

비대칭 분포를 따르는 공정에서 사분위수를 이용한 관리도법*

박성현¹⁾ 박희진²⁾

요약

본 연구에서는 분포를 알 수 없고 비대칭인 공정자료에 실용적으로 적용할 수 있는 간단한 관리도법을 제안하였다. 비대칭 분포를 따르는 공정자료에 정규성 가정에 기초한 슈하르트 관리도를 그대로 적용하면 비대칭성이 증가할수록 제 1종 오류를 범할 확률이 증가할 가능성이 높아지며 변동을 관리하는데 효율성이 떨어지게 된다. 이러한 문제를 해결하기 위해 본 연구에서 제시한 관리도는 관리한계선을 사분위수에 기초하여 정하는 방안을 제시하고 있다. 이러한 방법으로 관리한계선을 그릴 경우 제 1종 오류도 감소하게 되고, 비대칭분포를 하는 공정자료에 대하여 매우 실용적이라고 하겠다.

주요용어: 관리도, 비대칭분포, 사분위수, 상자그림, 제 1종 오류

1. 서론

관리도법은 공정의 안정상태 유무를 탐지하는 방법으로 벨 연구소의 슈하르트 박사에 의해 소개되어진 이후 통계적 공정관리에서 가장 기본적인 도구로 자리 잡았다. 이러한 원인에는 여러 가지가 있지만 가장 중요한 것은 관리도법이 가지고 있는 명확성과 단순성에 있다고 할 수 있다. 즉, 기술자들이 쉽게 이해할 수 있고 사용할 수 있다는 것이다.

공정을 관리하기 위해서는 공정의 상태를 나타내는 특성값들이 필요하며 이는 제품의 품질을 나타내는 여러 특성 중 관리하고자 하는, 또는 관심이 있는 특성을 품질특성치로 정의한다. 품질특성치는 일반적으로 공정에서 일정한 간격으로 랜덤추출하여 측정하며, 관리도법에서는 이러한 품질특성치가 목표값을 중심으로 특정한 수준 안에서 유지하는 지, 밖으로 벗어나는 지를 평가하여 공정의 안정상태 유무를 결정한다. 즉, 특정한 수준 안에서 유지하는 경우 공정이 안정상태에 있다고 하며, 이 때 발생하는 변동원인은 우연원인에 의한 것으로 확인과 제거가 어려운 반면, 특정한 수준 밖으로 벗어나는 경우 공정이 이상상태에 있다고 하며, 이때 발생하는 변동원인은 이상원인에 의한 것으로 확인과 제거가 가능하다.

이렇게 관리도는 이상원인을 탐지하여 제거할 수 있게 해줌으로써 공정이 안정상태를 유지할 수 있게 도움을 준다. 따라서 이상원인을 재빨리 탐지할 수 있으면 있을수록 더 효과적인 관리도가 되는 것이다. 이러한 관점에서 관리도에서 품질특성치의 목표값이 되는

* 본 연구는 한국과학재단 목적기초연구 지원으로 부분적으로 수행되었음. (R01-2003-000-10220-0)

1) (151-747) 서울시 관악구 신림동 산 56-1, 서울대학교 통계학과, 교수

E-mail: parksh@plaza.snu.ac.kr

2) (151-747) 서울시 관악구 신림동 산 56-1, 서울대학교 통계학과, 박사과정

E-mail : jini0306@snu.ac.kr

중심선과 공정에서의 품질특성치의 변동이 우연원인에 의한 것인지, 이상원인에 의한 것인지를 판단하는 기준이 되는 관리한계선을 결정하는 것은 관리도 작성의 가장 핵심적인 사항이라 할 수 있다. 이 밖에 관리도에 대한 자세한 내용은 박성현 외(1997), 박성현과 박영현(2003), 박창순(1999)을 참고하기 바란다.

슈하르트 관리도는 가장 많이 사용되는 관리도로 품질특성치가 정규분포 또는 근사적 정규분포를 따른다는 가정을 기본으로 하고 있다. 중심선은 품질특성치의 평균 μ 를 사용하고, 관리한계선은 평균을 중심으로 $\pm 3\sigma$ 수준을 일반적으로 사용한다. 이렇게 중심선과 관리한계선을 결정하면, 공정이 안정상태임에도 불구하고 관리한계선 밖으로 벗어나는 확률인 제 1종 오류가 0.0026에 불과하다.

그러나 많은 경우 품질특성치의 정규성 가정은 적절하지 못하다. 예를 들면, 반도체 제조공정에서의 와이퍼의 저항력의 분포는 비대칭 분포를 따른다. Bai와 Choi(1995)는 화학 공정에서의 측정치와 가속시험에서의 제품의 수명은 비대칭 분포를 따른다고 하였으며, Ryan(2000)과 Lewis, Montgomery와 Myers(2001)에는 정규분포를 따르지 않는 여러 품질 특성치들에 대해서 설명하고 있다.

비대칭 분포를 따르는 품질특성치에 대한 관리도에 정규성 가정을 기초한 슈하르트 관리도의 중심선과 관리한계선을 그대로 사용한다면 품질특성치의 변동을 관리하는데 효율성이 떨어져 비대칭성이 증가할수록 제 1종 오류를 범할 확률이 증가한다. 이러한 문제를 해결하기 위해 기존의 연구에서 제시한 해결책을 살펴보면, 표본의 크기를 늘려 중심극한 정리에 의해 품질특성치를 정규분포에 근사시키는 방법이 있다. 하지만 추가 비용의 문제가 발생하며, 특히 제조공정상 한번에 하나밖에 생산할 수 없는 경우에는 불가능하다. 또한, 품질특성치를 적절히 변환하여 정규성을 만족시키는 방법도 있지만 변환한 값을 이용한 관리도의 해석에 어려움이 따른다. 이러한 해결책은 좋은 특성들을 가지고 있지만 Bai와 Choi(1995)가 지적하였듯이 다소 복잡할 수 있으며, 정규성 가정을 기초로 한 기본적인 접근방법을 선호하는 기술자에 맞지 않을 수 있다. 즉, 관리도의 가장 큰 장점인 단순성과 명확성을 잃어버릴 수 있다는 것이다. 또한, Choobinch와 Ballard(1987), Bai와 Choi(1995)는 평균보다 높은 값과 낮은 값의 비율을 이용한 가중분산의 개념을 도입하여 자료의 비대칭의 정도와 방향을 고려한 관리한계선을 제안하였고, Janacek와 Meikle(1997)는 중앙값과 순서통계량을 이용한 중앙값 관리도를 제안하였으며, Grimshaw와 Alt(1997)는 분위수 합수값을 이용한 관리도를 제안하였다. 이러한 기존 연구들은 대부분 슈하르트 \bar{X} 관리도에 초점이 맞춰져 있다. 하지만 슈하르트 \bar{X} 관리도는 부분군의 평균을 사용함으로써 슈하르트 관리도에 비해 자료가 가지는 비대칭성에 영향을 덜 받는다.

