

Bootstrapping Vector-valued Process Capability Indices¹⁾

Joong-Jae Cho²⁾ and Byoung-Sun Park³⁾

Abstract

In actual manufacturing industries, process capability analysis often entails characterizing or assessing processes or products based on more than one engineering specification or quality characteristic. Since these characteristics are related, it is a risky undertaking to represent variation of even a univariate characteristic by a single index. Therefore, the desirability of using vector-valued process capability index(PCI) arises quite naturally.

In this paper, some vector-valued PCIs $C_p = (C_{px}, C_{py})$, $C_{pk} = (C_{pkx}, C_{pky})$ and $C_{pm} = (C_{pmx}, C_{pmy})$ considering univariate PCIs C_p , C_{pk} and C_{pm} are studied. First, we propose some asymptotic confidence regions of our vector-valued PCIs with bootstrap. And we examine the performance of asymptotic confidence regions of our vector-valued PCIs C_p and C_{pk} under the assumption of bivariate normal distribution $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ and bivariate chi-square distribution *Bivariate* $\chi^2(5, 5, \rho)$.

Keywords : Vector-valued Process Capability Index, Asymptotic Distribution, Consistency of Bootstrap, Confidence Region

1. 서론

다양한 고객의 요구를 만족시킬 수 있는 고품질과 정밀한 제품을 생산하기 위하여 자동설비의 이용이 확산되고 있다. 이러한 상황에서 단지 규격만을 만족시키는 수준이 아닌 보다 정밀한 제조공정의 품질관리가 중요시 되고 있다. 따라서 제품의 개발 및 제조단계에서 생산공정의 품질변동 정도를 측정하고 규격과 비교 분석하여 변동의 폭을 감소시키기 위하여 제반 통계적 방법들이 이용되고 있으며, 특히 규격과 관련하여 공정능력을 평가하는 척도로서 공정능력지수(process capability index)가 널리 이용되고 있다. 최근에 주목을 받고 있는 '6시그마'에서는 현재의 공정능

1) This work was supported by grant No. R05-2001-000-00082-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

2) Basic Science Research Institute, Chungbuk National University, Cheongju, 361-763, Korea.
E-mail : jjcho@chungbuk.ac.kr

3) Korea Research Institute of Standards and Science, Daejeon, 305-600, Korea.
E-mail : bspark@kriss.re.kr

력과 시그마 수준을 파악한 다음 현수준 대비 높은 이상목표를 설정하고 이를 토대로 개선목표 달성을 위한 활동 방향을 설정하기 위해 공정능력지수를 활용하고 있을 만큼 그 중요성은 크다고 할 수 있다.

공정 상의 변화를 효과적으로 감지함으로써 공정 변화에 따른 공정능력을 보다 정확하게 평가하기 위하여 다양한 지표들이 소개되어졌고 점차 보완되어 가고 있으며, 최근까지도 활발히 연구되고 있다. 이러한 지표들 중에서 공정능력지수 C_p , C_{pk} , C_{pm} 이 사용과 해석의 편리성으로 인하여 현재 널리 사용되고 있으며, 이 지표들은 모두 공정 변화의 원인이 되는 특성치를 한 개만 고려한 경우로 공정능력의 평가를 위한 일차원적인 척도들이라고 할 수 있다.

하지만 품질관리의 문제점 중에서는 공정 변화의 원인이 되는 품질 특성치들이 여러 개인 경우가 대부분이라는 것이다. 과거에는 이러한 특성치들을 동시에 수집하는 것이 힘들었지만, 지금은 FMS(Flexible Manufacturing System) 도입에 따라 이루어진 설비자동화로 인해, 자동 데이터 설비와 온라인 컴퓨터의 도움으로 동시에 여러 특성치들을 모니터링 하는 것이 일반화되고 있다. 특히 이들 특성치들은 하나 하나 측정되지만, 특성치들 간에 독립적이라기 보다는 오히려 관련이 있어 결합적으로 분포되어 있다고 가정하는 것이 보다 현실적일 것이다. 이러한 경우에 단지 일차원적인 측면에서 공정능력을 평가하는 것은 그 공정의 능력을 제대로 파악하지 못할 수 있을 뿐만 아니라, 일변량인 경우의 변동조차 하나의 지수로 표현하는 것은 적절하지 않을 수 있다. 이 때 벡터 공정능력지수를 사용하는 것이 매우 자연스럽고 바람직할 것이다.

벡터 공정능력지수는 기존의 일차원적인 지표들을 2차원 이상으로 확장한 개념의 공정능력지수(Kotz와 Johnson ; 1993)로, Kocherlakota와 Kocherlakota(1991)는 이변량 정규분포 $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ 하에서 가장 간단한 공정능력지수의 추정량들인 \widehat{C}_{px} 와 \widehat{C}_{py} 의 결합 확률분포함수를 계산하고 통계적 추론 방향을 제시하였다. 또한 최근에 Park *et al.*(2002)는 기본적인 2차원 벡터 공정능력지수 C_p 와 C_{pk} 에 대한 근사 신뢰영역을 제안하고, 이변량 정규분포 가정 하에서 모의실험을 통하여 제안된 신뢰영역의 정확성과 효율성을 연구하였다.

한편 공정능력지수들에 관한 로버스트한 추정 문제와 관련하여 일차원 공정능력지수에 대한 연구는 매우 활발하다고 할 수 있다. 특히 공정능력지수 분야에서 붓스트랩 기법을 응용한 대표적인 연구로 Franklin과 Wasserman(1992) 등을 들 수 있다. 반면 벡터 공정능력지수와 관련된 연구는 아직까지 미흡한 실정이라고 할 수 있다.

따라서 본 논문에서는 벡터 공정능력지수들에 대한 통계적 추론을 위하여 붓스트랩 방법을 이용한 효율적인 신뢰영역 설정문제에 대하여 연구하였다. 먼저 2절에서는 현장에서 많이 이용되는 일차원 공정능력지수 C_p , C_{pm} 과 C_{pk} 의 개념을 2차원 벡터 공정능력지수 $C_p = (C_{px}, C_{py})$,

$C_{pm} = (C_{pmx}, C_{pmy})$ 와 $C_{pk} = (C_{pkx}, C_{pky})$ 로 확장하여 이 지수의 플러그-인 추정량들과 관련된 극한 확률분포들을 소개하고, 이변량 정규분포 및 이변량 카이제곱 분포 하에서의 공분산 행렬을 계산하여 정규 근사(AN) 신뢰영역들을 제안하였다. 3절에서는 적당한 붓스트랩 알고리즘과 위의 지표들에 대한 붓스트랩 일치성을 소개하고, 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역들을 설정하여 제시하였다. 그리고 4절에서는 이변량 정규분포와 이변량 카이제곱 분포 하에서 앞서 제시한 2차원 벡터 공정능력지수 C_p 와 C_{pk} 에 대한 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역들의 정확성과 효율성 문제를 검토하기 위하여 다양한 모의실험 수

행 결과들을 제시하고 비교·분석하였다. 마지막으로 5절에서는 본 논문의 결론 및 향후 연구과제에 대하여 논의하였다.

2. 벡터 공정능력지수와 정규 근사 신뢰영역

2.1 추정량과 정규 근사 신뢰영역

이 절에서는 가장 기본적이고 많이 사용되고 있는 중요한 일차원 공정능력지수들인

$$C_p = \frac{USL - LSL}{6\sigma} \text{ 와 } C_{pm} = \frac{USL - LSL}{6\sqrt{E(X - T)^2}}$$

그리고

$$C_{pk} = \frac{\text{Min}(\mu - LSL, USL - \mu)}{3\sigma} = \frac{d - |\mu - M|}{3\sigma}$$

들을 기초로 하여 2차원 벡터 공정능력지수 $C_p = (C_{px}, C_{py})$ 와 $C_{pm} = (C_{pmx}, C_{pmy})$, 그리고

$C_{pk} = (C_{pkx}, C_{pky})$ 에 대하여 점추정량과 관련된 극한 분포이론 등을 연구하여, 이를 바탕으로 보다 간편하고 훨씬 유용한 근사 신뢰영역에 대하여 논의하게 될 것이다.

먼저 각각의 공정 평균 μ_x, μ_y 와 공정 분산 σ_x^2, σ_y^2 을 갖고 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$ 과 $\mu_{4y} = E(Y - \mu_y)^4$ 이 유한인 이변량 공정 모집단으로부터 2차원 벡터 공정능력지수 C_p 와 C_{pm} 그리고 C_{pk} 는 다음과 같이 정의된다(Kotz와 Johnson; 1993, Kocherlakota와 Kocherlakota; 1991).

$$C_p = (C_{px}, C_{py}) = \left(\frac{d_x}{3\sigma_x}, \frac{d_y}{3\sigma_y} \right)$$

$$C_{pm} = (C_{pmx}, C_{pmy}) = \left(\frac{d_x}{3\sqrt{E(X - T_x)^2}}, \frac{d_y}{3\sqrt{E(Y - T_y)^2}} \right)$$

$$C_{pk} = (C_{pkx}, C_{pky}) = \left(\frac{d_x - |\mu_x - M_x|}{3\sigma_x}, \frac{d_y - |\mu_y - M_y|}{3\sigma_y} \right)$$

단, USL_x, LSL_x 와 USL_y, LSL_y 는 각각 공정 변수 X 와 Y 에 대한 규격상한과 규격하한을 나타낸다. 또한 d_x 와 d_y 및 M_x 와 M_y 는 각각

$$d_x = \frac{USL_x - LSL_x}{2}, \quad d_y = \frac{USL_y - LSL_y}{2},$$

$$M_x = \frac{USL_x + LSL_x}{2}, \quad M_y = \frac{USL_y + LSL_y}{2}$$

로 정의되고, T_x 와 T_y 는 각각 품질특성치 X 와 Y 에 대한 목표값을 뜻한다.

