Robust Control Charts Based on Self-critical Estimation Process

원 형 규
한성대학교, 산업공학과

Abstract

Shewhart control chart is a basic technique to monitor the state of a process. We observe observations of a group of size four or five in a rational way and plot some statistics(e.g., means and ranges) on the chart. When setting up the control chart, the control limits are calculated based on preliminary 20-40 samples, which were supposedly obtained from stable operating conditions. But it may be hard to believe, especially at the beginning of constructing the chart for the first time, whether the process is stable and hence all samples were generated under the homogeneous operating conditions. In this report we suggest a mechanism to obtain robust control limits under self-criticism. When outliers are present in the sample, we obtain tighter control limits and hence increase the sensitivity of the chart. Examples will be given via simulation study.

1. 서론

Shewhart 관리도(control chart)는 제품의 품질을 관리하기 위하여 제조공정의 상태를 감시하는 기본 도구이다. 관리하고자 하는 공정으로 부터 샘플크기(sample size) 4나 5정도의 관찰치들을 합리적 방법으로 추출하여 평균이나 범위(또는 표준편차)등의 통계량(statistics)을 관리도상에 표시한다. 관리도를 작성하기 위해서는 관리한계값(control limits)들이 필요하며 이들은 보통 초기 20-40 개의 샘플들로 구성된 부분그룹들(subgroups)로 부터 계산된다. 이들 부분그룹들이 얻어질때 공정은 안정한 상태(stable state)에 놓여있으리라 여겨져야 한다. 그러나 새로운 공정을 구축하거나 또는 기존 공정에 대해 품질개선이 필요하여 관리도를 처음으로 작성하고자 할 때 공정의 안정성을 확신하기가 쉽지 않다. 공정이 안정되어 있지 않은 상황에서는, 소위 이상치(outliers)라 불리는 극단의 관찰치들(extreme observations)이 부분그룹들중에 내재할 가능성이 매우 높으며 이러한 현상은 실제로 제조공정상에서 자주 발생하고 있다.

이상치들이 부분그룹들중에 내재할 경우, Shewhart 관리도에서는 관리한계값들을 얻는데 필요한 모수 추정법(parameter estimation)이 이상치들에 의해 민감한 영향을 받아, 관리중심선(centerline)이 변하며 관리한계폭(width of control limits)이 확대된다. 그 결과 이상치들을 보유한 부분그룹들을 구별해 내기가 어렵게 됨은 물론 이후 이상원인(assignable cause)에 의해 공정상태가 불안정하게 될 때 이를 감지해내는 관리도의 능력이 저하된다. 따라서 초기 관리도 작성시 만일 있을수도 있는 이상치들에 의해 관리중심선과 관리한계값들이 크게 변하지 않기 위해서는 강건추정치들(robust estimates)의 이용이 제안되었다[1] [2] [3] [4] [6] [8] [9] [10].

본고에서는 강건추정치를 얻는 방법의 일종인 자기비평 추정법(self-critical estimation)을 소개함과 아울러 이를 평균-표준편차($\bar{x}$-S) 관리도에 적용하고자 한다. 먼저 정규분포 모형에서의 자기비평 추정치의 특성을 살펴본다. 그다음 이상치에 대한 둔감성(강건성)을 이용하여 새로운 자기비평 평균-표준편차(self-critical $\bar{x}$-S) 관리도를 제시한다. 마지막으로 제안된 관리도와 기존의 평균-범위($\bar{x}$-R) 및 평균-표준편차 관리도를 컴퓨터 모의실험을 통하여 비교분석한다.

2. 자기비평 추정법

자기비평 추정치(Self-critical estimate)는 다음의 목적함수, ℓ_c, 를 최대화(maximization) 함으로써 얻는다.

$$\ell_c(\theta) = (1/c)\sum_{i=1}^{n}\{f^c(x_i,\theta)/Q^{c/(1+c)} - 1\}$$

위식에서

$$Q = Q(\theta,c) = \int_{-\infty}^{\infty} f^{1+c}(x_i,\theta)dx_i$$

로 정의되며 c는 상수값으로 주어지는 자기비평 계수(coefficient)이다. 또 θ에 대하여 $\ell_c(\theta)$를 미분하여 0으로 놓으면 다음과 같은 추정방정식을 얻는다.