따라서 본 연구에서는 관리도의 장점인 단순성과 명확성을 잃지 않으면서, 비대칭분포를 따르는 품질특성치에 대해 슈하르트 X 관리도의 대안으로 사분위수를 이용한 관리도를 제시하고자 한다. 다음 장에서는 제안된 중심선과 관리한계선의 의미와 성질에 대해서 살펴본 후 실제 예제를 통하여 기존의 방법과 비교하고, 마지막으로 여러 비대칭분포에서 제안한 관리도의 제 1종 오류를 범할 확률을 계산해 보았다.

2. 중심선과 관리한계선

2.1. 중심선

정규분포의 특징상 평균은 중앙값과 최빈값과 동일한 값이다. 따라서 정규성 가정에 기초한 슈하르트 관리도에서 중심선으로 평균을 사용하여 품질특성의 평균을 관리하는 것은 중앙값을 관리하는 것과 같은 의미를 가진다. 그렇다면 품질특성치가 비대칭분포를 따라 평균과 중앙값이 동일하지 않을 때, 어떤 값을 관리도의 중심선으로 사용하여 관리해야 합리적인지를 판단하여야 한다. Janacek와 Meikle(1997)는 평균보다 중앙값을 사용할 경우 장점을 아래와 같이 3가지로 밝히고 있다.

- 1 중앙값은 평균에 비해 자료수집 시 발생하는 측정치나 기록오차에 덜 민감하다.
- 2 중앙값은 절대적 측정치일 뿐만 아니라 상대적 측정치이므로 정확하게 측정할 수 없는 특징을 가지고 있는 자료에 적합하며, 또한 순서를 이용하면 주관적 평가나 순서형 자료도 관리도에 적용시킬 수 있다.
- 3 계산에 필요한 노력을 줄일 수 있다. 왜냐하면, 중앙값을 계산하기 위해서 모든 관측치를 정확히 계산할 필요가 없다. 예를 들어, 자료의 개수가 홀수라면 개개의 관측치들의 순서를 파악한 후 가운데 있는 하나의 관측치만 정확히 측정하면 되기 때문이다.

위의 3가지 이유뿐만 아니라, 일반적으로 중앙값은 평균에 비해 이상치에 영향을 적게 받으며, 관리도의 관점에서는 중앙값을 기준으로 위와 아래에 같은 확률로 관측값이 분포되어 있기 때문에, 중심선으로 중앙값을 사용하는 것이 평균을 사용하는 것보다 타당할 것이다.

2.2. 관리한계선

슈하르트 관리도에서 관리상한과 관리하한은 중심선에서 3σ 만큼 동일한 거리에 있다. 이것은 정규분포는 대칭분포로 관측값은 평균을 중심으로 동일한 확률로 분포하고 있으며, 평균에서 동일한 거리의 구간에도 같은 확률로 분포하고 있기 때문에 평균을 중심으로 동일한 거리에서 관리상한과 관리하한을 정하는 것은 자연스러운 것이다. 하지만 비대칭분포는 이러한 두 가지 특징이 동시에 만족하지 않는다. 중심선을 중앙값으로 사용하여 관측값이 분포할 확률이 동일하게 할 수는 있지만, 중심선에서 관리상한, 관리하한과의 거리를 동일하게 두면 관측값이 분포할 확률이 두 구간에서 달라지기 때문에 비대칭의 방향과 정도에 따라 다르게 관리상한과 관리하한을 결정하여야 할 것이다.

기존 연구 Choobinch와 Ballard(1987), Bai와 Choi(1995)는 아래와 같이 중심선(CL)으로 평균을 사용하고 평균보다 높은 값과 낮은 값의 비율을 이용하여 관리상한(UCL)과 관

리하한(LCL)을 다르게 정의함으로써 자료가 가지는 비대칭의 방향과 정도를 해결하였다.

$$\begin{aligned} CL &= \mu \\ UCL &= \mu + 3\sigma\sqrt{2P} \\ LCL &= \mu - 3\sigma\sqrt{2(1-P)} \end{aligned}$$

여기에서 μ, σ 는 확률변수 X 의 평균과 표준편차이며, $P = \Pr(X \leq \mu)$ 이다.

본 연구에서는 중심선을 중앙값으로 사용하므로 비대칭분포에서도 중심선을 기준으로 자료가 분포할 확률은 동일하고, 범위의 문제는 자료의 비대칭의 방향과 정도에 의존하는 통계량을 찾고, 그것에 근거하여 관리한계선을 결정하는 방법을 연구하였다.

