변량추출비 관리도에서 이상원인 발생 시점의 추정

이재헌[1)], 박창순[2)]

요약

이 논문에서는 Samuel, Pignatiello와 Calvin(1998)이 제안한 $\overline{X}$ 관리도에서 이상원인 발생 시점에 대한 최대우도추정량에 기초하여 변량표본크기(VSS) $\overline{X}$ 관리도를 수행하는 경우에 사용할 수 있는 최대우도추정량을 제안한다. 또한 제안된 최대우도추정량을 이용하여 이상원인 발생 시점에 대한 신뢰구간을 설정하였다.

주요용어 : 변량표본크기, 이상원인 발생 시점, 최대우도추정량, $\overline{X}$ 관리도,

1. 서론

관리도에서 관리통계량이 관리한계를 벗어날 경우 이상원인이 발생했다는 신호를 주며, 신호가 발생할 경우 공정을 정지시킨 후 이상원인을 찾아 이를 규명하고 제거한 후 다시 공정을 가동시키는 것이 일반적이다. 이 때 이상원인이 발생한 시점(process change point)을 알 수 있다면 보다 빨리 이상원인을 제거하고 공정을 관리상태로 회복시킬 수 있을 것이다.

CUSUM 관리도와 EWMA 관리도에서는 신호가 발생한 후 관리통계량 값의 변화를 통하여 자체적으로 이상원인의 발생시점에 대한 추정량(built-in estimator)을 제공해 준다(Nishina (1992) 참조). Nishina(1992)는 CUSUM, EWMA, 그리고 MA 관리도에서 이상신호 후 자체적으로 제공하는 추정량을 서로 비교하였다. Samuel, Pignatiello와 Calvin(1998)은 Shewhart의 $\overline{X}$ 관리도를 수행할 때 이상신호 후 발생시점에 대한 최대우도추정량을 제안하였다. Pignatiello와 Samuel(2001)은 CUSUM과 EWMA 관리도에서도 신호 후 최대우도추정량을 사용하는 것이 자체적으로 제공하는 추정량에 비하여 더 효율적임을 모의실험을 통하여 보였다.

이상에서 언급한 관리도에서 표본을 추출하는 방법은 고정표본추출간격에서 고정표본크기를 추출하는 고정표본추출비(FSR)를 사용하는 것이다. 이에 반하여 현재의 관리통계량 값에 기초하여 다음 시점의 표본추출비를 변화시키는 관리도를 변량추출비(VSR) 관리도라 한다. VSR 관리도에서 다음 시점의 표본추출간격만을 변화시키는 관리도를 변량표본추출간격(VSI) 관리도라 하며, 다음 시점의 표본크기만을 변화시키는 관리도를 변량표본크기(VSS) 관리도라 한다. VSS와 VSI의 방법을 모두 적용시키는 관리도를 VSSVSI 또는 VSR 관리도라 한다.

공정 평균의 작은 변화를 탐지하는 경우에는 비효율적인 것으로 알려진 $\overline{X}$ 관리도에서 VSR 방법을 적용할 경우 그 단점을 충분히 보완할 수 있으며, 특히 $\overline{X}$ 관리도에서는 VSS 방법을 사용하는 것이 VSI 방법에 비하여 더 효율적인 것으로 알려져 있다. 이 논문에서는 VSS $\overline{X}$ 관리도를 수행하는 경우 이상신호 후 그 발생시점에 대한 최대우도추정량을 제안하고 그

1) (503-703) 광주광역시 남구 진월동 592-1 광주대학교 산업정보공학과 부교수
2) (156-756) 서울특별시 동작구 흑석동 221 중앙대학교 수학통계학부 교수

효율에 대하여 살펴보았다. 이 추정량은 FSS $\overline{X}$ 관리도에서 Samuel, Pignatiello와 Calvin(1998)이 제안한 최대우도추정량을 표본크기가 동일하지 않은 VSS 관리도에 사용할 수 있도록 확장시킨 것이다. 또한 최대우도추정량을 이용하여 이상원인의 발생시점에 대한 신뢰구간을 설정하는 방법을 제안한다.

