

와이블분포하에서의 최소값 및 최대값 관리도의 설계

조은경¹ · 이민구^{2*}

¹통계청 / ²충남대학교 정보통계학과

Design of Minimum and Maximum Control Charts under Weibull Distribution

Eun-Kyung Jo¹ · Minkoo Lee²

¹Statistics Korea

²Department of Information and Statistics, Chungnam National University

Statistical process control techniques have been greatly implemented in industries for improving product quality and saving production costs. As a primary tool among these techniques, control charts are widely used to detect the occurrence of assignable causes. In most works on the control charts it considered the problem of monitoring the mean and variance, and the quality characteristic of interest is normally distributed. In some situations monitoring of the minimum and maximum values is more important and the quality characteristic of interest is the Weibull distribution rather than a normal distribution. In this paper, we consider the statistical design of minimum and maximum control charts when the distribution of the quality characteristic of interest is Weibull. The proposed minimum and maximum control charts are applied to the wind data. The results of the application show that the proposed method is more effective than traditional methods.

Keywords: Weibull distribution, Minimum control chart, Maximum control chart

1. 서론

품질관리의 중요한 목적 중의 하나는 공정에 이상요인이 발생하였을 경우 가능한 빨리 이를 탐지하여 수정조치를 취함으로써 불량제품의 발생을 사전에 억제하는 것이다. 이러한 목적으로 가장 널리 사용되는 대표적인 기법 중의 하나가 관리도이다. 관리도는 품질의 변동 상황을 그래프로 나타낸 것으로, 평균, 산포, 불량률 등과 같이 공정에 대한 정보들을 추정하거나 공정능력을 결정하는데 이용될 수 있고, 공정 개선을 위한 유의한 정보를 제공하기도 한다.

관리도는 1920년대 미국 Bell 연구소의 Shewhart에 의해 최초로 제안된 이후 많은 연구와 발전을 거듭하여 통계적 공정관리에서 중요한 도구로 사용되고 있다. 관리도는 관리 상태에서의 타점되는 통계량의 기댓값을 나타내는 중심선(central line; CL)

과 관리상한선(upper control limit; UCL) 및 관리하한선(lower control limit; LCL)으로 구성된다. 관리한계선의 폭은 계산상의 편의를 고려하여 타점되는 통계량의 표준편차에 3배수를 사용하며, 이를 3 σ 관리한계선이라 부른다. 관리한계선을 결정하는 또 다른 방법은 공정이 관리 상태에 있을 때 관리도상의 한 점(타점되는 통계량의 값)이 관리한계선을 벗어날 확률(유의 수준)을 지정하는 것이다. 이러한 방법에 의해 구한 관리한계선을 확률한계선이라 한다. 관리하고자 하는 품질 특성치가 관리 상태일 때 관리도에 타점되는 통계량을 W 라 하면 $P\{W \geq UCL\} = P\{W \leq LCL\} = \alpha$ 를 만족하는 UCL, LCL을 α 확률한계선이라 하며 α 값으로는 보통 0.001을 사용한다.

관리도의 대표적인 형태로서 Shewhart 관리도, Page(1954)가 제안한 누적합(cumulative sum; CUSUM) 관리도, Roberts(1959)에 의해 시작된 지수가중이동평균(exponentially weighted mov-

* 연락처: 이민구 교수, 34134 대전광역시 유성구 대학로 99 충남대학교 정보통계학과 Tel : 042-821-5409, Fax : 042-822-0260,
E-mail : sixsigma@cnu.ac.kr

2015년 2월 27일 접수; 2015년 7월 7일 수정본 접수; 2015년 8월 21일 게재 확정.

ing average; EWMA) 관리도 등이 있다. Shewhart 관리도는 현재의 데이터에 중점을 두고 공정의 이상상태 여부를 판단하는 반면 CUSUM 관리도는 현재와 과거의 모든 데이터에 동일한 가중치를 주고 공정의 이상상태 여부를 판단한다. EWMA 관리도는 데이터에 주는 가중치 면에서 CUSUM 관리도를 수정한 것이다. EWMA는 시계열의 한 모형으로서 방해요인(disturbance)이 있는 공정을 표현하는데 사용될 뿐만 아니라, 다음 관측값을 예측하고 이 예측값을 이용하여 공정을 조정하는데 널리 이용되고 있다(Box and Kramer, 1992). 또한 관리도는 하나의 품질특성치를 관리하는 일변량 관리도와 여러 개의 품질특성치를 동시에 고려하는 다변량 관리도로 구분할 수 있으며 Hotelling(1947)이 제안한 Hotelling's T^2 관리도는 대표적인 다변량 관리도이다. 데이터가 다변량 정규분포를 따른다는 가정하에 T^2 통계량은 F분포를 따르게 되고, 이를 바탕으로 관리한계선을 설정하여 공정관리를 한다. 관리도의 설계모수인 부분군의 크기와 부분군의 채취 간격 및 관리한계선을 결정하는데 있어서 관련 비용을 최소화하는 경제적 설계모형은 Duncan(1956)에 의해 처음 소개된 이후 Taylor(1968), Jones and Case(1981), Montgomery(2004), Lee *et al.*(2013), Chang(2015), Kim *et al.*(2015) 등 많은 연구가 진행되어 왔다.

