최고의 정규 모집단을 뽑기 위한 부분집합선택절차론의 운용특성에 관한 연구+

손 중 권* Shanti S. Gupta**

1. 서 론

지난 30여년간 급격히 발전해 온 다중결정이론 중 부분집합선택론은 매우 중요한 위치를 차지하고 있다. 특히 여러가지 형태의 부분집합선택절차론 중에서 최초로 소개된 Gupta형 선택절차론은 모든 절차론들의 기본이 되어 오고 있음으로 그 중요성은 널리 인식되고 있다. 더우기 응용부문에 있어서도 가장 많이 사용되고 있는 선택절차론들 중의 하나이기도 하다. 따라서 Gupta형 선택절차론에 대한 일반적인 성질들도 많이 규명되어 왔다. 특히 결정론적 측면에서나 베이스 이론적 측면에서의 최적성 및 점근적 효율성에 있어서는 Gupta와 Hsu (1978), Bjørnstad(1980), 그리고 Bickel과 Yahav(1982)가 성질 들을 규명내지는 다른 형의 부분집합선택절차론들과 특정분포에 대해 비교 검토하였다. 또한 수집된 자료가 선택절차론의 근본 가정들을 위반할 경우가 실제로 다반사로 일어난다. 따라서 근본가정이 위배될 경우 선택절차론의 강건력에 대해서도 연구가 부분적으로 진행되었다. Gupta형 선택절차론과 중앙값 선택절차론과의 비교도 Gupta와 Singh(1980)과 Sohn(1985)에 의해 진행되었으며, 특히 스리피지 배치에서 점근적 효율성을 연구하였다. 하지만 부분집합선택절차론이 차지하는 중요성에 비해 그 자체에 대한 여러 측면에 있어서의 성질 및 운용특성에 대한 포괄적이고 일반적인 연구는 미흡한 편이다.

본 연구에서는 부분집합선택절차론의 운용특성에 대해 조사하였다.

2절에서는 절차론의 정의와 기존의 결과들 및 본 논문에서 사용되는 기호들을 소개하였다.

또 3절에서는 데이타가 기본 가정을 전혀 위배하지 않는 경우 주어진 실험조건에서 절차론을 적용할 때 실험자가 정해야 할 조건에 대해 조사하였다. 즉, 주어진 조건이 바뀜에 따라 절차론의 운용특성인 효율과 기타 성취도의 변화에 대해 알아 보았다.

4절에서는 데이타가 오염이 되었을 때 절차론의 효율 및 성취도 등에 미치는 영향에 대해 조사하였다. 이러한 조사는 시뮬레이션을 통해 진행되었다.

+ 본 연구는 1987년도 한국과학재단 연구비 지원에 의한 결과임. (과제번호 : 871-0105-00921)

* 경북대학교 자연과학대학 통계학과, 대구시 북구 산격동 1370.

** Dept. of Statistics, Purdue University, West Lafayette, IN, U. S. A.

2. 선택절차 R의 정의 및 기본 정리

$\Pi_1, \Pi_2, \cdots, \Pi_k$는 $k(\geqq 2)$개의 독립인 정규모집단으로 미지의 모평균 μ_i와 동일하면서 미지인 모분산 σ^2을 갖는다고 하자. 모수공간 Ω는

$$\Omega = \{\underset{\sim}{\mu} \in R^k \mid \underset{\sim}{\mu} = (\mu_1, \mu_2, \cdots, \mu_k),\ -\infty < \mu_i < \infty,\ 0 < \sigma^2 < \infty\}$$