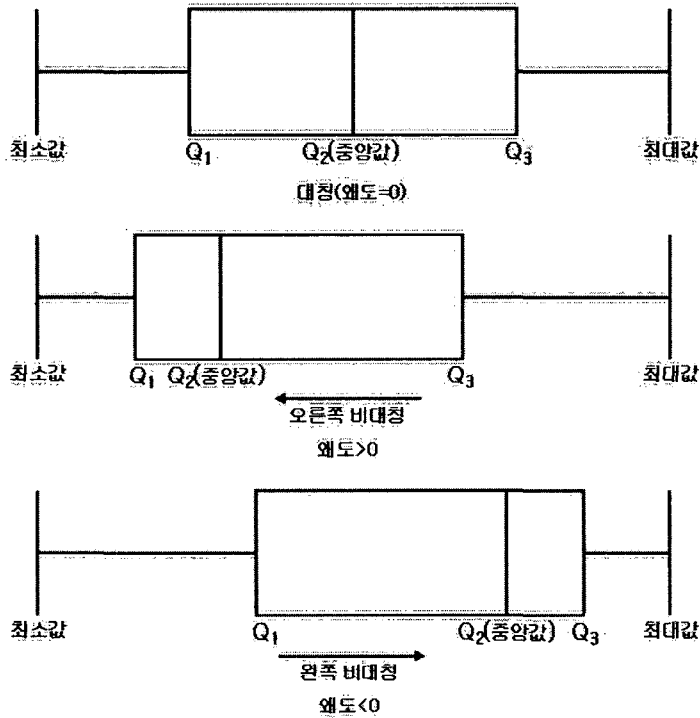


그림 2.1: 기대되는 상자그림(Expected Box-plot)

이를 위해 그림 2.1의 기대되는 상자그림(Expected Box-plot)의 모양을 통해서 접근하도록 하겠다. 각 통계량은 다음과 같다.

$$Q_1 = Q(0.25), \quad Q_2 = Q(0.50), \quad Q_3 = Q(0.75)$$

여기에서

$$\begin{aligned} Q(p) &= F^{-1}(p) \\ &= \min\{x : F(x) \geq p\}, \quad 0 \leq p \leq 1 \end{aligned}$$

이며, $F(x)$ 는 분포함수이다. 즉, Q_1, Q_2, Q_3 는 1사분위수, 중앙값, 3사분위수에 해당된다.

대칭인 분포의 경우 중앙값에서 1사분위수(Q_1)까지의 거리와 3사분위수(Q_3)까지의 거리는 동일하다. 마찬가지로 최소값과 최대값까지의 거리도 같다. 하지만 비대칭분포로 바뀌게 되면 이러한 관계는 성립하지 않는다. 오른쪽 비대칭분포의 경우, 중앙값(Q_2)은 1사분위수 방향으로 치우치게 되어 중앙값에서 3사분위수까지의 거리가 1사분위수까지의 거리보다 길고($Q_3 - Q_2 > Q_2 - Q_1$), 전체 자료의 범위에서도 중앙값에서 최대값까지의 거리가 최소값까지의 거리보다 길다. 이것은 비대칭의 정도가 심해질수록 차이도 더 커지게 된다. 왼쪽 비대칭분포의 경우는 반대의 관계가 성립된다($Q_3 - Q_2 < Q_2 - Q_1$). 이러한 사실을 이용하여 관리하한은 중앙값에서 1사분위수까지의 거리, 관리상한은 중앙값에서 3사분위수까지의 거리에 기초하여 결정하면 간단하면서도 비대칭분포의 방향과 정도의 문제를 해결할 수 있다.

따라서, 사분위수에 기초한 관리도의 중심선과 관리한계선을 다음과 같이 제안한다.

$$\begin{aligned} CL &= Q_2 \\ UCL &= Q_2 + c(Q_3 - Q_2) \\ LCL &= Q_2 - c(Q_2 - Q_1) \end{aligned}$$

여기에서, c 는 적당한 상수이다.

이것을 정규분포일 때 슈하르트 관리도와 비교해보면, 중심선은 중앙값으로 평균과 일치하고, 관리상한과 관리하한을 결정하는 중앙값에서 1사분위수와 3사분위수까지의 거리가 동일하므로 상수 c 값만 조절하면 슈하르트 관리도와 동일한 관리도가 된다. 이 때의 c 값은 아래의 식을 만족하는 c 값으로 수치적인 방법으로 찾아본 결과 $c \simeq 4.45$ 이다.

$$\Pr[\mu - 3\sigma \leq X \leq \mu + \sigma] = \Pr[Q_2 - c(Q_2 - Q_1) \leq X \leq Q_2 + c(Q_3 - Q_2)]$$

2.3. X관리도의 중심선과 관리한계선 추정

실제 관리도를 사용하기 위해서는 표본에서 각 통계량을 추정한 후 중심선과 관리한계선을 결정해야 한다. 본 연구에서는 슈하르트 X관리도에 대응되는 관리도에 관심이 있으므로 부분군의 크기가 1인 경우를 먼저 알아보고, 부분군의 크기가 2이상인 경우에는 어떻게 확장되는지 살펴보겠다.

품질특성치 $x_i, i = 1, 2, \dots, n$ 는 i 번째 시점에서 관측된 값이라고 하자. 중심선인 중앙값은 다음과 같이 정의할 수 있다.

$$\hat{Q}_2 = \text{median}_i(x_i), \quad i = 1, 2, \dots, n$$

관리한계선을 결정하는 Q_1, Q_3 의 추정치는 Parzen(1979)가 제시한 표본백분위수를 사용한다.

$$\hat{Q}_1 = \tilde{Q}(0.25), \quad \hat{Q}_3 = \tilde{Q}(0.75)$$

Parzen(1979)의 표본백분위수함수는 아래와 같은 방법으로 구할 수 있다.

$$\begin{aligned} \tilde{Q}(u) &= n \left(\frac{2i+1}{2n} - u \right) X(i;n) + n \left(u - \frac{2i-1}{2n} \right) X(i+1;n) \\ \frac{2i-1}{2n} &\leq u \leq \frac{2i+1}{2n}, \quad i = 1, 2, \dots, n-1 \end{aligned}$$

여기에서 $X(1;n) \leq X(2;n) \leq \dots \leq X(n;n)$ 은 순서통계량이며, $\tilde{Q}(u)$ 는 $u < \frac{1}{2n}$ 또는 $u > 1 - \frac{1}{2n}$ 일 때는 정의되지 않는다. Parzen(1979)의 표본백분위수 추정방법을 이용한 이유는 소표본에서 효율적으로 알려져 있어 관리도에서 사용할 수 있는 적당한 방법이기 때문이다.(Grimshaw와 Alt, 1997)

따라서, 표본에서 구한 통계량을 이용하여 본 연구에서 제안한 사분위수 관리도의 중심선과 관리한계선을 나타내면 다음과 같다.

$$\begin{aligned} CL &= \hat{Q}_2 \\ UCL &= \hat{Q}_2 + c(\hat{Q}_3 - \hat{Q}_2) \\ LCL &= \hat{Q}_2 - c(\hat{Q}_2 - \hat{Q}_1) \end{aligned}$$

여기에서, c 는 적당한 상수이다.