$$\sum_{i=1}^{n} f^c(x_i,\theta)\{(1+c)\frac{\partial \log f(x_i,\theta)}{\partial\theta} - \frac{\partial \log Q(\theta,c)}{\partial\theta}\} = 0$$

위식에서 c의 값이 0로 주어질때 목적함수 $\ell_c(\theta)$는 가능성함수(또는 우도함수, likelihood function)와 일치하며 또한 추정방정식도 서로 일치한다. 추정방정식으로 부터 얻은 추정자(estimator), $\hat{\theta}(c)$, 는 강건추정자(robust estimator)이며 불편 스코아(unbiased score)를 갖는다. 밀도함수

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

를 갖는 정규분포인 경우 자기비평 추정 방정식은 다음과 같다.

$$\sum_{i=1}^{n}(x_i-\mu)e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2} = 0$$

$$\sum_{i=1}^{n}\{(1+c)(x_i-\mu)^2 - \sigma^2\}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2} = 0\cdot$$

위의 두식으로 부터 자기비평 추정치 $\hat{\mu}(c)$와 $\hat{\sigma}^2(c)$는 초기 추정치(initial estimates)로 부터 반복해서(iteratively) 대입하여 아래와 같이 구한다.

$$\mu = \frac{\sum x_i v_{ic}}{\sum v_{ic}}$$

$$\sigma^2 = (1+c)\frac{\sum (x_i-\mu)^2 v_{ic}}{\sum v_{ic}}$$

여기서

$$v_{ic} = e^{-\frac{c}{2}(\frac{x_i-\mu}{\sigma})^2}$$

로 정의된다. 초기 추정치로는 모멘트 추정치(moment estimate)나 최대가능성 추정치(maximum likelihood estimate)들이 사용된다. c=0 근방에서 추정치 $\hat{\mu}(c)$와 $\hat{\sigma}^2(c)$는 M-추정치(M-estimates)이고, x_1, x_2, …, x_n 가 평균(mean) μ와 분산(variance) σ^2인 정규분포로부터 얻은 임의표본(random sample)일 경우 이들은 각각 μ, σ^2에 일치(consistent)한다. 또한 $\hat{\mu}(c)$ and $\hat{\sigma}^2(c)$는 asymptotically 독립적으로 분포한다. Asymptotic efficiencies도 c의 0근방에서 높게 나타나고 있다. $\hat{\mu}(c)$와 $\hat{\sigma}^2(c)$의 influence function은 c>0 일때 bounded되고 0로 재감소(redecended)하여 이상치들에 대한 강건성을 보인다[7].

3. 자기비평 평균-표준편차 관리도

Shewart가 제시한 관리도의 일반적 형태는

UCL = $\mu_W + k\sigma_W$

Cl = μ_W

LCL = $\mu_W - k\sigma_W$

의 모습을 갖고 있으며, 여기서 W는 공정의 변화를 측정하기 위한 하나의 샘플통계량(a sample statistic)이고 μ_W는 W의 기대값(expected value)이고 σ_W는 W의 표준편차(standard deviation)이다. 위의 일반형에서 W를 어떻게 구하느냐에 따라 여러종류의 관리도가 얻어진다. 예를 들면, 공정중심 의 변화를 파악하기 위해서 샘플평균 $\bar{x}$를 이용하거나, 공정산포의 변화는 샘플범위(sample range) R 또는 샘플표준편차 S를 사용한다. 이들은 모두 공정상태의 변화나 이상치에 민감하게 반응하는 특성을 지니고 있다. 따라서 이들은 공정상태의 변화를 측정하기 위한 수단으로서는 매우 적절하나 관리한계값을 정할때 이들을 그대로 사용하면 한계값이 이상치에 민감하게 반응하여 한계값들 사이의 폭이 증가되는 현상이 발생한다. 바람직하게는 이상치에 둔감하게 반응하는 추정자를 이용하여 관리상한(upper control limit)과 하한(lower limit) 사이의 폭을 가깝게(tightly) 유지시키는 방법이다. 관리한계값들이 가까이 설정된 관리도에 공정변화에 민감하게 반응하는 샘플통계량을 표시하면(plotting) 공정의 변화상태를 더욱 쉽게 파악할수 있다.

평균-표준편차 관리도는 컴퓨터가 일반화되어 계산의 어려움이 그다지 심각하지 않은 현장에서 평균-범위 관리도 못지 않게 많이 쓰이고 있다. 평균-표준편차 관리도에서의 평

균 관리도($\bar{x}$ chart)의 관리한계값들은 다음식으로 부터 계산된다.

