.0	1.0	4.33	3.84^1	6.02	3.87^3	6.73	3.87^2
	1.5	3.51	3.32^1	3.52	3.58	3.41^2	3.46^3
	2.0	2.44^1	2.92	2.54^3	2.76	2.53^2	2.59
	3.0	1.50^1	2.32	1.66^2	1.90	1.75	1.71^3
1.5	1.0	2.29^3	1.80^1	3.48	2.44	2.10^2	2.44
	1.5	2.17^3	1.89^2	2.59	2.45	1.83^1	2.42
	2.0	1.86^2	1.94^3	2.11	2.22	1.72^1	2.14
	3.0	1.38^1	1.89	1.55^3	1.76	1.52^2	1.61
2.0	1.0	1.15^1	1.22^3	2.52	1.86	1.21^2	1.86
	1.5	1.24^1	1.35^3	2.07	1.88	1.28^2	1.87
	2.0	1.25^1	1.45^3	1.80	1.83	1.32^2	1.79
	3.0	1.19^1	1.56	1.44^3	1.61	1.33^2	1.50
3.0	1.0	1.02^3	1.00^2	1.73	1.30	1.00^1	1.30
	1.5	1.08^3	1.03^2	1.54	1.33	1.01^1	1.33
	2.0	1.12^3	1.08^2	1.41	1.35	1.04^1	1.34
	3.0	1.11^2	1.20^3	1.26	1.34	1.11^1	1.28
관리한계(h)		8.695	2.804	1.732	1.030	1.762	$h_\mu = 1.030$ $h_{\sigma^2} = 0.532$

이 조금 증가하는 경우가 일부 있었다. 이것은 공정 평균과 분산이 동시에 변하는 특수성 때문에, 평균의 변화가 어느 정도 큰 경우 분산의 변화는 관리통계량에 긍정적인 기여를 하지 못한 이유라고 판단된다.

4. 결론

이 논문에서는 하나의 관리도를 사용해서 공정 평균과 분산의 변화를 동시에 탐지하는 관리도로서 GLR 관리도, Omnibus EWMA 관리도, MaxMin EWMA 관리도, Max EWMA 관리도 그리고 Interval 관리도를 소개하였고, 관리도의 효율을 모의실험을 통하여 EWMA-$\bar{X}$와 EWMA-$\ln S^2$ 관리도와 함께 서로 비교하여 각 관리도의 강점과 약점을 살펴보았다.

그 결과 하나의 관리도로서 EWMA-$\bar{X}$와 EWMA-$\ln S^2$ 관리도를 병행하는 절차와 효율이 유사하거나 더 좋은 경우가 많이 있음을 알 수 있었다. 특히 GLR 관리도는 전반적으로 효율이 좋았으며, 공정 평균의 변화가 큰 경우 Omnibus EWMA 관리도와 Interval 관리도도 충분

표 3.2: 관리도의 ARL_1 비교($\mathrm{ARL}_0 = 500$)

δ	γ	GLR	Omnibus EWMA	MaxMin EWMA	Max EWMA	Interval	EWMA-$\bar{X}$ & EWMA-$\ln S^2$
0.0	1.0	499.85	499.72	500.54	499.79	502.97	501.93
	1.5	9.74	17.04	8.36^2	8.90^3	18.11	8.31^1
	2.0	3.52^2	6.45	3.53^3	3.74	5.37	3.42^1
	3.0	1.63^1	3.06	1.83^2	2.10	2.15	1.87^3
0.5	1.0	14.74^3	25.30	23.40	12.10^2	58.12	12.07^1
	1.5	6.65	8.43	5.81^1	6.36^3	9.22	5.98^2
	2.0	3.18^1	4.94	3.22^3	3.47	4.21	3.19^2
	3.0	1.60^1	2.85	1.79^2	2.06	2.05	1.85^3
1.0	1.0	4.52	4.10^3	6.39	4.03^2	7.81	4.02^1
	1.5	3.65^3	3.49^1	3.67	3.72	3.68	3.61^2
	2.0	2.50^1	3.03	2.65^3	2.87	2.64^2	2.70
	3.0	1.52^1	2.39	1.70^2	1.96	1.81	1.77^3
1.5	1.0	2.38^3	1.88^1	3.63	2.51	2.28^2	2.51
	1.5	2.27^3	1.96^2	2.68	2.53	1.93^1	2.49
	2.0	1.92^2	1.99^3	2.18	2.30	1.78^1	2.22
	3.0	1.41^1	1.93	1.59^3	1.81	1.55^2	1.67
2.0	1.0	1.17^1	1.25^2	2.61	1.91	1.25^3	1.91
	1.5	1.26^1	1.38^3	2.14	1.94	1.31^2	1.92
	2.0	1.28^1	1.48^3	1.85	1.88	1.35^2	1.85
	3.0	1.20^1	1.59	1.47^3	1.66	1.35^2	1.54
3.0	1.0	1.02^3	1.00^2	1.77	1.33	1.00^1	1.33
	1.5	1.09^3	1.03^2	1.58	1.36	1.02^1	1.36
	2.0	1.13^3	1.09^2	1.44	1.38	1.05^1	1.37
	3.0	1.13^2	1.21^3	1.28	1.37	1.11^1	1.31
관리한계(h)		9.097	2.932	1.759	1.062	1.809	$h_\mu = 1.060$ $h_{\sigma^2} = 0.552$

히 좋은 효율을 가짐을 알 수 있었다. 이 결과를 실제 공정의 상황에 맞게 적용시킨다면 좀 더 효율적으로 공정을 관리할 수 있을 것이라 판단한다.

