응용통계연구 제 19권 3호, 2006년, pp. 535–544

분산 성분 모형에 대한 붓스트랩 보정 신뢰구간*

이용희[1)]

요 약

분산 성분 모형 하에서 분산 성분들의 함수에 대한 통계적인 추론, 특히 소표본 하에서의 신뢰구간에 대한 방법들은 오랜 기간에 걸쳐서 여러 가지 방법들이 개발되어져 왔다. 그 대표적인 방법이 Graybill and Wang(1980)에 의해 제안된 수정 대표본 방법에 의거한 신뢰구간 추정법이며 현재까지 다양한 실험계획 방법 하에서 분산 성분들의 여러 가지 형태의 함수들에 대하여 확장과 개량이 이루어져 왔다. 본 연구에서는 분산 성분 모형의 균형 실험 가정 하에서 분산 성분들의 선형 결합이 관심있는 모수일 때 분산분석에 의해 얻어진 수정 대표본 신뢰구간의 실제 포함확률을 붓스트랩 보정을 이용하여 개선하는 방법에 대하여 논의한다. 붓스트랩 보정을 이용함으로서 신뢰구간의 포함확률의 정도는 점근적 이차 차수까지 개선되며 특히 선형 결합의 계수들이 모두 양수이고 결합의 수가 증가할 경우 수정 대표본 신뢰구간의 포함확률이 주어진 신뢰계수보다 항상 커지게 되는 단점을 개선할 수 있음을 보인다. 제안된 붓스트랩 보정 신뢰구간의 효율을 소표본의 경우에 모의실험을 통하여 평가한다.

주요용어: 분산 성분 모형, 신뢰구간, 수정 대표본 방법, 붓스트랩 보정

1. 서론

분산 성분 모형(variance components model) 하에서 분산 성분들의 함수에 대한 통계적 추론은 품질관리, 유전학, 의학 등 많은 분야에서 실험이나 관측 연구에서 얻어진 관측치의 총변이를 관심 있는 요소들로 분해하여 전체적인 변이의 구조를 이해하는데 쓰인다. 흔히 이러한 분산 성분들의 함수에 대한 통계적인 추론, 특히 정규분포 분산 성분 모형의 가정 하에서 소표본인 경우에 분산 성분들의 선형 결합에 대한 신뢰구간의 통계적 추정 방법은 그 응용 범위가 넓어서 많은 연구가 진행되어 왔다. 분산 성분 모형은 주어진 실험이나 연구의 형태가 다양하지만 일반적으로 많은 경우에 관심있는 모수는 분산 성분들의 선형 결합 형태로 나타낼 수 있다.

$$\eta = c_1\sigma_1^2 + \cdots + c_p\sigma_p^2 = \mathbf{c}'\boldsymbol{\theta} \tag{1.1}$$

여기서 $\mathbf{c}' = (c_1, c_2, \ldots, c_p)$는 계수들의 벡터이고 $\boldsymbol{\theta}' = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$는 분산 성분들의 벡터이다. 이러한 분산 성분들의 선형 결합에 대한 여러 가지 신뢰구간들 중 수정 대표본 방

* 이 연구는 2005학년도 이화여자대학교 교내연구과제 지원에 의한 연구임
1) (120-750) 서울시 서대문구 대현동 11-1, 이화여자대학교 통계학과, 전임 강사
E-mail: ylee@ewha.ac.kr

법(Modified Large Sample Method)에 의한 신뢰구간이 널리 쓰인다. 수정 대표본 방법은 Graybill and Wang(1980)에 의해 식 (1.1)에서 선형 계수 c_i가 모두 양수인 경우에 제안되었고 오랜 기간에 걸쳐 계수 c_i의 부호에 제약이 없는 경우나 분산 성분들의 비(ratio)에 대해 지속적으로 확장되었다(Lu, et al., 1988, 1989; Ting, et al, 1991; Burdick and Graybill, 1992; Gui, et al., 1995; Lee, et al., 2004). 수정 대표본 방법에 대한 자세한 방법과 적용은 Burdick and Graybill(1992)에 의해 잘 개관되어 있다.

