

존슨변환을 이용한 공정능력분석 알고리즘 개발

김종걸* · 박은하** · 정연승*

성균관대학교 시스템경영공학부*
동양매직(주)**

Process Capability Analysis Algorithm Using Johson Transformation

Jong-Gurl Kim · Eun-Ha Park · Youn-seang Jung

Abstract

This paper considers an algorithm using Johnson transformation to calculate process capability index for non-normal distribution. Johnson transformation is well known as one of methods transforming the data with non-normal distribution to normal data.

1. 서론

정규분포를 기본 가정으로 개발된 방법을 비정규분포에서 사용하였을 경우 제대로 된 평가를 할 수 없게 된다. 이런 경우에는 비정규분포에 맞는 공정능력분석방법을 개발하여 사용해야 하지만 시간과 비용측면에서 많은 어려움이 있다. 따라서, 비정규분포를 따르는 공정에서 정규분포를 기본 가정으로 개발된 방법을 이용하는 방안 중의 한 가지는 비정규분포를 정규분포로 변환하는 것이다. 정규분포로 변환하여 정규분포를 기본가정으로 개발된 기법들을 이용하여 공정능력분석을 한다. 본 연구에서는 비정규분포를 정규분포로 변환시키는 변환방법중의 한가지인 존슨변환을 이용한 공정능력분석 알고리즘을 소개하고자 한다.

2. 변수변환방법

비정규분포를 따르는 공정을 정규분포를 따르는 공정으로 변환하여 변환된 정규공정에서 정규분포를 기본가정으로 하고 있는 공정능력분석방법을 적용하는데, 존슨에 의해 제시된 존슨분포의 변환식을 이용하였다.[3] 존슨(Johnson)분포는 공정의 형태에 따라서 세 가지 분포족을 갖기 때문에 Slifker와 Shapiro의 표본 백분위수를 이용하여 분포족의 형태를 결정하였다.[1]

2.1 존슨분포

존슨분포는 비정규분포를 모형화하는데 사용되었다. 주어진 공정에 가장 좋은 분포를 선택하는데 단순한 구조를 보여주기 때문에 다른 방법보다 일반적으로 쉽다고 Farnum은 지적하였다.[2]

또한 가장 현저한 특징 중의 하나는 표준정규곡선으로 변환하여 사용하는 확률계산이다. 이러한 존슨분포는 식(1)의 변환식과 식(2)~식(4)인 3가지 분포족을 갖는다.

$$X' = \gamma + \eta K_i(X, \lambda, \epsilon) \quad i = 1, 2, 3 \quad (1)$$

$$K_1(X, \lambda, \epsilon) = \sin h^{-1} \left(\frac{X - \epsilon}{\lambda} \right) \quad (2)$$

$$K_2(X, \lambda, \epsilon) = \ln \left(\frac{X - \epsilon}{\lambda + \epsilon - X} \right) \quad (3)$$

$$K_3(X, \lambda, \epsilon) = \ln \left(\frac{X - \epsilon}{\lambda} \right) \quad (4)$$

특성치 X 를 $\eta, \gamma, \lambda, \epsilon$ 의 적절한 모수선택에 의해 X' 로 변환시킬 수 있다. ϵ, γ 는 위치모수이며, λ, η 는 척도모수이다.

2.2 존슨변환

존슨분포의 세 가지 분포족은 S_B, S_L, S_U 이다. 식 (2) 는 S_U 곡선으로 정의역이 전 실직선이 되며, 식 (3) 은 S_B 곡선으로 개구간 $(\epsilon, \epsilon + \lambda)$ 에 대해 정의되는 곡선이다. 식 (4) 는 S_L (대수정규)곡선을 나타낸다. <표 1>은 특성치 X 를 변환시키는 변수 X' 에 대한 변환식과 모수 조건들이 포함되어 있다. 존슨의 변환식을 이용하기 위하여 <표 1>의 세 개의 분포족 중 하나를 선택해야 한다. 또한, 모수도 추정해야 한다. 분포족 중 한가지를 선택하는 많은 방법 중 적용방법이 간단한 분위수 방법을 이용한다.

