응용통계연구 (2010)
23(2), 375–382

붓스트랩을 활용한 최적 절사공간중위수 추정량

이동희[1] · 정병철[2]

[1]경기대학교 경영학과, [2]서울시립대학교 통계학과

(2010년 1월 접수, 2010년 3월 채택)

요 약

본 논문에서는 다변량 자료의 위치모수에 대한 로버스트 추정량으로 공간중위수에 대한 절사 추정량을 제안하였다. 최적절사율은 붓스트랩 방법을 이용하여 결정하였으며, 이중붓스트랩을 활용하여 추정된 절사공간중위수의 공분산행렬을 추정하였다. 모의실험 결과 붓스트랩 방법에 의한 절사공간중위수는 자료가 다변량 코시분포를 따르는 경우 기존 공간중위수에 비하여 작은 평균제곱오차를 보여 효율적인 추정량으로 나타났다. 아울러 이중붓스트랩을 이용한 절사추정량의 공분산행렬 추정량은 단순붓스트랩 방법에 의하여 추정된 공분산행렬이 갖는 과소추정의 문제를 해결하는 방법으로 나타났다.

주요용어: 공간중위수, 다변량위치모수, 붓스트랩, 절사공간중위수, 절사추정.

1. 서론

위치모수에 대한 추론에 있어서 자료에 대한 절단방법을 사용하는 것은 로버스트 성질을 유지시키기 위하여 흔히 사용되는 방법이다. 특히 다변량 자료에서 절사추정량에 대한 연구는 다양한 연구자에 의하여 연구되고 있다. Arcones (1995)는 다변량 절사평균의 점근적 정규성에 대하여 연구하였다. Gordaliza (1991)는 오염된 다변량 자료에 대한 절사공간중위수(trimmed spatial median; TSM) 추정량을, Vandev (1995)는 이를 계산하기 위한 알고리즘을 제안하였다. 최근 들어 Massé (2004, 2009)와 Zuo (2002)는 데이터뎁스(data depth)에 근거한 다변량 절사평균을 제안하였고 Massé와 Plante (2003)는 이변량 위치모수 추정문제에서 모의실험을 통하여 뎁스에 근거한 다변량 절사평균의 효율성을 파악하였다. 본 연구에서는 Gordaliza (1991) 및 Vandev (2005)에 의하여 제안된 TSM 추정량의 추론에 대하여 다루고자 한다. 사실 공간중위수(spatial median)는 일변량의 중위수를 다변량으로 확장한 형태의 위치모수에 대한 추정량으로 이상치에 로버스트하다는 성질을 갖고 있다 (Gower, 1974; Brown, 1983; Hettmansperger과 Randles, 2002). 그러므로 TSM은 기존의 공간중위수에서 일부 관측치에 대한 절사를 통해 로버스트한 위치모수에 대한 추정량을 얻기 위해 제안된 방법이다. 사실 TSM은 Gordaliza (1991) 및 Vandev (2005)에 의하여 제안되었지만 TSM의 통계적 성질에 대한 연구는 매우 미진한 상태이다. 본 연구에서는 붓스트랩 방법을 활용하여 이 추정량에 근거한 통계적 추론을 다루고자 한다. 먼저 단순붓스트랩을 활용하여 TSM 추정량에서 필요한 절사모수(trimming parameter)의 선택문제에 대하여 살펴보고자 한다. 사실 일변량의 경우 자료에 근거한 최적절사추정량에 대한 결과는 Jhun 등 (1993)에 의해 제시되었다. 본 연구에서는 Jhun 등 (1993)의 연구를 다변량으로 확장하고자 한다. 아

이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-314-C00039).

[2]교신저자: (130-743) 서울시 동대문구 전농동, 서울시립대학교 통계학과, 부교수. E-mail: bcjung@uos.ac.kr

울러 최적절사추정량의 공분산행렬은 이중붓스트랩 방법을 사용하여 추정하고자 한다. 이때 단순붓스트랩 대신 이중붓스트랩을 사용하여 공분산행렬을 추정하는 이유는 단순붓스트랩 방법에 의한 추정량은 실제 공분산행렬을 과소추정하기 때문이다 (Jhun 등, 1993).