지금까지 연구된 계량형 관리도의 대부분은 평균이나 산포를 관리하는데 사용하는 \bar{X} 관리도 및 R 관리도에 초점을 맞추고 있다. 그러나 경제적이거나 물리적인 제한 등으로 모든 데이터를 관측하지 못하고 최소값이나 최대값만을 얻는 경우나 관심 있는 품질특성치의 최소값이나 최대값의 관리가 더 중요한 경우가 있다. 예를 들어 합성수지를 생산하는 공정에서는 비용이나 시간적인 제약 등으로 생산된 수지의 인장강도와 분자량의 최대값과 최소값만을 측정하고 있다. 또한 기상 정보 같은 경우는 풍속, 강수량, 파고 등과 같은 특성치의 최대값이나 최소값의 관리가 더 중요하다. Howell(1949)은 품질특성치가 정규분포를 따를 때 표본의 최대값과 최소값을 하나의 관리도에 타점하는 Shewhart 형태의 관리도를 제안하였다. 또한 Amin and Wolff(1995)는 표본의 관측 값들의 절대값 중 가장 큰 값을 사용하는 EXTV(extreme value) EWMA 관리도를 제안하였다. 앞에서 연구한 관리도는 품질 특성치의 분포로 정규분포를 가정하고 있다. 그러나 기상정보의 특성치나 합성수지의 인장강도나 분자량 등과 같은 경우는 최소값과 최대값의 관리가 중요하며 이들 데이터의 분포는 대부분 와이블분포를 따른다. 따라서 기존의 정규분포를 가정하고 설계한 Shewhart 형태의 최소값 및 최대값 관리도는 이런 상황에서 더 이상 관리도의 사용목적을 충족시키지 못한다. 따라서 본 논문에서는 품질특성치가 와이블분포를 따를 때 최소값 및 최대값을 관리할 수 있는 관리도를 개발하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 와이블분포에서의 최소값 관리도를 설계하고, 제 3장에서는 와이블분포하에서의 최대값 관리도를 설계한다. 그리고 제 4장에서는 제 2장과 제 3장에서 설계한 최소값 관리도 및 최대값 관리도에 대한 풍속데이터 사례를 다룬다.

2. 최소값 관리도의 설계

관리도상에 타점되는 통계량에 따라서 여러 가지 종류의 관리도를 설계할 수 있다. 지금까지 개발된 계량형 관리도는 대부분 평균이나 산포를 관리하는 관리도들이다. 그러나 풍속이나 강수량, 파고 등과 같은 기상 정보나 합성수지의 인장강도나 분자량을 모니터링 하는 문제에서는 최소값이나 최대값의 관리가 매우 중요하다. 예를 들어 풍속, 강수량, 파고는 계량형 데이터이며 확률분포는 대부분 와이블분포를 따른다. 이 중에서 풍속은 풍력발전소를 운영하는데 중요한 지표가 되며 건물의 안전진단평가 기준으로 사용될 수 있다. 또한 강수량은 홍수나 가뭄을 관리하는데 이용하고, 파고는 기상특보를 발령하는데 이용된다. 따라서 풍속이나 강수량, 파고 등 기상자료는 평균보다 최소값 및 최대값을 관리하는 것이 더 효과적이다. 그러나 지금까지 품질 특성치가 와이블분포를 따르는 경우 최대값 및 최소값을 이용하는 관리도의 개발은 이루어지지 않아 이를 활용하기 위한 최소값 관리도 및 최대값 관리도의 개발이 시급한 상황이다. 따라서 본 장에서는 품질 특성치가 와이블분포를 따르는 경우 최소값 관리도의 설계문제를 다루고자 한다.

2.1 확률한계선을 이용한 최소값 관리도의 설계

관리도를 설계하기 위해서는 부분군의 크기(sample size), 부분군의 채취 간격(sampling interval), 중심선(center line), 관리한계선(control limits)을 결정해야한다. 이를 관리도의 설계모수라 하는데, 관리한계선은 관리도를 설계하는데 중요한 설계모수이며, 이를 결정하는 방법은 관리도가 추구하는 목적에 따라 경제적 설계와 통계적 설계로 구분된다. 통계적 설계방법중의 하나인 확률한계선을 이용한 관리한계선은 통계량이 따르는 확률분포가 치우침이 있는 경우, 관리도의 오경보율을 만족하는 관리도를 설계하는 용도로 사용한다. 정규분포와 같이 좌우 대칭인 분포의 경우는 일반적으로 3σ 관리한계선을 이용하며, 와이블분포와 같이 비대칭인 분포의 경우는 확률한계선을 이용하는 것이 효과적인 방법이다. 본 연구에서는 품질 특성치가 와이블분포를 따르므로 관리한계선으로 확률한계선을 이용한다.

와이블분포는 연속 확률 분포이며, 극치분포(extreme value distribution)중의 하나로서 Fisher and Tippett(1928)에 의해 제안되었고, 그 후 스웨덴의 물리학자 Weibull(1939)이 재료의 파괴강도를 분석하면서 고안한 확률분포로 와이블분포가 수명 분포 모형으로서 등장하는 계기가 되었다. 와이블분포는 유연하기 때문에 풍속, 강수량, 파고 등과 같은 날씨 관련 데이터 분석과 수명 데이터 분석에 자주 쓰이는데 정규분포나 지수분포 같은 다른 통계적인 분포를 묘사할 수도 있다.

품질특성치 X 가 척도모수(scale parameter)가 θ 이고, 형상모수(shape parameter)가 m 인 와이블분포를 따르면 확률밀도 함수는

$$f(x; \theta, m) = \begin{cases} \frac{m}{\theta} \left(\frac{x}{\theta}\right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

이 되고 최소값 $X_{(1)}$ 의 확률밀도함수는

$$\begin{aligned} f_1(x) &= n[1 - F(x)]^{n-1} f(x) \\ &= n \left(e^{-\left(\frac{x}{\theta}\right)^m} \right)^{n-1} \left(\frac{m}{\theta} \right) \left(\frac{x}{\theta} \right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m} \\ &= \frac{m}{\theta n^{\frac{1}{m}}} \left(\frac{x}{\theta n^{\frac{1}{m}}} \right)^{m-1} e^{-\left(\frac{x}{\theta n^{\frac{1}{m}}}\right)^m} \end{aligned} \quad (2)$$

이 된다. 여기서 $F(x)$ 는 X 의 누적분포함수이다. 확률밀도함수 $f_1(x)$ 를 살펴보면 와이블분포의 확률밀도함수임을 알 수 있다. 즉, 품질특성치가 와이블분포를 따르는 모집단에서 채취된 부분군의 최소값의 분포는 척도모수(scale parameter)가 $\theta n^{-\frac{1}{m}}$ 이고, 형상모수(shape parameter)가 m 인 와이블분포를 따른다. 따라서 최소값 $X_{(1)}$ 의 평균과 분산은 다음과 같다.