UCL = $\bar{\bar{x}} + A_3\bar{S}$

CL = $\bar{\bar{x}}$

LCL = $\bar{\bar{x}} - A_3\bar{S}$

위식에서 $\bar{\bar{x}}$는 부분그룹평균들의 평균이며 $\bar{S}$는 부분그룹 표준편차들의 평균이다. A_3는 상수로 샘플크기의 함수로 주어진다. 평균-표준편차 관리도에서의 표준편차 관리도(S chart)에서는 관리 한계값들이 아래와 같이 정의된다.

UCL = $B_4\bar{S}$

CL = $\bar{S}$

LCL = $B_3\bar{S}$

B_3, B_4는 샘플크기의 함수로 주어지는 상수값이다. 그러나 각 부분그룹의 정보를 요약하기 위해 사용된 샘플평균이나 샘플표준편차는 부분그룹내의 이상치에 민감하게 반응하는 추정치들이며, 관리한계값들을 구하는데 사용되는 이들의 평균도 역시 민감하게 반응한다.

본고에서는 관리한계값들이 자기 비평추정자에 의해 결정되는 평균-표준편차 관리도의 새로운 형태, 즉 자기비평 평균-표준편차 관리도를 제안한다. 자기비평 평균-표준편차 관리도에서의 평균 관리도 관리한계는 다음식으로 정의된다.

UCL = $\bar{\mu}_{(c)} + A_3\bar{\sigma}_{(c)}$

CL = $\bar{\mu}_{(c)}$

LCL = $\bar{\mu}_{(c)} - A_3\bar{\sigma}_{(c)}$

위의 관리 한계값들을 갖는 관리도에 개개의 부분그룹의 평균을 표시한다. 표준편차 관리도에서는 관리 한계값들이 아래와 같이 정의되는 관리도에 부분그룹의 표준편차를 표시한다.

UCL = $B_4\bar{\sigma}_{(c)}$

CL = $\bar{\sigma}_{(c)}$

LCL = $B_3\bar{\sigma}_{(c)}$

위에 제안된 관리도들에서 c=0 일 때는 자기비평 추정자는 최대가능성추정자와 일치하여 기존의 평균-표준편차 관리도가 된다.

4. 모의 실험 예

Monte Carlo 모의 실험을 통하여 새로이 제안된 자기비평 평균-표준편차 관리도를 기존의 관리도들과 비교하였다. 보통 관리도를 작성하기 위해서는 샘플크기 4나 5인 관찰치들을 부분그룹(subgroup)으로 하여 20내지 25정도 반복해서 부분그룹들을 모아 관리도 작성에 필요한 모수의 추정값들을 계산한다. 본 연구에서는 초기 씨값(initial seed) 123457을 이용하여 IMSL[5]로 부터 표본크기 5인 부분그룹들을 모두 25개 생성하여 관리도들을 비교하였다.

<u>예제 1.</u> 관찰치 모두 표준정규분포로 부터 생성된 경우

<표 1>은 관찰치 모두 N(0,1)에서 모의 생성되었을 때, 각 관리도의 관리한계값들과 관리한계를 넘어가는 부분그룹의 수를 보여준다. 이때 관리 한계값을 넘어가는 부분그룹이 하나도 없음을 알수 있다. 주목할 점은 자기비평 관리도의 관리한계값들은 $\bar{x}$-S 관리도의 한계값들을 포함하고 있다는 점이다.

<표 1> 각 관리도의 관리한계값

종류		상한	중심	하한	빈도수
$\bar{x}$-R	$\bar{x}$	1.272	0.003	-1.226	0
	R	4.652	2.200	0.000	0
$\bar{x}$-S	$\bar{x}$	1.303	0.003	-1.297	0
	S	1.903	0.911	0.000	0
self-critical(c=0.4)	$\bar{x}$	1.410	0.038	-1.333	0
	S	2.008	0.961	0.000	0