한편 다음과 같이 플러그-인 방법(Plug-in method)에 의해 정의되어지는 추정량 \widehat{C}_p 와 \widehat{C}_{pm} , 그리고 \widehat{C}_{pk} 을 생각해보자.

$$\begin{aligned} \widehat{\mathbf{C}}_p &= (\widehat{C}_{px}, \widehat{C}_{py}) = \left(\frac{d_x}{3S_x}, \frac{d_y}{3S_y} \right) \\ \widehat{\mathbf{C}}_{pm} &= (\widehat{C}_{pmx}, \widehat{C}_{pmy}) = \left(\frac{d_x}{3\sqrt{S_x^2 + (\bar{X} - T_x)^2}}, \frac{d_y}{3\sqrt{S_y^2 + (\bar{Y} - T_y)^2}} \right) \\ \widehat{\mathbf{C}}_{pk} &= (\widehat{C}_{pkx}, \widehat{C}_{pky}) = \left(\frac{d_x - |\bar{X} - M_x|}{3S_x}, \frac{d_y - |\bar{Y} - M_y|}{3S_y} \right) \end{aligned}$$

단, 표본평균 \bar{X} 와 \bar{Y} 는

$$\widehat{\mu}_x = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \widehat{\mu}_y = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

로 정의하고, 표본분산 S_x^2 와 S_y^2 은 다음과 같이 정의한다.

$$\widehat{\sigma}_x^2 = S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \widehat{\sigma}_y^2 = S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

만일 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$ 과 $\mu_{4y} = E(Y - \mu_y)^4$ 이 존재하는 적당한 이변량 확률분포로부터 확률표본 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 을 얻었다고 가정할 때, 최근에 연구된 Park et al.(2002)의 극한분포 결과를 이용한 기본적인 2차원 벡터 공정능력지수 \mathbf{C}_p 와 \mathbf{C}_{pk} 에 대한 $100(1-\alpha)\%$ 정규 근사 신뢰영역을 살펴보면 다음과 같다.

첫째, 벡터 공정능력지수 \mathbf{C}_p 에 대한 $100(1-\alpha)\%$ 정규 근사 신뢰영역은 다음과 같다.

$$n(\widehat{\mathbf{C}}_p - \mathbf{C}_p)'(\widehat{\mathbf{V}}_p)^{-1}(\widehat{\mathbf{C}}_p - \mathbf{C}_p) \leq \chi^2(2; \alpha)$$

단, $\chi^2(2; \alpha)$ 는 자유도가 2인 χ^2 -분포의 하위 $100(1-\alpha)\%$ 백분위수를 나타내고, 또한 추정행렬 $\widehat{\mathbf{V}}_p$ 는 행렬 \mathbf{V}_p 의 플러그-인 추정량을 나타낸다.

예를 들어, 공정 품질특성치들이 이변량 정규분포 $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ 로 잘 적합된다면, 다음의 극한 확률변수(Park et al.(2002)의 [정리1])

$$\left(\frac{-d_x Z_3}{6\sigma_x^3}, \frac{-d_y Z_4}{6\sigma_y^3} \right)$$

를 고려하여 추정행렬 $\widehat{\mathbf{V}}_p = (\widehat{\sigma}_{pjj})_{2 \times 2} (i, j=1, 2)$ 의 각 성분들은 다음과 같이 계산된다.

$$\widehat{\sigma}_{p11} = \widehat{Var}\left(-\frac{d_x Z_3}{6\sigma_x^3}\right) = \frac{d_x^2}{36\sigma_x^6} (\widehat{\mu}_{4x} - \widehat{\sigma}_x^4) = \frac{d_x^2}{36\sigma_x^6} \cdot 2\widehat{\sigma}_x^4 = \frac{d_x^2}{18\sigma_x^2},$$

$$\widehat{\sigma}_{p22} = \widehat{Var}\left(-\frac{d_y Z_4}{6\sigma_y^3}\right) = \frac{d_y^2}{36\sigma_y^6} (\widehat{\mu}_{4y} - \widehat{\sigma}_y^4) = \frac{d_y^2}{36\sigma_y^6} \cdot 2\widehat{\sigma}_y^4 = \frac{d_y^2}{18\sigma_y^2},$$

$$\widehat{\sigma}_{p12} = \widehat{\sigma}_{p21} = \frac{d_x d_y}{36\sigma_x^3 \sigma_y^3} \widehat{Cov}((X - \mu_x)^2, (Y - \mu_y)^2) = \frac{d_x d_y}{36\sigma_x^3 \sigma_y^3} \cdot 2\widehat{\rho}^2 \widehat{\sigma}_x^2 \widehat{\sigma}_y^2 = \frac{d_x d_y \widehat{\rho}^2}{18\sigma_x \sigma_y}$$

둘째, 벡터 공정능력지수 \mathbf{C}_{pk} 에 대한 $100(1-\alpha)\%$ 정규 근사 신뢰영역은 다음과 같다.

$$n(\widehat{\mathbf{C}}_{pk} - \mathbf{C}_{pk})'(\widehat{\mathbf{V}}_{pk})^{-1}(\widehat{\mathbf{C}}_{pk} - \mathbf{C}_{pk}) \leq \chi^2(2; \alpha)$$

단, 추정 행렬 $\widehat{\mathbf{V}}_{pk}$ 는 행렬 \mathbf{V}_{pk} 의 플러그-인 추정량으로 μ_x 와 M_x 의 크기 및 μ_y 와 M_y 의 크기에 따라 구체적으로 계산될 수 있을 것이다.

예를 들어, 공정 품질특성치들이 이변량 정규분포 $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ 로 잘 적합된다면, $\mu_x = M_x$ 이고 $\mu_y < M_y$ 인 경우에 다음의 극한 확률변수(Park et al.(2002)의 [정리2])

$$\left(-\frac{d_x Z_3}{6\sigma_x^3} - \frac{|Z_1|}{3\sigma_x}, \frac{(|M_y - \mu_y| - d_y)Z_4}{6\sigma_y^3} + \frac{Z_2}{3\sigma_y} \right)$$

를 고려하여 추정행렬 $\widehat{\mathbf{V}}_{pk} = (\widehat{\sigma}_{pkij})_{2 \times 2} (i, j = 1, 2)$ 의 각 성분들은 다음과 같이 계산된다.

$$\begin{aligned} \widehat{\sigma}_{pk11} &= \widehat{Var}\left(-\frac{d_x Z_3}{6\sigma_x^3} - \frac{|Z_1|}{3\sigma_x}\right) = \frac{d_x^2}{36\widehat{\sigma}_x^6}(\widehat{\mu}_{4x} - \widehat{\sigma}_x^4) + \frac{\pi-2}{9\pi} \\ &= \frac{d_x^2}{36\widehat{\sigma}_x^6} \cdot 2\widehat{\sigma}_x^4 + \frac{\pi-2}{9\pi} = \frac{d_x^2}{18\widehat{\sigma}_x^2} + \frac{\pi-2}{9\pi} \end{aligned}$$

$$\begin{aligned} \widehat{\sigma}_{pk22} &= \widehat{Var}\left(\frac{(|M_y - \mu_y| - d_y)Z_4}{6\sigma_y^3} + \frac{Z_2}{3\sigma_y}\right) = \frac{(M_y - \bar{Y} - d_y)^2}{36\widehat{\sigma}_y^6}(\widehat{\mu}_{4y} - \widehat{\sigma}_y^4) + \frac{1}{9} \\ &= \frac{(M_y - \bar{Y} - d_y)^2}{36\widehat{\sigma}_y^6} \cdot 2\widehat{\sigma}_y^4 + \frac{1}{9} = \frac{(M_y - \bar{Y} - d_y)^2}{18\widehat{\sigma}_y^2} + \frac{1}{9} \end{aligned}$$

$$\begin{aligned} \widehat{\sigma}_{pk12} &= \widehat{\sigma}_{pk21} = -\frac{d_x(M_y - \bar{Y} - d_y)}{36\widehat{\sigma}_x^3\widehat{\sigma}_y^3} \widehat{Cov}((X - \mu_x)^2, (Y - \mu_y)^2) - \frac{d_x}{18\widehat{\sigma}_x^3\widehat{\sigma}_y} \widehat{Cov}((X - \mu_x)^2, Y) \\ &= -\frac{d_x(M_y - \bar{Y} - d_y)}{36\widehat{\sigma}_x^3\widehat{\sigma}_y^3} \cdot 2\widehat{\sigma}_x^2\widehat{\sigma}_y^2\rho^2 = -\frac{d_x(M_y - \bar{Y} - d_y)\widehat{\rho}^2}{18\widehat{\sigma}_x\widehat{\sigma}_y} \end{aligned}$$

단, 4차 중심적률 μ_{4x} 와 μ_{4y} 의 추정량은 각각

$$\widehat{\mu}_{4x} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4, \quad \widehat{\mu}_{4y} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^4$$

으로 정의되고, 또한 상관계수 ρ 의 추정량은 표본상관계수 $\widehat{\rho}$ 을 의미한다.

이변량 정규분포 가정 하에서의 보다 구체적인 공분산 행렬에 대해서는 Park et al.(2002)을 참조하길 바란다.

2.2 이변량 카이제곱 분포하에서의 분산-공분산 행렬

이 절에서는 공정의 분포가 이변량 카이제곱 분포인 경우, 앞 절의 신뢰영역 추정 문제에서 중요한 역할을 하는 분산-공분산 행렬의 성분들을 구체적으로 살펴보기로 하자. 물론 공분산 행렬의 추정량은 플러그-인 방법에 의해 쉽게 추정될 것이다.

만일 4차 중심적률 $\mu_{4z} = E(Z - \mu_z)^4$ 과 $\mu_{4w} = E(W - \mu_w)^4$ 이 존재하는 이변량 정규분포 $BN(0, 0, 1, 1, \rho_{zw})$ 로부터 확률표본 $(Z_1, W_1), (Z_2, W_2), \dots, (Z_n, W_n)$ 을 얻었다고 하고, 확률변수 U 와 V 를 각각 다음과 같이 정의하도록 하자.

$$U = \sum_{i=1}^n Z_i^2, \quad V = \sum_{i=1}^n W_i^2$$

그러면 U 와 V 는 각각 자유도가 n 인 카이제곱 분포를 따르고, (U, V) 는 이변량 카이제곱 분포 $Bivariate \chi^2(n, n, \rho)$ 를 따르게 된다. 이 때 U 와 V 의 상관계수는 $\rho = \rho_{zw}^2$ 이 성립한다.

