한국통계학회논문집
2009, 16권, 3호, 449–461

이변량 지수 공정 하에서 위험함수와 공정능력지수에 대한 통계적 추정

조중재[1,a], 강수묵[a], 박병선[b]

[a]충북대학교 정보통계학과, [b]통계청 분석통계팀

요약

최근의 생산 공정은 공정의 자동화, 고객 요구의 다양화 등으로 많은 품질 특성치들을 갖는 다변량 공정의 형태가 일반적이며, 벡터 공정능력지수는 이러한 다변량 공정의 능력을 평가하기 위한 대표적인 측도라 할 수 있다. 한편 공정의 분포에 대한 정보를 정확히 파악하기 어려운 실제 현장의 상황에서 보다 정확한 공정능력을 평가할 수 있는 통계적 추정 문제는 현실적으로 중요한 문제라고 할 수 있다. 본 논문에서는 특정한 이변량 지수 공정 하에서 이변량 벡터 공정능력지수 C_{pkl}에 대한 신뢰영역의 추정 문제에 관하여 연구하였다. 먼저 지수분포의 특성을 고려하여 실제 현장에서 널리 사용되고 있는 가장 기본적인 일변량 공정능력지수들 중에서 규격 하한만을 고려한 형태인 C_{pkl}에 관하여 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly})$로 확장·정의하고, 이 지수의 플러그-인 추정량 및 관련 극한 확률분포를 유도하였다. 또한 이 지수에 대해 Marshall과 Olkin (1967)의 이변량 지수분포 모형을 기초로 근사 신뢰영역을 제시하였으며, 모의실험을 통하여 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 95% 정규 근사(Asymptotic Normality: AN) 신뢰영역에 대한 이용가능성 및 효율성을 비교·분석하였다.

주요용어: 벡터 공정능력지수, 이변량 지수 공정, 근사 신뢰영역, 극한분포.

1. 서론

오늘날 산업의 경쟁력을 유지하고 강화함은 물론 빠르게 변화하고 다양해진 고객의 니즈를 충족하기 위하여 제품 품질의 중요성이 더욱 강조되고 있다. 더욱이 6시그마의 급격한 확대 등으로 인해 실제 현장의 공정에서 고객이 요구하는 일정한 품질을 가진 제품을 생산하고 있는지의 여부, 즉 공정능력을 정확히 파악하고 이를 수치화한 공정능력지수에 대한 관심이 지속적으로 증대되고 있다. 현재 널리 사용되고 있는 C_p, C_{pk}, C_{pm} 등과 같은 공정능력지수들은 단일 특성치를 갖는 공정의 능력을 평가하기 위한 척도들로 일변량 공정능력지수라고 할 수 있다.

하지만 최근의 생산 공정들은 생산 기술의 발전, 제품 구조의 복잡성, 공정의 자동화, 고객 요구의 다양화 등에 따라 관리해야 할 품질 특성치들이 증가하게 됨으로써 단일 특성치에서 2개 이상의 특성치를 갖는 다변량 공정의 형태를 갖는 것이 일반적일 것이다. 이러한 다변량 공정에서는 수많은 특성치들이 발생하고, 제품의 품질은 이런 많은 특성치들의 결합에 의해 결정된다. 그런데 이들 특성치들은 개별적으로 측정될 지라도 특성치들 간에는 서로 독립적이기 보다 복잡한 관계를 가지고 제품의 품질에 영향을 미칠 수 있으므로 결합적으로 분포되어 있다고 가정하는 것이 보다 현실적일 것이다. 이와 같은 특성을 갖는 다변량 공정의 능력을 평가하기 위하여 현실적으로 일변량 공정능력지수를 직접

이 논문은 2007년도 충북대학교 학술연구 지원사업의 연구비 지원에 의하여 연구되었음.
[1] 교신저자: (361-763) 충북 청주시 흥덕구 성봉로 410, 충북대학교 정보통계학과, 교수.
E-mail: jjcho@chungbuk.ac.kr

적용할 수는 없을 것이다. 따라서 다변량 공정능력을 올바로 평가하는 문제가 중요한 이슈로 대두되었으며, 다변량 공정능력분석과 공정능력지수에 대한 연구가 진행되어 왔다.

Alt와 Smith (1988)는 현재 이용 가능한 다변량 공정관리 기법들을 정리·연구하였으며, Kocherlakota와 Kocherlakota (1991)는 상관계수 ρ를 갖는 이변량 정규분포 BN(μ_x, μ_y, σ_x^2, σ_y^2, ρ) 하에서 가장 간단한 공정능력지수의 추정량들인 $\hat{C}_{px}$와 $\hat{C}_{py}$의 결합 확률분포함수를 계산하였다. 그리고 Hubele 등 (1991)은 3가지 성분을 가지고 정의된 또 다른 형태의 벡터 공정능력지수에 관한 연구를 수행하였다. 하지만 이와 같은 다변량 공정능력지수에 관한 연구는 계산과정이 복잡한 관계로 제한된 연구결과만이 발표되고 있다. 최근에 Park (2003)이 주로 이변량 정규공정 하에서 벡터 공정능력지수들에 대한 효율적인 신뢰영역 추정문제에 대하여 연구하였으나 아직 미흡한 실정이라 할 수 있다. 따라서 벡터 공정능력지수들에 관한 로버스트한 추정 문제는 응용적인 측면에서의 중요성을 고려할 때 현실적으로 매우 중요한 과제라고 할 수 있다.

본 논문에서는 수명시험, 신뢰도를 비롯하여 많은 응용 분야에서 중요한 역할을 하고 있는 Marshall과 Olkin (1967)의 이변량 지수분포 모형과 그 확률 및 통계적 성질에 대하여 조사·연구하고, 공정능력분석에서 규격 하한이 중요시되는 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 통계적 추정문제를 연구하였다. 먼저 2절에서는 본 논문에서 주로 응용하고자하는 Marshall과 Olkin (1967)의 이변량 지수분포 모형에 대한 확률 및 통계적 성질에 대해서 살펴보았다. 또한 3절에서는 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$의 플러그-인 추정량 $\hat{\mathbf{C}}_{pkl}$ 및 관련 극한분포 이론을 유도하고 이들의 분산-공분산 행렬 $\mathbf{V}_{pkl}$를 구체적으로 계산하였으며, 이를 기초로 이변량 지수 공정 하에서의 근사 신뢰영역을 설정·제시하였다. 마지막으로 모의실험을 통하여 Marshall과 Olkin (1967)의 이변량 지수 공정 하에서 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 정규 근사(Asymptotic Normality: AN) 신뢰영역의 정확성과 효율성을 비교·분석하였다.

2. 이변량 지수분포 모형과 위험함수

지수분포는 양의 실수영역에서 확률을 갖는 분포로서 시간의 흐름에 따라 발생하는 사건에 대하여 적용할 수 있는 확률분포들 중의 하나이다. 예를 들면, 지진이 발생하고 난 뒤 다음 지진이 발생할 때까지의 경과시간이라든지 또는 기계의 작동시간과 관련하여 기계 고장 수리 후 다음 고장 때까지의 작동시간 등에 관한 분포 등은 지수분포로 가정할 수 있다. 즉, 확률변수 X가 모수 $\lambda(\geq 0)$인 지수분포를 따른다고 하면 확률밀도함수 $f(x)$, $x \in R^1$은 다음과 같이 정의된다.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & X > 0, \\ 0, & \text{그 외}. \end{cases}$$

이러한 지수분포의 확률밀도함수 $f(x)$는 $x = 0$에서 최대값을 갖는 감소함수라는 특징을 갖으며, 평균과 분산은 각각 $E(X) = 1/\lambda$, $\mathrm{Var}(X) = 1/\lambda^2$이다.