수정 대표본 방법에 의한 신뢰구간은 그 실제 포함확률(coverage probability)의 개선을 위해 지속적으로 확장되어 왔지만 Graybill and Wang(1980)에 의해 최초로 제안된 간단한 형태의 신뢰구간에서 더욱 복잡한 형태를 가지게 되어 수정 대표본 방법의 큰 장점인 단순성을 잃게 되었다. 복잡한 형태로의 확장에도 불구하고 분산 성분의 수가 증가하면 수정 대표본 방법에 의한 신뢰구간의 실제 포함확률이 주어진 신뢰계수보다 크게 되는 단점은 크게 개선하지 못하였다(Burdick and Graybill, 1992).

붓스트랩(Bootstrap) 방법은 주어진 자료를 재표본 추출(resampling)하는 모의실험 방법을 통하여 더욱 정확한 실제 포함확률을 가진 신뢰구간을 구하는 방법으로서 이 방법 또한 단순한 형태에서 시작하여 실제 포함확률의 정확성을 높이기 위하여 오랫동안 진화되어 왔다. 붓스트랩의 확장된 많은 방법 중에 Loh(1987, 1991)에 의해 제안된 붓스트랩 보정법(Bootstrap Calibration)은 한 번의 재표본 추출을 통하여 주어진 임의의 신뢰구간의 실제 포함확률을 이차 정확도(second-order correct)를 가지게 보정해 주는 방법이다. 일반적으로 이차 정확도를 갖는 붓스트랩 신뢰구간이 재표본 추출뿐만 아니라 부가적인 보정 상수를 추정해야 하는데 반하여 붓스트랩 보정법은 단순한 신뢰구간을 이용하여 한번의 재표본 추출로서 그 정확도를 한 차수(order)를 향상시킬 수 있는 일반적인 방법이다.

본 연구에서는 분산 성분들의 선형 결합에 대한 신뢰구간을 구하는 문제에서 수정 대표본 방법의 가장 간단한 형태로부터 출발하여 붓스트랩 보정법을 적용하여 그 실제 포함확률을 향상시켜 이차 정확도를 갖는 새로운 신뢰구간을 제안한다. 다음 장에서는 균형 설계 하에서 분산 성분들의 선형 결합에 대한 추정량의 일반적인 분포 이론을 살펴보고 3장에서는 붓스트랩 보정에 대하여 논의하며 4장에서 수정 대표본 신뢰구간에 대한 붓스트랩 보정을 제안한다. 5장에서는 제안된 방법의 효율을 모의실험을 통하여 확인하고 마지막 장에서는 본 연구에 대한 결론과 그 확장에 대해 논의한다. 본 연구에서 두 확률변수 x와 y의 확률 분포가 같으면 $\mathcal{L}(x) = \mathcal{L}(y)$와 같이 표시한다.

2. 분산 성분들의 선형 결합에 대한 분산분석 추정량의 분포

분산 성분 모형의 일반적인 실험계획 하에서 관측 벡터 $\mathbf{y}$가 평균이 $\mathbf{0}$이고 분산이 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$인 다변량 정규분포를 따른다고 할 때 분산분석(Analysis of Variance: ANOVA)을 통하여 각 요인들에 대한 평균 제곱합(mean squared error)이 구해지면 관심있는 분산 성분의 추정량은 그 평균 제곱합들의 선형 결합으로 표현된다. 즉 분산분석에서 p개의 평균 제곱합 MSE_i가 구해진다고 가정하고 각각의 제곱합에 대한 사영행렬 (Projection matrix) 을 $\mathbf{P}_i$라 하자. 평균 제곱합들은 정규분포의 관측 벡터 $\mathbf{y}$ 의 이차형식이므로 최종적으로 관심있는 모수 η의

추정량도 이차형식의 형태를 가진다. 벡터 $\mathbf{q}$ 를 각 분산분석의 요인의 평균 제곱합으로 이루어진 p 차원 벡터라 하면

$$\mathbf{q}' = (MSE_1,\ MSE_2,\ \ldots, MSE_p) = (\mathbf{y}'\mathbf{P}_1\mathbf{y},\ \mathbf{y}'\mathbf{P}_2\mathbf{y}, \ldots,\ \mathbf{y}'\mathbf{P}_p\mathbf{y})$$

제곱합 벡터 $\mathbf{q}$의 기대값은 분산 성분들의 선형 결합들로 나타내어 지므로 그 방정식을 분산 성분의 벡터 $\boldsymbol{\theta}$에 대해 풀면 각 분산 성분에 대한 분산분석 추정량 $\hat{\boldsymbol{\theta}}$을 구할 수 있고

$$E(\mathbf{q}) = \mathbf{A}\boldsymbol{\theta} \quad \Rightarrow \quad \hat{\boldsymbol{\theta}} = \mathbf{A}^{-1}\mathbf{q} \tag{2.1}$$