2. 이상원인의 발생시점에 대한 최대우도추정량

공정에서 관측하는 품질특성치가 $N(\mu, \sigma^2)$을 따를 때 공정 평균 μ가 목표값 μ_0에서 $\mu_1 = \mu_0 + \delta\sigma$로 변하는 것을 탐지하는 $\overline{X}$ 관리도에 대하여 생각해 보자. 먼저 관리도는 시점 T에서 이상신호를 주며 그 신호는 오경보가 아님을 가정하고, τ는 관리상태에서의 마지막 시점을 나타낸다고 하자. 즉 $\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_\tau$는 관리상태에서 관측된 표본평균이고, $\overline{X}_{\tau+1}, \overline{X}_{\tau+2}, \cdots, \overline{X}_T$는 이상상태에서 관측된 표본평균임을 가정한다.

Samuel, Pignatiello와 Calvin(1998)은 FSS $\overline{X}$ 관리도를 사용하여 공정 평균에 대한 계단변화(step shift)를 탐지하는 경우, 신호 후 이상원인의 발생시점 τ (실제는 시점 τ와 $\tau+1$ 사이에서 이상원인이 발생하는 것임)에 대한 최대우도추정량을 다음과 같이 제안하였다.

$$\hat{\tau}_F = \max_{0 \le t < T} \left\{ (T-t)\left(\overline{\overline{X}}_{T,t} - \mu_0 \right)^2 \right\}. \tag{1}$$

단 $\overline{\overline{X}}_{T,t} = \sum_{j=t+1}^{T} \overline{X}_j / (T-t)$는 $T-t$개의 부그룹에 기초한 μ_1의 추정량을 나타낸다.

이제 관측시점에 따라 표본크기가 변화하는 VSS $\overline{X}$ 관리도의 절차에 대하여 알아보자. N_t를 시점 t에서의 표본크기라 할 때, 표준화된 표본평균 Z_t는 다음과 같이 정의한다.

$$Z_t = \sqrt{N_t}\left(\frac{\overline{X}_t - \mu_0}{\sigma} \right) \tag{2}$$

VSS $\overline{X}$ 관리도의 절차는 미리 설정된 관리한계 c에 대하여 $|Z_t| > c$일 때 이상신호를 주게 된다. 본 논문에서는 편의상 2개의 표본크기를 사용하는 VSS $\overline{X}$ 관리도를 고려하기로 한다. 이 경우 표본크기 N_t는 다음과 같이 결정한다.

$$N_t = \begin{cases} n_1 & \text{만일 } |Z_{t-1}| < c_S \\ n_2 & \text{만일 } c_S \le |Z_{t-1}| < c \end{cases}.$$

여기서 $n_1 < n_2$이고, c_S는 표본크기를 결정짓는 영역의 분계선(threshold limit)이다.

$N_t = n_0$ 인 FSS $\overline{X}$ 관리도에서 Samuel, Pignatiello와 Calvin(1998)이 제안한 식 (1)의 최대우도추정량을 표준화된 통계량 Z_t로 표현하고 정리하면

$$\hat{\tau}_F = \max_{0 \le t < T} \left\{ (T-t)\left(\overline{Z}_{T,t} \right)^2 \right\} \tag{3}$$

이 되는 것을 알 수 있다. 여기서 $\overline{Z}_{T,t} = \sum_{j=t+1}^{T} Z_j / (T-t)$이다.

관측시점에 따라 표본크기가 달라질 수 있는 VSS $\overline{X}$ 관리도에서, 로그우도함수는 식 (2)를 이용하면

$$\ln L(\tau) = \sum_{i=1}^{T} \ln\left(\frac{\sqrt{N_i}}{\sqrt{2\pi}\sigma} \right) - \frac{1}{2}\left[\sum_{i=1}^{T} Z_i^2 - 2(\mu_1 - \mu_0) \sum_{i=\tau+1}^{T} \frac{\sqrt{N_i}}{\sigma} Z_i + (\mu_1 - \mu_0)^2 \sum_{i=\tau+1}^{T} \frac{N_i}{\sigma^2} \right]$$