$$E(X_{(1)}) = \theta n^{-\frac{1}{m}} \Gamma\left(1 + \frac{1}{m}\right) \quad (3)$$

$$Var(X_{(1)}) = (\theta n^{-\frac{1}{m}})^2 \left(\Gamma\left(1 + \frac{2}{m}\right) - \Gamma^2\left(1 + \frac{1}{m}\right) \right) \quad (4)$$

최소값 관리도의 중심선 및 관리한계선은 다음의 관계식으로 구할 수 있다.

$$\Pr(X_{(1)} \leq LCL) = \int_0^{LCL} f_1(x) dx = \alpha \quad (5)$$

$$CL = \int_0^{\infty} x f_1(x) dx = E(X_{(1)}) = \theta n^{-\frac{1}{m}} \Gamma\left(1 + \frac{1}{m}\right) \quad (6)$$

$$\Pr(X_{(1)} \geq UCL) = \int_{UCL}^{\infty} f_1(x) dx = \alpha \quad (7)$$

따라서 품질 특성치의 분포가 와이블분포일 때 척도모수와 형상모수를 알면 식 (5)~식 (7)을 이용하여 중심선 및 관리한계선을 구한다.

$\alpha = 0.001$ 이고 부분군의 크기가 각각 3, 4, 5일 때 모수값의 변화에 따른 중심선(CL) 및 관리한계선(LCL, UCL)은 <Table 1>과 같다. 표를 작성하는데 있어 MATLAB 프로그램을 이용하였으며, 최소값이 와이블분포를 따르므로 프로그램 내장함수 'wblinv'를 이용하여 관리한계선을 구하였다.

2.2 최소값 관리도의 ARL

본 절에서는 척도모수와 형상모수의 변화에 따른 ARL(aver-

age run length; 평균 런의 길이)의 변화를 살펴보고자 한다. ARL은 다점 통계량이 관리한계선을 벗어날 때까지 뽑는 부분군 채취회수의 기댓값이다. 이는 설계 모수인 부분군의 크기 및 관리한계선의 값에 영향을 받는다. 공정이 관리상태일 때의 ARL을 ARL_0 라 하고, 이상 상태 일때의 ARL을 ARL_1 이라 하면, ARL_0 와 ARL_1 은 다음과 같다.

$$ARL_0 = \frac{1}{2\alpha}, \quad (8)$$

$$ARL_1 = \frac{1}{1-\beta}, \quad (9)$$

여기서 2α 는 공정이 관리상태일 때 이상 상태라고 판정할 확률, 즉 제1종 오류를 범할 확률이고, β 는 공정이 이상상태일 때 관리상태 라고 판정할 확률, 즉 제2종 오류를 범할 확률이다. 그러므로 효율적인 관리도는 공정이 관리상태일 때의 ARL_0 이 크고, 공정이 이상 상태일 때의 ARL_1 이 작은 값을 가져야 한다.

공정의 변화를 빨리 탐지하기 위해, 제 1종 오류를 범할 확률을 늘리면 제 2종 오류를 범할 확률이 작아지고, 제 2종 오류를 범할 확률을 늘리면 제 1종 오류를 범할 확률이 작아진다. 그러므로 본 논문에서는 ARL_0 를 고정시킨 후, ARL_1 을 비교하는 방법을 사용한다.

제 2.1절에서 제안한 확률한계선($\alpha = 0.001$)을 이용한 최소값 관리도에서 공정의 변화(모수 값들의 변화)를 잘 탐지하는지를 조사하기 위해 $n = 3, \theta = 1, m = 2$ 을 기준으로 한 관리도의 ARL_0 에 고정시킨 후, θ 와 m 을 각각 0.5에서 3까지 0.5씩 증가시켜 모수 변화에 따른 ARL_1 의 변화를 살펴보았다.

$\alpha = 0.001$ 일 때 ARL_0 은 $ARL_0 = \frac{1}{2\alpha}$ 이므로 $ARL_0 = \frac{1}{0.002} = 500$ 이 된다. 형상모수 m 를 2로 고정시킨 후, 척도모수 θ 가 변화함에 따라 ARL_1 의 변화는 θ 가 기준값 $\theta = 1$ 에서 벗어날수록 평균 런(ARL_1)은 감소하는 경향을 보인다. 즉, 공정의 이상상태를 빨리 탐지함을 의미한다. 이는 식 (3)에서 알 수 있듯이 척도모수 θ 가 증가함에 따라 최소값 $X_{(1)}$ 의 평균이 증가하기 때문이다. 공정평균이 증가하면 공정이 안정 상태를 벗어나기 때문에 이를 빨리 탐지함을 나타낸다.

척도모수를 θ 를 1로 고정시킨 후, 형상모수 m 이 변화함에 따라 ARL_1 의 변화는 m 이 기준값 2에서 작아질수록 ARL_1 은 감소하나, m 이 3에 가까워질수록 품질특성치의 분포가 정규분포에 가까워지면서 ARL_1 은 증가한다. 즉, 형상모수가 기준값 2보다 작으면, ARL_1 은 작은 값을 가지나, 형상모수가 3에 가까워지면 ARL_1 은 500보다 큰 값을 갖는다. 이는 와이블분포 형태가 한쪽으로 치우친 모양을 갖고 있으나, $\theta = 1$ 이고 m 의 값이 3근방 일 때는 정규분포와 유사한 형태의 분포를 갖게 되며 데이터들이 평균 근처에 몰려있어 관리한계선 밖으로 타점될 확률이 적어지기 때문이다. 본 논문에서 다루고자 하는