2.4. \bar{X} 관리도의 중심선과 관리한계선 추정

본 연구에서 제안한 관리도를 부분군의 크기가 2이상으로 슈하르트 \bar{X} 관리도에 대응되는 경우, 다음과 같은 방법으로 중심선과 관리한계선에 사용하는 Q_1, Q_2 와 Q_3 를 추정한다.

품질특성치 x_{ij} , $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$ 는 i 번째 부분군의 j 번째 관측치를 나타내고, 표본은 n 크기의 부분군이 k 개로 구성되어 있다고 가정하자.

중심선인 중앙값의 추정치는 각 부분군에서 구한 중앙값들의 중앙값을 사용한다. 즉, 아래와 같이 정의할 수 있다.

$$\hat{Q}_2 = \text{median}_i \{ \text{median}_j (x_{ij}) \}$$

관리한계선을 결정하는 Q_1, Q_3 의 추정치는 Parzen(1979)가 제시한 표본백분위수를 이용하여 각 부분군에서 구한 1, 3 사분위수의 평균을 사용한다.

$$\hat{Q}_1 = \frac{1}{k} \sum_{i=1}^k \tilde{Q}_i(0.25), \quad \hat{Q}_3 = \frac{1}{k} \sum_{i=1}^k \tilde{Q}_i(0.75)$$

이와 같은 방법으로 구한 추정치를 사용하여 중심선과 관리한계선을 결정하면 부분군이 2인 경우에도 쉽게 확장가능하다.

표 3.1: 화학공정 자료

부분군	1	2	3	4	5	6	7	8	9	10
침전물1	5	10	35	51	11	8	34	47	3	8
침전물2	29	30	13	25	1	12	4	3	36	22
침전물3	2	3	4	5	10	54	38	110	6	24
침전물4	21	68	17	10	9	16	32	11	27	78
침전물5	65	11	14	4	4	14	9	45	10	17
부분군	11	12	13	14	15	16	17	18	19	20
침전물1	13	4	5	12	56	7	12	6	4	62
침전물2	13	18	5	10	15	37	4	28	4	4
침전물3	8	11	16	24	2	9	37	4	20	5
침전물4	9	2	6	12	17	8	7	93	11	3
침전물5	25	19	10	2	15	5	20	21	4	8
부분군	21	22	23	24	25	26	27	28	29	30
침전물1	4	36	4	26	7	6	6	11	9	6
침전물2	15	31	4	14	2	3	18	11	41	6
침전물3	6	135	7	22	16	6	2	18	14	49
침전물4	15	10	24	3	6	14	32	33	8	72
침전물5	10	30	3	20	5	12	11	19	13	16

3. 예제

예제로 사용할 자료는 비대칭분포에서 슈하르트 \bar{X} 관리도와 본 연구에서 제시한 관리도를 비교해 볼 목적으로 Cowden(1957)의 책에서 가져온 화학공정에 관한 자료이다. 구체적인 자료는 표 3.1에 나타내었다. 이 자료는 기존의 Bae와 Choi(1995)에서도 같은 목적으로 사용하고 있음을 밝혀둔다.

자료에 대해서 간단하게 설명을 하면, 안정상태에 있는 화학공정에서 나온 5가지 침전물을 연속적으로 30번 랜덤추출한 것으로 0보다 큰 값을 가지며, 오른쪽 비대칭분포를 하고 있다.

먼저, 슈하르트 X관리도와 본 연구에서 제시한 관리도를 비교하기 위해서 침전물1만의 데이터 (5, 10, 35, ..., 6)를 이용하였다.

슈하르트 X관리도의 중심선과 관리한계선은 아래와 같다.

$$\begin{aligned}
 CL &= \bar{X} = 16.93 \\
 UCL &= \bar{X} + 2.66\bar{R} \\
 &= 16.93 + 2.66 \times 17.33 = 63.02 \\
 LCL &= \bar{X} - 2.66\bar{R} \\
 &= 16.93 - 2.66 \times 17.33 = -29.16 \rightarrow 0
 \end{aligned}$$

본 연구에서 제안한 관리도의 중심선과 관리한계선은 아래와 같다.

$$\begin{aligned}
 CL &= \hat{Q}_2 = 8.50 \\
 UCL &= \hat{Q}_2 + c(\hat{Q}_3 - \hat{Q}_2) \\
 &= 8.50 + c(26.00 - 8.50) = 8.50 + 17.50c \\
 LCL &= \hat{Q}_2 - c(\hat{Q}_2 - \hat{Q}_1) \\
 &= 8.50 - c(8.50 - 6.00) = 8.50 - 2.50c
 \end{aligned}$$

여기서 c 를 4.45로 사용하면 관리한계선은 다음과 같다.

$$\begin{aligned}
 UCL &= 8.50 + 4.45 \times 17.50 \\
 &= 86.37 \\
 LCL &= 8.50 - 4.45 \times 2.50 \\
 &= -2.625 \rightarrow 0
 \end{aligned}$$

그 결과를 간단히 설명하면, 두 관리도의 중심선이 서로 다르다. 슈하르트 X관리도에 비해 본 연구에서 제안한 관리도의 중심선이 낮게 나타나며, 슈하르트 X관리도는 중심에서 관리상한과 관리하한의 거리가 같지만, 본 연구에서 제안한 관리도는 중심선에서 관리상한까지의 거리와 관리하한까지의 거리의 가중치가 다름을 알 수 있다.