로 표현할 수 있다. 이 식에서 μ_1의 추정량으로, $N_{T,\tau}=\sum_{j=\tau+1}^{T} N_j$라 할 때, 표본크기로 가중한 가중평균 $\hat{\mu}_1=\sum_{j=\tau+1}^{T} N_j \overline{X}_j / N_{T,\tau}$를 사용하며, 이 추정량은 표준화된 통계량 Z_t를 이용하여 $\hat{\mu}_1=\mu_0+\sigma\ \overline{Z}_{T,\tau}^{\ *}$로 나타낼 수 있다. 여기서 $\overline{Z}_{T,\tau}^{\ *}=\sum_{j=\tau+1}^{T}\sqrt{N_j}Z_j / N_{T,\tau}$이다. 이 추정량을 로그우도함수에 대입하여 정리하면

$$\ln L(\tau)=\sum_{i=1}^{T}\ln\left(\frac{\sqrt{N_i}}{\sqrt{2\pi\sigma}}\right)-\frac{1}{2}\sum_{i=1}^{T}Z_i^2+\frac{1}{2}N_{T,\tau}(\ \overline{Z}_{T,\tau}^{\ *})^2 \tag{4}$$

가 되고, 이 로그우도함수를 최대로 만드는 τ의 최대우도추정량은

$$\hat{\tau}_V=\max_{0\le t<T} N_{T,t}(\ \overline{Z}_{T,t}^{\ *})^2 \tag{5}$$

가 된다. 식 (5)에서 $N_t=n_0$인 경우에는 식 (3)의 FSS $\overline{X}$ 관리도의 최대우도추정량과 동일해지므로, 식 (5)의 추정량은 표본크기에 관계없이 사용할 수 있는 일반화된 형태의 추정량이라 할 수 있다.

이제 모의실험을 통하여 VSS $\overline{X}$ 관리도에서 제안된 식 (5)의 최대우도추정량의 정밀도(precision)를 알아보자. 먼저 관측시점 1에서 100까지의 표본은 정규분포에서 평균이 μ_0인 관리상태에서 추출하고 시점 101부터의 표본은 평균이 $\mu_1=\mu_0+\delta\sigma$인 이상상태에서 추출하는 모의실험을 고려해 보자. 즉 $\tau=100$인 경우이다.

일반적으로 사용하는 FSS 관리도와 이 논문에서 사용하고자 하는 VSS 관리도를 서로 비교하기 위해서 관리상태에서의 수행 능력을 동일하게 하는 것이 일반적이다. 이것은 공정이 관리상태일 때, VSS 관리도의 평균 표본크기를 FSS 관리도의 고정된 표본크기와 동일한 값으로 설정하며, 관리상태에서의 평균런길이(ARL), ARL_0,를 동일하게 함으로 달성할 수 있다. 즉 $E[N_t|-c<Z_{t-1}<c,\ \delta=0]=n_0$이고, 주어진 상수 A_0에 대하여 $ARL_0=A_0$라는 제약조건 하에서 서로 비교하는 것이다. 이 논문에서는 $A_0=370.4$, $n_0=3$과 5, 그리고 $\delta=0.5,\ 0.75,\ 1.0,\ 1.5,\ 2.0$을 고려하였다. 각 경우에 대한 VSS $\overline{X}$ 관리도의 관리모수(n_1, n_2, c_S, 그리고 c)는 주어진 δ에 대하여 이상상태에서의 평균런길이, ARL_1,을 최소로 하도록 선택하였으며, 이 값들은 Lee(2003)의 결과를 참조하였다.

<표 1>에서 첫째 행은 FSS $\overline{X}$ 관리도의 경우이고 둘째 행은 VSS $\overline{X}$ 관리도의 경우로서 모두 독립적으로 100,000번 반복한 모의실험의 결과이다. 여기서 $E(T)$는 처음 시작부터 이상신호(오경보 제외)까지 추출한 평균 표본수이며, $\hat{\tau}_F$와 $\hat{\tau}_V$는 각각 식 (3)과 식 (5)에 의하여 얻어진 최대우도추정량들의 평균값을 나타낸다. 또한 추정량의 정밀도를 알아보기 위하여 추정량과 참값($\tau=100$)과의 차이가 주어진 상수 이내일 확률을 계산하였다.