이 논문의 구성은 다음과 같다. 2장에서는 TSM에 대하여 설명하고 이의 붓스트랩 과정에 대하여 상세하게 설명하고 모의실험을 통하여 단순 및 이중붓스트랩 방법의 효율성을 파악하였다. 3장에서는 실제 자료에 본 연구에서 제안한 TSM를 적용하고, 4장에서는 결론을 나타내었다.

2. 절사공간중위수

먼저 $X_1, X_2, \ldots, X_n$을 $p \times 1$ 벡터라 정의하자. Vandev (2005)에 의하면 TSM 추정량, $\hat{\theta}_\alpha$은 다음과 같이 정의된다.

$$\hat{\theta}(\alpha) = \arg_\theta \min \sum_{i=1}^{[n(1-\alpha)]} ||X_{(i)} - \theta||, \tag{2.1}$$

여기서 $[c]$는 c보다 큰 가장 작은 정수로서 α는 0과 0.5 사이의 값을 갖는 절사율을 나타내고, $||X_{(1)} - \theta|| \leq ||X_{(2)} - \theta|| \leq \cdots$는 크기순으로 정렬된 값을 나타낸다. TSM 추정량의 붕괴점(breakdown point)에 대한 연구는 Gordaliza (1991)에 의하여 연구되었고, Vandev (1995)는 이를 얻기 위한 알고리즘을 제안하였다. TSM에서 절사율이 0에 가까울수록 추정량의 효율성(efficiency)은 좋아지지만, 반면 붕괴점과 같은 로버스트성은 떨어지는 문제점이 발생하며, 마찬가지로 반대로 절사율 α가 0.5에 가까울수록 로버스트성은 높아지지만 추정량의 효율성은 떨어지게 된다. 따라서 최적의 절사율을 찾아내는 것이 중요하다. 이때 추정량의 공분산행렬이 이러한 절사율을 찾아내는 과정에 필요하지만, TSM 추정량에 대한 점근성(asymptotic property)은 아직 알려져 있지 않기 때문에, 본 연구에서는 붓스트랩 방법을 이용한 다음과 같은 방법을 통해 TSM 추정량의 표본분포를 추정하고 공분산행렬의 행렬식을 최소로 하는 최적의 절사율을 제안하고자 한다

2.1. 단순붓스트랩을 이용한 최적 절사공간중위수 추정

식 (2.1)과 같이 정의된 TSM 추정량의 표본분포를 추정하기 위한 붓스트랩 방법은 Jhun 등 (1993)의 연구를 다변량으로 확장하면 다음과 같다.

Step 1. 주어진 표본 $X_1, X_2, \ldots, X_n$으로부터 붓스트랩 표본 $X_1^*, X_2^*, \ldots, X_n^*$을 추출하여, 절사율 α에 대응되는 TSM의 추정량 $\hat{\theta}^*(\alpha)$를 구한다.

Step 2. Step 1을 독립적으로 B_1번 반복시행하여 B_1개의 $\hat{\theta}_j^*(\alpha)$를 구하여, 이로부터 $\hat{\theta}^*(\alpha)$의 공분산행렬 추정량을 다음과 같이 구한다.

$$\hat{\Sigma}_\theta^*(\alpha) = \frac{1}{B_1 - 1} \sum_{j=1}^{B_1} \left(\hat{\theta}_j^*(\alpha) - \bar{\hat{\theta}}^*(\alpha)\right) \left(\hat{\theta}_j^*(\alpha) - \bar{\hat{\theta}}^*(\alpha)\right)',$$

여기서 $\bar{\hat{\theta}}^*(\alpha) = \sum_{j=1}^{B_1} \hat{\theta}_j^*(\alpha)/B_1$이다.

Step 3. 모든 가능한 α에 대하여 $|\hat{\Sigma}_\theta^*(\alpha)|$을 구하고, 이를 최소로 하는 절사율 $\alpha = \alpha^*$를 얻는다. 여기서 $|A|$는 행렬 A의 행렬식(determinant)이다.