여기서 행렬 $\mathbf{A}$는 고정된 값을 가지는 행렬로 그 구성원소는 주어진 실험계획과 자료의 수에 의존한다. 다시 식 (1.1)의 관심있는 모수에 대한 분산분석 추정량 $\hat{\eta}$도 이차형식으로 나타내어 질 수 있다 (Mathai and Provost, 1992, p. 29).

$$\hat{\eta} = \mathbf{c}'\mathbf{A}^{-1}\mathbf{q} \equiv \mathbf{y}'\mathbf{P}\mathbf{y} \tag{2.2}$$

이 때 분산 성분들의 선형 결합의 추정치에 관련된 행렬 $\mathbf{P}$ 는 다음과 같이 표현 된다.

$$\mathbf{P} = b_1\mathbf{P}_1 + b_2\mathbf{P}_2 + \cdots + b_p\mathbf{P}_p$$

여기서 $(b_1, b_2, \ldots, b_p) = \mathbf{c}'\mathbf{A}^{-1}$ 이다.

분산 성분 모형의 실험계획에 따라 모수 η 의 추정량 $\hat{\eta} = \mathbf{y}'\mathbf{P}\mathbf{y}$ 의 분포가 결정되어 진다. 일반적으로 $\hat{\eta}$ 의 분포는 서로 독립인 카이제곱 확률변수들의 선형 결합의 분포와 같다.

$$\mathcal{L}(\hat{\eta}) = \mathcal{L}(\mathbf{y}'\mathbf{P}\mathbf{y}) = \mathcal{L}(\sum_{i=1}^{p} \lambda_i \chi^2_{n_i}/n_i) \tag{2.3}$$

여기서 λ_i들은 행렬 $\mathbf{P}\boldsymbol{\Sigma}$의 각각 n_i개의 반복을 갖는 서로 다른 고유값이며 $\chi^2_{n_i}$ 은 자유도 n_i를 가지는 서로 독립인 카이제곱 분포를 따르는 확률변수이다. 식 (2.3)은 이차형식 $\mathbf{y}'\mathbf{P}\mathbf{y}$의 분포를 카이제곱 분포에 선형 결합으로 표현함으로서 모수 η 의 표현을 간단하게 고유치의 합으로 나타낼 수 있다는데 그 유용성이 크다.

$$\eta = \sum_{i=1}^{p} c_i\sigma_i^2 = \sum_{i=1}^{r} \lambda_i$$

고유치 λ_i는 실제 모르는 모수이며 모분산 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ 에 대한 임의의 일치추정량 $\hat{\boldsymbol{\Sigma}}$가 주어지면 $\mathbf{P}\hat{\boldsymbol{\Sigma}}$ 의 고유치 $\hat{\lambda}_i$로서 추정할 수 있다.

분산 성분 모형에서 각 요인의 수준의 반복수가 다른 불균형 자료 설계하에서는 일반적으로 식 (2.3)의 이차형식 $\mathbf{y}'\mathbf{P}\mathbf{y}$의 분포가 자유도 1인 많은 카이제곱 확률변수을 포함하는 선형 결합의 분포로 표시될 수 있다. 또한 불균형의 정도가 증가하면 식 (2.3)에서 카이제곱 확률변수의 수 r도 같이 늘어나게 되어 특별한 경우를 제외하면 모수 η 에 대한 간단한 신뢰구간을 얻기 힘든 경우가 대부분이다. 그러나 각 요인의 수준의 반복수가 모두 같은 균형 설계하에서는 식 (2.3)에 주어진 추정량의 분포는 분산 성분의 수와 같은 카이제곱

확률변수들의 선형 결합으로 나타나며 ($r = p$) 각 확률변수의 자유도는 분산 분석표의 자유도와 동일하다. 분산 분석 하에서 각각의 이차형식이 독립이 아닐 경우에도 자료가 균형 배치되면 간단한 카이제곱 확률변수들의 선형 결합으로 추정량의 분포를 표시할 수 있음도 알려 졌다(Lee, et al., 2004).