그림 3.1은 위의 결과를 이용해서 두 관리도를 나타낸 것이다. 개개의 값을 사용했기 때문에 공정형태는 같게 나타난다. 하지만 중심선을 중앙값으로 사용하면 공정이 안정상태일 때 런의 길이와 수의 형태를 보이므로 평균값을 사용했을 때보다 안정상태의 공정을 잘 표현하고 있다고 볼 수 있다. 여기서 사용한 c 값 4.45는 정규분포를 따른다면, 슈하르트 관리도와 같아지는 값이다.

다음으로 부분군의 크기가 2 이상인 경우 슈하르트 \bar{X} 관리도와 본 연구에서 제시한 관리도를 비교하기 위해서 5가지 칩전물을 모두 이용하였다.

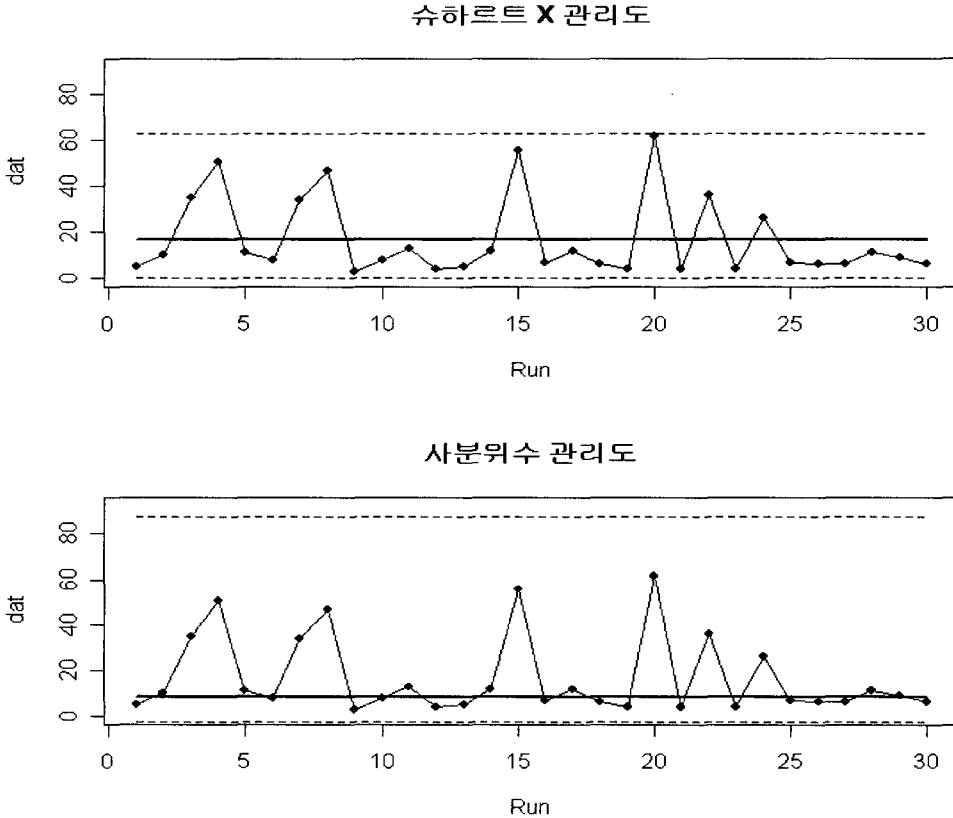


그림 3.1: 슈하르트 X관리도와 대응되는 사분위수 관리도

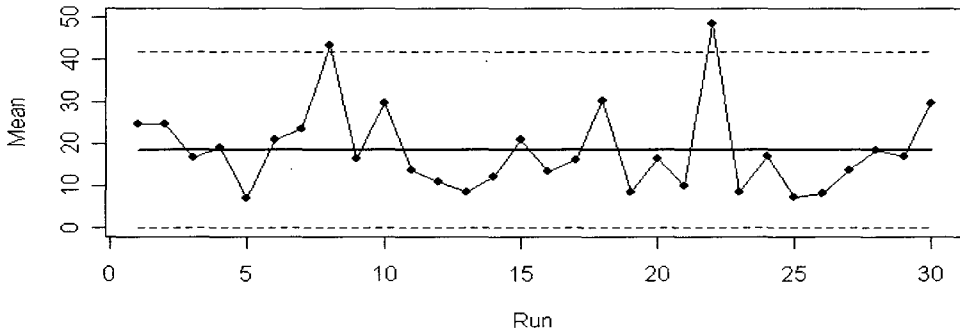
먼저 슈하르트 \bar{X} 관리도의 중심선과 관리한계선은 아래와 같다.

$$\begin{aligned}
 CL &= \bar{\bar{X}} = 18.45 \\
 UCL &= \bar{\bar{X}} + A_2\bar{R} \\
 &= 18.45 + 0.58 \times 40.37 = 41.86 \\
 LCL &= \bar{\bar{X}} - A_2\bar{R} \\
 &= 18.45 - 0.58 \times 40.37 = -4.96 \rightarrow 0
 \end{aligned}$$

본 연구에서 제시한 관리도의 중심선과 관리한계선은 다음과 같다. ($c = 4.45$)

$$\begin{aligned} CL &= \hat{Q}_2 = 12.50 \\ UCL &= \hat{Q}_2 + c(\hat{Q}_3 - \hat{Q}_2) \\ &= 12.50 + 4.45(20.13 - 12.50) \approx 46.45 \\ LCL &= \hat{Q}_2 - c(\hat{Q}_2 - \hat{Q}_1) \\ &= 12.50 - 4.45(12.50 - 8.66) \approx -4.59 \rightarrow 0 \end{aligned}$$