<표 1>의 결과를 살펴볼 때, δ가 작은 경우 VSS $\overline{X}$ 관리도는 FSS $\overline{X}$ 관리도에 비하여 훨씬 빨리 이상원인을 탐지하는 것을 알 수 있으며, VSS 관리도에서 제안된 식 (5)의 최대우도추정량은 FSS 관리도에서와 마찬가지로 평균적으로 거의 정확하게 이상원인의 발생시점($\tau=100$)을 추정하는 것으로 나타났다. 최대우도추정량과 참값과의 차이가 주어진 상수 이내일 확률도 두 관리도가 서로 유사하며 δ가 큰 경우 ($\delta=1.5$와 2.0) 최대우도추정량의 정밀도는 매우 높은 것을 알 수 있다.

3. 이상원인의 발생시점에 대한 신뢰구간

이제 $\overline{X}$ 관리도를 사용하는 경우 이상신호 후 이상원인의 발생시점에 대한 신뢰구간을 설정하는 방법에 대하여 논의하고자 한다. Box와 Cox(1964)는 로그우도함수를 이용하여 이산형 모수값을 갖는 모수에 대하여 신뢰구간을 설정하는 방법을 제안하였다. 이 방법에 따라 VSS $\overline{X}$ 관리도에서 이상원인의 발생시점에 대한 신뢰구간을 설정하면 주어진 상수 D에 대하여 $\{t: \ln L(\hat{\tau}_V) - \ln L(t) < D\}$과 같은 형태가 된다. 여기서 $\ln L(t)$는 식 (4)의 로그우도함수이고, $\ln L(\hat{\tau}_V)$은 τ에 최대우도추정량 $\hat{\tau}_V$를 대입한 로그우도함수의 최대값을 나타낸다. 위의 식을 정리하면 다음과 같이 표현된다.

$$\{t: N_{T,t}(\overline{Z}_{T,t}^{*})^2 > N_{T,\hat{\tau}_V}(\overline{Z}_{T,\hat{\tau}_V}^{*})^2 - 2D\}. \qquad (6)$$

Box와 Cox(1964)는 $100(1-\alpha)\%$ 신뢰수준의 신뢰구간을 얻기 위하여 $D_{BC} = (1/2)\chi^2_{1,\alpha}$를, Siegmund(1986)는 근사이론을 이용하여 $D_S = -\ln[1-(1-\alpha)^{1/2}]$를 제안하였다.

이제 <표 1>과 동일한 경우에 대하여 모의실험(100,000번 반복)을 통하여 식 (6)의 신뢰구간을 설정하고, 이 신뢰구간들의 포함확률(coverage probability)과 신뢰구간의 평균 길이를 <표 2>에 제시하였다. <표 2>의 포함확률을 살펴볼 때 Box와 Cox(1964)의 D_{BC}를 사용할 경우 작은 δ 값에 대하여 신뢰구간의 길이가 작게 형성되어 주어진 신뢰수준에 크게 미치지 못하였고, Siegmund (1986)의 D_S를 사용할 경우 일반적으로 신뢰구간의 길이가 크게 형성되어 주어진 신뢰수준 보다 더 큰 포함확률을 갖는 것으로 나타났다. 이와 같은 문제점을 해결하기 위하여 Siegmund (1986)의 D_S를 수정하여 다음과 같은 상수 D_{LP}를 제안한다.