한편 여러 가지 이변량 지수분포는 수명시험, 신뢰도를 비롯하여 많은 응용 분야에서 중요한 역할을 하고 있다. 두 개의 부품 (C_1, C_2)로 구성되어 있는 어떤 시스템을 생각하여 보자. 충격을 받는 두 부품 (C_1, C_2)의 수명시간을 (X, Y)로 둔다면 X와 Y는 일반적으로 서로 상관관계가 있는 종속적인 확률변수가 되는 경우가 많이 있다. 예를 들어 쌍발 제트엔진 비행기의 제트엔젠 하나가 고장이 나면 인접한 나머지 제트엔진도 영향을 받게 된다. Marshall과 Olkin (1967)은 이와 같이 서로 상관관계가 있는 두 부품으로 이루어진 어떤 시스템의 수명시간에 대한 모형으로서 이변량 지수분포를 제안하였고 그 분포의 여러 가지 중요한 성질에 대하여 연구하였다. 이 절에서는 본 논문에서 이용하고자 하는 Marshall과 Olkin (1967)의 이변량 지수분포 모형 및 그 확률과 통계적 성질에 대해서 알아보고자 한다.

확률변수 (X, Y)가 다음과 같은 분포함수로 주어진다면 모수 λ_1, λ_2, λ_3를 갖는 이변량 지수분포

BVE(λ_1, λ_2, λ_3)를 따른다고 한다.

$$\begin{aligned}\bar{F}(x,y) &= P(X > x, Y > y)\\ &= \exp\left[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)\right],\end{aligned}$$

여기서 $x > 0$, $y > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$이다.

이 때 X와 Y의 결합 확률밀도 함수는 다음과 같다.

$$f(x,y) = \begin{cases} \lambda_1(\lambda_2+\lambda_3)\exp\{-\lambda_1 x - (\lambda_2+\lambda_3)y\}, & x < y\text{인 경우},\\ \lambda_2(\lambda_1+\lambda_3)\exp\{-\lambda_2 x - (\lambda_1+\lambda_3)y\}, & y < x\text{인 경우},\\ \lambda_3\exp\{-(\lambda_1 x + \lambda_2 + \lambda_3)y\}, & x = y\text{인 경우}.\end{cases}$$

또한 X와 Y의 주변 확률분포함수는 각각 모수 $\lambda_1 + \lambda_3$과 $\lambda_2 + \lambda_3$을 갖는 지수분포이며, X와 Y의 기대값은 다음과 같다.

$$E(X) = \frac{1}{\lambda_1 + \lambda_3}, \qquad E(Y) = \frac{1}{\lambda_2 + \lambda_3}.$$

편의상, 위의 이변량 지수분포 모형에서 위험함수(hazard function)를 다음과 같이 정의하기로 하자.

$$\lambda_x = \lambda_1 + \lambda_3, \qquad \lambda_y = \lambda_2 + \lambda_3.$$

한편 Marshall과 Olkin (1967)의 이변량 지수분포 모형에서 $\lambda = \lambda_1 + \lambda_2 + \lambda_3$으로 놓으면 확률변수 X와 Y의 상관계수 ρ는 다음과 같다.

$$\rho = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{\lambda_3}{\lambda}.$$

마지막으로 이 지수분포 모형의 특징으로 확률변수 (X, Y)가 이변량 지수분포를 따르기 위한 필요충분조건은

$$X = \min(U, W), \qquad Y = \min(V, W)$$

을 만족하면서 지수분포를 따르는 서로 독립인 확률변수 U, V, W가 존재한다는 것이다. 본 논문에서는 이러한 Marshall과 Olkin (1967)의 연구 결과를 바탕으로 이변량 지수분포를 따르는 두 품질 특성치와 관련하여 보다 효율적인 통계적 추정 방안을 모색하고자 한다.

3. 이변량 벡터 공정능력지수 C_{pkl}와 통계적 추론

공정능력지수는 공정에서 생산되는 제품에 대해 통계적으로 관리해야 할 공정의 품질을 평가할 목적으로 개발되었으며, 공정이 규격을 만족하는 제품을 생산할 능력을 갖고 있는지를 나타내는 지수이다. 이러한 공정능력지수는 사용자가 공정의 성능을 쉽게 파악할 수 있고, 다수의 서로 다른 공정의 수행도(performance)를 동시에 감시할 수 있어 생산 분야에서 다양한 형태로 널리 사용되고 있다.

Juran (1974)에 의해 최초로 소개된 공정능력지수 C_p는 공정 특성치들에 대해 실제 공정의 산포와 허용 가능한 공정의 산포의 비율을 바탕으로 정의되었다. 이는 제품의 규격 상한(Upper Specification Limit: USL)과 규격 하한(Lower Specification Limit: LSL)이 주어지고 공정분산 σ^2이 고려된 경우로 다음과 같다.

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sqrt{E(X-\mu)^2}} = \frac{\text{USL} - \text{LSL}}{6\sigma},$$

여기서 μ는 공정평균이고 $d = (\mathrm{USL} - \mathrm{LSL})/2$이다.

그리고 Kane (1986)은 공정능력을 평가하는데 있어 공정평균 μ의 값을 공정 모수들 중의 하나로 고려하는 것으로 C_p를 변형한 공정능력지수 C_{pk}를 제안하였다. 이는 공정 평균 μ와 공정 분산 σ^2이 동시에 고려된 경우로 다음과 같다.

$$C_{pk} = \frac{\min(\mu - \mathrm{LSL}, \mathrm{USL} - \mu)}{3\sigma} = \frac{d - |\mu - M|}{3\sigma} = \left(1 - \frac{\mu - M}{d}\right)C_p = (1 - k)C_p,$$

단, $M = (\mathrm{USL} + \mathrm{LSL})/2$, $k = |\mu - M|/d$.

본 논문에서는 품질 특성치의 분포가 치우침이 있고 망대특성을 갖으며 한쪽 규격인 경우를 고려할 것이며, 이 경우에는 공정능력지수 C_p 혹은 C_{pk}에 대하여 규격 하한(LSL)만 있는 경우를 활용하는 것이 타당할 것이다 (Kotz와 Johnson, 1993). 즉, 규격 하한만 있는 경우의 공정능력지수 C_{pk}에 대한 공정능력지수 C_{pkl}은 다음과 같다.

$$C_{pkl} = \frac{\mu - \mathrm{LSL}}{3\sigma}.$$

한편 공정 평균이 각각 μ_x와 μ_y이고 분산이 σ_1^2과 σ_2^2이며 상관계수가 ρ인 이변량 공정에서 관심 있는 두 품질특성을 각각 X와 Y라 할 때, 2차원으로 확장한 이변량 벡터 공정능력지수는 다음과 같다.

$$\mathbf{C}_{pkl} = \left(C_{pklx}, C_{pkly}\right) = \left(\frac{\mu_x - \mathrm{LSL}_x}{3\sigma_x}, \frac{\mu_y - \mathrm{LSL}_y}{3\sigma_y}\right),$$

여기서 LSL_x, LSL_y는 각각 이변량 품질특성 X와 Y에 대한 규격 하한을 나타낸다.