3. 붓스트랩 보정

붓스트랩 보정(Bootstrap Calibration)을 이용한 신뢰구간은 Loh(1989, 1991)에 의해 제안되었으며 임의의 근사 신뢰구간의 실제 포함확률을 한번의 재표본 추출로서 그 정확도를 한 차수(order) 향상시킬 수 있는 일반적인 방법이다. 모수 η에 대한 임의의 $100\alpha\%$ 신뢰상한을 $\hat{\eta}_u[\alpha]$ 라고 하고 그 실제 포함확률을 $\beta(\alpha)$라 하자.

$$\beta(\alpha) \equiv P(\eta < \hat{\eta}_u[\alpha])$$

여기서 임의의 근사 신뢰구간은 자료의 수가 커질 때 그 포함확률이 계획한 신뢰계수 (confidence coefficient) α로 수렴하는 신뢰구간을 말한다(예로서 단순 정규 근사에 의한 신뢰구간).

일반적으로 자료의 수가 크지 않으면 실제 포함확률 $\beta(\alpha)$는 계획한 신뢰계수 α와 같지 않다. 붓스트랩 보정의 착상은 실제 포함확률 함수 $\beta(\alpha)$를 붓스트랩 방법으로 추정하여 신뢰계수를 보정하는 것이다. 예를 들어 $\beta(0.98) = 0.95$ 라면 95% 신뢰 상한을 $\hat{\eta}_u[0.95]$ 대신 $\hat{\eta}_u[0.98]$로 사용할 수 있다. 실제 포함확률 함수 $\beta(\alpha)$는 모르는 함수이므로 붓스트랩 방법을 통하여 추정할 수 있다.

$$\hat{\beta}(\alpha) = P_*\left(\hat{\eta} < \hat{\eta}_u^*[\alpha]\right) \tag{3.1}$$

여기서 P_* 는 붓스트랩 방법에 의한 확률분포를 가리키며 $\hat{\eta}_u^*[\alpha]$ 는 붓스트랩 자료에서에 구한 $100\alpha\%$ 신뢰상한이다. 위의 식 (3.1)에서 실제 포함확률 $\beta(\alpha)$를 추정하기 위하여 각 α값에 대하여 개별적인 붓스트랩을 적용해야 할 것처럼 보이지만 만약 $\hat{\eta}_u[\alpha]$가 α에 대한 증가함수라면 (많은 경우에 이 가정이 성립한다) 한번의 붓스트랩으로 $\beta(\alpha)$의 추정이 가능하다. 주어진 붓스트랩 자료에서 $\hat{\alpha}^*$를 신뢰 상한 $\eta_u^*[\alpha]$이 추정량 $\hat{\eta}$과 같게 되는 값이라고 하자. 즉,

$$\hat{\eta}_u^*[\hat{\alpha}^*] = \hat{\eta} \tag{3.2}$$

그렇다면 $\hat{\eta}_u[\alpha]$가 α에 대한 증가함수이기 때문에 사건 $\{\hat{\eta} < \hat{\eta}_u^*[\alpha]\}$ 는 사건 $\{\hat{\alpha}^* < \alpha\}$와 같으므로 다음과 같이 식 (3.1)를 표현 할 수 있다.

$$\hat{\beta}(\alpha) = P_*\left(\hat{\alpha}^* < \alpha\right)$$

다시 말하면 $\hat{\beta}(\alpha)$는 $\hat{\alpha}^*$의 분포함수로 표시할 수 있고 이는 B개의 붓스트랩 자료를 생성하여 각각의 자료에 대해 $\hat{\alpha}^*(b)$를 구하고 다음과 같이 $\hat{\beta}(\alpha)$를 추정할 수 있다.

$$\hat{\beta}(\alpha) = \frac{1}{B}\sum_{b=1}^{B} I[\hat{\alpha}^*(b) < \alpha]. \tag{3.3}$$

여기서 $I(x)$ 는 표시함수(indicator function)이다. Loh(1989, 1991)은 임의의 주어진 근사 신뢰구간에 붓스트랩 보정법을 적용하면 그 신뢰구간의 포함정확도 (coverage accuracy)가 한 차수 증가함을 보였다. 예를 들어 수정 대표본 방법에 의한 근사 신뢰구간의 포함정확도는 일반적으로 $O(n^{-1/2})$으로 알려졌으므로(Lee, et al., 2004) 붓스트랩 보정법에 의해 보정된 수정 대표본 방법에 의한 근사 신뢰구간의 포함정확도는 이차 정확도 $O(n^{-1})$를 가지게 된다. 붓스트랩 보정에 대한 자세한 내용과 다른 붓스트랩 방법들과의 비교는 Diciccio and Efron(1996)에 의해 잘 개관되어있다.