Table 1. CL, LCL, and UCL of Minimum Control Chart for Parameters θ, m, n

θ	m	$n = 3$			$n = 4$			$n = 5$		
		CL	LCL	UCL	CL	LCL	UCL	CL	LCL	UCL
0.5	0.5	0.111	0.000	2.651	0.063	0.000	1.491	0.040	0.000	0.954
0.5	1	0.167	0.000	1.151	0.125	0.000	0.863	0.100	0.000	0.691
0.5	1.5	0.217	0.002	0.872	0.179	0.002	0.720	0.154	0.002	0.620
0.5	2	0.256	0.009	0.759	0.222	0.008	0.657	0.198	0.007	0.588
1	0.5	0.222	0.000	5.302	0.125	0.000	2.982	0.080	0.000	1.909
1	1	0.333	0.000	2.303	0.250	0.000	1.727	0.200	0.000	1.382
1	1.5	0.434	0.005	1.744	0.358	0.004	1.439	0.309	0.003	1.240
1	2	0.512	0.018	1.517	0.443	0.016	1.314	0.396	0.014	1.175
1.5	0.5	0.333	0.000	7.953	0.188	0.000	4.473	0.120	0.000	2.863
1.5	1	0.500	0.001	3.454	0.375	0.000	2.590	0.300	0.000	2.072
1.5	1.5	0.651	0.007	2.616	0.537	0.006	2.159	0.463	0.005	1.861
1.5	2	0.768	0.027	2.276	0.665	0.024	1.971	0.595	0.021	1.763
2	0.5	0.444	0.000	10.604	0.250	0.000	5.965	0.160	0.000	3.817
2	1	0.667	0.001	4.605	0.500	0.001	3.454	0.400	0.000	2.763
2	1.5	0.868	0.010	3.487	0.717	0.008	2.879	0.617	0.007	2.481
2	2	1.023	0.037	3.035	0.886	0.032	2.628	0.793	0.028	2.351
2.5	0.5	0.556	0.000	13.255	0.313	0.000	7.456	0.200	0.000	4.772
2.5	1	0.833	0.001	5.756	0.625	0.001	4.317	0.500	0.001	3.454
2.5	1.5	1.085	0.012	4.359	0.896	0.010	3.599	0.772	0.009	3.101
2.5	2	1.279	0.046	3.794	1.108	0.040	3.285	0.991	0.035	2.938
3	0.5	0.667	0.000	15.906	0.375	0.000	8.947	0.240	0.000	5.726
3	1	1.000	0.001	6.908	0.750	0.001	5.181	0.600	0.001	4.145
3	1.5	1.302	0.014	5.231	1.075	0.012	4.318	0.926	0.010	3.721
3	2	1.535	0.055	4.552	1.329	0.047	3.942	1.189	0.042	3.526

Table 2. ARL_1 for $n = 3, ARL_0 = 500$

ARL_1 as a function of $\theta(m = 2)$				ARL_1 as a function of $m(\theta = 1)$			
θ	mean	ARL_0	ARL_1	m	mean	ARL_0	ARL_1
0.5	0.26	-	250.4	0.5	0.22	-	2.8
1	0.51	500	-	1	0.33	-	15.7
1.5	0.77	-	21.3	1.5	0.43	-	90.5
2	1.02	-	5.6	2	0.51	500	-
2.5	1.28	-	3.0	2.5	0.57	-	2969.2
3	1.53	-	2.2	3	0.62	-	21595.0

품질특성치의 분포는 와이블분포이기 때문에 m 이 3에 가까워지는 경우의 ARL_1 은 의미가 없다.

3. 최대값 관리도의 설계

기상 관련 데이터인 풍속, 강수량 및 파고 등은 계량형 데이터

이면서 확률분포는 와이블분포를 따른다. 예를 들어 풍력 발전을 위해서는 적절한 풍속이 필요하다. 따라서 풍속의 최소값뿐만 아니라 최대값을 모니터링 하는 것이 필요하다. 강수량은 홍수를 대비해 최대 강수량을 모니터링 해야 하고, 파고는 해일의 예방을 위해 파고의 최대값을 관리할 필요가 있다. 또한 합성수지의 분자량 같이 품질 특성치의 최대값을 관리해야 하는 경우가 있다. 이러한 경우 품질 특성치의 평균을 관리

하기 보다는 최대값을 관리하는 것이 필요하다.

3.1 확률한계선을 이용한 최대값 관리도의 설계

확률변수 X 가 와이블분포를 따를 때 순서통계량의 확률밀도함수를 이용하면 최대값 $X_{(n)}$ 의 확률밀도함수는

$$\begin{aligned} f_n(x) &= n[F(x)]^{n-1}f(x) \\ &= n\left(1 - e^{-\left(\frac{x}{\theta}\right)^m}\right)^{n-1} \frac{m}{\theta} \left(\frac{x}{\theta}\right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m} \end{aligned} \quad (10)$$

이 된다. 품질 특성치가 와이블분포를 따를 때 부분군의 최대값 $X_{(n)}$ 의 평균과 분산은 다음과 같다.

$$\begin{aligned} E(X_{(n)}) &= \int_{-\infty}^{\infty} x f_n(x) dx \\ &= \int_{-\infty}^{\infty} x n \left(1 - e^{-\left(\frac{x}{\theta}\right)^m}\right)^{n-1} \frac{m}{\theta} \left(\frac{x}{\theta}\right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m} dx \end{aligned} \quad (11)$$

$$Var(X_{(n)}) = E(X_{(n)}^2) - E(X_{(n)})^2 \quad (12)$$

여기서 $E(X_{(n)}^2) = \int_{-\infty}^{\infty} x^2 f_n(x) dx = \int_{-\infty}^{\infty} x^2 n \left(1 - e^{-\left(\frac{x}{\theta}\right)^m}\right)^{n-1} \frac{m}{\theta} \left(\frac{x}{\theta}\right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m} dx$ 이다.

최대값 관리도의 중심선 및 관리한계선은

$$\Pr(X_{(n)} \leq LCL) = \int_0^{LCL} f_n(x) dx = \alpha \quad (13)$$

$$CL = \int_0^{\infty} x f_n(x) dx = E(X_{(n)}) \quad (14)$$

$$\Pr(X_{(n)} \geq UCL) = \int_{UCL}^{\infty} f_n(x) dx = \alpha \quad (15)$$

의 관계식으로부터 구할 수 있다. 따라서 품질특성치의 분포가 와이블분포일 때 척도모수와 형태모수를 알면 식 (13)~식 (14)를 이용하여 중심선 및 관리한계선을 구한다. 위의 식 (13)~식 (14)를 만족하는 중심선과 관리한계선은 컴퓨터를 활용하여 수치적으로 밖에 구할 수 없다.