3.1. 플러그-인 추정량 및 관련 극한분포

공정 평균과 분산이 각각 μ_x와 μ_y, σ_1^2과 σ_2^2이고 상관계수가 ρ이며, 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$과 $\mu_{4y} = E(Y - \mu_y)^4$이 유한인 이변량 모집단으로부터 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$은 현장에서의 응용성을 고려하여 보다 자연스러운 플러그-인 방법(plug-in method)에 의해 다음과 같은 추정량을 고려할 수 있을 것이다.

$$\hat{\mathbf{C}}_{pkl} = \left(\hat{C}_{pklx}, \hat{C}_{pkly}\right) = \left(\frac{\bar{X} - \mathrm{LSL}_x}{3\sigma_x}, \frac{\bar{Y} - \mathrm{LSL}_y}{3\sigma_y}\right),$$

단, 표본평균 $\bar{X}$와 $\bar{Y}$는 각각 $\bar{X} = 1/n\sum_{i=1}^n X_i$, $\bar{Y} = 1/n\sum_{i=1}^n Y_i$로 정의하고, 표본분산 S_x^2과 S_y^2은 각각 다음과 같이 정의한다.

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2, \qquad S_y^2 = \frac{1}{n-1}\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2.$$

그리고 편의상, 분산-공분산 행렬 $\mathbf{\Sigma}_{4\times4}$과 기대값 μ_{ixjy}를 다음과 같이 정의하기로 하자.

$$\mathbf{\Sigma}_{4\times4} = (\sigma_{ij})_{4\times4} = \begin{pmatrix} \sigma_x^2 & \mu_{1x1y} & \mu_{3x} & \mu_{1x2y} \\ & \sigma_y^2 & \mu_{2x1y} & \mu_{3y} \\ & \text{symm.} & \mu_{4x} - \sigma_x^4 & \mu_{2x2y} - \sigma_x^2\sigma_y^2 \\ & & & \mu_{4y} - \sigma_y^4 \end{pmatrix},$$

$$\mu_{ixjy} = E\left[(X - \mu_x)^i (Y - \mu_y)^j\right], \quad \mu_{ix0y} = \mu_{ix}, \quad \mu_{0xjy} = \mu_{jy} \quad (i, j = 1, 2, 3, 4).$$

또한 확률벡터 $(Z_1, Z_2, Z_3, Z_4) \sim \mathrm{MN}(\mathbf{0}, \mathbf{V}_{4\times 4})$라고 정의하면, Park (2003)에서 제시된 보조정리를 기초로 이변량 벡터 공정능력지수의 추정량 $\hat{\mathbf{C}}_{pkl}$과 관련된 극한 확률분포를 유도하면 다음의 정리 1과 같다.

정리 1. 만일 4차 중심적률 μ_{4x}와 μ_{4y}이 존재하면, 표본의 크기 n이 $n \to \infty$일 때 다음의 결과가 성립한다.

$$\begin{aligned}\sqrt{n}\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right) &= \sqrt{n}\left(\hat{C}_{pklx} - C_{pklx},\ \hat{C}_{pkly} - C_{pkly}\right)\\ &\xrightarrow{d} \left(\frac{1}{3\sigma_x^2}\left[\sigma_x Z_1 + \frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x} Z_3\right],\ \frac{1}{3\sigma_y^2}\left[\sigma_y Z_2 + \frac{\mathrm{LSL}_y - \mu_y}{2\sigma_y} Z_4\right]\right),\end{aligned}$$

단, 극한 확률변수의 분산-공분산행렬 $\mathbf{V}_{pkl} = (\sigma_{pklij})_{2\times 2}$의 각 성분은 다음과 같다.

$$\begin{aligned}\sigma_{pkl11} &= \frac{1}{9\sigma_x^4}\left[\sigma_x^2\sigma_x^2 + \left(\frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x}\right)^2\left(\mu_{4x} - \sigma_x^4\right) + (\mathrm{LSL}_x - \mu_x)\mu_{3x}\right],\\ \sigma_{pkl12} &= \sigma_{pkl21}\\ &= \frac{1}{9\sigma_x^2\sigma_y^2}\left[\sigma_x^2\sigma_x^2\mu_{1x1y} + \frac{\sigma_X(\mathrm{LSL}_y - \mu_y)}{2\sigma_y}\mu_{1x2y}\right.\\ &\quad \left. + \frac{(\mathrm{LSL}_x - \mu_x)(\mathrm{LSL}_y - \mu_y)}{4\sigma_x\sigma_y}\left(\mu_{2x2y} - \sigma_x^2\sigma_y^2\right) + \frac{\sigma_y(\mathrm{LSL}_y - \mu_x)}{2\sigma_x}\mu_{2x1y}\right],\\ \sigma_{pkl22} &= \frac{1}{9\sigma_y^4}\left[\sigma_y^2\sigma_y^2 + \left(\frac{\mathrm{LSL}_y - \mu_y}{2\sigma_y}\right)^2\left(\mu_{4y} - \sigma_y^4\right) + (\mathrm{LSL}_y - \mu_y)\mu_{3y}\right].\end{aligned}$$

증명: $\sqrt{n}(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl})$의 극한분포를 유도하면 다음과 같다.

$$\begin{aligned}\sqrt{n}\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right) &= \sqrt{n}\left(\hat{C}_{pklx} - C_{pklx},\ \hat{C}_{pkly} - C_{pkly}\right)\\ &= \sqrt{n}\left(\frac{\bar{X} - \mathrm{LSL}_x}{3S_x} - \frac{\mu_x - \mathrm{LSL}_x}{3\sigma_x},\ \frac{\bar{Y} - \mathrm{LSL}_y}{3S_y} - \frac{\mu_y - \mathrm{LSL}_y}{3\sigma_y}\right)\\ &= \sqrt{n}\left(\frac{1}{3S_x\sigma_x}\left[\bar{X}\sigma_x - \mu_x S_x + \mathrm{LSL}_x(S_x - \sigma_x)\right],\ \frac{1}{3S_y\sigma_y}\left[\bar{Y}\sigma_y - \mu_y S_y + \mathrm{LSL}_y(S_y - \sigma_y)\right]\right)\\ &= \sqrt{n}\left(\frac{1}{3S_x\sigma_x}\left[\sigma_x\left(\bar{X} - \mu_x\right) - \frac{\mu_x\left(S_x^2 - \sigma_x^2\right)}{S_x + \sigma_x} + \frac{\mathrm{LSL}_x\left(S_x^2 - \sigma_x^2\right)}{S_x + \sigma_x}\right],\right.\\ &\qquad \left.\frac{1}{3S_y\sigma_y}\left[\sigma_y\left(\bar{Y} - \mu_y\right) - \frac{\mu_y\left(S_y^2 - \sigma_y^2\right)}{S_y + \sigma_y} + \frac{\mathrm{LSL}_y\left(S_y^2 - \sigma_y^2\right)}{S_y + \sigma_y}\right]\right)\\ &\xrightarrow{d} \left(\frac{1}{3\sigma_x^2}\left[\sigma_x Z_1 - \frac{\mu_x}{2\sigma_x}Z_3 + \frac{\mathrm{LSL}_x}{2\sigma_x}Z_3\right],\ \frac{1}{3\sigma_y^2}\left[\sigma_y Z_2 - \frac{\mu_y}{2\sigma_y}Z_4 + \frac{\mathrm{LSL}_y}{2\sigma_y}Z_4\right]\right)\\ &\sim \mathrm{MN}(\mathbf{0}, \mathbf{V}_{pkl}), \quad \mathbf{V}_{pkl} = \begin{pmatrix} \sigma_{pkl11} & \sigma_{pkl12}\\ \sigma_{pkl21} & \sigma_{pkl22}\end{pmatrix}.\end{aligned}$$