<표 1> 존슨 변환

존슨분포족	변환식	모수조건	X의 범위
S_B	$X' = \gamma + \eta \ln\left(\frac{X - \epsilon}{\lambda + \epsilon - X}\right)$	$\eta, \lambda > 0, -\infty < \gamma < \infty$ $-\infty < \epsilon < \infty$	$\epsilon < X < \epsilon + \lambda$
S_L	$X' = \gamma + \eta \ln(X - \epsilon)$	$\eta > 0, -\infty < \gamma < \infty$ $-\infty < \epsilon < \infty$	$X > \epsilon$
S_U	$X' = \gamma + \eta \sinh^{-1}\left(\frac{X - \epsilon}{\lambda}\right)$	S_B 와 같음	$-\infty < X < \infty$

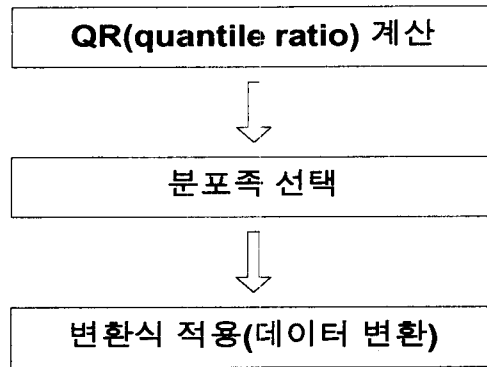
존슨변환은 형태에 따라서 세 가지 변환식을 갖음으로서 다른 변환방법에 비해 정확한 변환을 할 수 있으며 대부분의 비정규분포에서 적용가능하기 때문에 범용성을 갖는다는 장점이 있다.

다음표는 지금까지 개발된 변수변환식과 그것을 개발한 사람들이 나타나 있다.

<표 2> 변수변환식

변환식 개발자	변환식
Tukey (1959)	$y = X^\lambda$
Box & Cox (1964)	$y = \frac{X^\lambda - 1}{\lambda}, \lambda \neq 0$ $= \log X, \lambda = 0$
Draper & Cox (1969)	$y = X^{0.268}$
Taylor (1985)	$y = X^{0.2654}$
Nelson (1994)	$y = X^{0.2777}$
Rodriguez (1992)	존슨변환
Kittlitz (1999)	$y = X^{1/4}$

여기서, 1999년에 Kittlitz 가 개발한 변환식은 지수분포를 따르는 공정에서 적용가능한 것이다. 마찬가지로, 존슨 변환식을 제외한 대부분의 변환식은 일부 비정규분포에서만 적용가능하다. 따라서, 본 연구에서는 거의 모든 비정규분포에 적용할 수 있는 존슨 변환을 이용하였다. <그림 1>은 존슨변환의 적용절차를 대략적으로 나타낸 것이다.



<그림 1> 변환식 적용절차

2.3 분위수 방법

모수 추정과 분포족 중 한가지를 선택하는 방법 중에서 백분위수 방법을 이용하는 절차는 다음과 같다.[3]

2.3.1 QR(quantile ratio) 정의

표준정규곡선상의 $-3z, -z, z, 3z$ 에 대한 누적확률은 $q_1 = \Phi(-3z), q_2 = \Phi(-z), q_3 = \Phi(z), q_4 = \Phi(3z)$ 이고, x_i 는 데이터 집합 중 q_i 번째 있는 데이터이라 하면, QR(quantile ratio)은 다음과 같다.

$$QR = \frac{(x_4 - x_3)(x_2 - x_1)}{(x_3 - x_2)^2} \quad (5)$$

2.3.2 QR 의 추정

특성치 X 가 어떤 분포인지 알 수 없다면 QR 도 구할 수 없다는 것에 주의하여야 한다. 즉, 데이터가 어떤 분포인지 알 수 없으면 QR의 정확한 값을 구할 수 없으므로 식 (7) 의 기준을 적용시킬 수 없다. 그러나, 예가 주어지면, QR의 추정식에 의해 QR 을 추정할 수 있다. 추정식은 다음과 같다.

$$\widehat{QR} = \frac{(\widehat{x}_4 - \widehat{x}_3)(\widehat{x}_2 - \widehat{x}_1)}{(\widehat{x}_3 - \widehat{x}_2)^2} \quad (6)$$

여기서, \widehat{x}_i 는 표본데이터에서 q_i 번째 분위수이다. 보통 \widehat{x}_i 는 데이터 집합 중에서 j 번째 관측치가 된다. 여기서, $j = [nq_i] + 1$ 또는 $j = [(n+1)q_i]$, 여기서 $[\cdot]$ 는 정수함수이다.

2.3.3 존슨 분포족 선택

Bowman과 Shenton은 QR을 기초로 하여 S_B 와 S_U 분포 중에서 선택하는 일반적 기준을 제시하였다.[4]

계산된 QR 값이 1보다 작으면 S_B 분포의 변환식을 이용하고, 1과 같으면 S_L 분포

의 변환식을 이용한다. 그리고, 1보다 크면 S_U 분포의 변환식을 이용한다.[1]

$$\begin{aligned} S_B \text{ 분포, } QR < 1 \\ S_L \text{ 분포, } QR = 1 \\ S_U \text{ 분포, } QR > 1 \end{aligned} \quad (7)$$

2.4 정규성 검사 방법

정규성을 검사하기 위한 방법으로 Shapiro-Wilk 에 의한 통계치 W 가 있다. 이것을 이용하여 최적분포를 추정하였다.