이와 같은 붓스트랩 과정을 통하여 최적절사율 α^*를 얻게 되면, 이를 이용한 식 (2.1)과 같이 정의된 최적 절사공간중위수 $\hat{\theta}(\alpha^*)$를 얻을 수 있다. 반면 이에 대한 공분산행을 알려져 있지 않기 때문에 본 연구

표 2.1. 각 추정량의 평균제곱오차

n	Dist.	ρ	절사공간중위수 μ_1	절사공간중위수 μ_2	절사공간중위수 μ_3	공간중위수 μ_1	공간중위수 μ_2	공간중위수 μ_3	표본평균 μ_1	표본평균 μ_2	표본평균 μ_3
50	Normal	0.3	0.0261	0.0273	0.0271	0.0249	0.0259	0.0254	0.0209	0.0211	0.0199
		0.6	0.0267	0.0254	0.0271	0.0242	0.0229	0.0247	0.0196	0.0187	0.0197
		0.9	0.0343	0.0329	0.0327	0.0273	0.0258	0.0257	0.0194	0.0191	0.0187
	Cauchy	0.3	0.0363	0.0368	0.0389	0.0401	0.0422	0.0414	286.116	1422.23	1337.17
		0.6	0.0354	0.0375	0.0373	0.0392	0.0392	0.0392	4848.02	7606.74	8990.80
		0.9	0.0439	0.0460	0.0438	0.0477	0.0487	0.0478	1109.99	775.268	1158.06
100	Normal	0.3	0.0122	0.0120	0.0127	0.0119	0.0117	0.0123	0.0101	0.0103	0.0107
		0.6	0.0130	0.0126	0.0130	0.0122	0.0118	0.0123	0.0104	0.0099	0.0102
		0.9	0.0163	0.0166	0.0173	0.0133	0.0134	0.0139	0.0098	0.0093	0.0097
	Cauchy	0.3	0.0170	0.0170	0.0179	0.0187	0.0201	0.0191	421.633	157.875	340.925
		0.6	0.0179	0.0198	0.0187	0.0187	0.0217	0.0193	471.594	235.692	464.003
		0.9	0.0209	0.0208	0.0205	0.0216	0.0220	0.0219	268.799	246.397	158.421

에서는 이에 대응되는 공분산행렬의 추정량으로써 Step 2에서 산출된 공분산 추정량 가운데 α^*에 대응되는 $\hat{\Sigma}_\theta^*(\alpha^*)$를 사용하고자 한다.

최적절사율 α^*를 이용한 위치모수에 대한 TSM의 효율성을 알아보기 위하여 간단한 모의실험을 실시하였다. 각 모의실험은 표본수 $n = 50$과 100에서 시행되었으며, 먼저 $p = 3$으로 고정시킨 상태에서 $X_1, X_2, \ldots, X_n$은 평균벡터가 모두 0이고 상관계수가 0.3, 0.6, 0.9인 다변량 정규분포와 코시분포에서 생성시켰다. 각 모수에서 총 1000번의 반복을 통하여 표본평균, 공간중위수 및 최적 절사모수를 이용한 절사공간중위수를 각각 계산하였다. 이때 절사율 α의 값은 0.0에서 0.4까지 0.05단위로 변화시켰으며 최적 절사모수를 찾기위한 붓스트랩 반복은 500번의 반복을 사용하였다. 다음 표 2.1은 각 모수조합에서 얻어진 각 추정량의 평균제곱오차(mean squared errors; MSE)를 나타낸다.

표 2.1의 결과를 살펴보면 먼저 $X_1, X_2, \ldots, X_n$가 다변량정규분포를 따르는 경우 표본평균의 MSE가 변수간 상관계수에 관계없이 가장 작게 나타나 평균벡터에 대한 가장 좋은 추정량으로 나타났다. 이 경우 공간중위수(SM)는 최적 절사모수를 이용한 절사공간중위수(TSM)보다는 약간 작은 MSE를 보이고 있다. 반면 $X_1, X_2, \ldots, X_n$가 다변량 코시분포를 따르는 경우 표본평균의 MSE는 아주 크게 나타나 평균모수에 대한 추정량으로 효율적이지 않음을 알 수 있다. 더불어 이 분포에서는 TSM추정량의 MSE가 SM추정량의 MSE보다 작게 나타나 SM추정량에 비하여 효율적인 추정량으로 판단된다.