$\alpha = 0.001$ 이고 부분군의 크기가 각각 3, 4, 5일 때 모수값 변화에 따른 중심선(CL) 및 관리한계선(LCL, UCL)은 <Table 3>과 같다. <Table 3>을 살펴보면, 같은 θ , m 에 대하여 부분군의 크기 n 이 커질수록 관리한계선의 폭이 좁아지는 현상을 보이지 않고 있다. 위에서 설계된 최대값 관리도는 형상모수 $m \leq 1$ 일 때는 n 이 커질수록 관리한계선의 폭이 넓어지며, 형상모수 $m > 1$ 일 때 n 이 커질수록 관리한계선의 폭이 좁아진

다. 이러한 현상은 수학적으로 증명할 수는 없지만 수치적인 분석을 통해 형상모수 m 과 부분군의 크기 n 의 변화에 따라 최대값의 확률밀도함수가 $m \leq 1$ 일 때는 n 이 클수록 함수 모양이 퍼진 모양을 가지며, $m > 1$ 이면 n 이 커질수록 확률밀도함수는 더 좁아지는 모양을 가짐을 알 수 있었다.

3.2 최대값 관리도의 ARL

$n = 3, \theta = 1, m = 2$ 을 기준으로 한 관리도의 ARL_0 에 고정시킨 후, θ 와 m 을 각각 0.5에서 3까지 0.5씩 증가시켜 모수 변화에 따른 ARL_1 의 변화를 조사하였다. $\alpha = 0.001$ 일 때 $ARL_0 = \frac{1}{2\alpha} = \frac{1}{0.002} = 500$ 이 된다. 형상모수 m 를 2로 고정시킨 후, 척도모수 θ 가 변화함에 따라 ARL_1 의 변화는 θ 가 기준값 $\theta = 1$ 에서 벗어날수록 ARL_1 은 감소하는 경향을 보인다. 이는 공정이 안정상태에서 많이 벗어날수록 공정의 이상 상태를 빨리 탐지함을 의미한다. 척도모수 θ 를 1로 고정시킨 후, 형상모수 m 이 변화함에 따라 ARL_1 의 변화는 m 이 기준값 2에서 작아질수록 ARL_1 은 감소하나, m 이 3에 가까워질수록 품질특성치의 분포가 정규분포에 가까워지면서 ARL_1 은 증가한다. 즉, 형상모수가 기준값 2보다 작으면, ARL_1 은 작은 값을 가지나, 형상모수가 3에 가까워지면 ARL_1 은 500보다 큰 값을 갖는다. 이러한 결과가 나온 이유와 의미는 제 2.2절에서 기술한 내용과 유사하다.

4. 사례연구

본 장에서는 와이블분포를 따르는 풍속데이터를 적용하여 제 2장과 제 3장에서 개발한 최소값 및 최대값 관리도의 적용방법을 소개한다. 기상자료인 풍속, 강우량 및 파고 등은 데이터의 분포 형태가 정규분포 보다는 와이블분포에 가까운 경우가 많이 있다. 본 장에서는 실제 예제로 풍속을 이용하였다. 풍속 데이터는 와이블분포를 따를 뿐만 아니라 풍력발전소를 운영 시 최소값과 최대값이 모두 중요변수가 되어 실제 예제로 선택하게 되었다.

4.1 원시자료

본 자료는 대관령에서 측정한 2011년 3월 1일부터 2012년 3월 31일까지 분별 풍속 데이터를 기상청에서 받아 원시자료로 사용하였다. 기상청의 분별 풍속자료는 매 1분 평균 풍속(0.1m/s) 측정치이다. <Table 5>와 <Table 6>은 원시자료의 일부를 나타낸 것이다. 본 사례에서는 2011년도 3월의 데이터를 이용하여 분포 및 모수를 추정하고 이를 활용하여 관리도를 작성한 다음 2012년 3월의 데이터를 작성된 관리도에 타점해 보았다.

Table 3. CL, LCL, and UCL of Maximum Control Chart for Parameters θ, m, n

θ	m	n = 3			n = 4			n = 5		
		CL	LCL	UCL	CL	LCL	UCL	CL	LCL	UCL
0.5	0.5	2.361	0.006	32.048	2.882	0.019	34.393	3.339	0.042	36.268
0.5	1	0.917	0.053	4.003	1.042	0.098	4.147	1.142	0.145	4.258
0.5	1.5	0.718	0.112	2.001	0.788	0.169	2.049	0.842	0.219	2.085
0.5	2	0.645	0.162	1.415	0.694	0.221	1.440	0.731	0.269	1.459
1	0.5	4.722	0.011	64.097	5.764	0.038	68.785	6.677	0.084	72.536
1	1	1.833	0.105	8.006	2.083	0.196	8.294	2.283	0.289	8.517
1	1.5	1.436	0.223	4.002	1.577	0.337	4.097	1.684	0.437	4.170
1	2	1.290	0.325	2.829	1.389	0.443	2.880	1.462	0.538	2.918
1.5	0.5	7.083	0.017	96.145	8.646	0.058	103.178	10.016	0.126	108.804
1.5	1	2.750	0.158	12.009	3.125	0.294	12.441	3.425	0.434	12.775
1.5	1.5	2.154	0.335	6.003	2.365	0.506	6.146	2.526	0.656	6.256
1.5	2	1.936	0.487	4.244	2.083	0.664	4.320	2.193	0.807	4.378
2	0.5	9.444	0.022	128.193	11.528	0.077	137.570	13.354	0.167	145.072
2	1	3.667	0.211	16.012	4.167	0.392	16.587	4.567	0.579	17.034
2	1.5	2.872	0.446	8.004	3.153	0.674	8.195	3.368	0.875	8.341
2	2	2.581	0.649	5.659	2.777	0.885	5.760	2.924	1.076	5.837
2.5	0.5	11.806	0.028	160.242	14.410	0.096	171.963	16.693	0.209	181.339
2.5	1	4.583	0.263	20.015	5.208	0.490	20.734	5.708	0.723	21.292
2.5	1.5	3.590	0.558	10.005	3.941	0.843	10.243	4.210	1.093	10.426
2.5	2	3.226	0.811	7.074	3.471	1.106	7.200	3.655	1.345	7.296
3	0.5	14.167	0.033	192.290	17.292	0.115	206.355	20.032	0.251	217.607
3	1	5.500	0.316	24.018	6.250	0.587	24.881	6.850	0.868	25.550
3	1.5	4.308	0.669	12.006	4.730	1.012	12.292	5.053	1.312	12.511
3	2	3.871	0.974	8.488	4.166	1.328	8.640	4.386	1.614	8.755