이제 확률변수 U 와 V 를 기초로 각각 다음과 같은 변환을 고려하자.

$$X = \frac{\sigma_x}{\sqrt{2n}} U + \mu_x - \frac{n\sigma_x}{\sqrt{2n}}, \quad Y = \frac{\sigma_y}{\sqrt{2n}} V + \mu_y - \frac{n\sigma_y}{\sqrt{2n}}$$

위와 같이 변환된 이변량 카이제곱분포 $Bivariate \chi^2(n, n, \rho)$ 로부터 확률표본 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 을 얻었다고 하자. 그러면 Cho et al.(2001)의 [보조정리1]을 이용하여, 표본의 크기 n 이 $n \rightarrow \infty$ 일 때, 다음의 결과를 얻는다.

$$\begin{aligned} & (Z_{1n}, Z_{2n}, Z_{3n}, Z_{4n}) \\ &= (\sqrt{n}(\bar{X} - \mu_x), \sqrt{n}(\bar{Y} - \mu_y), \sqrt{n}(S_x^2 - \sigma_x^2), \sqrt{n}(S_y^2 - \sigma_y^2)) \\ &\xrightarrow{d} (Z_1, Z_2, Z_3, Z_4) = MN(\mathbf{0}, \Sigma_{4 \times 4}) \end{aligned}$$

단,

$$\Sigma_{4 \times 4} = (\sigma_{ij})_{4 \times 4} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y & \frac{4\sigma_x^3}{\sqrt{2n}} & \frac{4\sigma_x\sigma_y^2\rho}{\sqrt{2n}} \\ & \sigma_y^2 & \frac{4\sigma_x^2\sigma_y\rho}{\sqrt{2n}} & \frac{4\sigma_y^3}{\sqrt{2n}} \\ & & \text{symm.} & \\ & & & \frac{2(n+6)\sigma_x^4}{n} & \frac{\{(n+5)\rho^2 + 8\rho - n + 1\}\sigma_x^2\sigma_y^2}{n} \\ & & & & \frac{2(n+6)\sigma_y^4}{n} \end{pmatrix}$$

위의 결과를 바탕으로, 벡터 공정능력지수들의 추정량들의 극한분포와 관련된 분산-공분산 행렬 구조를 파악하면 다음과 같다.

• 이변량 카이제곱분포 하에서 분산-공분산 행렬 V_ρ 의 구조

신뢰영역 추정시 필요한 분산-공분산 행렬 V_ρ 의 각 성분은 다음과 같이 계산된다.

$$\sigma_{\rho 11} = \frac{(n+6)d_x^2}{18n\sigma_x^2}, \quad \sigma_{\rho 22} = \frac{(n+6)d_y^2}{18n\sigma_y^2}, \quad \sigma_{\rho 12} = \sigma_{\rho 21} = \frac{d_x d_y \{(n+5)\rho^2 + 8\rho - n + 1\}}{36n\sigma_x\sigma_y}$$

• 이변량 카이제곱분포 하에서 분산-공분산 행렬 V_{pm} 의 구조

신뢰영역 추정시 필요한 분산-공분산 행렬 V_{pm} 의 각 성분은 다음과 같이 계산된다.

$$\begin{aligned} \sigma_{pm11} &= \frac{d_x^2 \sigma_x^2}{18 \tau_x^6} \left[\frac{(n+6) \sigma_x^2}{n} + 2(\mu_x - T_x)^2 + \frac{2\sigma_x}{\sqrt{2n}} \right], \\ \sigma_{pm22} &= \frac{d_y^2 \sigma_y^2}{18 \tau_y^6} \left[\frac{(n+6) \sigma_y^2}{n} + 2(\mu_y - T_y)^2 + \frac{2\sigma_y}{\sqrt{2n}} \right], \\ \sigma_{pm12} &= \sigma_{pm21} = \frac{d_x d_y \sigma_x \sigma_y}{36 \tau_x^3 \tau_y^3} \left[\frac{\{(n+5) \rho^2 + 8\rho - n + 1\} \sigma_x \sigma_y}{n} + \frac{8\rho(\mu_y - T_y) \sigma_x}{\sqrt{2n}} \right. \\ &\quad \left. + \frac{8\rho(\mu_x - T_x) \sigma_y}{\sqrt{2n}} + 4\rho(\mu_x - T_x)(\mu_y - T_y) \right] \end{aligned}$$

• 이변량 카이제곱분포 하에서 분산-공분산 행렬 V_{pk} 의 구조

μ_x 와 M_x 의 크기 및 μ_y 와 M_y 의 크기에 따른 여러 조건들 중에서 $\mu_x = M_x$ 이고 $\mu_y \neq M_y$ 인 경우를 중심으로 논하는 경우에, 신뢰영역 추정시 필요한 분산-공분산행렬 V_{pk} 의 각 성분들은 다음과 같이 계산된다.

$$\begin{aligned} \sigma_{pk11} &= \frac{(n+6) d_x^2}{18 n \sigma_x^2} + \frac{\pi-2}{9\pi}, \\ \sigma_{pk22} &= \frac{(n+6) (|M_y - \mu_y| - d_y)^2}{18 n \sigma_y^2} + \frac{4(|M_y - \mu_y| - d_y) \text{sgn}(M_y - \mu_y)}{9\sqrt{2n} \sigma_y} + \frac{1}{9}, \\ \sigma_{pk12} &= \sigma_{pk21} = -\frac{d_x (|M_y - \mu_y| - d_y) \{(n+5) \rho^2 + 8\rho - n + 1\}}{36 n \sigma_x \sigma_y} - \frac{2 d_x \text{sgn}(M_y - \mu_y) \rho}{9\sqrt{2n} \sigma_x} \end{aligned}$$

3. 붓스트랩과 신뢰영역

붓스트랩 방법은 여러 분야의 중요한 통계문제에서 이론적으로 혹은 응용적인 측면에서 쓰임새가 많은 효율적인 강력한 도구이다. Hall(1988)은 여러 가지 붓스트랩 신뢰구간들을 이론적으로 폭넓게 비교·연구하였다. 그리고 붓스트랩 관련 공정능력지수들에 대한 구간추정 연구결과들로서, Franklin과 Wasserman(1992)은 공정능력지수 C_p , C_{pk} , C_{pm} 에 대한 하한 신뢰한계에 대해서 3가지 형태의 붓스트랩 방법 즉, SB(Standard Bootstrap), PB(Percentile Bootstrap), BCPB(Bias-Corrected Percentile Bootstrap)를 기초로 포괄적으로 비교·연구하였다. 최근 들어, Cho, Park과 Kim(1999)은 생산 현장에서 가장 많이 사용되고 있는 공정능력지수 C_{pk} 에 대하여 보다 효율적인 붓스트랩 신뢰구간 추정문제를 심도있게 연구하였다.

3.1 붓스트랩 알고리즘

우선, 본 논문에서 사용하게 될 벡터 공정능력지수들을 보다 효과적으로 추정하기 위한 붓스트랩 알고리즘을 소개하면 다음과 같다.

[1단계] 주어진 확률표본 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 으로부터 복원추출방법에 의해 m 개의 붓스트랩 표본 $(X_1^*, Y_1^*), (X_2^*, Y_2^*), \dots, (X_m^*, Y_m^*)$ 을 얻는다.

[2단계] [1단계]의 붓스트랩 표본들로부터 붓스트랩 표본평균 \bar{X}^* 와 \bar{Y}^* , 그리고 표본분산 S_x^{*2} 와 S_y^{*2} 를 다음과 같이 계산한다.

$$\bar{X}^* = \frac{1}{m} \sum_{i=1}^m X_i^*, \quad \bar{Y}^* = \frac{1}{m} \sum_{i=1}^m Y_i^*,$$

$$S_x^{*2} = \frac{1}{m-1} \sum_{i=1}^m (X_i^* - \bar{X}^*)^2, \quad S_y^{*2} = \frac{1}{m-1} \sum_{i=1}^m (Y_i^* - \bar{Y}^*)^2$$

[3단계] 벡터 공정능력지수들의 붓스트랩 추정량 \widehat{C}_p^* 와 \widehat{C}_{pm}^* , 그리고 \widehat{C}_{pk}^* 를 얻는다.

$$\widehat{C}_p^* = (\widehat{C}_{px}^*, \widehat{C}_{py}^*) = \left(\frac{d_x}{3S_x^*}, \frac{d_y}{3S_y^*} \right)$$

$$\widehat{C}_{pm}^* = (\widehat{C}_{pmx}^*, \widehat{C}_{pmy}^*) = \left(\frac{d_x}{3\sqrt{S_x^{*2} + (X^* - T_x)^2}}, \frac{d_y}{3\sqrt{S_y^{*2} + (Y^* - T_y)^2}} \right)$$

$$\widehat{C}_{pk}^* = (\widehat{C}_{pkx}^*, \widehat{C}_{pky}^*) = \left(\frac{d_x - |\bar{X}^* - M_x|}{3S_x^*}, \frac{d_y - |\bar{Y}^* - M_y|}{3S_y^*} \right)$$

이 때 붓스트랩 표본 크기 m 이 원래의 표본 크기 n 과 같은 경우에 n^n 개의 붓스트랩 추정량들이 계산 가능하며, 이들은 추정량 \widehat{C}_p , \widehat{C}_{pm} 와 \widehat{C}_{pk} 의 (완전한) 붓스트랩 확률분포의 구성요소가 될 것이다. 이러한 붓스트랩 알고리즘을 바탕으로 강력한 컴퓨터 모의실험을 통하여 2차원 벡터 공정능력지수 C_p , C_{pm} 와 C_{pk} 에 대한 효율적인 붓스트랩 신뢰영역을 설정하게 될 것이다.