2.2. 이중붓스트랩을 활용한 절사공간중위수의 공분산행렬 추정

이제 $\hat{\theta}(\alpha^*)$를 이용한 모평균벡터의 추론 문제를 생각해보자. 앞서 언급했듯이 TSM에 대한 점근성은 아직 알려져 있지 않기 때문에 붓스트랩에 기반한 $\hat{\Sigma}_\theta^*$를 통한 다음과 같은 이차형식(quadratic form)을 이용하여 귀무가설하의 모평균벡터 θ_0에 대해서 추론(구간추정 및 가설검정)을 고려해 볼 수 있다.

$$\left(\hat{\theta}(\alpha^*) - \theta_0\right)' \left(\hat{\Sigma}_\theta^*(\alpha^*)\right)^{-1} \left(\hat{\theta}(\alpha^*) - \theta_0\right) \sim \chi_p^2. \tag{2.2}$$

그러나 여기서 사용된 $\hat{\theta}(\alpha^*)$에 대한 공분산행렬의 추정량 $\hat{\Sigma}_\theta^*(\alpha^*)$는 과소추정하는 경향이 있다. 왜냐하면 최적절사율 α^*는 $|\hat{\Sigma}_\theta^*(\alpha))|$를 최소화하는 α이기 때문이다 (Jhun 등, 1993). 이와 같은 문제는 다음과 같은 이중붓스트랩을 통하여 해결할 수 있다.

먼저 앞 절의 단순붓스트랩 방법의 알고리즘에 이어서 다음과 같은 이중붓스트랩 방법을 고려해보자.

Step 4. Step 1에서 구한 각 붓스트랩표본으로부터 이중붓스트랩표본 $X_1^{**}, X_2^{**}, \ldots, X_n^{**}$을 추출하여, 절사율 α의 TSM 추정량 $\hat{\theta}^{**}(\alpha)$를 구한다.

Step 5. Step 4를 독립적으로 반복시행하여 B_2개의 TSM 추정량 $\hat{\theta}^{**}(\alpha)$들을 가능한 α의 값에 대하여 구하고, 이로부터 $\hat{\theta}^{**}(\alpha)$의 공분산행렬의 행렬식을 최소화하는 $\alpha = \alpha^{**}$를 Step 1의 각 붓스트랩표본 B_1개에 대하여 모두 구한다.

Step 6. Step 5에서 구한 α^{**}를 해당하는 붓스트랩 표본의 절사율로 사용한 B_1개의 $\hat{\theta}_j^{**}(\alpha^{**})$를 구한다. 이와 같이 얻어진 B_1개의 $\hat{\theta}_j^{**}(\alpha^{**})$를 이용하여 Step 3에서 얻어진 최적 절사량에 대응되는 TSM 추정량 $\hat{\theta}(\alpha^*)$의 공분산행렬은 다음과 같이 추정된다.

$$\hat{\Sigma}_{\theta}^{**}(\alpha^*) = \frac{1}{B_1 - 1} \sum_{j=1}^{B_1} \left(\hat{\theta}_j^{**}(\alpha^{**}) - \bar{\hat{\theta}}^{**}(\alpha^{**})\right)\left(\hat{\theta}_j^{**}(\alpha^{**}) - \bar{\hat{\theta}}^{**}(\alpha^{**})\right)',$$

여기서 $\bar{\hat{\theta}}^{**}(\alpha^{**}) = \sum_{j=1}^{B_1} \hat{\theta}_j^{**}(\alpha^{**})/B_1$이다.

이상과 같은 과정을 통하여 얻어진 최적 TSM 추정량 $\hat{\theta}(\alpha^*)$의 공분산행렬 추정량 $\hat{\Sigma}_{\theta}^{**}(\alpha)$이 적절히 추정되는가에 대해서는 단일 모집단에서 위치모수벡터에 대한 공동신뢰구간이나 가설검정을 통하여 알아볼 수 있다.