2.4.1 Shapiro-Wilk 의 통계치

Shapiro-Wilk의 통계치는 다음과 같다.[5]

$$W = \frac{(\sum_{i=1}^n a_i v_i)^2}{\sum_{i=1}^n (v_i - \bar{v})^2} \quad (8)$$

v_i 는 데이터를 오름차순으로 정리한 것이다. $\bar{v} = \sum_{i=1}^n v_i / n$, a 는 Sarhan and Greenberg이 나타낸 계수로서 불편성을 갖는 최적선형계수이다.[6] Royston은 W 를 표준정규분포에 가까운 변수, Z 로 변환시켜 사용하였다.[7] 변환식은 다음과 같다.

$$Y = (1 - W)^\lambda \quad (9)$$

$$Z = (Y - \mu_y) / \sigma_y \quad (10)$$

μ_y, σ_y 는 Y 의 평균과 표준편차이다. Z 의식은 본래의 W 값의 유의수준을 계산하는데 이용된다. 큰 Z 값(W 값은 작은 경우가 된다.)은 정규성에서 멀어진다는(비정규분포)는 것을 가리킨다. λ, μ_y, σ_y 는 모두 n 에 의한 함수이다. $(1 - W)^\lambda$ 에 대한 실험 수 (0.1, 0.5, 1, 2, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 99 퍼센트의 점) 사이의 상관 계수를 이용하여 간단한 라인탐색법으로 λ 값을 찾는다.

<표 3> $\lambda, \log_e(\mu_y), \log_e(\sigma_y)$ 을 계산하기 위한 계수 표.(통계치 W 를 $Z = ((1 - W)^\lambda - \mu_y) / \sigma_y$ 에 의해 표준정규분포로 변환. 큰 Z 값은 비정규분포를 따름. 다항식은 $\sum c_i (\log n - d)^i$ 에 의해 계산된다). 여기서, $n \leq 20$ 인 경우 $d=3$, $21 \leq n \leq 2000$ 인 경우 $d=5$. [8]

Parameter	Range of n	Coefficients						
		0	1	2	3	4	5	6
λ	7-20	0.118898	0.133414	0.327907				
	21-2000	0.480385	0.318828	0	-0.0241665	0.00879701	0.002989646	
$\log_e(\mu_y)$	7-20	-37542	-0.492145	-1.124332	-0.199422			
	21-2000	-1.91487	-1.37888	-0.04183209	0.1066339	-0.03513666	-0.01504614	
$\log_e(\sigma_y)$	7-20	-3.152805	0.729399	3.01855	1.558776			
	21-2000	-3.73538	-1.015807	-0.331885	0.1773538	-0.01638782	-0.03215081	0.003852646

수치예제

표본의 크기는 $n = 137$ 이고 $W = 0.9620$ 이라고 하자. <표 3>에서 $x = \log(137) - 5 = -0.080$ 로 되어 있다.

$$\lambda = 0.480385 + (0.318828)(-0.080) + \dots + (0.002989646)(-0.080)^5 = 0.4549$$

$$\log \mu_y = -1.91487 + (-1.37888)(-0.080) + \dots + (-0.01504614)(-0.080)^5 = -1.8049$$

$$\mu_y = 0.1645$$

$$\log \sigma_y = -3.73538 + (-1.015807)(-0.080) + \dots + (0.003852646)(-0.080)^6 = -3.6563$$

$$\sigma_y = 0.02583$$

$$Z = (Y - \mu_y) / \sigma_y = [(1 - 0.9620)^{0.4549} - 0.1645] / 0.02583 = 2.38 \quad (p < 0.01)$$

예제는 1 퍼센트 수준에서 상당한 비정규성을 갖는다.