왜냐하면, 슬루츠키 정리(Slutsky's theorem)와 Park (2003)의 보조정리에 의해 극한분포 결과를 얻을 수 있기 때문이다. 그리고 분산-공분산 행렬 $\mathbf{V}_{pkl}$에서의 3가지 성분들을 구체적으로 계산하면 주어진 내용을 얻게 될 것이다. □

3.2. 분산-공분산 행렬 $\mathbf{V}_{pkl}$과 정규 근사(AN) 신뢰영역

본 절에서는 이변량 확률 공정 하에서 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 근사 신뢰영역의 설정에 관하여 살펴보고자 한다. 일반적으로 공정의 분포에 대한 정보가 없다면 앞 절의 정리 1에서 제시된 분산-공분산 행렬의 각 성분들은 적률 추정을 이용한 추정량들을 이용할 수 있을 것이다. 하지만 공정의 분포가 특정한 형태의 이변량 분포를 따른다고 할 수 있다면, 앞 절의 정리 1에서 제시된 분산-공분산 행렬의 각 성분들은 보다 구체적으로 계산될 수 있을 뿐만 아니라 이러한 결과들을 통계적인 추론에 이용할 수 있을 것이다. 따라서 본 논문에서는 이변량 지수분포와 일반적인 이변량 확률분포 상황 하에서 다음의 두 가지 방법에 의한 분산-공분산 행렬 $\mathbf{V}_{pkl}$의 성분들을 계산하여 정규 근사(AN) 신뢰영역을 설정하고자 한다.

[방법1] 이변량 지수 공정 하에서, 구체적으로 계산 가능한 분산-공분산 행렬의 추정량을 이용

[방법2] 이변량 확률 공정 하에서, 분산-공분산 행렬의 적률 추정을 이용한 추정량을 이용

3.2.1. 이변량 지수 공정 하에서의 근사 신뢰영역: [방법1]

만일 4차 중심적률 $\mu_{4x} = E(X-\mu_x)^4$과 $\mu_{4y} = E(X-\mu_y)^4$이 존재하는 이변량 지수 공정분포로부터 확률표본 $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$을 얻었다고 가정할 때, 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 $100(1-\alpha)\%$ 정규 근사(AN) 신뢰영역은 다음과 같이 설정할 수 있다.

$$n\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right)'\left(\hat{\mathbf{V}}_{pkl}\right)^{-1}\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right) \le \chi^2(2;\alpha),$$

여기서 $\chi^2(2;\alpha)$는 자유도가 2인 χ^2-분포의 하위 $100(1-\alpha)$ 백분위수를 나타내고, 또한 추정행렬 $\hat{\mathbf{V}}_{pkl}$은 분산-공분산 행렬 $\mathbf{V}_{pkl}$의 추정량으로 이변량 지수 공정 하에서 구체적으로 제시하면 다음과 같다.

$$\begin{aligned}
\hat{\sigma}_{pkl11} &= \frac{1}{9}\left[1 - \left(1 - \hat{\lambda}_x \mathrm{LSL}_x\right)^2 + \left(\hat{\lambda}_x \mathrm{LSL}_x\right)^2\right], \\
\hat{\sigma}_{pkl22} &= \frac{1}{9}\left[1 - \left(1 - \hat{\lambda}_y \mathrm{LSL}_y\right)^2 + \left(\hat{\lambda}_y \mathrm{LSL}_y\right)^2\right], \\
\hat{\sigma}_{pkl12} &= \hat{\sigma}_{pkl21} \\
&= \frac{\left(\hat{\lambda}_x + \hat{\lambda}_y - \hat{\lambda}\right)}{9\hat{\lambda}}\left[1 + \frac{\hat{\lambda}_y\left(\hat{\lambda}_y \mathrm{LSL}_y - 1\right) + \hat{\lambda}_x\left(\hat{\lambda}_x \mathrm{LSL}_x - 1\right)}{\hat{\lambda}} + \frac{2\hat{\lambda}_x\hat{\lambda}_y\left(\hat{\lambda}_x \mathrm{LSL}_x - 1\right)\left(\hat{\lambda}_y \mathrm{LSL}_y - 1\right)}{\hat{\lambda}^2}\right].
\end{aligned}$$

단, λ_x의 추정량 $\hat{\lambda}_x$는 이변량 지수분포로부터 $1/\bar{X}$이고, λ_y의 추정량 $\hat{\lambda}_y$는 $1/\bar{Y}$이며, λ의 추정량 $\hat{\lambda}$는 다음과 같다.

$$\hat{\lambda} = \hat{\lambda}_x + \hat{\lambda}_y - \hat{\lambda}_3 = \frac{1}{\bar{X}} + \frac{1}{\bar{Y}} - \frac{1}{\bar{W}}.$$

다음으로 분산-공분산 행렬 $\mathbf{V}_{pkl} = (\sigma_{pklij})_{2\times 2}$의 각 성분 계산 과정은 다음과 같다.

먼저 모수 λ인 일변량 지수분포함수에 대한 확률변수 X의 적률모함수 $M(t)$로부터 1차 적률 $E(X)$, 2차 적률 $E(X^2)$, 3차 적률 $E(X^3)$ 및 4차 적률 $E(X^4)$은 각각 다음과 같이 계산된다.

$$\begin{aligned}
E(X) &= M'(t)|_{t=0} = \lambda(\lambda - t)^{-2}|_{t=0} = \frac{1}{\lambda}, \\
E(X^2) &= M''(t)|_{t=0} = 2\lambda(\lambda - t)^{-3}|_{t=0} = \frac{2}{\lambda^2}, \\
E(X^3) &= M'''(t)|_{t=0} = 3!\lambda(\lambda - t)^{-4}|_{t=0} = \frac{6}{\lambda^3},
\end{aligned}$$

$$E(X^4) = M''''(t)|_{t=0} = 4!\lambda(\lambda - t)^{-5}|_{t=0} = \frac{24}{\lambda^4}.$$

또한 적률모함수 $M(t)$로부터 확률변수 X의 분산 σ_x^2과 4차 중심적률 μ_{4x}은 다음과 같이 쉽게 계산되어 진다.

$$\sigma_x^2 = E(X-\mu_x)^2 = E\left(X^2 - \mu_x^2\right) = \frac{2}{\lambda_x^2} - \frac{1}{\lambda_x^2} = \frac{1}{\lambda_x^2},$$

$$\begin{aligned}\mu_{4x} &= E(X-\mu_x)^4 = E\left(X^4 - 4X^3\mu_x + 6X^2\mu_x^2 - 4X\mu_x^3 + \mu_x^4\right)\\ &= \frac{24}{\lambda_x^4} - 4\frac{6}{\lambda_x^3}\cdot\frac{1}{\lambda_x} + 6\frac{2}{\lambda_x^2}\cdot\frac{1}{\lambda_x^2} - 4\frac{1}{\lambda_x^4} + \frac{1}{\lambda_x^4} = \frac{9}{\lambda_x^4}.\end{aligned}$$