$$D_{LP} = 1.181\,D_S - 0.896\,\delta\sqrt{n_0}.$$

제안된 상수 D_{LP}는 FSS $\overline{X}$ 관리도에서 주어진 신뢰수준을 잘 만족하는 상수를 경험적으로 찾아내어 이 값들을 종속변수로 설정하고, D_S 그리고 n_0와 δ를 독립변수로 하는 회귀분석을 실시하여 얻어낸 식이다. 상수 D_{LP}는 FSS $\overline{X}$ 관리도에서 주어진 신뢰구간을 잘 만족하게 하며, VSS $\overline{X}$ 관리도에서도 D_{BC}와 D_S에 비하여 주어진 신뢰구간을 잘 만족하게 하는 것으로 나타났다. 다만 n_0와 δ가 큰 경우 D_{LP} 값은 음수가 되는 단점이 있다. (<표 2>에서 신뢰수준은 90%이고 $n_0 = 5$ 와 $\delta = 2.0$인 경우임.) 그러나 이런 경우에는 신뢰구간의 길이가 크지 않기 때문에 최대우도추정량으로 점추정한 결과로도 이상원인의 발생시점을 정확하게 추정할 수 있으며, 불가피하게 신뢰구간이 필요한 경우에는 상수 D_{BC}를 사용할 것을 권장한다.

4. 결론

이 논문은 매 시점마다 표본크기를 변화시키는 VSS $\overline{X}$ 관리도에서 신호 후 이상원인의 발생시점에 대한 최대우도추정량을 제안하였다. 이 추정량은 Samuel, Pignatiello와 Calvin(1998)이 제안한 추정량을 표본크기가 일정하지 않은 VSS $\overline{X}$ 관리도에서도 사용할 수 있도록 일반화시킨 것이다. 또한 제안된 최대우도추정량을 이용하여 FSS와 VSS $\overline{X}$ 관리도에서 신뢰구간을 설정하는 방법을 제시하였다. 모의실험 결과 VSS $\overline{X}$ 관리도에서 제안된 최대우도추정량과

신뢰구간을 이용하면 공정 평균의 작은 변화를 빨리 탐지하면서 그 발생시점 또한 정확하게 추정하는 것으로 나타났다.

참고문헌

[1] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society B*, Vol. 26, 211-243.

[2] Lee, J. (2003). $\overline{X}$ control charts with variable sample sizes and variable sampling intervals, *Journal of the Korean Data & Information Science Society*, Vol. 14, 429-440.

[3] Nishina, K. (1992). A comparison of control charts from the viewpoint of change-point estimation, *Quality and Reliability Engineering International*, Vol. 8, 537-541.

[4] Pignatiello, J. J., Jr. and Samuel, T. R. (2001). Estimation of the change point of a normal process mean in SPC applications, *Journal of Quality Technology*, Vol. 33, 82-95.

[5] Samuel, T. R., Pignatiello, J. J., Jr., and Calvin, J. A. (1998). Identifying the time of a step change with $\overline{X}$ control charts, *Quality Engineering*, Vol. 10, 521-527.

[6] Siegmund, D. (1986). Boundaries crossing probabilities and statistical applications, *Annals of Statistics*, Vol. 14, 361-404.

<표 1> FSS와 VSS $\overline{X}$ 관리도에서 τ에 대한 최대우도추정량과 정밀도

δ	n_0 / n_1	n_2	c_S	c / c	$E(T)$	$\hat{\tau}_F$ / $\hat{\tau}_V$	$\widehat{\Pr}(\lvert\hat{\tau}-\tau\rvert \le \varepsilon)$ $\varepsilon=0$	1	2	3
0.5	3			3.0	160.61	100.53	0.21	0.39	0.51	0.60
	1	34	1.86	3.0	114.71	101.32	0.16	0.32	0.43	0.52
0.75	3			3.0	122.47	100.01	0.37	0.61	0.74	0.81
	1	17	1.52	3.0	105.94	100.34	0.30	0.54	0.68	0.76
1.0	3			3.0	109.78	99.76	0.53	0.77	0.87	0.92
	2	12	1.63	3.0	103.55	99.87	0.50	0.76	0.87	0.92
1.5	3			3.0	102.91	99.59	0.76	0.92	0.96	0.97
	2	8	1.38	3.0	101.97	99.78	0.72	0.93	0.97	0.98
2.0	3			3.0	101.48	99.63	0.87	0.96	0.98	0.98
	2	4	0.67	3.0	101.43	99.70	0.86	0.97	0.98	0.99
0.5	5			3.0	133.36	100.19	0.30	0.53	0.66	0.74
	2	36	1.69	3.0	108.19	100.53	0.25	0.47	0.60	0.69
0.75	5			3.0	110.78	99.77	0.51	0.75	0.86	0.91
	3	21	1.58	3.0	103.72	99.91	0.47	0.74	0.85	0.91
1.0	5			3.0	104.50	99.65	0.68	0.88	0.94	0.97
	3	15	1.38	3.0	102.38	99.81	0.63	0.88	0.95	0.97
1.5	5			3.0	101.57	99.62	0.86	0.96	0.97	0.98
	4	9	1.28	3.0	101.46	99.73	0.86	0.97	0.98	0.99
2.0	5			3.0	101.08	99.79	0.94	0.98	0.99	0.99
	4	6	0.67	3.0	101.10	99.81	0.94	0.98	0.99	0.99