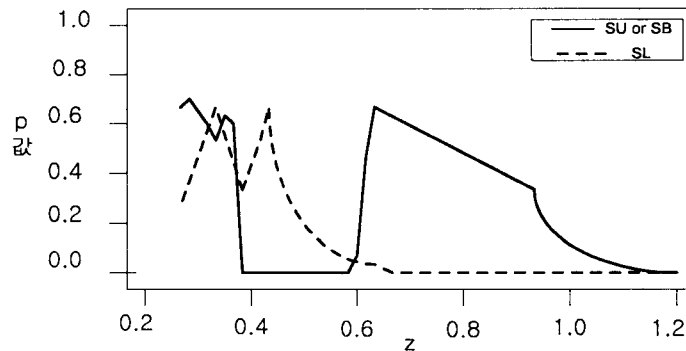
2.5 최적분포 추정

이 장에서는 존슨분포 중에서 데이터에 가장 적합한 분포를 추정하는 절차를 제시하였다. Shapiro와 Wilk에 의한 통계치 W는 정규성을 검사하는데 사용된다.[5] 작은 W 값은 비정규성을 가리킨다. 정규성이라는 가정은 계산된 W 값이 w_α 값보다 적으면 유의수준 α 에서 거절된다. w_α 는 검사 통계치의 α 번째 분위수가 된다. 주어진 데이터집합에서 p 값은 실제 계산된 W 값보다 더 작은 값을 얻을 확률이 된다. 작은 p 값은 데이터가 비정규성을 갖고 있음을 나타내고, 큰 p 값은 정규분포를 따르고 있음을 나타낸다. 비정규 데이터에 적합한 존슨분포는 z 값에 의해 선택됨으로 존슨분포에 적합

한 최대 W 값을 구하기 위한 최적 z 값은 한정되어 있다. 통계치 W의 최대 값은 0.90 이상이 되어야 하지만, p 값은 정해져 있지 않다.

z 와 p 값의 관련성을 나타낸 그래프는 <그림 2>에서 보여준다. 여기서 z 값과 p 값 사이의 관련성은 발견할 수 없었다. Mandraccia, Halverson, Chou는 미국에 있는 소니 반도체회사에서 얻은 데이터를 이용하여 모든 z 값에 대한 QR값을 구하는 시뮬레이션 연구와 사례연구를 하였다. 이 연구에서 z 의 최적가능 범위는 { z : z = 0.25, 0.26, ..., 1.25 }로 나타났다.[9]

이런 범위를 벗어나는 z 값에서는 QR 의 값을 제대로 구할 수 없게 된다. 예를 들어, z=1.28 일 때. -3z, -z, z, 3z 의 표준정규곡선의 값은 0.000062, 0.100273, 0.899727, 0.999939 이다. 표본의 첫 번째, 10번째, 90번째, 100번째 데이터가 된다. 즉, 표본크기가 100일 때, \widehat{QR} 은 데이터의 가장 큰값과 가장작은 값으로 구해진다. 10번째와 90번째 사이의 데이터는 무시된다.



<그림 2> p 와 z 값의 변화

어떤 분포를 따르는지 알 수 없는 X_1, \dots, X_n 데이터의 최적분포 추정절차는 다음과 같다.[10]

- 단계 1 : 특성치의 정규성을 검사할 수 있는 W 값과 p 값은 컴퓨터 프로그램으로 계산된다.[7] 정규분포를 따르면 멈추고, 정규분포가 아니면 step 2 로 간다.
- 단계 2 : { z : z= 0.25, 0.26, ..., 1.25 }에서, z 값에 대한 식(6)의 QR 값을 구한다. $QR < 1$ 이면 S_B 분포 또는 S_L 분포, $QR > 1$ 이면 S_U 분포 또는 S_L 분포가 된다. <표 1>에서 제시된 변환을 선택하기 위하여 Slifker와 Shapiro가 제시한 식을 이용하여 모수를 추정한다.[1]
- 단계 3 : 데이터가 S_B 또는 S_L 분포일 때, X 의 범위를 벗어나는 데이터인지 검사한다. 일부 데이터가 X 의 범위를 벗어나면 분포는 데이터에 잘 맞는 분포가 아니다. 따라서, 이 분포를 제거한다.
- 단계 4 : 변환된 데이터에서 W 와 p 값을 계산한다.
- 단계 5 : 존슨분포에 가장 적합한 변환은 가장 큰 W 값을 갖는 경우이다. 변환된 데이터는 정규분포가 되어, 통계적 절차는 바로 적용될 수 있다.

추정치 식에서 $\widehat{QR} \leq 1$ 이면 S_U 의 추정치 λ, γ 는 존재하지 않는다는 것을 알 수 있다. 유사하게 $\widehat{QR} \geq 1$ 이면, S_B 의 추정치 $\lambda, \gamma, \epsilon$ 은 존재하지 않는다. 그러므로, 표본 분위비(QR)는 S_U 와 S_B 분포족 중 한가지를 선택하는 것이 적당하다.