마찬가지로, 확률변수 Y의 분산 σ_y^2과 4차 중심적률 μ_{4y}은 다음과 같다.

$$\sigma_y^2 = E(Y-\mu_y)^2 = \frac{1}{\lambda_y^2}, \qquad \mu_{4y} = E(Y-\mu_y)^4 = \frac{9}{\lambda_y^4}.$$

이제 분산-공분산 행렬 $\mathbf{V}_{pkl}$의 각 성분을 보다 구체적으로 계산하기로 하자. 먼저 σ_{pkl11}를 구하면 다음과 같다.

$$\begin{aligned}\sigma_{pkl11} &= \mathrm{Var}\left(\frac{1}{3\sigma_x^2}\left[\sigma_x Z_1 + \frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x}Z_3\right]\right)\\ &= \frac{1}{9\sigma_x^4}\left[\sigma_x^2\sigma_x^2 + \left(\frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x}\right)^2\left(\mu_{4x} - \sigma_x^4\right) + (\mathrm{LSL}_x - \mu_x)\mu_{3x}\right]\\ &= \frac{\lambda_x^4}{9}\left[\frac{1}{\lambda_x^4} + \left(\frac{\mathrm{LSL}_x - 1/\lambda_x}{2(1/\lambda_x)}\right)^2\left(\frac{9}{\lambda_x^4} - \frac{1}{\lambda_x}\right) + \left(\mathrm{LSL}_x - \frac{1}{\lambda_x}\right)\frac{2}{\lambda_x^3}\right]\\ &= \frac{1}{9}\left[1 - (1-\lambda_x\mathrm{LSL}_x)^2 + (\lambda_x\mathrm{LSL}_x)^2\right].\end{aligned}$$

이와 마찬가지로 유사한 계산 절차에 의해 σ_{pkl22}를 구하면 다음과 같다.

$$\sigma_{pkl22} = \mathrm{Var}\left(\frac{1}{3\sigma_y^2}\left[\sigma_y Z_2 + \frac{\mathrm{LSL}_y - \mu_y}{2\sigma_y}Z_4\right]\right) = \frac{1}{9}\left[1 - \left(1-\lambda_y\mathrm{LSL}_y\right)^2 + \left(\lambda_y\mathrm{LSL}_y\right)^2\right].$$

다음으로 σ_{pkl12}와 σ_{pkl21}을 구하면 다음과 같다.

$$\begin{aligned}\sigma_{pkl12} &= \sigma_{pkl21}\\ &= \frac{1}{9\sigma_x^2\sigma_y^2}\mathrm{Cov}\left(\left[\sigma_x Z_1 + \frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x}Z_3\right], \left[\sigma_y Z_2 + \frac{\mathrm{LSL}_y - \mu_y}{2\sigma_y}Z_4\right]\right)\\ &= \frac{1}{9\sigma_x^2\sigma_y^2}\left[\sigma_x\sigma_y\mathrm{Cov}(Z_1, Z_2) + \sigma_x\frac{\mathrm{LSL}_y - \mu_y}{2\sigma_y}\mathrm{Cov}(Z_1, Z_4) + \sigma_y\frac{\mathrm{LSL}_x - \mu_x}{2\sigma_x}\mathrm{Cov}(Z_2, Z_3)\right.\\ &\quad\left. + \frac{(\mathrm{LSL}_x - \mu_x)(\mathrm{LSL}_y - \mu_y)}{4\sigma_x\sigma_y}\mathrm{Cov}(Z_3, Z_4)\right]\\ &= \frac{\lambda_x^2\lambda_y^2}{9}\left[\frac{\mu_{1x1y}}{\lambda_x\lambda_y} + \frac{\lambda_y(\mathrm{LSL}_y - 1/\lambda_y)}{2\lambda_x}\mu_{1x2y} + \frac{\lambda_x(\mathrm{LSL}_x - 1/\lambda_x)}{2\lambda_y}\mu_{2x1y}\right.\\ &\quad\left. + \frac{\lambda_x\lambda_y(\mathrm{LSL}_x - 1/\lambda_x)(\mathrm{LSL}_y - 1/\lambda_y)}{4}\left(\mu_{2x2y} - \frac{1}{\lambda_x^2\lambda_y^2}\right)\right].\end{aligned}$$

위의 공분산 항을 구체적으로 계산하기 위하여 Marhall과 Olkin (1967) 이변량 지수분포의 고차적률에 관한 다음과 같은 식을 이용하고자 한다.

$$E\left(X^iY^j\right) = j\Gamma(i+1)\sum_{k=0}^{i-1}\frac{\Gamma(j+k)}{\Gamma(k+1)\gamma_1^{i-k}\lambda^{j+k}} + i\Gamma(j+1)\sum_{k=0}^{j-1}\frac{\Gamma(i+k)}{\Gamma(k+1)\gamma_2^{j-k}\lambda^{i+k}},$$

여기서 $\gamma_i = \lambda_i + \lambda_3 (i = 1, 2)$이고 $\lambda = \lambda_1 + \lambda_2 + \lambda_3$이다. 또한 이러한 기본 가정으로부터 γ_1과 γ_2는 각각 λ_x와 λ_y와 같으므로, 윗 식은 다시 다음과 같이 바꿔 쓸 수 있다.

$$E\left(X^iY^j\right) = j\Gamma(i+1)\sum_{k=0}^{i-1}\frac{\Gamma(j+k)}{\Gamma(k+1)\lambda_x^{i-k}\lambda^{j+k}} + i\Gamma(j+1)\sum_{k=0}^{j-1}\frac{\Gamma(i+k)}{\Gamma(k+1)\lambda_y^{j-k}\lambda^{i+k}}.$$

따라서 이 식으로부터 공분산 항의 $\mu_{1x1y}, \mu_{1x2y}, \mu_{2x1y}, \mu_{2x2y}$을 계산하는 과정에서 필요한 확률변수 X와 Y에 관련된 기대값들은 다음과 같다.

$$E(XY) = \Gamma(2)\frac{\Gamma(1)}{\Gamma(1)\lambda_x\lambda} + \Gamma(2)\frac{\Gamma(1)}{\Gamma(1)\lambda_y\lambda} = \frac{1}{\lambda_x\lambda} + \frac{1}{\lambda_y\lambda} = \frac{\lambda_x+\lambda_y}{\lambda_x\lambda_y\lambda},$$

$$E\left(X^2Y\right) = \Gamma(3)\sum_{k=0}^{1}\frac{\Gamma(1+k)}{\Gamma(k+1)\lambda_x^{2-k}\lambda^{1+k}} + 2\Gamma(2)\frac{\Gamma(2)}{\Gamma(1)\lambda_y\lambda^2}$$

$$= 2\left(\frac{1}{\lambda_x^2\lambda} + \frac{1}{\lambda_x\lambda^2}\right) + \frac{2}{\lambda_y\lambda^2} = \frac{2\left(\lambda_y\lambda + \lambda_x\lambda_y + \lambda_x^2\right)}{\lambda_x^2\lambda_y\lambda^2},$$