이를 위하여 앞 절에서 시행한 모의실험을 동일하게 실시하였다. 즉, 표본수 $n = 50$과 100을 고려한 상태에서 $p = 3$으로 고정시키고 $X_1, X_2, \ldots, X_n$은 평균벡터가 모두 0이고 상관계수가 0.3, 0.6, 0.9인 다변량 정규분포와 코시분포에서 생성시켰다. 각 모수에서 총 1000번의 반복을 통하여 표본평균, 공간중위수 및 최적 절사모수를 이용한 절사공간중위수를 각각 계산하였다. 이때 절사율 α의 값은 0.0에서 0.4까지 0.05단위로 변화시켰으며 단순붓스트랩과 이중붓스트랩에 필요한 붓스트랩 반복은 각각 $B_1 = 500$과 $B_2 = 200$을 사용하였다. TSM의 공분산행렬은 각각 단순붓스트랩과 이중붓스트랩을 사용하여 추정한 $\hat{\Sigma}_{\theta}^{*}(\alpha^*)$과 $\hat{\Sigma}_{\theta}^{**}(\alpha^*)$를 이용하였다. 이제 다음과 같은 가설을 고려해보자.

$$H_0 : \vec{\theta} = 0 \qquad H_1 : \vec{\theta} \neq 0. \tag{2.3}$$

이와 같은 가설에 대한 검정통계량으로 표본평균을 이용하는 Hotelling의 T^2 통계량, 공간중위수를 이용한 검정통계량 M^2 (Somorcik, 2006), 식 (2.2)를 이용한 최적 절사공간중위수를 이용한 검정통계량을 각각 구하고 이들의 명목유의수준을 구하였다. 이때 최적 절사공간중위수를 이용하는 경우 단순붓스트랩에 의한 공분산행렬 추정치와 이중붓스트랩에 의한 공분산행렬 추정치를 각각 사용하여 검정통계량을 구성하였다. 다음 표 2.2는 가설 (2.3)에 대한 각 검정통계량의 추정된 유의수준을 나타낸다. 본 연구에서는 검정력에 대한 모의실험은 실시하지 않았다. 그 이유는 이중붓스트랩에 의한 공분산행렬 추정치의 적절성에 모의실험의 초점이 있기 때문이다.

표 2.2를 살펴보면 먼저 Hotelling의 T^2 통계량은 표본수에 관계없이 다변량 정규분포를 따르는 경우 명목유의수준을 제대로 유지한 반면 코시분포를 따르는 경우 명목유의수준을 과소추정하는 것으로 나타났다. 공간중위수를 이용한 M^2 통계량은 표본수 및 자료의 분포에 관계없이 명목유의수준을 과대추정하는 것으로 나타났다. 최적 TSM에서 단순붓스트랩을 이용하여 공분산행렬을 추정하는 경우 명목유의수준을 약간 과대추정하는 것으로 나타나 공분산행렬에 대하여 과소추정하는 것으로 나타났다. 반면 이중붓스트랩을 이용하여 공분산행렬을 추정하는 경우 자료의 분포와 표본수에 관계없이 명목유의수준을 제대로 유지하여 단순붓스트랩에서 나타나는 공분산행렬의 과소추정 현상을 개선시키고 있음을 알 수 있다.

표 2.2. 고려된 각 검정의 추정된 유의수준(명목유의수준 0.05)

n	Dist.	ρ	표본평균 T^2	공간중위수 M^2	최적 절사공간중위수 단순붓스트랩	최적 절사공간중위수 이중붓스트랩
50	Normal	0.3	0.053	0.107	0.112	0.070
		0.6	0.041	0.089	0.086	0.056
		0.9	0.043	0.127	0.089	0.059
	Cauchy	0.3	0.015	0.086	0.085	0.061
		0.6	0.011	0.092	0.080	0.064
		0.9	0.013	0.160	0.091	0.059
100	Normal	0.3	0.053	0.082	0.076	0.062
		0.6	0.048	0.069	0.075	0.061
		0.9	0.048	0.095	0.090	0.058
	Cauchy	0.3	0.018	0.069	0.089	0.057
		0.6	0.017	0.078	0.097	0.064
		0.9	0.016	0.104	0.090	0.063

표 3.1. Hawkins-Bradu-Kass 자료 가운데 사용된 변수의 평균과 표준편차

	x_1		x_2		x_3	
	평균	표준편차	평균	표준편차	평균	표준편차
전체	3.21	3.653	5.60	8.239	7.23	11.740
1–14	10.48	0.838	22.23	3.854	31.39	2.665
1–14 제외	1.54	1.064	1.78	1.073	1.69	1.034