3.2 붓스트랩 방법의 일치성

Park *et al.*(2002)의 중요한 연구결과인 [정리1], [정리2]와 붓스트랩 알고리즘 하에서 다음의 [정리1], [정리2]에 대응하는 극한분포들이 각각 서로 정확하게 일치하게 되는데, 이는 우리의 벡터 공정능력지수들에 대한 붓스트랩 통계적 추론을 정당화해 주는 중요한 조건으로서, 이러한 성질을 붓스트랩 방법의 일치성이라 한다.

[정리1] 만일 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$ 과 $\mu_{4y} = E(Y - \mu_y)^4$ 이 존재한다면, 원래의 표본의 크기 n 과 붓스트랩 표본의 크기 m 이 $m, n \rightarrow \infty$ 일 때, 다음 결과가 성립한다.

$$(a) \sqrt{m} (\widehat{C}_p^* - \widehat{C}_p) \mid \mathbf{x}_n = \sqrt{m} (\widehat{C}_{px}^* - \widehat{C}_{px}, \widehat{C}_{py}^* - \widehat{C}_{py}) \mid \mathbf{x}_n$$

$$\xrightarrow{d} \left(-\frac{d_x Z_3}{6\sigma_x^3}, -\frac{d_y Z_4}{6\sigma_y^3} \right) = MN(\mathbf{0}, \mathbf{V}_p)$$

$$(b) \sqrt{m} (\widehat{C}_{pm}^* - \widehat{C}_{pm}) \mid \mathbf{x}_n = \sqrt{m} (\widehat{C}_{pmx}^* - \widehat{C}_{pmx}, \widehat{C}_{pmy}^* - \widehat{C}_{pmy}) \mid \mathbf{x}_n$$

$$\begin{aligned} &\xrightarrow{d} \left(-\frac{d_x}{6\tau_x^3} \{Z_3 + 2(\mu_x - T_x)Z_1\}, -\frac{d_y}{6\tau_y^3} \{Z_4 + 2(\mu_y - T_y)Z_2\} \right) \\ &= MN(\mathbf{0}, \mathbf{V}_{pm}) \end{aligned}$$

단, \mathbf{x}_n 는 주어진 이변량 자료인 $\mathbf{x}_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ 이다.

증명 Cho et al.(2001) 참조 □

[정리2] 만일 4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$ 과 $\mu_{4y} = E(Y - \mu_y)^4$ 이 존재한다면, 원래의 표본의 크기 n 과 붓스트랩 표본의 크기 m 이 $m, n \rightarrow \infty$ 일 때, 다음 결과가 성립한다.

$$\begin{aligned} &\sqrt{m} (\widehat{\mathbf{C}}_{pk}^* - \widehat{\mathbf{C}}_{pk}) \mid \mathbf{x}_n = \sqrt{m} (\widehat{\mathbf{C}}_{pkx}^* - \widehat{\mathbf{C}}_{pkx}, \widehat{\mathbf{C}}_{pky}^* - \widehat{\mathbf{C}}_{pky}) \mid \mathbf{x}_n \\ &\xrightarrow{d} \left(-\frac{d_x Z_3}{6\sigma_x^3} - \frac{|Z_1|}{3\sigma_x}, -\frac{d_y Z_4}{6\sigma_y^3} - \frac{|Z_2|}{3\sigma_y} \right) = MN(\mathbf{0}, \mathbf{V}_{pk}), \mu_x = M_x \text{ \& } \mu_y = M_y \\ &\left(-\frac{d_x Z_3}{6\sigma_x^3} - \frac{|Z_1|}{3\sigma_x}, \frac{(|\mu_y - M_y| - d_y)Z_4}{6\sigma_y^3} + \frac{\text{sgn}(\mu_y - M_y)Z_2}{3\sigma_y} \right) \\ &\quad \xrightarrow{d} MN(\mathbf{0}, \mathbf{V}_{pk}'), \mu_x = M_x \text{ \& } \mu_y \neq M_y \\ &\left(\frac{(|\mu_x - M_x| - d_x)Z_3}{6\sigma_x^3} + \frac{\text{sgn}(\mu_x - M_x)Z_1}{3\sigma_x}, -\frac{d_y Z_4}{6\sigma_y^3} - \frac{|Z_2|}{3\sigma_y} \right) \\ &\quad \xrightarrow{d} MN(\mathbf{0}, \mathbf{V}_{pk}''), \mu_x \neq M_x \text{ \& } \mu_y = M_y \\ &\left(\frac{(|\mu_x - M_x| - d_x)Z_3}{6\sigma_x^3} + \frac{\text{sgn}(\mu_x - M_x)Z_1}{3\sigma_x}, \frac{(|\mu_y - M_y| - d_y)Z_4}{6\sigma_y^3} + \frac{\text{sgn}(\mu_y - M_y)Z_2}{3\sigma_y} \right) \\ &\quad \xrightarrow{d} MN(\mathbf{0}, \mathbf{V}_{pk}'''), \mu_x \neq M_x \text{ \& } \mu_y \neq M_y \end{aligned}$$

단, \mathbf{x}_n 는 주어진 이변량 자료인 $\mathbf{x}_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ 이다.

증명 Cho,J.J.와 Park,B.S.(2003) 참조 □

3.3 붓스트랩 신뢰영역

4차 중심적률 $\mu_{4x} = E(X - \mu_x)^4$ 과 $\mu_{4y} = E(Y - \mu_y)^4$ 이 존재하는 적당한 이변량 공정분포하에 서, 붓스트랩 방법의 일치성을 기초로 벡터 공정능력지수 \mathbf{C}_p , \mathbf{C}_{pm} 과 \mathbf{C}_{pk} 에 대한 3가지 형태의 $100(1 - \alpha)\%$ 붓스트랩 신뢰영역들을 설정할 수 있을 것이다.

앞으로 전개될 2차원 벡터 공정능력지수들에 대한 붓스트랩 신뢰영역을 설정함에 있어서, \mathbf{C} 를 벡터 공정능력지수 $\mathbf{C}_p, \mathbf{C}_{pm}$ 와 \mathbf{C}_{pk} 에 대한 대표적인 표기로 사용하자. 그리고 $\widehat{\mathbf{C}}$ 는 원래의 표본 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 로부터 얻어지는 벡터 공정능력지수 \mathbf{C} 의 플러그-인 추정량을 나타내고, $\widehat{\mathbf{C}}^*$ 는 붓스트랩 표본 $(X_1^*, Y_1^*), (X_2^*, Y_2^*), \dots, (X_n^*, Y_n^*)$ 로부터 얻어지는 벡터 공정능력지수 \mathbf{C} 의 붓스트랩 추정량을 나타낸다고 하자.

(1) 표준 붓스트랩 방법(Standard Bootstrap : SB)

이 방법을 이용한 신뢰영역을 설정하기 위하여, 먼저 2차원 벡터 공정능력지수 \mathbf{C} 에 대한 B 개의 붓스트랩 추정치 $\widehat{\mathbf{C}}^* = (\widehat{C}_{xi}^*, \widehat{C}_{yi}^*), i=1, 2, \dots, B$ 들로부터 다음과 같이 붓스트랩 표본 평균 벡터 $\overline{\mathbf{C}}^*$ 와 표본 분산-공분산 행렬 $\mathbf{S}_{\mathbf{C}}^*$ 을 구한다.

$$\overline{\mathbf{C}}^* = \begin{pmatrix} \overline{C}_x^* \\ \overline{C}_y^* \end{pmatrix} = \begin{pmatrix} \frac{1}{B} \sum_{i=1}^B \widehat{C}_{xi}^* \\ \frac{1}{B} \sum_{i=1}^B \widehat{C}_{yi}^* \end{pmatrix},$$

$$\mathbf{S}_{\mathbf{C}}^* = \begin{pmatrix} s_{11}^* & s_{12}^* \\ s_{21}^* & s_{22}^* \end{pmatrix} = \begin{pmatrix} \frac{1}{B-1} \sum_{i=1}^B (\widehat{C}_{xi}^* - \overline{C}_x^*)^2 & \frac{1}{B-1} \sum_{i=1}^B (\widehat{C}_{xi}^* - \overline{C}_x^*)(\widehat{C}_{yi}^* - \overline{C}_y^*) \\ \text{symm.} & \frac{1}{B-1} \sum_{i=1}^B (\widehat{C}_{yi}^* - \overline{C}_y^*)^2 \end{pmatrix}$$

그러면 2차원 벡터 공정능력지수 \mathbf{C} 에 대한 $100(1-\alpha)\%$ 표준 붓스트랩(SB) 신뢰영역은 다음과 같이 설정할 수 있다.

$$(\widehat{\mathbf{C}} - \mathbf{C})'(\mathbf{S}_{\mathbf{C}}^*)^{-1}(\widehat{\mathbf{C}} - \mathbf{C}) \leq \chi_{(2;\alpha)}^2$$

여기서 $\chi_{(2;\alpha)}^2$ 는 자유도가 2인 χ^2 -분포의 하위 $100(1-\alpha)$ 백분위수를 나타낸다.

(2) 스튜던트화 붓스트랩 방법(Studentized Bootstrap : STUD)

이 방법은 $m(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})'(\widehat{\mathbf{V}}^*)^{-1}(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})$ 의 분포로부터 백분위 지점을 찾아 신뢰영역을 설정하는 것으로, 이론적인 근거는 앞서 논의한 붓스트랩 방법의 일치성이다. 2차원 벡터 공정능력지수 \mathbf{C} 에 대한 $100(1-\alpha)\%$ 스튜던트화 붓스트랩(STUD) 신뢰영역은 다음과 같이 설정할 수 있다.

$$n(\widehat{\mathbf{C}} - \mathbf{C})'(\widehat{\mathbf{V}})^{-1}(\widehat{\mathbf{C}} - \mathbf{C}) \leq \widehat{y}_\alpha$$

여기서 \widehat{y}_α 는

$$\Pr\{m(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})'(\widehat{\mathbf{V}}^*)^{-1}(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}}) \leq \widehat{y}_\alpha \mid \mathbf{x}_n\} = 1 - \alpha$$

를 만족하는 값이고, $\widehat{\mathbf{V}}$ 와 $\widehat{\mathbf{V}}^*$ 는 벡터 공정능력지수 \mathbf{C} 의 분산-공분산 행렬 \mathbf{V} 에 대응하는 추정량과 붓스트랩 추정량을 나타낸다.