데이터를 존슨분포에 맞도록 변환시키는 절차에서, 통계치 W 값은 데이터에 잘 맞는 분포를 결정하는데 사용한다. 표본데이터 X_1, \dots, X_n 에서 S_B 분포를 따르면, 데이터는 $(\epsilon, \epsilon + \lambda)$ 의 범위에 있어야 한다. S_L 분포를 따르면, 데이터는 $(\epsilon, -\infty)$ 의 범위에 있어야 한다. 따라서 데이터가 S_B 또는 S_L 분포를 따를 때, 랜덤변수 X 의 범위 내에 있는지 검사할 필요가 있다.[4]

추정된 모수를 바탕으로 제시된 변환이 만족되지 않을 수도 있다. 하지만, 시뮬레이션 연구와 실제사례연구에서, 존슨변환은 정규성 변환에 충분한 유연성을 갖고 있음을 알 수 있다.

3. 공정능력지수

3.1 공정능력지수의 정의

규격과 관련하여 공정능력을 평가하는 척도로서 공정능력지수(process capability index)가 있다. 공정능력지수는 공정능력(6σ)과 규격의 폭과의 비율로서 공정이 규격에 맞는 제품을 생산할 능력을 가지고 있는지를 나타내는 지수이다.

3.2 공정능력지수의 종류

① C_p

오늘날 산업계에서 가장 먼저 사용된 공정능력지수는 전통적 6σ 개념을 기초로 하여 개발된 C_p 이다. 이것은 식(11)으로 정의하여 나타낸다.

$$C_p = \frac{USL - LSL}{6\sigma} \quad (11)$$

USL(Upper Specification Limit) : 규격상한

LSL(Lower Specification Limit) : 규격하한

σ : 특성치 X 의 표준편차

이 지수의 큰 결점은 공정의 평균을 고려하지 않고 실제 공정 산포로 정의된 잠재능력을 측정한다는 사실이다. 그러므로 C_p 는 실제 공정성과를 나타내질 못한다. 이것은 규격내 제품을 생산하는 공정능력이 갖는 공정평균이 변하는 영향을 반영하지 못하고 있다. 이런 문제 해결을 위해 C_{pk} 지수가 개발되었다. C_p 와 C_{pk} 를 동시에 사용하게 되면 공정범위와 위치에 관해서 올바르게 공정능력을 제시하여 주게 된다.

② C_{pk}

공정능력지수인 C_{pk} 는 식(12)으로 정의된다.

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right) \quad (12)$$

여기서, μ 는 공정평균이다. 공정이 완전히 규격중심에 위치할 때, $C_p = C_{pk}$ 이다.

③ C_{pm}

다구찌 손실함수를 기초하여 개발된 공정능력지수 C_{pm} 은 식(13)과 같이 정의된다.

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \quad (13)$$

여기서, T는 목표치이다. C_{pm} 지수는 더욱 강건한 개선으로, 공정변동뿐만 아니라 공정위치에 관해서도 공정능력을 평가하는 능력을 갖는다.

④ C_{pm}^*

목표치 T가 규격의 중심을 벗어나는 경우에는 C_{pm} 지수를 C_{pm}^* 으로 표현한다.

$$C_{pm}^* = \frac{\min(T - LSL, USL - T)}{3\sqrt{\sigma^2 + (\mu - T)^2}} \quad (14)$$

⑤ C_{pmk}

C_{pmk} 지수는 공정평균이 목표치에 충분히 근접하고 있는지의 여부와 공정평균에 대한 변동이 충분히 작은 지의 여부를 동시에 결정할 수 있다. C_{pmk} 를 식(15)과 같이 정의하고있다.

$$C_{pmk} = \frac{\min(USL - \mu - |\mu - T|, \mu - LSL - |\mu - T|)}{3\sqrt{\sigma^2 + (\mu - T)^2}} \quad (15)$$

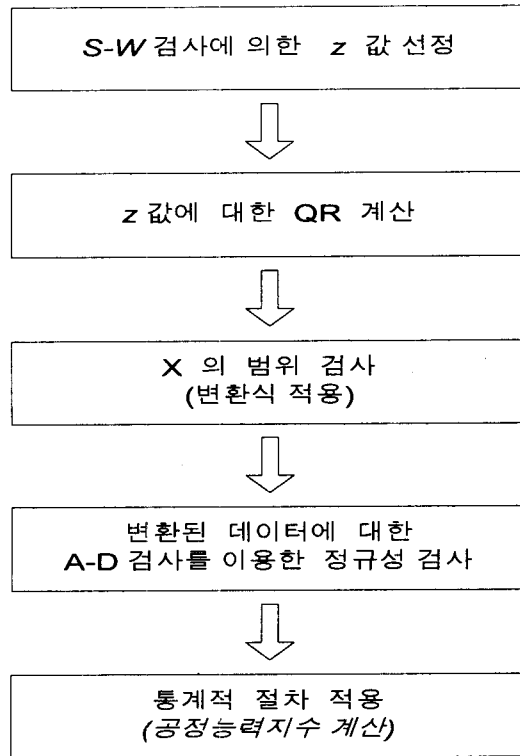
⑥ C_{psk}

C_{psk} 지수는 기설정된 목표치 T에 대한 공정변동을 나타내는 인자 $|\mu - T|$ 를 분자에 도입함으로써, C_{pmk} 로부터 만들어졌다.