<표 2> FSS와 VSS $\overline{X}$ 관리도에서 신뢰구간에 대한 포함확률과 신뢰구간의 길이

신뢰수준	n_0	δ	포함확률			신뢰구간의 길이		
			D_{BC}	D_S	D_{LP}	D_{BC}	D_S	D_{LP}
90%	3	0.5	0.7050	0.9203	0.9025	10.92	23.26	21.30
			0.7764	0.9485	0.9366	15.73	29.51	27.48
		0.75	0.7658	0.9432	0.8998	5.63	12.40	9.48
			0.8119	0.9647	0.9337	7.63	15.34	12.16
		1.0	0.8210	0.9588	0.8940	3.79	8.83	5.32
			0.8270	0.9661	0.9031	4.03	9.31	5.67
		1.5	0.9102	0.9787	0.8970	2.72	7.07	2.46
			0.8985	0.9788	0.8820	2.49	6.14	2.25
		2.0	0.9527	0.9876	0.9039	2.28	6.02	1.33
			0.9497	0.9876	0.8947	2.13	5.41	1.29
	5	0.5	0.7391	0.9334	0.8994	7.05	15.46	12.78
			0.7850	0.9580	0.9331	9.36	18.68	15.86
		0.75	0.8142	0.9580	0.8928	3.95	9.14	5.69
			0.8218	0.9661	0.9055	4.32	9.84	6.27
		1.0	0.8813	0.9722	0.8954	3.00	7.42	3.28
			0.8709	0.9741	0.8883	2.97	7.16	3.26
		1.5	0.9495	0.9870	0.9045	2.39	6.26	1.45
			0.9472	0.9877	0.8968	2.13	5.45	1.35
		2.0	0.9764	0.9935	-	1.65	3.54	-
			0.9775	0.9941	-	1.62	3.51	-
95%	3	0.5	0.8126	0.9574	0.9531	15.04	29.79	28.70
			0.8699	0.9720	0.9692	20.65	35.98	34.91
		0.75	0.8532	0.9709	0.9527	7.73	16.43	13.47
			0.8976	0.9837	0.9722	10.31	19.64	16.55
		1.0	0.8902	0.9795	0.9511	5.24	12.48	8.14
			0.9005	0.9853	0.9607	5.65	12.92	8.65
		1.5	0.9457	0.9893	0.9496	3.77	10.81	3.99
			0.9404	0.9900	0.9456	3.43	9.18	3.61
		2.0	0.9707	0.9933	0.9491	3.21	9.59	2.20
			0.9694	0.9940	0.9464	2.89	8.53	2.01
	5	0.5	0.8357	0.9668	0.9539	9.72	20.06	17.69
			0.8802	0.9782	0.9690	12.52	23.60	21.10
		0.75	0.8862	0.9789	0.9510	5.40	12.62	8.51
			0.8977	0.9839	0.9598	6.00	13.50	9.34
		1.0	0.9268	0.9865	0.9496	4.10	11.11	5.19
			0.9245	0.9880	0.9498	4.11	10.29	5.13
		1.5	0.9685	0.9932	0.9489	3.20	9.75	2.28
			0.9694	0.9938	0.9486	2.96	8.84	2.13
		2.0	0.9855	0.9964	0.9521	2.11	5.44	1.14
			0.9850	0.9964	0.9499	2.11	5.56	1.13