부록 A. 식 (2.1)의 유도

공정 평균 μ와 분산 σ^2에 변화가 있을 경우, $\bar{X}_t$의 분포는 $t \le \tau$인 경우 $N(\mu_0, \sigma_0^2/n)$이고 $t \ge \tau + 1$인 경우 $N(\mu_a, \sigma_a^2/n)$이며, S_t^2의 분포는 $t \le \tau$인 경우 $\mathrm{Gamma}\left((n-1)/2, 2\sigma_0^2/(n-1)\right)$이고 $t \ge \tau + 1$인 경우 $\mathrm{Gamma}\left((n-1)/2, 2\sigma_a^2/(n-1)\right)$임을 쉽게 알 수 있다.

이 경우 Z_t의 분포는

$$Z_t = \frac{\sqrt{n}\,(\bar{X}_t - \mu_0)}{\sigma_0} = \gamma\,\zeta_t + \sqrt{n}\,\delta$$

이므로, $t \le \tau$인 경우 $N(0, 1)$이고 $t \ge \tau + 1$인 경우 $N(\sqrt{n}\,\delta, \gamma^2)$가 된다. 여기서 ζ_t는 서로

독립인 $N(0,\,1)$ 확률변수이다. 또한 V_t의 분포는

$$V_t = \frac{(n-1)\,S_t^2}{\sigma_0^2} = \gamma^2\,\eta_t$$

이므로, $t \le \tau$인 경우 $\chi^2(n-1)$이고 $t \ge \tau+1$인 경우 $\mathrm{Gamma}\left((n-1)/2,\, 2\gamma^2\right)$가 된다. 여기서 η_t는 서로 독립인 $\chi^2(n-1)$ 확률변수이다.

이를 이용하면, 가설 H_0와 H_a하에서 $Z_1,\ldots,Z_t$와 $V_1,\ldots,V_t$에 대한 우도함수(likelihood function) L_0와 L_a는 각각

$$\begin{aligned}
&L_0(\delta,\gamma,\tau\,|\,Z_1,\ldots,Z_t,\,V_1,\ldots,V_t)\\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^t \left\{\frac{1}{\Gamma((n-1)/2)\,2^{\frac{n-1}{2}}}\right\}^t \prod_{k=1}^{t}\left(V_k^{\frac{n-1}{2}-1}\right)\\
&\quad \times \exp\left[-\frac{1}{2}\left(\sum_{k=1}^{t} Z_k^2 + \sum_{k=1}^{t} V_k\right)\right],\\
&L_a(\delta,\gamma,\tau\,|\,Z_1,\ldots,Z_t,\,V_1,\ldots,V_t)\\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^t \left\{\frac{1}{\Gamma((n-1)/2)\,2^{\frac{n-1}{2}}}\right\}^t \left(\frac{1}{\gamma^2}\right)^{\frac{n(t-\tau)}{2}} \prod_{k=1}^{t}\left(V_k^{\frac{n-1}{2}-1}\right)\\
&\quad \times \exp\left[-\frac{1}{2}\left(\sum_{k=1}^{\tau} Z_k^2 + \sum_{k=1}^{\tau} V_k\right)\right] \exp\left[-\frac{1}{2\gamma^2}\left\{\sum_{k=\tau+1}^{t}(Z_k-\sqrt{n}\delta)^2 + \sum_{k=\tau+1}^{t} V_k\right\}\right]
\end{aligned}$$

가 된다. 따라서

$$\begin{aligned}
\ln LR_t &= \ln \frac{L_a(\delta,\gamma,\tau\,|\,Z_1,\ldots,\,Z_t,\,V_1,\ldots,V_t)}{L_0(\delta,\gamma,\tau\,|\,Z_1,\ldots,\,Z_t,\,V_1,\ldots,V_t)}\\
&= -\frac{1}{2}\,n\,(t-\tau)\,\ln\gamma^2 + \frac{1}{2}\left(\sum_{k=\tau+1}^{t} Z_k^2 + \sum_{k=\tau+1}^{t} V_k\right)\\
&\quad -\frac{1}{2\gamma^2}\left\{\sum_{k=\tau+1}^{t}(Z_k-\sqrt{n}\delta)^2 + \sum_{k=\tau+1}^{t} V_k\right\}
\end{aligned}$$

가 되며, 이를 정리하면 식 (2.1)이 유도된다.