<u>예제 2.</u> 각 부분그룹의 5번째 관찰치가 일반정규분포에서 모의 생성된 복합분포(stable mixtures)인 경우

각 부분그룹에서 5번째 관찰치가 일반 정규분포로 부터 생성되었다. 따라서 전체적으로 20%가 표준정규분포가 아닌 다른 일반정규분포들로 부터 생성되었을 경우 관리도에 미치는 영향을 살펴보았다. <표 2>는 복합모형의 경우에 따라 자기비평 평균-표준편차 관리도의 관리한계를 벗어나는 부분그룹의 수를 보여준다. 기존의 평균-범위 관리도의 관리한계는 전혀 일반 정규분포의 존재를 보여주지 못하고 있으며, 기존의 평균-표준편차 관리도에서는 S관리도에서 일반정규 분포의 존재를 간헐적으로 보여준다. 그러나 자기비평 평균-표준편차 관리도에서는 c가 증가할수록 이상징후를 더 많이 검출해 내고 있음을 알 수 있

다. 그러나 c가 너무크면 efficiency가 낮아지므로 c의 값은 0.3에서 0.5사이가 적절하다.

<표 2> 복합분포의 경우 각 관리도의 관리한계를 넘어가는 부분그룹들의 수

종류		N(0,1)	N(0,3)	N(3,1)	N(3,3)
$\bar{x}$-R	$\bar{x}$	0	0	0	0
	R	0	0	0	0
$\bar{x}$-S	$\bar{x}$	0	0	0	0
	S	0	1	0	1
self-critical(c=0.4)	$\bar{x}$	0	0	0	4
	S	0	4	0	7
self-critical(c=0.5)	$\bar{x}$	0	1	0	4
	S	0	4	1	8

5. 결론

초기 관리도 작성에서 관리한계를 설정할 때, 자기비평 추정법을 사용하면, 이상치들에 의한 영향을 최소화 내지는 둔감시킬수 있다. 그 결과 관리한계 폭이 축소되며 관리도의 민감도는 증대되어 이상치의 유무를 쉽게 판별할수 있게 된다. 따라서 공정상태에 대한 보다 정확한 진단을 가능케하고 이상 원인이 존재할 경우 이에 대한 파악을 용이케 하여줌으로써 관리도의 효용성을 증대시켜 생산현장에서의 품질향상에 적극 활용케 할수 있다. 본 고에서는 강건추정치의 일종인 자기비평 추정치를 이용한 평균-S 관리도가 평균-R 및 평균-S 관리도보다 이상치가 있을 경우 이들을 지닌 샘플들을 보다 잘 구별해 낼수 있음을 보였다. 그리고 본 방법론은 p-관리도와 같은 이산형 관리도에도 적용이 가능하며 실험계획, 회귀모형 및 시계열모형등의 분석에도 적용될 수 있으리라 본다.

6. 참고문헌

[1] Alloway, J. A. JR, and M. Raghavachari, "Control Charts Based on the Hodges-Lehmann Estimator", Journal of Quality Technology, vol. 23, no. 4, 1991, pp.336-347.

[2] Amin, R. W., and R. W. Miller, "A Robustness Study of bar X Charts with Variable Sampling Intervals", Journal of Quality Technology, vol. 25, no. 1, 1993, pp.36-44.

[3] Ferrel, E. B., "Control Charts Using Midrange and Medians", Industrial Quality Control, vol.9, pp.30-34.

[4] Hawkins, D. M, "Robustification of Cumulative Sum Charts by Winsorization", Journal of Quality Technology, vo. 25, no. 4, 1993, pp.248-261.

[5]IMSL, User's manual(C functions for scientific programming), Visual Numerics, Inc., 1994.

[6] Langenberg P. and B. Iglewicz, "Trimmed Mean bar X and R Charts", Journal of Quality Technology, vol. 18, no. 3, 1986, pp.152-161.

[7] Paulson, A. S., Presser, M. A., and E. H., Nicklin, "Self-Critical And Robust Procedures For the Analysis of Univariate Complete Data.", unpublished research paper, Rensselaer Polytechnic Institute.

[8] Quesenberry C. P., "Screening Outliers in Normal Process Control Data with Uniform Residuals", Journal of Quality Technology, vo. 18, no. 4, 1986, pp.226-233.

[9] Rocke, D. M., "Robust Control Charts", Technometrics, vol. 31, no. 2, 1989, pp.173-184.

[10] Schilling, E. G., and P. R. Nelson, "The Effect of Non-Normality on the Control Limits of bar X Charts", Journal of Quality Technology, vol. 8, no. 4, 1976, pp.183-188.