로 정의된다. 여기서 $\mu_{[1]} \leqq \mu_{[2]} \leqq \cdots \leqq \mu_{[k]}$는 μ_i들을 크기 순서대로 나열한 것으로 $\mu_{[i]}$에 해당되는 정규모집단을 $\Pi_{(i)}$라 할 때 $\mu_{[k]}$의 값을 모평균으로 갖는 정규모집단 $\Pi_{(k)}$를 최고의 정규모집단이라 정의한다. 단 $\mu_{[i]}$와 $\Pi_{(i)}$ 사이의 관계는 미지라 가정한다. 이렇게 모집단이 주어졌을 때 우리의 목적은 최고의 정규모집단 $\Pi_{(k)}$를 뽑는 것이다. 이 목적을 달성하기 위해 X_{ij}, $j = 1, 2, \cdots, n$를 모집단 Π_i로부터 얻어진 n개의 독립인 확률표본이라 하자. 그러면 각 정규모집단에서의 표본평균은 $\overline{X}_i = \sum_{j=1}^{n} X_{ij}/n$으로 정의된다. 이렇게 얻어진 $\overline{X}_i$를 기초로 해서 Gupta에 의해 제안된 부분집합선택절차(이하 R이라 표기한다)는 다음과 같다.

R : Π_i를 뽑기 위한 필요충분조건은 $\overline{X}_i \geqq \overline{X}_{[k]} - dS/\sqrt{n}$이다.

여기서 $X_{[k]}$는 $\overline{X}_i$들 중에서 제일 큰 값이며 $S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2/k(n-1)$은 동일한 미지의 분산 σ^2의 합동추정량이다. 이 절차 R에 의해 $\Pi_1, \Pi_2, \cdots, \Pi_k$ 중에서 뽑힌 부분집합을 I로 표시했을 때 만약 $\Pi_{(k)}$가 I에 포함되어지면 우리는 이를 올바른 선택(Correct Selection, CS)이라 하며 올바른 선택이 이루어질 확률, 즉 P(CS)에 대한 조건, 즉 $\inf_{\Omega} P(CS) \geqq P^*$, $(1/k < P^* < 1)$ 이 제약조건으로 주어진다. 따라서 절차 R에서 사용되는 d는 음이 아닌 상수로써 절차 R이 P^*－조건을 만족하도록 고안되어지며, 정의된 절차 R에 대해 다음의 정리가 성립한다.

정리 1. Gupta형 부분집합선택절차 R에 대해 P(CS)의 최소값은 다음과 같이 얻어지며 따라서 이 식으로부터 P^*－조건을 만족하는 상수 d의 값을 구할 수 있다. 이 때 d는 n, k 그리고 P^*의 함수이다.

$$P(CS \mid R) \geqq \inf_{\mu \in \Omega} P(CS \mid R) = \int_0^{\infty}\int_{-\infty}^{\infty} \Phi^{k-1}(x + du)\, d\Phi(x)\, dQ_\nu(u) \tag{1}$$

여기서 $Q_\nu(\cdot)$는 자유도가 $\nu = k(n-1)$인 $\chi(\nu)/\sqrt{\nu}$ 분포의 누적확률밀도함수이다.

위의 정리 1로부터 n, k 그리고 P^*의 여러 값에 대한 d값은 Gupta, Panchapakesan 그리고 Sohn(1985)의 논문에 실려 있다. 만약 모집단 $\Pi_{(i)}$가 선택되어질 확률을 $p_{(i)}$이라 하면 $p_{(i)} = P(\Pi_{(i)} \in I)$이며 $p_{(i)}$의 값은

$$p_{(i)} = \int_0^{\infty}\int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq i}}^{k} \Phi\left(x + du + \frac{\mu_{[i]} - \mu_{[j]}}{\sigma/\sqrt{n}}\right) d\Phi(x)\, dQ_\nu(u) \tag{2}$$

임을 알 수 있다. 따라서, 절차 R에 의해 뽑혀진 부분집합 I의 크기를 $|I|$로 표시하면 $|I|$의 기대값 $E(|I|)$는 다음과 같이 얻어진다.

$$\begin{aligned} E(|I|) &= \sum_{i=1}^{k} p_{(i)} \\ &= \sum_{i=1}^{k} \int_0^{\infty} \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq i}}^{k} \Phi\left(x + du + \frac{\mu_{[i]} - \mu_{[j]}}{\sigma/\sqrt{n}}\right) d\Phi(x)\, dQ_\nu(u) \end{aligned} \tag{3}$$

또한 다음과 같은 정리가 얻어진다.