부록 B. 식 (2.3)의 유도

$Z_1,\ldots,Z_t$와 $V_1,\ldots,V_t$가 주어진 경우, 로그우도비함수(log likelihood function)는

$$\begin{aligned}
&\ln L\,(\delta,\gamma^2,\tau\,|\,Z_1,\ldots,Z_t,V_1,\ldots,V_t)\\
&= c - \frac{1}{2}\left(\sum_{k=1}^{\tau} Z_k^2 + \sum_{k=1}^{\tau} V_k\right) - \frac{1}{2}\,n\,(t-\tau)\,\ln\gamma^2 - \frac{1}{2\gamma^2}\left\{\sum_{k=\tau+1}^{t}(Z_k-\sqrt{n}\delta)^2 + \sum_{k=\tau+1}^{t} V_k\right\}
\end{aligned}$$

로 표현할 수 있으며, 여기서 c는 δ, γ^2 그리고 τ를 포함하지 않는 부분을 나타낸다. 만일 τ를 알고 있다고 가정하면, δ와 γ^2의 MLE는 $\bar{Z}_{a,b} = \sum_{k=a}^{b} Z_k \,/\, (b-a+1)$에 대하여

$$\hat{\delta} = \frac{1}{\sqrt{n}} \bar{Z}_{\tau+1,t}, \qquad \hat{\gamma}^2 = \frac{\sum_{k=\tau+1}^{t} (Z_k - \bar{Z}_{\tau+1,t})^2 + \sum_{k=\tau+1}^{t} V_k}{n\,(t-\tau)}$$

가 된다. 이 추정량들을 로그우도비함수에 대입하면

$$\begin{aligned} &\ln LR\,(\tau \mid Z_1, \ldots, Z_t,\, V_1, \ldots, V_t) \\ &= c - \frac{1}{2} \left[\sum_{k=1}^{\tau} Z_k^2 + \sum_{k=1}^{\tau} V_k + n(t-\tau) \left\{ \ln \left(\frac{\sum_{k=\tau+1}^{t} (Z_k - \bar{Z}_{\tau+1,t})^2 + \sum_{k=\tau+1}^{t} V_k}{n\,(t-\tau)} \right) + 1 \right\} \right] \end{aligned}$$

이 되므로, 공정의 변화시점 τ의 MLE 및 δ와 γ^2의 MLE는 식 (2.3)과 같이 유도된다.

참고문헌

Amin, R. W. and Wolff, H. (1995). The behavior of EWMA–type quality control schemes in the case of mixture alternatives, *Sequential Analysis*, **14**, 157–177.

Amin, R. W., Wolff, H., Besenfelder, W. and Baxley, R., Jr. (1999). EWMA control charts for the smallest and largest observations, *Journal of Quality Technology*, **31**, 189–206.

Apley, D. W. and Shi, J. (1999). The GLRT for statistical process control of autocorrelated processes, *IIE Transactions*, **31**, 1123–1134.

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

Crowder, S. V. and Hamilton, M. D. (1992). An EWMA for monitoring a process standard deviation, *Journal of Quality Technology*, **24**, 12–21.

Domangue, R. and Patch, S. C. (1991). Some omnibus exponentially weighted moving average statistical process monitoring schemes, *Technometrics*, **33**, 299–313.

Gan, F. F., Ting, K. W. and Chang, T. C. (2004). Interval charting schemes for joint monitoring of process mean and variance, *Quality and Reliability Engineering International*, **20**, 291–304.

Howell, J. M. (1949). Control chart for largest and smallest values, *The Annals of Mathematical Staistics*, **20**, 305–309.

Sarkadi, K. and Vincze, I. (1974). *Mathematical methods of statistical quality control*, Academic Press, New York.

Vander Wiel, S. A. (1996). Monitoring processes that wander using integrated moving average models, *Technometrics*, **38**, 139–151.

[2008년 1월 접수, 2008년 2월 채택]

Procedures for Monitoring the Process Mean and Variance with One Control Chart*

Sang Hyun Jung[1] Jaecheon Lee[2]

ABSTRACT

Two control charts are usually required to monitor both the process mean and variance. In this paper. we introduce control procedures for jointly monitoring the process mean and variance with one control chart, and investigate efficiency of the introduced charts by comparing with the combined two EWMA charts. Our numerical results show that the GLR chart, the Omnibus EWMA chart, and the Interval chart have good ARL properties for simultaneous changes in the process mean and variance.

Keywords: Process control, GLR chart, EWMA chart, interval chart, average run length.

* This Research was supported by the Chung-Ang University Research Grants in 2007.

1) Graduate Student, Dept. of Statistics, Chung-Ang University, Seoul 156-756, Korea.
E-mail: kami1001@nate.com

2) Corresponding author. Professor, Dept. of Statistics, Chung-Ang University, Seoul 156-756, Korea.
E-mail: jaeheon@cau.ac.kr