$$E\left(XY^2\right) = 2\Gamma(2)\frac{\Gamma(2)}{\Gamma(1)\lambda_x\lambda^2} + \Gamma(3)\sum_{k=0}^{1}\frac{\Gamma(1+k)}{\Gamma(k+1)\lambda_y^{2-k}\lambda^{1+k}}$$

$$= \frac{2}{\lambda_x\lambda^2} + 2\left(\frac{1}{\lambda_y^2\lambda} + \frac{1}{\lambda_y\lambda^2}\right) = \frac{2\left(\lambda_y^2 + \lambda_x\lambda + \lambda_x\lambda_y\right)}{\lambda_x\lambda_y^2\lambda^2},$$

$$E\left(X^2Y^2\right) = 2\Gamma(3)\sum_{k=0}^{1}\frac{\Gamma(2+k)}{\Gamma(k+1)\lambda_x^{2-k}\lambda^{2+k}} + 2\Gamma(3)\sum_{k=0}^{1}\frac{\Gamma(2+k)}{\Gamma(k+1)\lambda_y^{2-k}\lambda^{2+k}}$$

$$= 4\left(\frac{1}{\lambda_x^2\lambda^2} + \frac{2}{\lambda_x\lambda^3}\right) + 4\left(\frac{1}{\lambda_y^2\lambda^2} + \frac{2}{\lambda_y\lambda^3}\right) = \frac{4\left(\lambda_x^2\lambda + 2\lambda_x\lambda_y^2 + \lambda_x^2\lambda + 2\lambda_x^2\lambda_y\right)}{\lambda_x^2\lambda_y^2\lambda^3}.$$

그러므로 $\mu_{1x1y}, \mu_{1x2y}, \mu_{2x1y}, \mu_{2x2y}$는 각각 다음과 같다.

$$\mu_{1x1y} = E\left[(X-\mu_x)(Y-\mu_y)\right] = E(XY) - \mu_x\mu_y = \frac{\lambda_x+\lambda_y}{\lambda_x\lambda_y\lambda} - \frac{1}{\lambda_x\lambda_y} = \frac{\lambda_x+\lambda_y-\lambda}{\lambda_x\lambda_y\lambda},$$

$$\mu_{1x2y} = E\left[(X-\mu_x)(Y-\mu_y)^2\right] = E\left[\left(X-\mu_x\right)\left(Y^2 - 2\mu_yY + \mu_y^2\right)\right] = \frac{2(\lambda_x+\lambda_y-\lambda)}{\lambda_x\lambda_y\lambda^2},$$

$$\mu_{2x1y} = E\left[(X-\mu_x)^2(Y-\mu_y)\right] = E\left[\left(X^2 - 2\mu_xX + \mu_x^2\right)\left(Y-\mu_y\right)\right] = \frac{2(\lambda_x+\lambda_y-\lambda)}{\lambda_x\lambda_y\lambda^2},$$

$$\mu_{2x2y} = E\left[(X-\mu_x)^2(Y-\mu_y)^2\right] = E\left[\left(X^2 - 2\mu_xX + \mu_x^2\right)\left(Y^2 - 2\mu_yY + \mu_y^2\right)\right] = \frac{8(\lambda_x+\lambda_y-\lambda)}{\lambda_x\lambda_y\lambda^3} + \frac{1}{\lambda_x^2\lambda_y^2},$$

여기서 $\lambda_x = \lambda_1 + \lambda_3$, $\lambda_y = \lambda_2 + \lambda_3$이고 $\lambda = \lambda_1 + \lambda_2 + \lambda_3$이다.

따라서 분산-공분산 행렬 $\mathbf{V}_{pkl}$의 공분산 항인 $\sigma_{pkl12} = \sigma_{pkl21}$은 다음과 같다.

$$\begin{aligned}
\sigma_{pkl12} &= \sigma_{pkl21} \\
&= \frac{\lambda_x^2\lambda_y^2}{9}\left[\frac{\mu_{1x1y}}{\lambda_x\lambda_y} + \frac{\lambda_y(\mathrm{LSL}_y - 1/\lambda_y)}{2\lambda_x}\mu_{1x2y} + \frac{\lambda_x(\mathrm{LSL}_x - 1/\lambda_x)}{2\lambda_y}\mu_{2x1y}\right. \\
&\quad \left. + \frac{\lambda_x\lambda_y(\mathrm{LSL}_x - 1/\lambda_x)(\mathrm{LSL}_y - 1/\lambda_y)}{4}\left(\mu_{2x2y} - \frac{1}{\lambda_x^2\lambda_y^2}\right)\right] \\
&= \frac{(\lambda_x + \lambda_y - \lambda)}{9\lambda}\left[1 + \left\{\frac{\lambda_y(\lambda_y\mathrm{LSL}_y - 1) + \lambda_x(\lambda_x\mathrm{LSL}_x - 1)}{\lambda}\right\} + \frac{2\lambda_x\lambda_y(\lambda_x\mathrm{LSL}_x - 1)(\lambda_y\mathrm{LSL}_y - 1)}{\lambda^2}\right].
\end{aligned}$$

3.2.2. 이변량 확률 공정 하에서의 근사 신뢰영역: [방법2]

만일 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$과 $\mu_{4y} = E(X - \mu_y)^4$이 존재하는 이변량 공정분포로부터 확률표본 $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$을 얻었다고 가정할 때, 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 $100(1-\alpha)\%$ 정규 근사(AN) 신뢰영역은 다음과 같이 설정할 수 있다.

$$n\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right)'\left(\hat{\mathbf{V}}_{pkl}\right)^{-1}\left(\hat{\mathbf{C}}_{pkl} - \mathbf{C}_{pkl}\right) \le \chi^2(2;\alpha),$$

여기서 $\chi^2(2;\alpha)$는 자유도가 2인 χ^2-분포의 하위 $100(1-\alpha)$ 백분위수를 나타내고, 또한 추정행렬 $\hat{\mathbf{V}}_{pkl}$은 분산-공분산 행렬 $\mathbf{V}_{pkl}$의 추정량으로 이변량 확률 공정 하에서 구체적으로 제시하면 다음과 같다.

$$\hat{\sigma}_{pkl11} = \frac{1}{9\hat{\sigma}_x^4}\left[\hat{\sigma}_x^2\hat{\sigma}_x^2 + \left(\frac{\mathrm{LSL}_x - \hat{\mu}_x}{2\hat{\sigma}_x}\right)^2\left(\hat{\mu}_{4x} - \hat{\sigma}_x^4\right) + \left(\mathrm{LSL}_x - \hat{\mu}_x\right)\hat{\mu}_{3x}\right],$$

$$\hat{\sigma}_{pkl22} = \frac{1}{9\hat{\sigma}_y^4}\left[\hat{\sigma}_y^2\hat{\sigma}_y^2 + \left(\frac{\mathrm{LSL}_y - \hat{\mu}_y}{2\hat{\sigma}_y}\right)^2\left(\hat{\mu}_{4y} - \hat{\sigma}_y^4\right) + \left(\mathrm{LSL}_y - \hat{\mu}_y\right)\hat{\mu}_{3y}\right],$$

$$\begin{aligned}
\hat{\sigma}_{pkl12} &= \hat{\sigma}_{pkl21} \\
&= \frac{1}{9\hat{\sigma}_x^2\hat{\sigma}_y^2}\left[\hat{\sigma}_x\hat{\sigma}_y\hat{\mu}_{1x1y} + \frac{\hat{\sigma}_x\left(\mathrm{LSL}_y - \hat{\mu}_y\right)}{2\hat{\sigma}_y}\hat{\mu}_{1x2y} + \frac{\left(\mathrm{LSL}_x - \hat{\mu}_x\right)\left(\mathrm{LSL}_y - \hat{\mu}_y\right)}{4\hat{\sigma}_x\hat{\sigma}_y}\left(\hat{\mu}_{2x2y} - \hat{\sigma}_x^2\hat{\sigma}_y^2\right)\right. \\
&\quad \left. + \frac{\hat{\sigma}_y\left(\mathrm{LSL}_x - \hat{\mu}_x\right)}{2\hat{\sigma}_x}\hat{\mu}_{2x1y}\right].
\end{aligned}$$