4. 균형 실험하에서 수정 대표본 신뢰구간의 붓스트랩 보정

이 연구에서는 균형 설계의 경우에 모든 선형 계수가 양수인 분산 성분의 선형 결합에 대한 신뢰상한을 고려한다.

$$\eta = c_1\sigma_1^2 + c_2\sigma_2^2 + \cdots + c_p\sigma_p^2 \quad \text{그리고} \quad c_i \geq 0 \ \text{ for all } i = 1, 2, \ldots, p \tag{4.1}$$

여기서 모든 선형 계수가 양수인 경우를 고려한 이유는 수정 대표본 신뢰구간이 가장 보수적이라는 사실이 경험적으로 또한 이론적으로 알려져 있기 때문이다(Burdick and Graybill, 1992; Lee, et al., 2004). 많은 균형 설계에서 총 분산 (total variance)을 추정하는 경우가 이에 해당된다.

이러한 가정하에서 분산 분석 추정량 $\hat{\eta}$의 분포는 식 (2.3)와 같이 표현할 수 있으며 독립인 카이제곱 변수의 수는 분산 성분의 수와 같으며 $(r = p)$ 모수는 $\lambda_i = c_i\sigma_i^2$ 로 표시되며 자유도 n_i는 분산 분석표의 자유도와 같다. 본 연구에서 모든 선형 계수가 양수를 가지는 경우에 대해 고려하는 이유는 선형 계수가 양수일 분산 성분의 수가 증가하면 수정 대표본 방법에 의한 신뢰구간의 포함확률이 빠르게 증가하여 성분의 수가 3개 이상인 경우에는 일반적으로 그 실제 포함확률이 1에 아주 가깝게 되어 그 유용성이 떨어지는 경우를 고려하기 위함이다.

위의 조건하에서 모수 η에 대한 수정 대표본 100α% 신뢰상한(confidence upper bound)의 형태는 다음과 같다(Graybill and Wnag, 1980; Burdick and Graybill, 1992).

$$\hat{\eta}_u^{ML}[\alpha] = \hat{\lambda}_1 + \cdots + \hat{\lambda}_p + \sqrt{\hat{\lambda}_1^2\left(\frac{n_1}{u_1} - 1\right)^2 + \cdots + \hat{\lambda}_q^2\left(\frac{n_p}{u_p} - 1\right)^2} \tag{4.2}$$

여기서 $u_i = \chi^2_{1-\alpha, n_i}$ 는 자유도 n_i를 가지는 카이제곱 분포의 $100(1-\alpha)$ 하위 백분위수이다.

분산 성분 모형의 균형 실험계획 하에서는 관심있는 모수 η가 식 (4.1)과 같을 때 식 (4.2)에서 주어진 수정 대표본 신뢰상한의 포함정확도를 다음과 같이 모수적 붓스트랩 보정 방법를 적용하여 보정할 수 있다.

(1) 식 (2.1)를 이용하여 분산 성분의 추정치 $\hat{\boldsymbol{\theta}}' = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_p^2)$ 구하고 고유치를 $\hat{\lambda}_i = c_i\hat{\sigma}_i^2$로 추정한다.

(2) 추정된 고유치 $\hat{\lambda}_i$를 가지고 모수 η를 추정한다.

$$\hat{\eta} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

(3) 추정된 고유치 $\hat{\lambda}_i$를 가지고 식 (2.3)의 표현을 이용하여 b 번째 붓스트랩 자료를 생성한다. 여기서 모수적 붓스트랩에 의해 생성된 자료는 p개의 서로 독립인 확률변수 $y_1, y_2, \ldots, y_p$ 이며 각각의 분포는 다음과 같다.

$$\mathcal{L}(y_i) = \mathcal{L}(\hat{\lambda}_i \chi^2_{n_i}/n_i), \quad i = 1, 2, \ldots, p$$

(4) 주어진 b 번째 붓스트랩 자료 $y_1, y_2, \ldots, y_p$를 이용하여 구한 수정 대표본 신뢰상한이 $\hat{\eta}$ 과 같아지는 신뢰계수 $\alpha^*(b)$를 구한다

$$\begin{aligned} \hat{\eta} &= \hat{\eta}_u^{ML}[\alpha^*(b)] \\ &\equiv y_1 + \cdots + y_p + \sqrt{y_1^2\left(\frac{n_1}{u_1} - 1\right)^2 + \cdots + y_p^2\left(\frac{n_p}{u_p} - 1\right)^2} \end{aligned}$$

여기서 u_i 는 자유도 n_i를 가지는 카이제곱 분포의 $100(1-\alpha^*(b))$ 하위 백분위수이다.