슈하르트 Xbar 관리도



사분위수 관리도

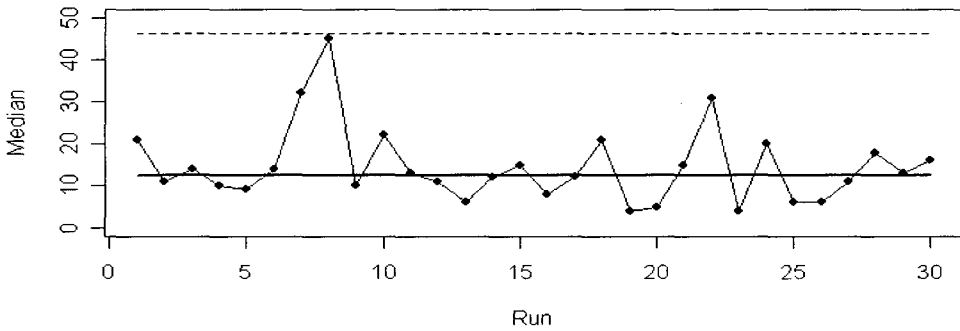


그림 3.2: 슈하르트 \bar{X} 관리도와 대응되는 사분위수 관리도

그 결과를 살펴보면, 두 관리도의 중심선은 다르다. 슈하르트 \bar{X} 관리도는 평균을 관리하고자 하는 것이고, 본 연구에서 제안한 관리도는 중앙값을 관리하고자 하는 것이기 때문에 나는 차이이고, 앞의 결과와 마찬가지로 슈하르트 \bar{X} 관리도는 중심에서 관리상한과 관리하한의 거리가 같지만, 본 연구에서 제안한 관리도는 중심선에서 관리상한까지의 거리와 관리하한까지의 거리가 다를 수 있다.

그림 3.2는 위의 결과를 이용해서 두 관리도를 나타낸 것이다. 각각 평균과 중앙값을 사용했기 때문에 공정형태가 다르게 나타나며, 중앙값을 사용했을 때가 좀 더 공정이 안정적임을 알 수 있다. 특히, 슈하르트 \bar{X} 관리도의 경우 8, 22번째 공정이 안정상태임에도 불구하고, 이상원인에 의한 이상상태임을 나타내고 있다. 하지만, 본 연구에서 제안된 관리도의 경우에는 이러한 오류를 범하고 있지 않다. 이 예제를 통해서 비대칭분포의 경우 본 연구에서 제안한 관리도가 슈하르트 \bar{X} 관리도에 비해 효율적임을 알 수 있다.

4. 제 1종 오류와 c 의 결정

본 연구에서 제안한 관리도를 실제 자료에 적용하기 위해서는 c 값을 결정해야하는 문제가 남아 있다. c 값을 결정하는 기준으로는 본 연구에서 제시한 관리도가 정규분포를 따르는 자료에서 슈하르트 관리도가 갖는 제 1종 오류와 같은 값을 갖게 하는 것이 타당할 것이다.

만약, 자료의 분포를 알고 있다면 자료를 통해 모수를 추정할 수 있고 이를 토대로 제 1종 오류를 구할 수 있으므로 c 값을 쉽게 결정할 수 있을 것이다. 하지만, 실제자료의 분포를 정확히 알거나 추론할 수 있는 경우는 많지 않으므로 c 값을 결정하는 데는 어려움이 따른다. 따라서 문제를 해결하기 위한 하나의 방법으로 자료의 비대칭정도에 따른 c 값 변화의 형태를 살펴보기로 한다.

표 4.1: 제 1종 오류

분포	모수	슈하르트 관 리 도	사분위수 관리도					
			$c = 4.0$	$c = 4.2$	$c = 4.4$	$c = 4.6$	$c = 4.8$	$c = 5.0$
정규	*	0.0027	0.0070	0.0046	0.0030	0.0019	0.0012	0.0007
지수	*	0.0183	0.0313	0.0272	0.0237	0.0206	0.0179	0.0156
카이 제 곱	$\alpha = 1$	0.0125	0.0475	0.0428	0.0387	0.0349	0.0315	0.0285
	$\alpha = 5$	0.0000	0.0195	0.0162	0.0135	0.0112	0.0093	0.0077
	$\alpha = 10$	0.0000	0.0143	0.0115	0.0092	0.0074	0.0059	0.0047
	$\alpha = 20$	0.0000	0.0108	0.0084	0.0065	0.0050	0.0038	0.0029
	$\alpha = 50$	0.0000	0.0085	0.0061	0.0045	0.0034	0.0024	0.0017
와이블	$a = 0.5$	0.0197	0.0822	0.0776	0.0734	0.0695	0.0659	0.0625
	$a = 1$	0.0183	0.0313	0.0272	0.0237	0.0206	0.0179	0.0156
	$a = 2$	0.0056	0.0075	0.0055	0.0040	0.0029	0.0021	0.0014
	$a = 5$	0.0021	0.0047	0.0028	0.0016	0.0009	0.0004	0.0002
	$a = 10$	0.0069	0.0115	0.0089	0.0068	0.0052	0.0039	0.0029

* : 모수에 관계없이 동일한 값

** 제 1종 오류 : $1 - \Pr(LCL \leq X \leq UCL)$

먼저, 정규분포, 지수분포, 카이제곱분포와 와이블 분포를 이용하여 대칭분포와 다양한 비대칭분포의 형태를 표현할 수 있는 모수의 값을 정하였다. 각 모수에 대하여 슈하르트 관리도와 c 값을 4.0에서 5.0까지 변화시켜가면서 본 연구에서 제안한 관리도의 제 1종 오류를 계산하였다. 표 4.1, 그림 4.1, 그림 4.2와 그림 4.3은 그 결과를 나타낸 것이다. 각 분포에서 제 1종 오류는 척도(scale)모수와 위치(location)모수에는 무관하며 형태(shape)모수에 따라 서로 달라진다.