Table 4. ARL_1 for n = 3, $ARL_0 = 500$

ARL_1 as a function of $\theta(m = 2)$			
θ	mean	ARL_0	ARL_1
0.5	0.65	-	24.6
1	1.29	500	-
1.5	1.94	-	12.0
2	2.58	-	2.8
2.5	3.23	-	1.6
3	3.87	-	1.3

ARL_1 as a function of $m(\theta = 1)$			
m	mean	ARL_0	ARL_1
0.5	4.72	-	1.8
1	1.83	-	5.3
1.5	1.44	-	33.0
2	1.29	500	-
2.5	1.22	-	4950.3
3	1.17	-	26312.4

4.2 분포 및 모수추정

풍속의 분포 및 모수의 추정은 다음과 같은 절차로 진행한다.

- 1단계 : 2011년 3월 분별 원시자료에서 하루에 5개의 데이터를 랜덤 추출한다.
- 2단계 : 1단계에서 추출한 데이터를 이용하여 풍속 데이터의 분포를 파악한다.

3단계 : 2단계에서 파악한 분포에 대하여 모수를 추정한다.

다음은 위의 분포 파악 및 모수 추정의 각 단계의 내용을 정리한 것이다.

[1단계] 분 자료에서 1일 5개 데이터를 랜덤 추출한 결과는 <Table 7>과 같다.

Table 5. Wind data from 1 March 2011 to 31 March 2011

Year	Month	Day	Hour	Minute	Wind speed (0.1m/s)
2011	3	1	0	1	5
2011	3	1	0	2	4
⋮	⋮	⋮	⋮	⋮	⋮
2011	3	31	23	59	21
2011	4	1	0	0	18

Table 6. Wind data from 1 March 2012 to 31 March 2012

Year	Month	Day	Hour	Minute	Wind speed (0.1m/s)
2012	3	1	0	1	12
2012	3	1	0	2	12
⋮	⋮	⋮	⋮	⋮	⋮
2012	3	31	23	59	30
2012	4	1	0	0	41

Table 7. Random sampling data from 1 March 2011 to 31 March 2011(n = 5)

Year	Month	Day	Hour	Minute	Wind speed (0.1m/s)
2011	3	1	6	57	3
2011	3	1	7	15	1
2011	3	1	12	49	37
2011	3	1	13	55	21
2011	3	1	16	56	20
2011	3	2	0	21	25
⋮	⋮	⋮	⋮	⋮	⋮
2011	3	30	22	16	20
2011	3	31	0	45	19
2011	3	31	2	10	14
2011	3	31	6	10	8
2011	3	31	7	45	3
2011	3	31	16	11	41

[2단계] 풍속분포를 파악하고자 미니탭을 이용한 결과는 <Figure 1>과 같으며, 자료는 정규분포 보다 와이블분포에 가까움을 알 수 있다.

[3단계] 미니탭을 이용하여 와이블분포의 모수를 추정하면 <Figure 2>와 같다.

2011년 3월 풍속 자료를 미니탭을 이용하여 분석한 결과, 풍속데이터는 자료는 와이블분포를 따르며 형상모수와 척도모수의 최우 추정값은 $\theta = 62.4695$, $m = 2.4123$ 로 각각 추정되었

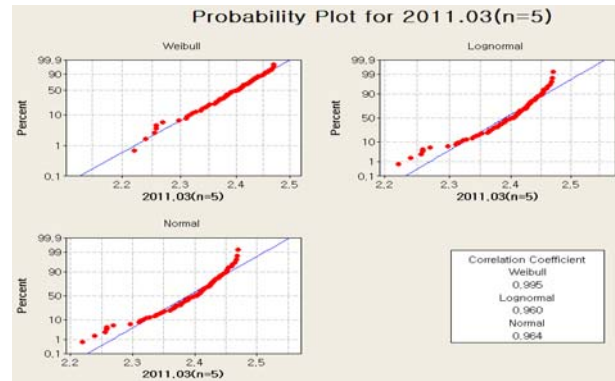


Figure 1. Probability plot for March 2011

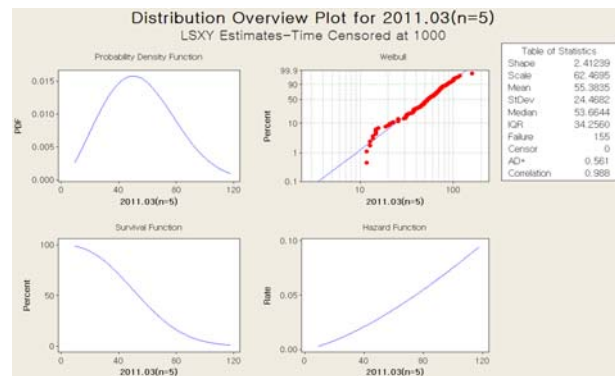


Figure 2. Distribution overview plot for March 2011

다. 본 절에서 추정된 형상모수와 척도모수는 충분한 크기의 표본으로부터 추정했기 때문에 참값이라고 가정한다.