C_{psk} 는 다음과 같이 정의된다

$$C_{psk} = \frac{\min(USL - \mu - |\mu - T|, \mu - LSL - |\mu - T|)}{3\sqrt{\sigma^2 + (\mu - T)^2}} \quad (16)$$

<그림 3>은 존슨변환을 이용한 공정능력분석 절차를 나타낸 것이다.



<그림 3> 존슨변환을 적용한 공정능력지수 계산 절차

공정능력지수	한쪽규격 불량률(PPM)	양쪽규격 불량률(PPM)
0.25	226628	453255
0.5	66807	133614
0.6	35931	71861
0.7	17865	35729
0.8	8198	16395
0.9	3467	6934
1	1350	2700
1.1	484	967
1.2	159	318
1.3	48	96
1.4	14	27
1.5	4	7
1.6	1	2
1.7	0.17	0.34
1.8	0.03	0.06
2	0.0009	0.0018

<표 4> 정규분포공정에서 공정능력지수 값에 대한 불량률 값

<표 4>는 정규분포공정에서 공정능력지수 값에 대한 불량률의 값을 나타낸 것이다. 표에서 알 수 있듯이 공정능력지수가 작으면 불량률은 높고, 공정능력지수가 크면 불량률은 낮게 나온다. 즉, 공정능력지수와 불량률은 반비례 관계에 있다.[13]

4. 변환방법 적용

4.1 공정능력분석

지금까지 설명한 존슨변환을 실제 데이터에 적용시켜 정규분포를 기본 가정으로 하고 있는 공정능력분석 방법중의 한가지인 공정능력지수를 계산하였다. 변환전의 데이터에서 계산한 공정능력지수와 변환된 데이터에서 계산한 공정능력지수를 비교함으로써, 공정능력분석시 변환된 데이터를 이용하면 실제 공정에 더 가깝게 평가할 수 있음을 보여준다.

4.2 공정능력지수 계산

Pearn과 Kotz의 자료를 이용하고자 한다.[11] 공정능력분석 절차를 이용하여 예제로 부터 신규생산라인의 품질개선을 하기 위해 고무전단 무게(g)에 대한 100개의 측정 데이터 <표 5>를 적용시켜 실제 공정능력을 분석하고자 한다. 목표치 T 는 8.7g 이며, USL=8.96, LSL=8.44 이다. 대표적인 비정규분포를 따르는 데이터이기 때문에 정규분포 가정 하에 공정능력을 추정하는 것은 좋지 않다.

<표 5> Pearn과 Kotz의 사례 인용 데이터

8.84	8.4	8.85	8.85	8.85	8.85	8.85	8.85	8.86	8.86
8.84	8.6	8.86	8.86	8.87	8.87	8.87	8.87	8.87	8.87
8.85	8.7	8.87	8.87	8.88	8.88	8.88	8.88	8.88	8.88
8.85	8.88	8.88	8.88	8.88	8.88	8.89	8.89	8.89	8.89
8.85	8.89	8.89	8.89	8.89	8.89	8.89	8.89	8.89	8.89
8.85	8.89	8.89	8.89	8.89	8.9	8.9	8.9	8.9	8.9
8.85	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9
8.85	8.9	8.9	8.91	8.91	8.91	8.91	8.91	8.91	8.81
8.85	8.91	8.91	8.91	8.91	8.91	8.91	8.91	8.91	8.92
8.85	8.92	8.92	8.92	8.92	8.92	8.93	8.93	8.93	8.93

앞에서 설명한 절차를 이용하여 데이터에 맞는 존슨분포를 선택한다. z의 범위 {0.25, 0.26, ..., 1.25}에서, z=0.524 에서 최적 분포가 된다. 따라서, $QR \geq 1.5$ 이므로 분포족은 S_U 이다. 모수추정식을 이용하여, $\hat{\gamma} = 0.524$, $\hat{\eta} = 1.512$, $\hat{\lambda} = 0.027$, $\hat{\epsilon} = 8.900$ 으로 계산된다.[1] 추정된 변환식은 다음과 같다.

$$Z = \gamma + \eta \sinh^{-1} \left(\frac{X - \epsilon}{\lambda} \right)$$

$$= 0.524 + 1.512 \times \sinh^{-1} \left(\frac{X - 8.900}{0.027} \right)$$

<표 6>은 변환전의 데이터에서 계산한 공정능력지수와 변환된 데이터에서 계산한 공정능력지수를 비교한 것이다. 표에서 알 수 있듯이 샘플 데이터들은 규격을 벗어나는 불량은 없으나, 목표치를 매우 벗어나고 있어 품질개선이 요구된다.