(3) 하이브리드 붓스트랩 방법(Hybrid Bootstrap : HYB)

이 방법은 $m(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})'(\widehat{\mathbf{V}})^{-1}(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})$ 의 분포로부터 백분위 지점을 찾아 신뢰영역을 설정하는 것으로, 이론적인 근거는 앞서 논의한 붓스트랩 방법의 일치성이다. 2차원 벡터 공정능력지수 \mathbf{C} 에 대한 $100(1-\alpha)\%$ 하이브리드 붓스트랩(HYB) 신뢰영역은 다음과 같이 설정할 수 있다.

$$n(\widehat{\mathbf{C}} - \mathbf{C})'(\widehat{\mathbf{V}})^{-1}(\widehat{\mathbf{C}} - \mathbf{C}) \leq \widehat{x}_\alpha$$

여기서 \widehat{x}_α 는

$$\Pr\{m(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}})'(\widehat{\mathbf{V}})^{-1}(\widehat{\mathbf{C}}^* - \widehat{\mathbf{C}}) \leq \widehat{x}_\alpha \mid \mathbf{x}_n\} = 1 - \alpha$$

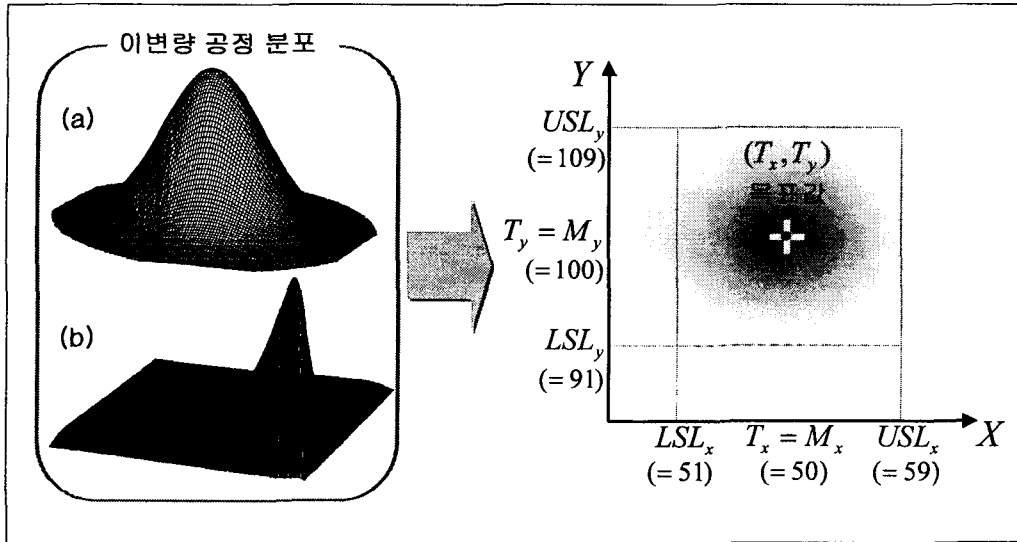
를 만족하는 값이고, $\widehat{\mathbf{V}}$ 는 벡터 공정능력지수 \mathbf{C} 의 분산-공분산 행렬 \mathbf{V} 에 대응하는 추정량을 나타낸다.

4. 모의실험 연구

이 절에서는 이변량 정규분포와 이변량 카이제곱분포 하에서, 앞 절에서 논의된 2차원 벡터 공정능력지수 \mathbf{C}_p 와 \mathbf{C}_{pk} 에 대한 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역에 대한 포괄적이고 다양한 모의실험을 통하여 정확성과 효율성을 비교·분석함으로써 2차원 벡터 공정능력지수에 대한 보다 효율적인 근사 신뢰영역 추정 방법에 관하여 논의하고자 한다.

4.1 모의실험 설계 및 절차

우선 모의실험을 수행함에 있어, 공정에서 관심을 두고 있는 두 품질특성 X 와 Y 에 대한 규격은 <그림 1>과 같이 설정하였다. 특히 X 와 Y 에 대한 목표값 T_x 와 T_y 는 각각의 품질특성에 대하여 설정된 규격의 중심 M_x 와 M_y 의 값과 동일하게 설정하였다.



<그림 1> 이변량 공정의 두 품질특성 X와 Y에 대한 규격

다음으로 공정의 분포는 Gunter(1989)의 연구 내용을 바탕으로 일반적으로 가정하고 있는 정규성을 지닌 이변량 정규분포 $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ (<그림 1> (a))와 치우침을 가진 분포로 각각의 자유도가 5인 이변량 카이제곱 분포 $Bivariate \chi^2(5, 5, \rho)$ (<그림 1> (b))를 고려하였다. 특히 이변량 카이제곱 분포에 대하여, 설정된 공정 평균 μ_x, μ_y 와 공정 표준편차 σ_x, σ_y , 그리고 상관계수 ρ 를 갖도록 하기 위하여 다음과 같은 변환을 실시하였다.

$$(X, Y) \sim Bivariate \chi^2(5, 5, \rho) \Rightarrow \left(\frac{\sigma_x}{\sqrt{10}} X + \mu_x - \frac{5\sigma_x}{\sqrt{10}}, \frac{\sigma_y}{\sqrt{10}} Y + \mu_y - \frac{5\sigma_y}{\sqrt{10}} \right)$$

그리고 공정 평균 μ_x, μ_y 와 공정 표준편차 σ_x, σ_y 의 값은 공정능력지수와 시그마 수준과의 관계를 고려하여 이변량 공정 (X, Y) 의 시그마 수준이 각각 3.0과 6.0 시그마 수준의 조합이 되도록 선택하였다. 또한 다양한 상관관계를 갖는 이변량 공정을 고려하기 위하여 X와 Y의 상관계수 ρ 의 값을 이변량 정규분포의 경우에는 -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9로, 이변량 카이제곱분포인 경우에는 0, 0.3, 0.6, 0.9를 선택하였다.

각 쌍의 이변량 공정 모수 $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ 에 대하여, 표본크기 n 은 30과 60을 사용하여 다양한 조건별로 모의실험을 수행하였다. 그리고 붓스트랩 표본크기 m 은 모의실험의 편리성을 위하여 n 과 동일한 값을 사용하였다.

먼저 2차원 벡터 공정능력지수 C_p 에 대한 모의실험에 사용된 공정 모수와 C_p 값들은 다음의 <표 1>과 같다. 다음으로 공정 평균 (μ_x, μ_y) 와 규격 중심 (M_x, M_y) 과의 차이를 고려한 2차원 벡터 공정능력지수 C_{pk} 에 대해서는 μ_x 와 M_x 의 크기 및 μ_y 와 M_y 의 크기에 따른 여러 조건들 중에서 대표적으로 $\mu_x < M_x$ 이고 $\mu_y < M_y$ 인 경우와 $\mu_x = M_x$ 이고 $\mu_y < M_y$ 인 경우를 선택하여 가능한 다양한 공정의 상태가 고려되도록 하였다. 구체적으로 모의실험에 사용된 공정 모수와 C_{pk} 값들은 다음의 <표 2>와 같다.

위의 모의실험 설계 하에서, 각 쌍의 이변량 공정 모수 $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ 와 표본 크기 n 에 의해서 이루어지는 한 번의 모의실험 절차는 다음과 같다.

[1단계] 상관계수가 ρ 인 이변량 공정 분포 $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ 로부터 크기가 $n(=30, 60)$ 인 원래의 이변량 표본들을 추출한 후, 이 원래의 표본으로부터 각각 붓스트랩 표본크기가 $m(=n)$ 인 $B(=1,000)$ 개의 붓스트랩 이변량 표본들을 생성한다.

[2단계] [1단계]에서 생성된 표본들을 이용하여, 2차원 벡터 공정능력지수 C 에 대한 극한분포를 활용한 95% 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역을 설정한다.

[3단계] [2단계]에서 설정된 신뢰영역 안에 2차원 벡터 공정능력지수 C 의 참값이 포함되는지를 결정한다.

위의 3단계로 이루어진 한 번의 모의실험 절차를 $N(=1,000)$ 번 반복한다. 이로부터 95% 신뢰영역 안에 실제 벡터 공정능력지수 C 의 참값이 포함되는 횟수의 비율을 계산하였다.

본 모의실험은 SAS Ver 8.1을 사용하여 수행하였으며, 특히 행렬연산을 위하여 SAS/IML를 이용하였다.

<표 1> 모의실험에 사용된 공정 모수와 C_p 값들

구 분	μ_x	μ_y	σ_x	σ_y	$C_p = (C_{px}, C_{py})$	시그마 수준
$\mu_x = M_x$ & $\mu_y = M_y$	50.0	100.0	3.0	3.0	(1.000, 1.000)	(3.0, 3.0)
			3.0	1.5	(1.000, 2.000)	(3.0, 6.0)
			1.5	1.5	(2.000, 2.000)	(6.0, 6.0)

<표 2> 모의실험에 사용된 공정 모수와 C_{pk} 값들

구 분	μ_x	μ_y	σ_x	σ_y	$C_{pk} = (C_{pkx}, C_{pky})$	시그마 수준
$\mu_x < M_x$ & $\mu_y < M_y$	45.5	95.5	3.0	3.0	(0.500, 0.500)	(3.0, 3.0)
			3.0	1.0	(0.500, 1.500)	(3.0, 6.0)
			1.0	1.0	(1.500, 1.500)	(6.0, 6.0)
$\mu_x = M_x$ & $\mu_y < M_y$	50.0	95.5	3.0	3.0	(1.000, 0.500)	(3.0, 3.0)
			3.0	1.0	(1.000, 1.500)	(3.0, 6.0)
			1.5	1.0	(2.000, 1.500)	(6.0, 6.0)

4.2 모의실험 결과 분석

앞 절에서 소개한 모의실험 설계 및 절차에 의하여 2차원 벡터 공정능력지수 C_p 와 C_{pk} 에 대한 95% 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역의 포함비율을 구한 결과가 <표 3>~<표 8>에 제시되어 있다. 먼저 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 모의실험 수행 결과표에서 사용된 ‘*’ 표시는 포함비율에 대한 99% 신뢰구간에 해당하는 위의 범위를 벗어나는 경우를 의미한다. 구체적으로 각각의 이변량 공정 분포에 대한 모의실험 수행결과를 살펴보자.