(5) (3)번과 (4)번의 단계를 B 번 반복하여 B개의 $\alpha^*(b)$ 값을 구한다. 여기서 구한 $\{\alpha^*(b), b = 1, 2, \ldots, B\}$를 이용하여 수정 대표본 신뢰상한의 포함확률의 추정치가 α 가 되는 보정 신뢰계수 $\hat{\alpha}^{BC}$를 구한다.

$$\alpha = \hat{\beta}(\alpha^{BC}) = \frac{1}{B}\sum_{b=1}^{B} I[\hat{\alpha}^*(b) < \alpha^{BC}].$$

(6) 위 (5)에서 구한 보정 신뢰계수 $\hat{\alpha}^{BC}$와 $\hat{\lambda}_i$를 이용하여 붓스트랩 보정된 수정 대표본 $100\alpha\%$ 신뢰상한을 구한다

$$\hat{\eta}_u^{ML}[\alpha^{BC}] = \hat{\lambda}_1 + \cdots + \hat{\lambda}_p + \sqrt{\hat{\lambda}_1^2\left(\frac{n_1}{u_1} - 1\right)^2 + \cdots + \hat{\lambda}_q^2\left(\frac{n_p}{u_p} - 1\right)^2} \qquad (4.3)$$

여기서 u_i 는 자유도 n_i를 가지는 카이제곱 분포의 $100(1-\hat{\alpha}^{BC})$ 하위 백분위수이다.

일반적으로 수정 대표본 $100\alpha\%$ 신뢰상한의 포함정확도 (coverage accuracy)가 $O(n^{-1/2})$ 으로 알려졌으므로 식 (4.3)에 주어진 붓스트랩 보정법에 의해 보정된 수정 대표본 $100\alpha\%$ 신뢰상한의 포함정확도는 이차 정확도 $O(n^{-1})$를 가지게 된다.

5. 모의실험

소표본하에서 붓스트랩에 의해 보정된 수정 대표본 신뢰상한 식 (4.3)의 포함정확도를 알아보기 위하여 간단한 모의실험을 실시하였다. 특히 분산 성분의 선형 결합에서 계수가

표 5.1: 수정 대표본 방법(MLS)과 붓스트랩 보정된 수정 대표본 방법(BC-MLS)에 의한 95% 신뢰상한의 포함정확도 ($\lambda_i = i$ for $i = 1, 2, \ldots, k$ 그리고 자유도(n)는 고정된 경우)

선형 결합의 수	자유도(n)=1		자유도(n)=5		자유도(n)=10	
k	MLS	BC-MLS	MLS	BC-MLS	MLS	BC-MLS
1	0.9570	0.9540	0.9430	0.9440	0.9400	0.9460
2	0.9930	0.9580	0.9880	0.9540	0.9780	0.9560
3	1.0000	0.9730	0.9840	0.9490	0.9780	0.9440
4	1.0000	0.9670	0.9920	0.9470	0.9920	0.9600
5	1.0000	0.9680	0.9990	0.9600	0.9880	0.9440
6	1.0000	0.9750	0.9980	0.9620	0.9940	0.9510
7	1.0000	0.9710	0.9980	0.9600	0.9910	0.9630
8	1.0000	0.9750	0.9970	0.9510	0.9920	0.9540
9	1.0000	0.9780	0.9980	0.9510	0.9970	0.9580
10	1.0000	0.9800	0.9990	0.9650	0.9960	0.9560

양수이며 그 성분의 수가 늘어날 때 수정 대표본 방법의 실제 포함확률이 주어진 신뢰 수준보다 크게 되는 단점을 보완하기 위하여 붓스트랩에 의한 보정을 고려하였으므로 모의실험도 이 경우에 맞추어 실시하였다. 앞의 2장에서 살펴보았듯이 균형실험하에서 분산 성분의 선형 결합에 대한 추정량의 분포는 식 (2.3)와 같이 카이제곱 확률변수의 간단한 선형 결합으로 나타나기 때문에 모의실험도 이러한 카이제곱 표현을 이용하여 실시하였다. 모의실험의 반복수와 붓스트랩 반복수는 모두 1000이다.