단, 중심적률 $\mu_{ix}, \mu_{iy}(i = 3, 4)$, $\mu_{ixjy}(i = 1, 2)$의 추정량 $\hat{\mu}_{ix}, \hat{\mu}_{iy}, \hat{\mu}_{ixjy}$는 각각 다음과 같다.

$$\hat{\mu}_{ix} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^i, \qquad \hat{\mu}_{iy} = \frac{1}{n}\sum_{k=1}^{n}\left(Y_k - \bar{Y}\right)^i, \quad i = 3, 4,$$

$$\hat{\mu}_{ixjy} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^i\left(Y_k - \bar{Y}\right)^j, \quad i, j = 1, 2.$$

4. 이변량 지수 공정 하에서의 모의실험

실제적으로 치우침을 가진 공정 분포의 상황은 표면의 거칠음, 코팅 두께 등의 측정에서와 같이 그 측정값이 어떤 특정한 값보다 작을 수 없는 경우에 많이 나타난다. 이러한 비슷한 상황은 오염의 정도,

수명실험에서 고장이 날 때까지의 시간을 관측할 때에도 종종 나타난다. 이 때 공정 데이터는 낮은 값에 많이 몰려있지만 큰 값이 종종 나타나게 된다. 이와 같은 상황은 구멍 뚫는 공정의 경우에도 발생할 수 있다. 즉, 구멍의 직경을 조사하는 경우 대부분의 직경은 드릴의 직경과 비슷하거나 그 보다 조금 클 것이다. 그러나 가끔 기계의 진동, 무딘 도구, 잘못된 설치 등이 원인이 되어 뚫린 구멍의 직경이 예상외로 아주 커질 수 있다. 이렇듯 실제 공정에서는 이러한 치우침을 가진 분포의 발생을 고려할 수 있으며, 따라서 이 절에서는 이변량 지수분포(BVE(λ_1, λ_2, λ_3))를 갖는 공정의 가정 하에서, 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly})$에 대한 정규 근사(AN) 신뢰영역의 정확성과 효율성을 비교·분석하기 위하여 모의실험을 수행하였다.

먼저 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly})$에 대한 모의실험을 수행하는데 있어 공정 평균 (μ_x, μ_y)와 표준편차 (σ_x, σ_y)의 값은 동일한 값을 주었으며, 또한 공정능력지수를 고려하여 적당한 조건의 규격 하한 (LSL_x, LSL_y)을 통해 모의실험을 수행하였다. 이변량 지수분포의 모수 λ_1과 λ_2는 동일한 값으로서 0.01, 0.02, 0.1, 0.2, 0.3을 사용하였으며, λ_3은 λ_1과 λ_2에 따라서 조건별로 달리하였다. 각 쌍의 이변량 지수 공정 모수 (λ_1, λ_2, λ_3)에 대하여, 표본 크기 n은 소표본과 대표본에서의 검정 결과를 비교하기 위하여 각각 20, 30, 40, 60, 80을 사용하여 다양한 조건별로 모의실험을 수행하였다.

각 쌍의 이변량 지수 공정 모수 (λ_1, λ_2, λ_3)와 표본크기 n에 의해 이루어지는 모의실험 절차는 다음과 같다.

- 단계1: BVE(λ_1, λ_2, λ_3)로부터 크기가 $n(=20, 30, 40, 60, 80)$인 원래의 이변량 표본들을 생성한다.
- 단계2: 단계1에서 생성된 표본들을 이용하여 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 95% 근사 신뢰영역을 설정한다.
- 단계3: 단계2에서 설정된 근사 신뢰영역 안에 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$의 참값이 포함되는지를 결정한다.

위의 3단계로 이루어진 한 번의 모의실험 절차를 $N(= 1{,}000)$번 반복한다. 이로부터 95% 신뢰영역 안에 실제 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$의 참값이 포함되는 횟수의 비율 즉, 포함비율을 계산할 것이다. 이 때 95% 신뢰영역에 대한 기대빈도는 $p = 0.95$이고, $N = 1000$인 이항확률변수이므로 포함비율에 대한 99% 신뢰구간을 설정하면 다음과 같다.

$$0.95 \pm 2.576 \times \sqrt{\frac{0.95 \times 0.05}{1000}} \rightarrow (0.933,\ 0.967).$$

따라서 앞으로 제시될 표 1의 모의실험 수행 결과에서, 위의 범위 안에 들어가는 포함비율에 대해서는 '*' 표시를 하였다.

앞의 3절에서 제시한 추정 분산-공분산 행렬들을 기초로 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 정규 근사(AN) 신뢰영역은 다음과 같이 두 가지 방법으로 구분하여 설정·제시하였다.

[방법1] 이변량 지수 공정 하에서, 3.2.1절에서 제시한 구체적으로 계산된 분산-공분산 행렬의 추정량을 이용하는 경우

[방법2] 이변량 확률 공정 하에서, 3.2.2절에서 제시한 분산-공분산 행렬의 적률 추정을 이용한 추정량을 이용하는 경우

각각의 공정 모수 조건에 따라 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 95% 정규 근사(AN) 신뢰영역의 포함비율을 구한 결과가 다음의 표 1에 제시되어 있다.

표 1: $\mathbf{C}_{pkl}$에 대한 95% 정규 근사(AN) 신뢰영역의 포함비율 결과

$\lambda_1 = \lambda_2$	λ_3	$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	구분	표본크기(n)				
				20	30	40	60	80
0.30	0.10	(2.5, 2.5, 2.5, 2.5)		$\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly}) = (0.200, 0.200)$				
			[방법1]	0.970	0.981	0.976	0.987	0.986
			[방법2]	0.968	0.949*	0.935*	0.942*	0.940*
0.20	0.20	(2.5, 2.5, 2.5, 2.5)		$\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly}) = (0.200, 0.200)$				
			[방법1]	0.974	0.990	0.975	0.989	0.985
			[방법2]	0.963*	0.964*	0.943*	0.946*	0.945*
0.10	0.30	(2.5, 2.5, 2.5, 2.5)		$\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly}) = (0.200, 0.200)$				
			[방법1]	0.973	0.981	0.975	0.985	0.987
			[방법2]	0.967*	0.944*	0.940*	0.933*	0.927
0.02	0.08	(10, 10, 10, 10)		$\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly}) = (0.300, 0.300)$				
			[방법1]	0.928	0.937*	0.943*	0.936*	0.943*
			[방법2]	0.896	0.886	0.885	0.884	0.885
0.01	0.03	(25, 25, 25, 25)		$\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly}) = (0.300, 0.300)$				
			[방법1]	0.917	0.925	0.928	0.935*	0.944*
			[방법2]	0.879	0.876	0.883	0.881	0.871