4.3 최소값 및 최대값 관리도 중심선 및 관리한계선

<Table 8>의 데이터는 2012년 3월 1일부터 2012년 3월 31일까지 $n = 5$ 의 랜덤 추출 한 데이터의 최소값 및 최대값을 나타낸 것이다. 제 4.2절에서 추정된 모수를 이용하여 최소값 관리도와 최대값 관리도를 작성하면 다음과 같다. 2011년도 3월 풍속 데이터가 Weibull(62.4695, 2.4123)를 따르므로, 부분군을 5로 하는 최소값의 분포는 Weibull($62.4695 \cdot 5^{-\frac{1}{2.4123}}$, 2.4123)를 따른다. 따라서 최소값 관리도의 중심선 및 관리한계선은 식 (5)~식 (7)를 이용하여 구하면 다음과 같다.

$$LCL = 1.8299$$

$$CL = 28.4209$$

$$UCL = 71.4256$$

위에서 구한 관리도에 2011년 3월 풍속 데이터의 최소값을 타점해본 결과 관리한계선을 벗어나 점이 없었다. 따라서 2011년 3월 풍속의 최소값은 관리상태에 있다. 2011년 3월 풍속 데이터로 작성된 최소값 관리도를 이용하여 2012년 3월 최

Table 8. Random sampling data from 1 March 2012 to 31 March 2012

Year	Month	Day	Minimum wind speed(0.1m/s)	Maximum wind speed(0.1m/s)
2012	3	1	11	60
2012	3	2	25	40
2012	3	3	25	41
2012	3	4	26	41
2012	3	5	42	60
2012	3	6	11	87
2012	3	7	11	57
2012	3	8	13	42
2012	3	9	12	48
2012	3	10	32	61
2012	3	11	81	106
2012	3	12	32	56
2012	3	13	31	61
2012	3	14	76	103
2012	3	15	14	38
2012	3	16	20	40
2012	3	17	13	41
2012	3	18	18	50
2012	3	19	16	68
2012	3	20	14	96
2012	3	21	68	107
2012	3	22	27	98
2012	3	23	13	54
2012	3	24	72	156
2012	3	25	74	93
2012	3	26	22	87
2012	3	27	53	123
2012	3	28	18	51
2012	3	29	24	105
2012	3	30	42	103
2012	3	31	47	76

소값 풍속 데이터를 모니터링 한 결과는 <Figure 3>과 같다.

관리도에서 알 수 있듯이 2012년 3월 11, 14, 24, 25일 데이터가 관리상한선을 벗어났다. 2012년 3월 풍속의 최소값이 2011년 3월 풍속의 최소값에 비해 크을 알 수 있다.

2011년도 3월 자료가 Weibull(62.4695, 2.4123)를 따를 때, 부분군의 크기가 5인 최대값의 확률밀도함수는 다음과 같다.

$$f_n(x) = n[F(x)]^{n-1}f(x) \quad (16)$$

$$= n(1 - e^{-\left(\frac{x}{\theta}\right)^m})^{n-1} \frac{m}{\theta} \left(\frac{x}{\theta}\right)^{m-1} e^{-\left(\frac{x}{\theta}\right)^m}$$

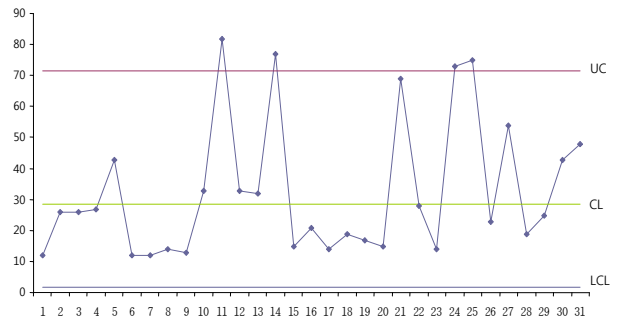


Figure 3. Minimum control chart for March 2012

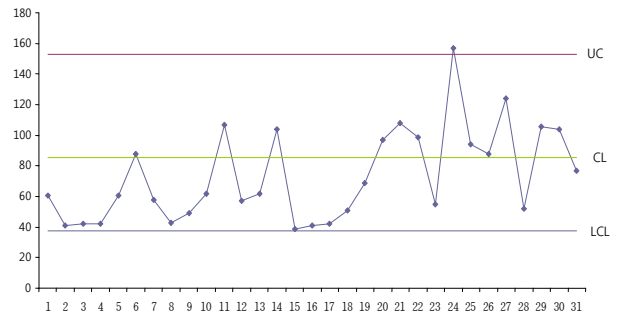


Figure 4. Maximum control chart for March 2012

$$= 5 \left(1 - e^{-\left(\frac{x}{62.4695}\right)^{2.4123}}\right)^{5-1} \frac{2.4123}{62.4695}$$

$$\times \left(\frac{x}{62.4695}\right)^{2.4123-1} e^{-\left(\frac{x}{62.4695}\right)^{2.4123}}$$

최대값 관리도의 중심선 및 관리한계선은 식 (13)~식 (15)을 이용하여 구하면 다음과 같다.

$$LCL = 37.3946$$

$$CL = 85.6378$$

$$UCL = 153.0674$$

2011년 3월 풍속 데이터의 최대값을 위의 관리도에 타점해 본 결과 관리한계선을 벗어나 점이 없었다. 따라서 2011년 3월 풍속의 최대값은 관리상태에 있다. 위와 같이 2011년 3월 풍속 데이터로 작성된 최대값 관리도를 이용하여 2012년 3월 최대값 풍속 데이터를 모니터링 한 결과는 <Figure 4>와 같다.

위의 관리도에서 알 수 있듯이 2012년 3월 24일 데이터가 관리상한선을 벗어났다. 24일의 경우 최대풍속이 관리상한선을 벗어났기 때문에 풍랑주의보를 내리거나 근해에서 조업 중인 어선들에게 정보를 알려주어 안전한 조업을 할 수 있도록 한다.

본 논문에서 제안한 관리도를 Howell(1949)이 제안한 관리도와 비교하면 다음과 같다. Howell은 최대값과 최소값을 하나의 관리도에 타점했지만 본 논문과의 비교를 위해 최대값과 최소값 관리도를 독립적으로 작성하면 <Figure 5>와 <Figure 6>과 같다. 작성된 관리도에서 알 수 있듯이 Howell(1949)이 제