(1) 공정의 분포가 이변량 정규분포인 경우

공정의 분포가 이변량 정규분포일 때, 2차원 벡터 공정능력지수 C_p 와 C_{pk} 에 대한 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 방법에 대한 모의실험 수행결과가 각각 <표 3>~<표 5>에 제시되어 있다.

<표 3>의 C_p 에 대한 95% 신뢰영역의 포함비율 결과를 살펴보면, 모든 표본크기 n 에 대하여 정규 근사(AN) 방법과 표준 붓스트랩 방법(SB)이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 두 방법에 대하여, 42개의 포함비율 중에서 7개만이 포함비율에 대한 99% 신뢰구간에 해당하는 범위를 벗어나고 있는 것으로 나타나, 이 두 방법에 대한 수행결과가 매우 좋음을 알 수 있다. 또한 예상할 수 있듯이, 표본크기 n 이 증가할수록 좋은 수행결과를 보여주고 있음을 알 수 있다. 다음으로 스튜던트화 붓스트랩(STUD) 방법에 대한 결과를 살펴보면, 42개의 포함비율 중 9개가 범위 0.933~0.967를 벗어나고 있는 것으로 나타났으며, 특히 그러한 포함비율들은 모두 기대값 0.95보다 유의하게 낮은 포함비율을 갖음을 알 수 있다. 이 방법 역시 표본크기 n 이 증가할수록 그 수행결과가 좋음을 알 수 있다. 마지막으로 하이브리드 붓스트랩(HYB) 방법에 대한 결과를 살펴보면, 42개의 포함비율 중 25개가 범위를 벗어나고 있는 것으로 나타났으며, 그러한 포함비율은 스튜던트화 붓스트랩(STUD) 방법에 대한 결과와는 달리 모두 기대값 0.95보다 유의하게 높은 포함비율을 갖음을 알 수 있다. 한편 이 방법은 스튜던트화 붓스트랩(STUD) 방법과는 달리 표본크기 n 이 증가할수록 기대값 0.95로 감소하는 경향이 있는 것으로 나타났다. <표 3>의 결과에서는 정규 근사(AN) 방법, 표준 붓스트랩(SB) 방법, 그리고 스튜던트화 붓스트랩(STUD) 방법이 좋은 수행결과를 보여주었다.

<표 4>는 $\mu_x < M_x$ 이고 $\mu_y < M_y$ 인 경우의 C_{pk} 에 대한 95% 신뢰영역의 포함비율 결과이다. 결과를 살펴보면, 모든 표본크기 n 에 대하여 정규 근사(AN) 방법이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 이 방법에 대한 42개의 포함비율 모두가 범위 0.933~0.967 내에 있음을 알 수 있다. 다음으로 기대값 0.95에 가까운 포함비율을 갖는 표준 붓스트랩(SB) 방법과 스튜던트화 붓스트랩(STUD) 방법에 대해서는 42개의 포함비율 중 7개가 범위 0.933~0.967를 벗어나고 있는 것으로 나타났다. 한편 표준 붓스트랩(SB) 방법은 n 이 증가할수록 좋은 결과를 보여주고 있지는 않은 것으로 나타났지만, 스튜던트화 붓스트랩(STUD) 방법은 n 이 증가할수록 그 수행결과가 좋음을 알 수 있다. 마지막으로 하이브리드 붓스트랩(HYB) 방법에 대한 결과를 살펴보면, 이 방법에 대한 42개의 포함비율 중 31개가 범위를 벗어나고 있는 것으로 나타나 네 방법 중에서 수행결과가 가장 좋지 않은 것으로 나타났다. <표 3>의 결과에서와 마찬가지로 범위를 벗어나는 포함비율은 모두 기대값 0.95보다 유의하게 높은 포함비율을 갖음을 알 수 있으며, 또한 표본크기 n 이 증가할수록 기대값 0.95로 감소하는 경향이 있는 것으로 나타났다.

<표 4>의 결과에서는 정규 근사(AN) 방법, 표준 붓스트랩(SB) 방법, 그리고 스튜던트화 붓스트랩(STUD) 방법이 좋은 수행결과를 보여주었다.

<표 5>는 $\mu_x = M_x$ 이고 $\mu_y < M_y$ 인 경우의 C_{pk} 에 대한 95% 신뢰영역의 포함비율 결과이다. 결과를 살펴보면, 모든 표본크기 n 에 대하여 표준 붓스트랩(SB) 방법과 하이브리드 붓스트랩(HYB) 방법이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 두 방법에 대한 42개의 포함비율 중 12개만이 범위 0.933~0.967를 벗어나고 있는 것으로 나타났다. 또한 두 방법 모두 표본크기 n 이 증가할수록 좋은 수행결과를 보여주고 있음을 알 수 있다. 다음으로는 정규 근사(AN) 방법이 기대값 0.95에 가까운 포함비율을 갖는 것으로 나타났다. 실제로 이 방법에 대한 42개의 포함비율 중 13개가 범위 0.933~0.967를 벗어나고 있는 것으로 나타났으며, 역시 표본크기 n 이 증가할수록 수행결과 좋아짐을 알 수 있다. 마지막으로 스튜던트화 붓스트랩(STUD) 방법에 대한 결과를 살펴보면, 이 방법에 대한 42개의 포함비율 중 19개가 범위 0.933~0.967를 벗어나고 있는 것으로 나타나 네 방법 중에서 수행결과가 가장 좋지 않은 것으로 나타났다. 하지만 이 방법 역시 표본크기 n 이 증가할수록 기대값 0.95로 증가하는 경향이 있는 것으로 나타났다. <표 5>의 결과에서는 표준 붓스트랩(SB) 방법, 하이브리드 붓스트랩(HYB) 방법, 그리고 정규 근사(AN) 방법이 좋은 수행결과를 보여주었다.

<표 3> C_p 에 대한 95% 신뢰영역의 포함비율 : 이변량 정규분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ							
			-0.9	-0.6	-0.3	0	0.3	0.6	0.9	
(50.0,100.0,3.0,3.0)	$C_p = (C_{px}, C_{py}) = (1.000, 1.000) : (3.0, 3.0)$ 시그마 수준									
	30	AN	0.945	0.949	0.955	0.958	0.957	0.944	0.943	
		SB	0.943	0.946	0.945	0.940	0.941	0.933	0.941	
		STUD	0.939	0.951	0.946	0.938	0.943	0.942	0.945	
		HYB	0.982*	0.985*	0.983*	0.982*	0.986*	0.986*	0.984*	
	60	AN	0.939	0.938	0.951	0.948	0.949	0.948	0.945	
		SB	0.939	0.946	0.945	0.945	0.942	0.941	0.949	
		STUD	0.936	0.938	0.937	0.945	0.937	0.941	0.934	
		HYB	0.965	0.963	0.966	0.970*	0.971*	0.968*	0.967	
	(50.0,100.0,3.0,1.5)	$C_p = (C_{px}, C_{py}) = (1.000, 2.000) : (3.0, 6.0)$ 시그마 수준								
		30	AN	0.927*	0.923*	0.926*	0.932*	0.928*	0.928*	0.928*
			SB	0.922*	0.916*	0.921*	0.935	0.931*	0.927*	0.926*
STUD			0.917*	0.912*	0.914*	0.918*	0.922*	0.921*	0.931*	
HYB			0.966	0.968*	0.963	0.964	0.967	0.966	0.973*	
60		AN	0.949	0.947	0.958	0.947	0.947	0.950	0.945	
		SB	0.944	0.941	0.945	0.943	0.938	0.947	0.942	
		STUD	0.941	0.948	0.944	0.942	0.937	0.947	0.941	
		HYB	0.962	0.966	0.974*	0.969*	0.964	0.968*	0.961	
(50.0,100.0,1.5,1.5)		$C_p = (C_{px}, C_{py}) = (2.000, 2.000) : (6.0, 6.0)$ 시그마 수준								
		30	AN	0.941	0.939	0.942	0.946	0.951	0.948	0.939
			SB	0.939	0.941	0.944	0.948	0.947	0.948	0.940
	STUD		0.943	0.941	0.932*	0.930*	0.941	0.941	0.937	
	HYB		0.978*	0.980*	0.977*	0.975*	0.978*	0.980*	0.975*	
	60	AN	0.941	0.945	0.955	0.952	0.958	0.955	0.948	
		SB	0.941	0.944	0.949	0.954	0.947	0.945	0.939	
		STUD	0.940	0.938	0.939	0.942	0.952	0.948	0.944	
		HYB	0.966	0.967	0.966	0.966	0.972*	0.977*	0.969*	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