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와 같다고 할 수 있음을 의미

전체적으로 [방법2]의 경우에는 25개 중 13개가 포함비율에 대한 99% 신뢰구간에 해당하는 범위 (0.933, 0.967) 내에 있는 반면, [방법1]의 경우에는 25개 중 단지 6개 만이 범위 내에 있는 것으로 나타났다. 이는 [방법2]에 의한 신뢰영역의 추정이 포함비율의 기대값인 0.95에 가까운 값을 갖음을 의미하는 것으로, 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 95% 신뢰영역의 추정에 있어서 [방법2] 즉, 분산-공분산 행렬의 적률추정을 이용한 추정량을 이용하는 것이 더 바람직하다고 할 수 있을 것이다. 이러한 결과는 이변량 지수분포로부터 생성된 난수들을 이용하였지만, 극한 확률분포를 이용하였기 때문에 충분히 효율성이 다를 수 있다고 판단된다.

결론적으로 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대한 95% 정규 근사 신뢰영역을 추정하는 경우에 [방법1]과 [방법2] 모두 합리적인 추정방법이라고 할 수 있지만, 두 가지 방법 모두 만족스럽다고 할 만한 실험결과는 아니라고 판단된다. 따라서 공정분포에 관계없이 이변량 벡터 공정능력지수에 대한 보다 효율적이고 로버스트한 신뢰영역의 추정을 위해 붓스트랩 등과 같은 기법의 적용과 더불어 포괄적인 모의실험 연구가 필요하다고 사료된다.

5. 결론

6시그마(Six Sigma)는 급격한 경영환경의 변화와 무한 경쟁시대 속에서 많은 기업들에게 품질에 대한 중요성을 더욱 크게 인식시켰으며, 최근 국내·외 많은 기업들이 도입·적용하여 상당한 경영성과를 보여주고 있다. 이에 따라 6시그마를 위한 공정능력분석은 매우 중요한 영역이라고 할 수 있으며, 관련된 많은 연구들이 현재까지도 활발히 수행되고 있다. 하지만 이러한 연구는 주로 일변량 공정능력지수들에 대한 것들로, 여러 개의 품질특성치를 갖는 다변량 공정능력을 위한 측도들에 대한 연구는 미흡한 편이라고 할 수 있다. 또한 실제 공정의 분포가 지수분포인 경우가 많음에도 불구하고 아직까지 공정분포를 지수분포로 가정한 경우의 공정능력지수에 관한 연구는 거의 없는 실정이라고 할 수 있다. 따라서 본 논문에서는 특정한 이변량 지수 공정 하에서 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대하여 점추정량과 관련된 극한분포 이론 등을 유도하여, 보다 유용한 근사 신뢰영역에 대하여 연구하였다.

본 논문에서의 주요 결론을 요약하면 다음과 같다.

첫째, 실제 현장에서 널리 사용되고 있는 공정능력지수 C_{pk}에서 지수분포의 특성을 고려하여 규격하한만을 고려한 형태인 공정능력지수 C_{pkl}을 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl} = (C_{pklx}, C_{pkly})$로 확장·정의하고, 이 지수의 플러그-인 추정량들과 관련된 극한 확률분포를 유도·연구하였다.

둘째, 이변량 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대해 Marshall과 Olkin (1967)의 이변량 지수분포 모형(BVE(λ_1, λ_2, λ_3))을 기초로 정규 근사 신뢰영역 추정문제에 관하여 두 가지 방법을 제안하였으며, 모의실험을 통하여 정규 근사(AN) 신뢰영역의 정확성과 효율성을 비교·분석하였다. 여러 공정 모수 조건을 통한 이변량 지수 공정 하에서의 모의실험을 수행한 결과, 대체적으로 분산-공분산 행렬의 적률추정에 의한 추정량을 이용한 경우의 수행결과가 비교적 좋은 것으로 나타났다. 두 가지 방법이 아주 만족스럽지는 않지만, 무난한 실험결과라고 사료된다. 따라서 벡터 공정능력지수 $\mathbf{C}_{pkl}$에 대하여 붓스트랩과 같은 방법을 적용함으로써 공정 분포에 관계없이 보다 효율적이고 로버스트한 신뢰영역의 추정문제를 추후 과제로 연구할 필요가 있다고 판단된다.

참고 문헌

Alt, F. B. and Smith, N. D. (1988). *Multivariate Process Control*, Handbook of Statistics, **7**, North-Holland, Amsterdam, 333–351.

Hubele, N. F., Shahriari, H. and Cheng, C. S. (1991). *A Bivariate Process Capability Vector, in Statistical Process Control in Manufacturing(J.B. Keats and D.C. Montgomery, eds.)*, Marcel Dekker, New York, 299–310.

Juran, J. M. (1974). *Juran's Quality Control Handbook*, 3rd ed., McGraw-Hill, New York.

Kane, V. E. (1986). Process capability indices, *Journal of Quality Technology*, **18**, 41–52.

Kocherlakota, S. and Kocherlakota, K. (1991). Process capability index: Bivariate normal distribution, *Communications in Statistics - Theory and Methods*, **20**, 2529–2547.

Kotz, S. and Johnson, N. L. (1993). *Process Capability Indices*, Chapman & Hall/CRC, London.

Marshall, A. W. and Olkin, I. (1967). A generalized bivariate exponential distribution, *Journal of Applied Probability*, **4**, 291–302.

Park, B. S. (2003). *More Efficient Confidence Regions of Vector-valued Process Capability Indices for Six Sigma*, A thesis for the degree of Doctor, Graduate School, Chungbuk National University, Cheongju.

2009년 3월 접수; 2009년 3월 채택

Statistical Estimation for Hazard Function and Process Capability Index under Bivariate Exponential Process

Joong-Jae Cho[1,a], Su-Mook Kang[a], Byoung-Sun Park[b]

[a]Department of Information & Statistics, Chungbuk National University
[b]Statistical Analysis Team, Korea National Statistical Office

Abstract

Higher sigma quality level is generally perceived by customers as improved performance by assigning a correspondingly higher satisfaction score. The process capability indices and the sigma level Z_{st} have been widely used in six sigma industries to assess process performance. Most evaluations on process capability indices focus on statistical estimation under normal process which may result in unreliable assessments of process performance. In this paper, we consider statistical estimation for bivariate VPCI(Vector-valued Process Capability Index) $\mathbf{C}_{pkl} = (C_{pklx}, C_{pkl x})$ under Marshall and Olkin (1967)'s bivariate exponential process. First, we derive some limiting distribution for statistical inference of bivariate VPCI $\mathbf{C}_{pkl}$. And we propose two asymptotic normal confidence regions for bivariate VPCI $\mathbf{C}_{pkl}$. The proposed method may be very useful under bivariate exponential process. A numerical result based on our proposed method shows to be more reliable.

Keywords: Vector-valued process capability index, bivariate exponential process, asymptotic confidence region, limiting distribution.

This work was supported by Chungbuk National University Research Grant in 2007.

[1] Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, Cheongju 361-763, Korea. E-mail: jjcho@chungbuk.ac.kr