정규분포, 지수분포

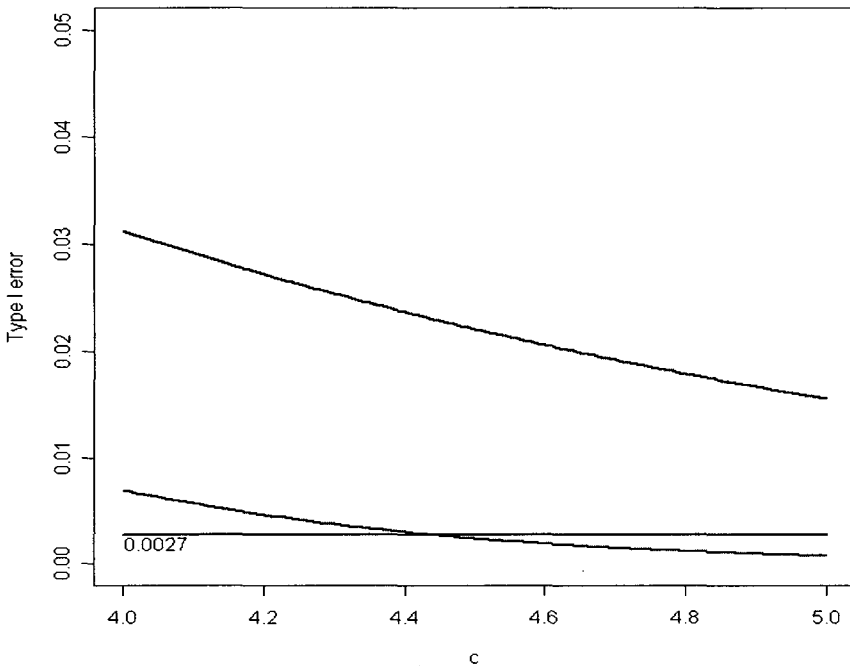


그림 4.1: 정규분포, 지수분포에서 사분위수 관리도의 제 1종 오류

* 위에서부터 지수분포, 정규분포에 해당

** 기준선 0.0027은 정규분포 가정시 슈하르트 관리도의 제 1종 오류에 해당(그림 4.2-4.3에도 해당)

대칭분포인 정규분포의 경우 앞에서 언급하였듯이 값이 4.45정도일 때, 제 1종 오류를 범할 확률이 슈하르트 관리도와 동일하게 나타나는 것을 알 수 있다. 비대칭분포에서는 지수분포를 슈하르트 관리도와 본 연구에서 제안한 관리도에 적용했을 때, 제 1종 오류는 c 값이 4.7정도 일 때 슈하르트 관리도와 비슷한 결과를 보이며, 정규분포일 때의 슈하르트 관리도와 비슷한 효율을 가지기 위해서는 c 값이 5보다 큰 값이 되어야 함을 알 수 있다. 이것은 지수분포의 비대칭정도가 아주 심해서 생기는 결과이다. 마찬가지로 카이제곱 분포와

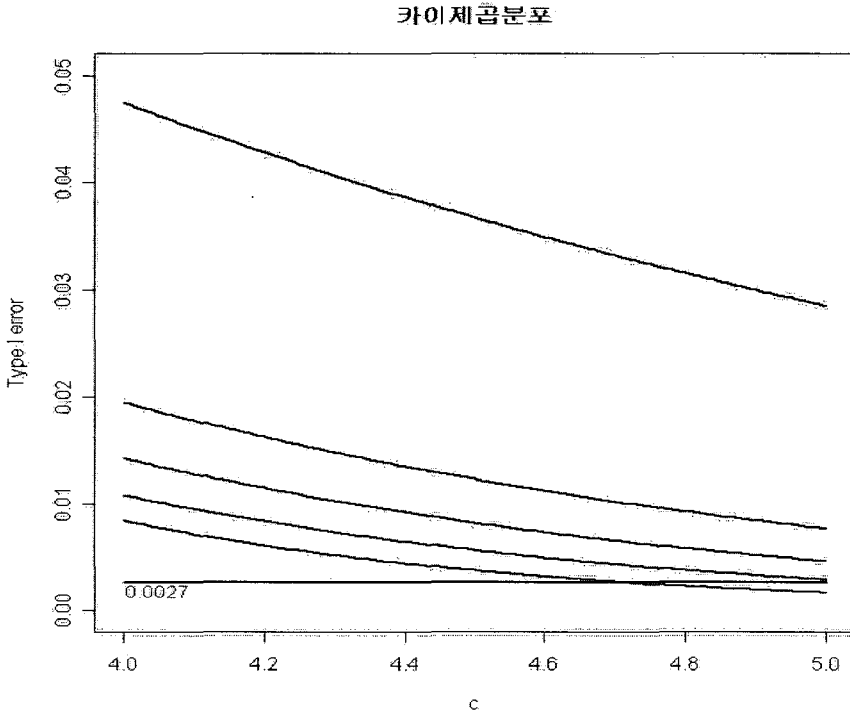


그림 4.2: 카이제곱분포에서 사분위수 관리도의 제 1종 오류

* 위에서부터 카이제곱분포의 형태모수(자유도)가 1, 5, 10, 20, 50에 해당

와이블 분포에서도 형태모수의 값이 적은 경우, 즉 비대칭이 심하면 c 값은 상대적으로 큰 값이 되어야 한다. 그러나 형태모수의 값이 높아 비대칭이 심하지 않을 때에는 c 값이 상대적으로 작아야 정규분포에서의 슈하르트 관리도와 비슷한 효율을 가지게 됨을 알 수 있다. 하지만 위의 결과에서 비대칭정도가 아주 심한 극단적인 경우를 제외하면 분포에 관계없이 4.5정도에서 c 값을 결정할 수 있다. 이것은 정규분포에서 슈하르트 관리도의 제 1종 오류와 가까운 값을 가지게 하므로 타당성 있는 값이라고 할 수 있다.

표 4.1에 사용한 각 분포의 확률밀도함수는 다음과 같다.