<표 4> $C_{pk}(\mu_x < M_x, \mu_y < M_y)$ 에 대한 95% 신뢰영역의 포함비율 : 이변량 정규분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ							
			-0.9	-0.6	-0.3	0	0.3	0.6	0.9	
(45.5, 95.5, 3.0, 3.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (0.500, 0.500) : (3.0, 3.0)$ 시그마 수준									
	30	AN	0.955	0.956	0.962	0.962	0.951	0.950	0.955	
		SB	0.939	0.944	0.945	0.956	0.952	0.943	0.951	
		STUD	0.932*	0.947	0.946	0.949	0.951	0.949	0.951	
		HYB	0.971*	0.968*	0.978*	0.977*	0.974*	0.976*	0.979*	
	60	AN	0.945	0.938	0.950	0.950	0.941	0.943	0.937	
		SB	0.932*	0.931*	0.937	0.932*	0.932*	0.931*	0.930*	
		STUD	0.938	0.933	0.935	0.938	0.932*	0.935	0.935	
		HYB	0.958	0.958	0.960	0.957	0.949	0.950	0.949	
	(45.5, 95.5, 3.0, 1.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (0.500, 1.500) : (3.0, 6.0)$ 시그마 수준								
		30	AN	0.944	0.949	0.947	0.944	0.951	0.949	0.955
			SB	0.937	0.931*	0.938	0.936	0.940	0.940	0.939
STUD			0.934	0.924*	0.926*	0.928*	0.931*	0.938	0.944	
HYB			0.968*	0.975*	0.970*	0.973*	0.980*	0.983*	0.980*	
60		AN	0.955	0.954	0.965	0.958	0.956	0.951	0.949	
		SB	0.947	0.941	0.957	0.960	0.949	0.947	0.944	
		STUD	0.943	0.948	0.956	0.958	0.950	0.948	0.940	
		HYB	0.969*	0.971*	0.975*	0.978*	0.965	0.966	0.970*	
(45.5, 95.5, 1.0, 1.0)		$C_{pk} = (C_{pkx}, C_{pky}) = (1.500, 1.500) : (6.0, 6.0)$ 시그마 수준								
		30	AN	0.952	0.937	0.946	0.945	0.953	0.942	0.939
			SB	0.940	0.938	0.941	0.942	0.949	0.942	0.933
	STUD		0.941	0.933	0.935	0.942	0.944	0.941	0.933	
	HYB		0.979*	0.983*	0.978*	0.974*	0.977*	0.979*	0.978*	
	60	AN	0.953	0.954	0.951	0.947	0.959	0.957	0.950	
		SB	0.948	0.942	0.951	0.953	0.956	0.948	0.942	
		STUD	0.946	0.930*	0.945	0.947	0.952	0.957	0.947	
		HYB	0.974*	0.966	0.966	0.971*	0.969*	0.972*	0.973*	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

<표 5> $C_{pk} (\mu_x = M_x, \mu_y < M_y)$ 에 대한 95% 신뢰영역의 포함비율 : 이변량 정규분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ							
			-0.9	-0.6	-0.3	0	0.3	0.6	0.9	
(50.0,95.5,3.0,3.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (1.000, 0.500) : (3.0, 3.0)$ 시그마 수준									
	30	AN	0.934	0.950	0.961	0.959	0.958	0.949	0.927*	
		SB	0.934	0.926*	0.933	0.945	0.942	0.939	0.947	
		STUD	0.924*	0.926*	0.937	0.943	0.934	0.938	0.930*	
		HYB	0.960	0.969*	0.971*	0.970*	0.967	0.967	0.967	
	60	AN	0.927*	0.938	0.948	0.949	0.944	0.945	0.920*	
		SB	0.936	0.936	0.941	0.938	0.940	0.936	0.936	
		STUD	0.940	0.933	0.941	0.938	0.938	0.938	0.935	
		HYB	0.953	0.956	0.958	0.955	0.954	0.956	0.951	
	(50.0,95.5,3.0,1.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (1.000, 1.500) : (3.0, 6.0)$ 시그마 수준								
		30	AN	0.907*	0.924*	0.926*	0.930*	0.929*	0.924*	0.915*
			SB	0.912*	0.903*	0.916*	0.919*	0.923*	0.914*	0.916*
STUD			0.916*	0.905*	0.910*	0.916*	0.913*	0.914*	0.914*	
HYB			0.960	0.955	0.952	0.952	0.956	0.954	0.957	
60		AN	0.930*	0.937	0.939	0.939	0.932*	0.937	0.922*	
		SB	0.933	0.931*	0.939	0.936	0.923*	0.922*	0.937	
		STUD	0.926*	0.936	0.934	0.930*	0.927*	0.932*	0.933	
		HYB	0.959	0.961	0.961	0.949	0.946	0.950	0.953	
(50.0,95.5,1.5,1.0)		$C_{pk} = (C_{pkx}, C_{pky}) = (2.000, 1.500) : (6.0, 6.0)$ 시그마 수준								
		30	AN	0.938	0.938	0.941	0.950	0.951	0.947	0.940
			SB	0.933	0.933	0.931*	0.942	0.943	0.937	0.940
	STUD		0.933	0.930*	0.930*	0.934	0.938	0.937	0.932*	
	HYB		0.975*	0.973*	0.975*	0.971*	0.970*	0.969*	0.973*	
	60	AN	0.945	0.948	0.943	0.948	0.950	0.948	0.939	
		SB	0.935	0.934	0.943	0.948	0.946	0.940	0.937	
		STUD	0.936	0.922*	0.931*	0.942	0.944	0.944	0.939	
		HYB	0.961	0.959	0.963	0.970*	0.966	0.969*	0.963	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

(2) 공정의 분포가 이변량 카이제곱분포인 경우

공정의 분포가 이변량 카이제곱 분포일 때, C_p 와 C_{pk} 에 대한 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 방법에 대한 모의실험 수행결과가 각각 <표 6>~<표 8>에 있다.

<표 6>의 C_p 에 대한 95% 신뢰영역의 포함비율 결과를 살펴보면, 모든 표본크기 n 에 대하여 하이브리드 붓스트랩(HYB) 방법이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 이 방법에 대한 24개의 포함비율 모두가 범위 0.933~0.967 내에 있는 것으로 나타나 수행결과가 매우 좋음을 알 수 있다. 다음으로는 정규 근사(AN) 방법에 대한 24개의 포함비율 중 6개만이 범위 0.933~0.967를 벗어나고 있는 것으로 나타나 수행결과 좋음을 알 수 있다. 마지막으로 표준 붓스트랩(SB) 방법과 스튜던트화 붓스트랩(STUD) 방법에 대해서는 24개의 포함비율 모두가 범위 0.933~0.967를 벗어나고 있는 것으로 나타나 그 수행결과가 좋지 않음을 알 수 있으며, 그러한 포함비율들은 모두 기대값 0.95보다 유의하게 낮은 포함비율을 갖음을 알 수 있다. 하지만 이 두 방법 모두 표본크기 n 이 증가할수록 기대값 0.95로 증가하는 경향이 있는 것으로 나타났다. <표 6>의 결과에서는 하이브리드 붓스트랩(HYB) 방법과 정규 근사(AN) 방법이 좋은 수행결과를 보여주었다.

<표 7>은 $\mu_x < M_x$ 이고 $\mu_y < M_y$ 인 경우의 C_{pk} 에 대한 95% 신뢰영역의 포함비율 결과이다. 결과를 살펴보면, 모든 표본크기 n 에 대하여 정규 근사(AN) 방법과 하이브리드 붓스트랩(HYB) 방법이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 두 방법에 대한 24개의 포함비율 중 단지 3개만이 범위 0.933~0.967를 벗어나고 있는 것으로 나타났으며, 또한 두 방법 모두 표본크기 n 이 증가할수록 수행결과가 좋은 것으로 나타났다. 다음으로 표준 붓스트랩(SB)은 24개의 포함비율 중 단지 3개만이 범위 0.933~0.967 내에 있는 것으로 나타났지만, 전반적으로는 기대값 0.95보다 유의하게 낮은 포함비율을 갖음을 알 수 있다. 한편 이 방법은 표본크기 n 의 증가에 따른 경향을 보이고 있진 않은 것으로 나타났다. 마지막으로 스튜던트화 붓스트랩(STUD) 방법에 대한 결과를 살펴보면, 이 방법에 대한 24개의 포함비율 모두가 범위 0.933~0.967를 벗어나고 있는 것으로 나타났으며, 특히 그러한 포함비율들은 모두 기대값 0.95보다 유의하게 낮은 포함비율을 갖음을 알 수 있다. 하지만 이 방법은 표본크기 n 이 증가할수록 기대값 0.95로 증가하는 경향이 있는 것으로 나타났다. <표 7>의 결과에서는 정규 근사(AN) 방법과 하이브리드 붓스트랩(HYB) 방법이 좋은 수행결과를 보여주었다.

<표 8>은 $\mu_x = M_x$ 이고 $\mu_y < M_y$ 인 경우의 C_{pk} 에 대한 95% 신뢰영역의 포함비율 결과이다. 결과를 살펴보면, 모든 표본크기 n 에 대하여 하이브리드 붓스트랩(HYB) 방법이 일관성 있게 기대값 0.95에 가까운 값을 갖는 것으로 나타났다. 실제로 이 방법에 대한 24개의 포함비율 중 단지 1개만이 범위 0.933~0.967를 벗어나고 있는 것으로 나타났으며, 또한 표본크기 n 이 증가할수록 수행결과가 좋음을 알 수 있다. 다음으로는 정규 근사(AN) 방법이 24개의 포함비율 중 3개만이 범위 0.933~0.967를 벗어나고 있는 것으로 나타나, 이 방법에 대한 수행결과 역시 좋음을 알 수 있다. 마지막으로 표준 붓스트랩(SB) 방법과 스튜던트화 붓스트랩(STUD) 방법에 대해서는 24개의 포함비율 모두가 범위 0.933~0.967를 벗어나고 있는 것으로 나타나 그 수행결과가 좋지 않음을 알 수 있으며, 그러한 포함비율들은 모두 기대값 0.95보다 유의하게 낮은 포함비율을 갖음을 알 수 있다. 하지만 이 두 방법 모두 표본크기 n 이 증가할수록 기대값 0.95로 증가하는 경향이 있는 것으로 나타났다. <표 8>의 결과에서는 하이브리드 붓스트랩(HYB) 방법과 정규 근사(AN) 방법이 좋은 수행결과를 보여주었다.