<표 6> 공정능력지수 계산 비교

변환전 데이터	변환후 데이터
평균(μ) = 8.8865 , 분산(σ^2) = 0.0006	평균(μ) = -0.094 , 분산(σ^2) = 1.0033
USL=8.96 , LSL=8.44 , T=8.7	USL=2.8678 , LSL=-4.831 , T=-3.5771
$C_p = 4.01235$	$C_p = 1.27891$
$C_{pk} = 1.0725$	$C_{pk} = 0.9216$
$C_{pm} = 0.45200$	$C_{pm} = 0.33720$
$C_{pm}^* = 0.45200$	$C_{pm}^* = 0.10980$
$C_{pmk} = 0.12080$	$C_{pmk} = 0.24290$
$C_{psk} = 0$	$C_{psk} = 0$

변환전의 데이터에서 계산된 공정능력지수는 변환된 데이터에서 계산된 공정능력지수보다 크게 나왔다. 이것은 정규성을 갖도록 변환시킨 데이터에서 정규분포를 기본 가정으로 사용되는 지수를 적용시켰을 때 비정규성을 갖는 실제 데이터를 더 잘 평가해줄 수 있다.

4.3 불량률 추정

여기서 C_{pmk} 와 C_{psk} 지수에서는 변환의 효과를 찾아볼 수 없으므로 데이터에서 규격을 벗어나는 불량률을 추정하여 변환의 효과를 비교해 보겠다. 먼저, (가) 실제 데이터가 이루고 있는 비정규분포의 불량률을 추정한다. (나) 변환전의 비정규분포를 정규분포로 가정하여 불량률을 추정한다. (다) 변환된 정규분포에서 불량률을 추정한다.

이렇게 구한 불량률을 서로 비교하여 변환후의 데이터에서 구한 불량률이 변환전의 비정규분포를 정규분포로 가정하여 계산한 불량률보다 실제 불량률에 가깝게 추정됨을 제시하였다.

<표 5>의 데이터를 이용하여 불량률을 추정한다. 평균(μ)은 8.8865 , 분산(σ)은 0.0006 이다. USL=8.96, LSL=8.44 이다.

4.3.1 실제 데이터가 이루고 있는 비정규분포에서의 불량률 추정

분포의 형태를 결정하기 위해 피어슨 시스템을 이용하였다.[14] K값을 구한 결과 형

태1인 베타분포로 확인되었다. 베타 분포의 불량률을 추정하는 절차로부터

- ① 표본통계량은 $\bar{X}=8.89$, $s=0.0215$, $S_K=-0.24$, $K_U=2.43$ 으로 구해졌다.
- ② S_K, K_U 로부터 구한 K 값이 $K=-2.20 < 0$ 이므로 베타분포를 한다.
- ③ Gruska et al.의 표에서 $S_K=0.2$ 와, $S_K=0.3$ 값으로부터 $S_K=0.24$ 의 값을 보간법으로 계산하여 Z 값을 구한다.[12]
- ④ $X=Z \cdot s + \bar{X}$ 로부터, $Z=-2.474$ 일 때, $X=8.8368$, $Z=-2.134$ 일 때, $X=8.8441$ 이다. 따라서 규격(LSL)을 벗어나는 불량률을 추정하면 다음과 같다.

$$P(X < 8.44) \approx 0.514\%$$

- ⑤ 규격(USL)을 벗어나는 불량률을 계산하기 위해 $USL=8.96$ 일 때 베타 함수식을 적용하면 다음과 같다.

$$\begin{aligned}
 y &= y_0(1 + Z/A_1)^{P_1}(1 - Z/A_2)^{P_2} & (17) \\
 &= 0.0104(1 + 3.2558/2.3274)^{1.5850}(1 - 3.2558/3.32)^{2.6875} \\
 &= 0.000008
 \end{aligned}$$

따라서, 규격을 벗어나는 불량률은 0.00514 가 된다.

4.3.2 변환전의 비정규분포를 정규분포로 가정하여 불량률 추정

$$\begin{aligned}
 P(X \leq LSL) &= P(Z \leq \frac{LSL - \mu}{\sigma}) = P(Z \leq \frac{8.84 - 8.8865}{0.0245}) \\
 &= P(Z \leq -1.898) = 1 - P(Z \leq 1.898) \approx 0.0294 \\
 P(X \geq USL) &= P(Z \geq \frac{USL - \mu}{\sigma}) = P(Z \geq \frac{8.96 - 8.8865}{0.0245}) \\
 &= P(Z \geq 3) = 0.0013
 \end{aligned}$$

따라서, 규격을 벗어나는 불량률은 0.0307 이 된다.