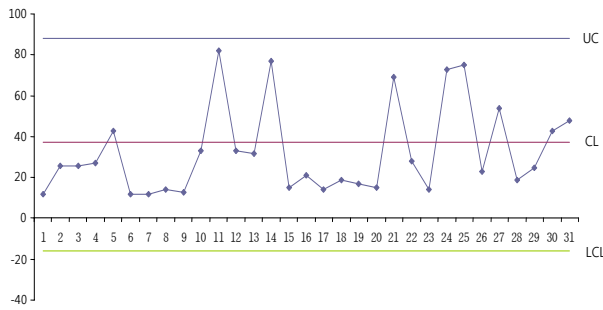


Figure 5. Howell's minimum control chart for March 2012

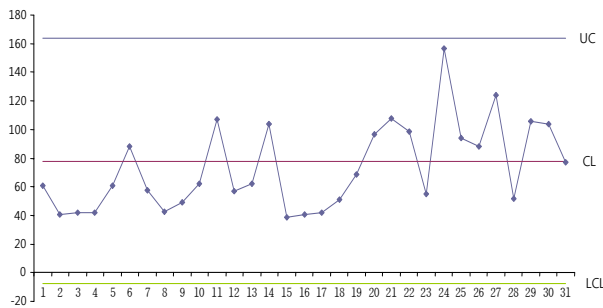


Figure 6. Howell's maximum control chart for March 2012

안한 방법으로 관리도를 작성하면 최소값 관리도이나 최대값 관리도에서 관리한계선을 벗어나는 점이 없음을 알 수 있다. 따라서 위와 같은 상황에서는 본 논문에서 제시한 방법과 차이가 있음을 알 수 있다. 기상청 자료를 살펴보면 2011년 3월 보다는 2012년 3월 풍속이 더 빨랐음을 알 수 있었다. 따라서 본 논문에서 제안한 방법이 Howell(1949)이 제안한 방법보다 합리적임을 알 수 있다. 또한 품질 특성치의 분포가 정규분포인 경우라 하더라도 Howell(1949)이 제시한 방법 보다는 본 논문에서 제시한 방법으로 최소값 및 최대값 관리도를 작성하는 것이 합리적이다.

5. 결론

본 논문에서는 품질 특성치가 정규분포가 아닌 와이블분포를 따를 때, 부분군의 최소값 및 최대값을 사용하여 프로세스를 모니터링 할 수 있는 관리도를 개발하였다. 관리도에 관련된 기존 연구는 대부분 품질 특성치의 분포가 정규분포를 따른다는 가정 하에 품질특성치의 평균이나 산포를 관리하는 것이다. 그러나 제품의 강도, 풍속, 강우량 및 파고와 같은 데이터는 정규분포 보다는 와이블분포를 따른다. 또한 와이블분포를 따르는 데이터의 경우 평균이나 산포보다는 최소값이나 최대값을 관리하는 것이 효과적인 경우가 많이 있다. 따라서 본 연구에서 개발한 최소값 및 최대값 관리도는 이러한 문제들을 극복하는데 많은 도움을 줄 것으로 기대된다.

확률한계선의 개념을 이용하여 최소값 관리도 및 최대값 관리도를 개발하였으며 개발한 관리도의 현장 활용을 높이기

위해 부분군의 크기와 모수값(척도모수와 형상모수)들에 따른 관리도의 중심선과 관리한계선을 표로 제시하였다. 개발한 관리도를 ARL 관점에서 볼 때 공정의 변화(모수 값들의 변화)를 잘 탐지함을 알 수 있었다. 대관령에서 측정한 2011년 3월 1일부터 2012년 3월 31일까지 분별 풍속데이터를 가지고 본 연구에서 개발한 관리도에 적용해 본 결과, 풍속 데이터는 와이블분포를 따르며 정규분포를 가정한 것 보다는 효과적임을 알았다. 추후 연구 과제로는 와이블분포의 모수를 모를 경우와 최소값 및 최대값의 산포를 관리할 수 있는 관리도의 설계 문제를 다루어 볼 수 있을 것이다.

참고문헌

- Amin, R. W. and Wolff, H. (1999), EWMA Control Charts for the Smallest and Largest Observations, *Journal of Quality Technology*, **31**(2), 189-206.
- Box, G. E. P. and Kramer, T. (1992), Statistical process monitoring and feedback adjustment-a discussion, *Technometrics*, **34**(3), 251-285.
- Chang, Y. S. (2014), Variable Sampling Interval \bar{X} Control Chart Using Weighted Standard Deviation Method, *Journal of the Korean Operations Research and Management Science Society*, **39**(1), 1-12.
- Duncan, A. J. (1956), The economic design of charts used to maintain current control of a process, *Journal of American Statistical Association*, **51**, 228-242.
- Fisher, R. A. and Tippett, L. M. C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. of the Cambridge Philosophical Society*, **24**, 180-190.
- Hotelling, H. (1947), Multivariate quality control-illustrated by the air testing of sample bombings, in *Techniques of Statistical Analysis*, Eisenhart C., Hastay M. W., and Walliw W. A., (eds) McGrawHill, New York, 111-184.
- Howell, J. M. (1949), Control chart for largest and smallest values, *The Annals of Mathematical Statistics*, **20**, 305-309.
- Jones, L. L. and Case, K. E. (1981), Economic Design of a Joint and R control Chart, *AIE Transactions*, **13**, 182-195.
- Kim, J. W., Park, C. S., Kim, J. S., Kim, S. S., Baek, J. G. (2014), Update Cycle Detection Method of Control Limits using Control Chart Performance Evaluation Model, *Journal of the Korean Institute of Industrial Engineers*, **40**(1), 43-51.
- Lee, T. H., Lee, J. H., Lee, M. K. (2013), Economic of variable sampling interval control chart using a surrogate variable, *Journal of the Korean Institute of Industrial Engineers*, **39**(5), 422-428.
- Montgomery, D. C. (1980), Economic design of control charts : A review and Literature survey, *Journal of Quality Technology*, **12**(2), 75-87.
- Page, E. S. (1954), Continuous Inspection Schemes, *Biometrika*, **41**, 100-115.
- Roberts, S. W. (1959), Control Chart Tests Based on Geometric Moving Averages, *Technometrics*, **1**, 239-250.
- Sarkadi, K. and Vincze, I. (1974), *Mathematical methods of statistical quality control*, Academic Press, New York.
- Taylor, H. M. (1968), The economic design of cumulation sum control chart, *Technometrics*, **10**(3), 479-488.
- Weibull, W. (1939), A statistical theory of the strength of materials, *Ing. Velenskaps Akad, Handl*, **151**, 1-45.