<표 6> C_p 에 대한 95% 신뢰영역의 포함비율 : 이변량 카이제곱 분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ				
			0	0.3	0.6	0.9	
(50.0,100.0,3.0,3.0)	$C_p = (C_{px}, C_{py}) = (1.000, 1.000) : (3.0, 3.0)$ 시그마 수준						
	30	AN	0.949	0.954	0.951	0.974*	
		SB	0.855*	0.846*	0.850*	0.855*	
		STUD	0.844*	0.849*	0.852*	0.859*	
		HYB	0.955	0.948	0.943	0.948	
	60	AN	0.956	0.953	0.960	0.976*	
		SB	0.891*	0.897*	0.891*	0.887*	
		STUD	0.898*	0.905*	0.913*	0.908*	
		HYB	0.951	0.947	0.951	0.950	
	(50.0,100.0,3.0,1.5)	$C_p = (C_{px}, C_{py}) = (1.000, 2.000) : (3.0, 6.0)$ 시그마 수준					
		30	AN	0.949	0.954	0.951	0.974*
			SB	0.862*	0.842*	0.854*	0.854*
STUD			0.840*	0.849*	0.848*	0.848*	
HYB			0.955	0.945	0.945	0.948	
60		AN	0.956	0.953	0.960	0.976*	
		SB	0.890*	0.903*	0.887*	0.889*	
		STUD	0.899*	0.901*	0.917*	0.903*	
		HYB	0.953	0.948	0.948	0.953	
(50.0,100.0,1.5,1.5)		$C_p = (C_{px}, C_{py}) = (2.000, 2.000) : (6.0, 6.0)$ 시그마 수준					
		30	AN	0.949	0.955	0.952	0.975*
			SB	0.862*	0.852*	0.854*	0.855*
	STUD		0.849*	0.861*	0.852*	0.858*	
	HYB		0.953	0.950	0.947	0.956	
	60	AN	0.956	0.954	0.941	0.977*	
		SB	0.889*	0.899*	0.890*	0.888*	
		STUD	0.899*	0.907*	0.920*	0.906*	
		HYB	0.951	0.950	0.950	0.954	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

<표 7> $C_{pk}(\mu_x < M_x, \mu_y < M_y)$ 에 대한 95% 신뢰영역의 포함비율 : 이변량 카이제곱 분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ				
			0	0.3	0.6	0.9	
(45.5, 95.5, 3.0, 3.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (0.500, 0.500) : (3.0, 3.0)$ 시그마 수준						
	30	AN	0.966	0.962	0.962	0.975*	
		SB	0.942	0.930*	0.936	0.933	
		STUD	0.916*	0.915*	0.921*	0.912*	
		HYB	0.979*	0.976*	0.965	0.964	
	60	AN	0.946	0.946	0.951	0.965	
		SB	0.930*	0.918*	0.925*	0.923*	
		STUD	0.922*	0.912*	0.913*	0.913*	
		HYB	0.961	0.950	0.943	0.949	
	(45.5, 95.5, 3.0, 1.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (0.500, 1.500) : (3.0, 6.0)$ 시그마 수준					
		30	AN	0.966	0.955	0.957	0.964
			SB	0.912*	0.905*	0.908*	0.908*
STUD			0.889*	0.890*	0.884*	0.886*	
HYB			0.972*	0.966	0.967	0.958	
60		AN	0.948	0.952	0.954	0.967	
		SB	0.903*	0.909*	0.900*	0.909*	
		STUD	0.908*	0.893*	0.897*	0.907*	
		HYB	0.952	0.947	0.938	0.944	
(45.5, 95.5, 1.0, 1.0)		$C_{pk} = (C_{pkx}, C_{pky}) = (1.500, 1.500) : (6.0, 6.0)$ 시그마 수준					
		30	AN	0.957	0.952	0.953	0.972*
			SB	0.871*	0.874*	0.863*	0.871*
	STUD		0.855*	0.843*	0.844*	0.845*	
	HYB		0.960	0.955	0.949	0.942	
	60	AN	0.954	0.958	0.960	0.972*	
		SB	0.893*	0.891*	0.885*	0.886*	
		STUD	0.898*	0.894*	0.894*	0.895*	
		HYB	0.951	0.943	0.941	0.935	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

<표 8> $C_{pk} (\mu_x = M_x, \mu_y < M_y)$ 에 대한 95% 신뢰영역의 포함비율 : 이변량 카이제곱 분포인 경우

$(\mu_x, \mu_y, \sigma_x, \sigma_y)$	표본크기 (n)	방 법	ρ				
			0	0.3	0.6	0.9	
(50.0,95.5,3.0,3.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (1.000, 0.500) : (3.0, 3.0)$ 시그마 수준						
	30	AN	0.946	0.948	0.938	0.952	
		SB	0.907*	0.911*	0.909*	0.902*	
		STUD	0.884*	0.876*	0.879*	0.874*	
		HYB	0.960	0.956	0.949	0.950	
	60	AN	0.947	0.954	0.957	0.960	
		SB	0.919*	0.906*	0.908*	0.923*	
		STUD	0.915*	0.903*	0.906*	0.907*	
		HYB	0.957	0.942	0.938	0.939	
	(50.0,95.5,3.0,1.0)	$C_{pk} = (C_{pkx}, C_{pky}) = (1.000, 1.500) : (3.0, 6.0)$ 시그마 수준					
		30	AN	0.944	0.940	0.947	0.957
			SB	0.868*	0.871*	0.871*	0.884*
STUD			0.853*	0.841*	0.849*	0.844*	
HYB			0.953	0.944	0.935	0.929*	
60		AN	0.950	0.957	0.962	0.972*	
		SB	0.896*	0.892*	0.887*	0.900*	
		STUD	0.909*	0.896*	0.898*	0.898*	
		HYB	0.953	0.941	0.934	0.935	
(50.0,95.5,1.5,1.0)		$C_{pk} = (C_{pkx}, C_{pky}) = (2.000, 1.500) : (6.0, 6.0)$ 시그마 수준					
		30	AN	0.951	0.949	0.957	0.969*
			SB	0.860*	0.865*	0.858*	0.870*
	STUD		0.847*	0.838*	0.842*	0.842*	
	HYB		0.954	0.949	0.945	0.936	
	60	AN	0.954	0.956	0.963	0.973*	
		SB	0.894*	0.891*	0.887*	0.902*	
		STUD	0.901*	0.904*	0.903*	0.898*	
		HYB	0.955	0.945	0.941	0.937	

* 표시는 유의수준 1%에서 신뢰영역의 포함확률이 95%와는 같다고 할 수 없음을 의미한다.

5. 결론

본 논문에서는 벡터 공정능력지수들에 대한 통계적 추론을 위하여 붓스트랩 방법을 이용한 효율적인 신뢰영역 설정문제에 대하여 연구하였다. 먼저 제2절에서는 벡터 공정능력지수 $C_p = (C_{px}, C_{py})$, $C_{pm} = (C_{pmx}, C_{pmy})$ 와 $C_{pk} = (C_{pkx}, C_{pky})$ 에 대한 플러그-인 추정량들과 관련된 극한 확률분포를 기초로 근사 신뢰영역을 제안하였다. 제3절에서는 붓스트랩 알고리즘에 의한 벡터 공정능력지수 C_p , C_{pm} 과 C_{pk} 의 붓스트랩 추정량들과 관련된 극한 확률분포를 소개하여 붓스트랩 방법의 일치성을 제시하고 3가지 형태의 붓스트랩 신뢰영역들을 제안하였다. 그리고 제4절에서는 소표본 모의실험연구로 이변량 정규분포와 이변량 카이제곱 분포 하에서 앞에서 제시한 2차원 벡터 공정능력지수 C_p 와 C_{pk} 에 대한 정규 근사(AN) 및 3가지 형태의 붓스트랩(SB, STUD, HYB) 신뢰영역들의 효율성 문제를 검토하기 위하여 포괄적이고 다양한 모의실험을 수행하였다. 각각의 신뢰영역 설정시 분산-공분산 행렬의 추정 문제가 매우 중요한 역할을 할 것인 바, 공정의 분포가 이변량 정규분포인 경우, 정규 근사 방법(AN)과 표준 붓스트랩 방법(SB)이 안정적으로 만족스러운 결과로 나타났다. 이와는 달리, 공정의 분포가 이변량 카이제곱분포인 경우에는 대체적으로 정규근사 방법(AN)과 하이브리드 방법(HYB)이 안정적으로 명목 포함확률 95% 수준을 만족스럽게 유지하고 있는 것으로 나타났다. 하지만 상관관계가 높아질수록 명목 포함확률 95%를 벗어나는 경우가 존재하는 것으로 드러났지만, 현장에서 적용하기에는 충분히 의미가 있으리라고 사료된다. 하지만 공정분포에 대해 구체적으로 가정하기 어려운 많은 경우에는 정규 근사 방법(AN)에 비해 붓스트랩 방법들이 보다 더 효율적인 신뢰영역이 될것이라 사료되며 추후 연구과제로 남긴다.

참고문헌

- [1] Cho,J.J. Kim,J.S. and Park,B.S.(1999). Better Nonparametric Bootstrap Confidence Interval for Process Capability Index C_{pk} , *Korean Journal of Applied Statistics*, 12(1), 45-65.
- [2] Cho,J.J. Park,B.S.and Lim,S.D.(2001). On Statistical Estimation of Multivariate (Vector-valued) Process Capability Indices with Bootstrap, *Korean Communications in Statistics*, 8(3), 697-709.
- [3] Cho,J.J. and Park,B.S.(2003). Unpublished manuscript.
- [4] Franklin,L.A. and Wasserman,G.S.(1992) . Bootstrap Lower Confidence Interval Limits for Capability Indices, *Journal of Quality Technology*, 24, 196-210.
- [5] Gunter,B.H.(1989). The Use and Abuse of C_{pk} : Part2, *Quality Progress*, 22(3), 108-109.
- [6] Hall,P.(1988). Theoretical Comparison of Bootstrap Confidence Intervals. *Annals of Statistics*, 16, 927-953.
- [7] Kocherlakota,S. and Kocherlakota,K.(1991). Process capability indices : Bivariate Normal distribution, *Communication in Statistics : Theory and Methods*, 20, 2529-2547.
- [8] Kotz,S. and Johnson,N.L.(1993). Process Capability Indices, 1st ed., Chapman & Hall.
- [9] Park,B.S., Lee,C.H.and Cho,J.J.(2002). On the Confidence Region of Vector-valued Process

422 Joong-Jae Cho, Byoung-Sun Park

Capability Indices C_p and C_{pk} . Journal of the Korean Society for Quality Management
Vol.30, No.4, 44-57.

[2003년 4월 접수, 2003년 6월 채택]