첫 번째 모의실험은 분산 성분의 선형 결합에서 그 계수가 각각 다르고 자유도는 고정되어 있는 경우를 고려하였다. 즉, 추정량의 분포가 k개의 서로 독립인 카이제곱 확률변수의 선형 결합이며 그 계수는 $\lambda_i = i$ 이고 자유도는 모두 같은 경우이다. 여기서 선형 결합의 수 는 하나인 경우에서부터 ($k = 1$) 열 개인 경우($k = 10$)까지 고려하고 자유도는 각각 $n = 1,\ 5,\ 10$ 인 경우를 고려하였다. 만약 선형 결합의 수가 k개이고 자유도가 n인 경우의 추정량은 다음과 같은 식에 의해 임의적으로 생성될 수 있다.

$$\mathcal{L}(\hat{\eta}) = \mathcal{L}(y_1 + y_2 + y_3 + \cdots + y_k)$$

여기서 y_i는 서로 독립인 확률변수이고 그 분포는 자유도가 n인 카이제곱 분포에 비례한다. 즉,

$$\mathcal{L}(y_i) = \mathcal{L}(i\chi^2(n)/n), \qquad \text{for } i = 1, 2, \ldots, k$$

표 5.1은 수정 대표본 방법 식 (4.2)와 붓스트랩 보정된 수정 대표본 방법 식 (4.3)에 의한 95% 신뢰상한의 포함정확도에 대한 모의실험 결과를 나타낸다. 주어진 자유도에 관계

없이 선형 결합의 수가 늘어남에 따라 수정 대표본 방법의 포함정확도는 주어진 신뢰계수 95%를 넘어서 100%에 빠르게 가까와진다. 반면 붓스트랩 보정된 수정 대표본 방법은 자유도가 아주 작은 경우($n = 1$)에 그 포함정확도가 선형 결합의 수에 따라 증가하지만 그 속도가 수정 대표본 방법에 비해 크게 느리고 더 나아가 자유도가 증가하면 ($n = 5,\ 10$) 선형 결합의 수에 거의 관계없이 그 포함정확도가 주어진 신뢰계수 95%에 가깝게 됨을 알 수 있다.

두 번째 모의실험은 분산 성분의 선형 결합의 추정량의 분포가 계수와 자유도가 같이 변하는 경우와

$$\mathcal{L}(\hat{\eta}) = \mathcal{L}(y_1 + y_2 + y_3 + \cdots + y_k) \quad \text{그리고} \quad \mathcal{L}(y_i) = \mathcal{L}[(k - i + 1)\chi^2(i)/i] \tag{5.1}$$

계수는 모두 같고 ($\lambda_i = 5$ for all i) 자유도만 변하는 경우를 고려하였다.

$$\mathcal{L}(\hat{\eta}) = \mathcal{L}(y_1 + y_2 + y_3 + \cdots + y_k) \quad \text{그리고} \quad \mathcal{L}(y_i) = \mathcal{L}(5\chi^2(i)/i) \tag{5.2}$$

표 5.2에 나타난 모의실험의 결과를 보면 선형 결합에서 계수와 자유도가 같이 변하는 식 (5.1)과 같은 경우에 붓스트랩 보정된 수정 대표본 방법은 선형 결합의 수에 크게 영향을 받지 않고 그 포함정확도가 주어진 신뢰계수 95%에 가깝게 됨을 알 수 있다. 또한 계수는 모두 같고 자유도만 변하는 식 (5.2)과 같은 경우 포함정확도가 선형 결합의 수에 따라 증가하는 경향이 있지만 그 속도가 보정하지 않은 수정 대표본 방법보다 훨씬 느린 것을 알수 있다. 위의 모의실험에서 본 결과로 붓스트랩 보정된 수정 대표본 방법은 선형 결합의 수

표 5.2: 수정 대표본 방법(MLS)과 붓스트랩 보정된 수정 대표본 방법(BC-MLS)에 의한 95% 신뢰상한의 포함정확도 (추정량의 분포가 식 (5.1)과 식 (5.2)의 경우)

선형 결합의 수	식 (5.1)의 경우		식 (5.2)의 경우	
k	MLS	BC-MLS	MLS	BC-MLS
1	0.9570	0.9540	0.9480	0.9500
2	0.9990	0.9510	0.9940	0.9680
3	0.9990	0.9660	0.9980	0.9720
4	1.0000	0.9530	1.0000	0.9780
5	0.9980	0.9570	1.0000	0.9660

에 크게 관계없이 그 포함정확도가 주어진 신뢰계수 95%와 매우 가까울뿐만 아니라 대부분의 경우에 신뢰계수보다 크다는 것을 알 수 있다.