표 3.2. Hawkins-Bradu-Kass 자료에 대한 위치모수 추정량

모수	표본평균($\bar{X}$)	공간중위수	절사공간중위수(최적절사율)
μ_1	3.2067	1.6769	1.5374
μ_2	5.5973	2.1414	1.8893 (0.15)
μ_3	7.2307	2.1195	1.7101

3. 사례연구

본 절에서는 사례를 통해 본 논문에서 제안한 최적 TSM의 추정과 이의 공분산행렬의 추정결과를 살펴보고자 한다. 이 자료는 Hawkins 등 (1984)에서 사용된 것 가운데 일부를 발췌한 것으로 소선하 등 (2009)은 이 자료를 이용하여 다변량 위치모수의 검정을 시도하였다. 원자료는 회귀분석에 사용되었으나 본 절에서는 독립변수로 사용된 부분만을 이용하여 모평균벡터 및 그의 공분산행렬에 대한 추정을 실시하였다. 분석에 사용된 자료는 3-변량 75개의 관찰벡터로 구성되었으며, 이 가운데 최초 14개의 관찰벡터는 이상치(18.7%)로 알려져 있다. 이 자료의 평균과 표준편차는 다음 표 3.1에 나타나 있다.

표 3.1은 각각 원자료에 대해서 각 변수별로 전체, 이상치로 알려진 14개의 관측치(1–14) 및 이들 이상치를 제외한 나머지 자료(1–14 제외)에 대한 평균과 표준편차를 나타낸 것이다. 14개의 이상치에 의해 전체 자료의 평균과 표준편차가 크게 나타나 이들에 의해 자료가 크게 왜곡돼 있음을 확인할 수 있다. 이 자료에 대한 평균벡터에 대한 추정치와 그의 공분산행렬 추정치는 다음과 같다.

표 3.2의 결과를 살펴보면, 먼저 표본평균은 이상치들로 인하여 크게 추정되고 있다. 이에 반해 SM과 TSM은 이러한 이상치의 영향이 축소된 형태로 나타났다. TSM의 경우 최적절사율은 15%로 나타나 데이터의 이상치 비율과 거의 비슷한 절사율을 보였다. 아울러 위치모수에 대한 추정치도 14개의 이상치

표 3.3. Hawkins-Bradu-Kass 자료에 대한 위치모수 추정량의 공분산행렬 추정치

모수	표본평균			공간중위수			절사공간중위수					
							단순붓스트랩			이중붓스트랩		
	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
μ_1	0.1778	0.3795	0.5499	0.0409	0.0062	0.0133	0.0458	−0.0001	0.0116	0.0597	0.0199	0.0370
μ_2	0.3795	0.9051	1.2622	0.0062	0.0319	0.0154	−0.0001	0.0357	0.0067	0.0199	0.0500	0.0370
μ_3	0.5499	1.2622	1.8377	0.0133	0.0154	0.0516	0.0116	0.0067	0.0473	0.0370	0.0370	0.0940

를 제외한 표본평균들과 비슷한 추정량을 보이고 있다.

표 3.3은 위치모수 추정량의 공분산행렬을 추정한 결과를 나타낸 것이다. 표본평균의 공분산행렬 추정치는 이상치들로 인하여 크게 나타나고 있는 반면, SM과 TSM의 공분산행렬 추정치는 이상치의 영향을 덜 받는것으로 보인다. TSM 추정량의 경우 단순붓스트랩과 이중붓스트랩을 이용한 공분산행렬 추정치를 비교해보면 단순붓스트랩을 이용한 공분산행렬은 이중붓스트랩을 이용한 것보다 비대각 원소에서는 큰 차이가 없지만 대각원소에서는 작게 추정되는 것으로 나타나 단순붓스트랩의 분산추정치가 과소추정되고 있음을 알 수 있다. 하지만 이중붓스트랩을 이용하는 경우 이와 같은 단순붓스트랩의 과소추정문제를 해결하고 있는것으로 판단된다.