4.3.3 변환된 정규분포에서 불량률 추정

규격을 벗어난 불량률을 계산하면 다음과 같다.

$$\begin{aligned}
 P(X < LSL) &= P(Z < \frac{LSL - \mu}{\sigma}) = P(Z < -4.456) = 0.0000008 \approx 0 \\
 P(X > USL) &= P(Z > \frac{USL - \mu}{\sigma}) = P(Z > 2.732) \approx 0.0032
 \end{aligned}$$

따라서, 규격을 벗어나는 불량률은 0.0032 가 된다.

변환전 베타분포	변환전의 데이터를 정규분포로 가정	변환후의 데이터를 정규분포로 계산
0.514%	3.07%	0.32%

위의 세 가지 경우에서 불량률을 추정해본 결과 변환전의 비정규분포에서 계산된 불량률은 정규분포로 변환시킨 후 계산한 불량률과 비슷하게 나왔으나, 변환전의 비정규분포를 정규분포로 가정하여 계산한 불량률과는 상당한 차이가 있음을 알 수 있다. 이것으로 비정규분포에서 정규분포를 기본가정으로 개발된 방법을 이용하는 것보다 정규분포로 변환하여 정규분포를 기본 가정으로 하고 있는 방법을 적용하면 실제 공정에 더 가깝게 평가할 수 있다는 것을 보여준다.

5. 결론

본 논문은 현장 엔지니어들이 사용하기에는 다소 어려운 점이 있을 것으로 생각되지만 현재 산업현장의 현실적인 문제를 해결하기 위한 방안을 제시하고자 했다. 앞으로의 과제는 공정의 분포 패턴과 관계없이 현장에서 쉽게 활용할 수 있고, 적용성이 뛰어난 공정능력지수의 산출방법을 모색하는 것이 매우 중요할 것으로 본다.

참고문헌

- [1] Slifker, J. F. and Shapiro, S. S., "The Johnson System: Selection and Parameter Estimation". *Technometrics* Vol. 22, pp. 39-246. (1980).
- [2] Farnum, N. R., "Using Johnson Curves to Describe Non-Normal Process Data". *Quality Engineering* Vol. 9(2), pp. 329-336. (1996-1997).
- [3] Johson, N. L., "System of Frequency Curves Generated by Methods of Translation". *Biometrika* Vol. 36, pp. 149-176. (1949).
- [4] Bowman, K. O. and Shenton, L. R., " S_B and S_U Distributions Fitted by Percentiles: A General Criterion". *Communications in Statistics-Simulation and Computation* Vol. 18, pp. 1-13. (1989).
- [5] Shapiro, S. S. and Wilk, M. B., "An Analysis of Variance Test for Normality". *Technometrics* Vol. 14, pp. 355-370. (1965).
- [6] Sarhan, A. E. and Greenburg, B. G., Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples. *Ann. Math. Statist* Vol. 27, pp. 427-451. (1965).
- [7] Royston, J. P., "The W Test for Normality". *Applied Statistics* Vol 31, pp. 176-180. (1982).
- [8] Anderson-Darling, T. W. and Darling, D. A., "A Test of Goodness of Fit". *Journal of the American Statistical Association* Vol. 49, pp. 765-969. (1954).
- [9] Mandraccia, S. T. and Halverson, G. D. and Chou, Y-M., "Control Chart Design Strategies for Skewed Data" in Process, Equipment, and Materials Control in Integrated Circuit Manufacturing II edited by A. Iturralde and T. Lin. *Proceedings of Society of Photo-Optical Instrumentation Engineers* Vol. 2876, pp. 196-205. (1996).
- [10] Shayib, M. A., "The Procedure for Selection of Transformations from the Johnson

- System". *Communications in Statistics-Simulation and Computation* Vol. 18, pp. 1457-1464. (1989).
- [11] Pearn, W. L. and Kotz, S., "Application of clement's Method for Calculating Second-and-Third-Generation Process Capability Indices Non-Normal Pearsonian Population". *Quality Engineering* Vol. 7, pp.139-145. (1994-1995).
- [12] Gruska, G. F. and Lamberson, L. R. and Mirkhani, K., Non-Normal Data Analysis. Multiface Publishing Co., Michgan (1989).
- [13] Douglas, C. and Montgomery, George. C. and Runger, Norma Faris Hubele., *Engineering statistics*. Wiley. New York, (1998).
- [14] 채규용, 비정규분포에 대한 공정능력평가에 관한 비교 연구. 건국대학교 대학원 박사학위 논문. (1999).