6. 결론

본 연구에서는 분산 성분 모형 가정하에서 분산 성분들의 선형 결합에 대한 수정 대표

본 신뢰구간의 포함확률의 정도를 높이는 방법에 대하여 논의하였다. 붓스트랩 보정을 이용함으로서 신뢰구간 포함확률의 정도는 점근적 이차 계수까지 개선되며 특히 선형 결합의 계수가 모두 양수인 경우에 대한 추론에서 수정 대표본 신뢰구간의 포함확률이 주어진 신뢰계수보다 항상 커지게 되는 단점은 개선할 수 있음을 보였다.

본 연구에서 논의된 붓스트랩 보정 방법은 좀 더 일반적인 경우들에 확장될 수 있다. 즉 제안된 붓스트랩 보정은 수정 대표본 방법하에서 선형 결합 계수의 부호에 대한 제한이 없는 경우뿐만 아니라 또한 수정 대표본 방법이 아닌 다른 일반적인 신뢰구간 방법에 대해서도 적용할 수 있다.

참고문헌

Burdick, R. K. and Graybill, F. A. (1992). *Confidence intervals on variance components*, volume 127 of *Statistics: Textbooks and Monographs*, Marcel Dekker Inc., New York.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals , *Statistical Science*, **11**, 189-228, With comments and a rejoinder by the authors.

Graybill, F. A. and Wang, C. M. (1980). Confidence intervals on nonnegative linear combinations of variances, *Journal of the American Statistical Association*, **75**, 869-873.

Gui, R., Graybill, F. A., Burdick, R. K., and Ting, N. (1995). Confidence intervals on ratios of linear combinations for non-disjoint sets of expected mean squares, *Journal of Statistical Planning and Inference*, **48**, 215-227.

Lee, Y., Shao, J., and Chow, S. C. (2004). Modified large-sample confidence intervals for linear combinations of variance components: extension, theory, and application, *Journal of the American Statistical Association*, **99**, 467-478.

Loh, W. Y. (1987). Calibrating confidence coefficients , *Journal of the American Statistical Association*, **82**, 155-162.

Loh, W. Y. (1991). Bootstrap calibration for confidence interval construction and selection, *Statistica Sinica*, **1**, 477-491.

Lu, T. F. C., Graybill, F. A., and Burdick, R. K. (1988). Confidence intervals on a difference of expected mean squares, *Journal of Statistical Planning and Inference*, **18**, 35-43.

Lu, T. F. C., Graybill, F. A., and Burdick, R. K. (1989). Confidence intervals on the ratio of expected mean squares $(\theta_1 - d\theta_2)/\theta_3$, *Journal of Statistical Planning and Inference*, **21**, 179-190.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables*, volume 126 of *Statistics: Textbooks and Monographs*, Marcel Dekker Inc., New York. , Theory and applications.

Ting, N., Burdick, R. K., and Graybill, F. A. (1991). Confidence intervals on ratios of positive linear combinations of variance components, *Statistics and Probability Letters*, **11**, 523-528.

Ting, N., Burdick, R. K., Graybill, F. A., Jeyaratnam, S., and Lu, T.-F. C.(1990). Confidence intervals on linear combinations of variance components that are unrestricted in sign, *Journal of Statistical Computation and Simulation*, **35**, 135-143.

[2006년 5월 접수, 2006년 8월 채택]

Bootstrap Calibrated Confidence Bound for Variance Components Model*

Yonghee Lee[1)]

ABSTRACT

We consider use of Bootstrap calibration in the problem of setting a confidence interval for a linear combination of variance components. Based on the the modified large sample(MLS) method by Graybill and Wang(1980), Bootstrap Calibration is applied to improve the coverage probability of the MLS confidence bound when the experiment is balanced and coefficients of a linear combination are positive. Performance of the proposed confidence bound in small sample is investigated by simulation studies.

Keywords: Variance components, Confidence bound, Modified large sample method, Bootstrap calibration

* This work was supported by the Ewha Womans University Research Grant of 2005
1) Instructor, Department of Statistics, Ewha Womans University, Seoul 120-750, Korea.
E-mail: ylee@ewha.ac.kr