1 정규분포

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

여기에서 μ 는 위치모수, σ 는 척도모수이다. ($-\infty < \mu < \infty, \sigma > 0$)

2 지수분포

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

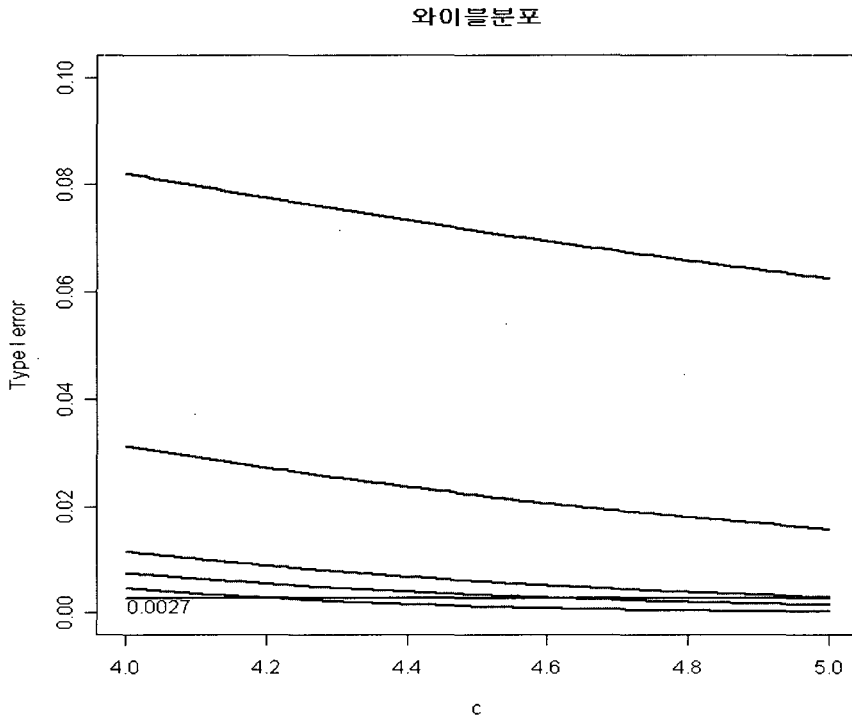


그림 4.3: 와이블분포에서 사분위수 관리도의 제 1종 오류

* 위에서부터 와이블분포의 형태모수가 0.5, 1, 10, 2, 5에 해당

여기에서 λ 는 척도모수이다. ($\lambda > 0$)

3 카이제곱분포

$$f(x) = \frac{1}{\Gamma(\alpha/2)2^{\alpha/2}} x^{\alpha/2-1} e^{-x/2}, \quad x > 0$$

여기에서 α 는 형태모수이다. ($\alpha > 0$)

4 와이블분포

$$f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a}, \quad x > 0$$

여기에서 a 는 형태모수, b 는 척도모수이다.

5. 결론

본 연구에서는 분포를 알 수 없는 비대칭인 자료에 쉽고 간단하게 적용할 수 있는 관리

도를 제안하였다. 이 관리도의 특징은 중심선은 중앙값으로 두었고 관리상한과 관리하한은 중앙값에서 3사분위수, 1사분위수의 거리에 비례하게 결정함으로써 자연스럽게 자료의 비대칭을 고려하는 특징을 가지고 있다. 관리도의 효율을 높이기 위해서 관리상한과 관리하한의 비례상수는 c 자료의 비대칭정도에 따라 결정해야하지만 여러 비대칭분포에서 조사한 결과 4.5가 적당한 것으로 판명되었다.

참고문헌

- 박성현, 박영현, 이명주 (1997). <통계적 공정관리>, 민영사.
 박성현, 박영현 (2003). <통계적 품질관리>, 개정판, 민영사.
 박창순 (1999). <통계적 품질관리>, 을곡출판사.
 Abel, V. (1989). Comment on : Control-Limits of QC charts for skewed distributions using weighted-variance, *IEEE Transactions on Reliability*, **38**, 265.
 Bai, D. S. and Choi, I. S. (1995). \bar{X} and \bar{R} control charts for skewed populations, *Journal of Quality Technology*, **27**, 120-131.
 Choobinch, F. and Ballard, J. L. (1987). Control-Limits of QC charts for skewed distributions using weighted-variance, *IEEE Transactions on Reliability*, **36**, 473-477.
 Cowden, D. J. (1957). *Statistical Methods in Quality Control*, Prentice-Hall, Englewood Cliffs.
 Grimshaw, Scott J. and Alt, F. B. (1997). Control charts for quantile function values, *Journal of Quality Technology*, **29**, 1-7.
 Janacek, G. J. and Meikle, S. E. (1997). Control charts based on medians, *The Statistician*, **46**, 19-31.
 Lewis, L. S., Montgomery, C.D. and Myers, H. R. (2001). Examples of Designed Experiments With Nonnormal Response, *Journal of Quality Technology*, **33**, 265-278.
 Parzen, E. (1979). Nonparametric statistical data modeling, *Journal of the American Statistical Association*, **74**, 105-121.
 Ryan, P. T. (2000). *Statistical Methods for Quality Improvement*, Second Edition, John Wiley and Sons, New York.

[2005년 7월 접수, 2005년 9월 채택]

A Control Chart Method Using Quartiles for Asymmetric Distributed Processes*

Sung Hyun Park¹⁾ Hee Jin Park²⁾

ABSTRACT

This paper proposes a simple control chart method which can be practically used for asymmetric process data where the distribution is unknown. If we use the Shewhart type control charts which are based on normality assumption for the asymmetric process data, the type I error could increase as the asymmetry increases and the effectiveness of control chart to control variation decreases. To solve such problems, this paper suggests to calculate the control limits based on the quartiles. If we obtain the control limits by such quartile method, the type I error could decrease and it looks much more practical for asymmetric distributed process data.

Keywords: Control chart, Asymmetric distribution, Quartile, Box plot, Type I error.

* This work was partially supported by a grant (R01-2003-000-10220-0) from the Korean Science and Engineering Foundation.

1) Professor, Department of Statistics, Seoul National University, San 56-1 Sillim-dong, Gwanak-gu, Seoul 151-747, Korea

E-mail : parksh@plaza.snu.ac.kr

2) Graduate student, Department of Statistics, Seoul National University, San 56-1 Sillim-dong, Gwanak-gu, Seoul 151-747, Korea

E-mail : jini0306@snu.ac.kr