4. 결론

본 논문에서는 다변량 자료의 위치모수에 대한 로버스트 추정량으로 공간중위수에 대한 절사추정량을 제안하였다. 최적절사율은 붓스트랩 방법을 이용하여 결정하였으며, 이중붓스트랩을 활용하여 TSM 추정량의 공분산행렬을 추정하였다. 이는 일변량에서 절사평균에 대한 Jhun 등 (1993)의 연구를 다변량으로 확장한 것이다. 모의실험 결과 본 연구에서 제안한 TSM은 자료에 이상치가 많이 포함되어 있는 다변량 코시분포를 따르는 경우 공간중위수에 비하여 작은 MSE를 보여 효율적인 추정량으로 나타났다. 아울러 이중붓스트랩을 이용한 절사추정량의 공분산행렬 추정량은 단순붓스트랩 방법에 의하여 추정된 공분산행렬이 갖는 공분산행렬의 과소추정 문제를 해결하는 방법으로 나타났다.

참고문헌

소선하, 이동희, 정병철 (2009). 다변량 자료에서 위치모수에 대한 로버스트 검정, <응용통계연구>, **22**, 1355–1364.

Arcones, M. A. (1995). Asymptotic normality of multivariate trimmed means, *Statistics and Probability Letters*, **25**, 43–53.

Brown, B. M. (1983). Statistical uses of the spatial median, *Journal of the Royal Statistical Society B*, **45**, 25–30.

Gordaliza, A. (1991). Best approximations to random variables based on trimming procedures, *Journal of Approximation Theory*, **64**, 162–180.

Gower, J. C. (1974). Algorithm AS 78: The mediancentre, *Applied Statistics*, **23**, 466–470.

Hawkins, D. M., Bradu, D. and Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets, *Technometrics*, **26**, 197–208.

Hettmansperger, T. P. and Randles, R. H. (2002). A practical affine equivariant multivariate median, *Biometrika*, **89**, 851–860.

Jhun, M., Kang, C. W. and Lee, J. C. (1993). Bootstrapping trimmed estimators in statistical inferences, *Proceedings of the Asian Conference on Statistical Computing*.

Massé, J-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean, *Bernoulli*, **10**, 379–419.

Massé, J-C. (2009). Multivariate trimmed means based on the Tukey depth, *Journal of Statistical Planning and Inference*, **139**, 366–384.

Massé, J-C. and Plante, J-F. (2003). A Monte Carlo study of the accuracy and robustness of ten bivariate location estimators, *Computational Statistics & Data Analysis*, **42**, 1–26.

Somorcik, J. (2006). Tests using spatial median, *Austrian Journal of Statistics*, **35**, 331–338.

Vandev, D. L. (1995). Computing of trimmed L_1 median, In *Multidimensional Analysis in Behavioral Sciences. Philosophic to technical*, 152–157.

Zuo, Y. (2002). Multivariate trimmed means based on data depth, In *Statistical Data Analysis Based on the L1-Norm and Related Methods*, (ed. by Y. Dodge), 313–322.

A Trimmed Spatial Median Estimator Using Bootstrap Method

Dong-Hee Lee[1] · Byoung Cheol Jung[2]

[1]Department of Business Administration, Kyonggi University
[2]Department of Statistics, University of Seoul

(Received January 2010; accepted March 2010)

Abstract

In this study, we propose a robust estimator of the multivariate location parameter by means of the spatial median based on data trimming which extending trimmed mean in the univariate setup. The trimming quantity of this estimator is determined by the bootstrap method, and its covariance matrix is estimated by using the double bootstrap method. This extends the work of Jhun *et al.* (1993) to the multivariate case. Monte Carlo study shows that the proposed trimmed spatial median estimator yields better efficiency than a spatial median, while its covariance matrix based on double bootstrap overcomes the under-estimating problem occurred on single bootstrap method.

Keywords: Bootstrap, multivariate location parameter, spatial median, trimming estimation, trimmed spatial median.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund)(KRF-2007-314-C00039).

[2]Corresponding author: Associate Professor, Department of Statistics, University of Seoul, Jeonnong-Dong 90, Dongdaemun-Gu, Seoul 136-743, Korea. E-mail: bcjung@uos.ac.kr