정리 2. 부분집합선택절차 R을 사용할 때 뽑혀지는 집합 I의 크기에 대한 기대치 $E(|I|)$에 대해

$$\sup_{\underline{\mu} \in \Omega} E(|I|) = kP^*$$

이다.

3. 절차 R의 운용특성

그러면 여기서 절차 R의 운용특성에 관해서 조사해 보기로 한다. 일반적으로 절차 R을 적용할 때 실험자가 정해야 할 값중에서 일반적으로 k와 n은 실험의 경제적 조건에 의해 정해진다고 보면 실험자는 P(CS)의 최소값인 P^*값을 정해야 함으로 따라서 여기에 대해 생각해 보자.

부분집합선택절차론 R에 대한 최소로 유리한 배열(least favorable configuration, LFC)은 정리 1로부터 $\mu_1 = \mu_2 = \cdots = \mu_k$이 됨을 알 수 있다. 따라서 주어진 조건하에 LFC상에서 얻은 d값은 실제의 모평균 μ_i들의 배열을 고려하면 상당히 조심스러운(conservative) 값이 됨으로 이러한 값을 사용하면 이 때 얻어지는 부분집합 I의 크기는 상당히 커지게 되어 궁극적으로 "최고"의 모집단을 구별하는데도 어려움이 따를 뿐만 아니라 절차 R 그 자체가 무의미해진다. 이러한 점을 고려해서 절차 R의 효율(effciency)을 P(CS)에 대한 $E(|I|)$의 비로 간주될 수 있기에

$$EFF(R) = \frac{kP(CS)}{E(|I|)} \tag{4}$$

와 같이 정의하기로 한다. 그러면 LFC상에서는 $kP(CS) = E(|I|)$임을 알 수 있으므로, 따라서 $EFF(R) = 1$이 되며, 정리 1과 정리 2로부터 모평균 μ_i들의 다른 배열상에서도 $kP(CS) \geq E(|I|)$이므로 $EFF(R) \geq 1$임을 알 수 있다. 여기서 우리는 모평균 μ_i들의 두가지 배열, 즉 스리피지배열(slippage configuration)인

$$\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \triangle, \quad \triangle > 0$$

과 등간격배열(equi-spaced configuration)인

$$\mu_{[1]} + (k-1)\triangle = \mu_{[2]} + (k-2)\triangle = \cdots = \mu_{[k-1]} + \triangle = \mu_{[k]},\ \triangle > 0$$

을 생각해 보기로 한다.

스리피지배열상에서

$$P(CS \mid R) = \int_0^\infty \int_{-\infty}^\infty \Phi^{k-1}\left(x + du + \frac{\triangle}{\sigma/\sqrt{n}}\right) d\Phi(x)\, dQ_\nu(u) \tag{5}$$

이고

$$E(\mid I \mid) = P(CS) + (k-1) \int_0^\infty \int_{-\infty}^\infty \Phi^{k-1}\left(x + du - \frac{\triangle}{\sigma/\sqrt{n}}\right) d\Phi(x) dQ_\nu(u) \tag{6}$$

가 되며 또한 등간격배열상에서는

$$P(CS \mid R) = \int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \Phi\left(x + du + \frac{(k-1)\triangle}{\sigma/\sqrt{n}}\right) d\Phi(x) dQ_\nu(u) \tag{7}$$

이고

$$E(\mid I \mid) = P(CS) + \sum_{i=1}^{k-1} \int_0^\infty \int_{-\infty}^\infty \prod_{\substack{i=1 \\ j\neq i}}^{k} \Phi\left(x + du - \frac{(k-j)\triangle}{\sigma/\sqrt{n}}\right) d\Phi(x) dQ_\nu(u) \tag{8}$$

가 된다.

이상의 식 (5), (6), (7), (8)의 값들은 수치적 적분(numerical integration)으로도 가능하지만 몬테칼로 시뮬레이션(Monte Carlo simulation)을 통해 구했다.

시뮬레이션 및 결과

모평균 μ_i들의 배치에 상관없이 P* = 0.90, P* = 0.95와 k = 2, 3, 5, 8, 10, 15에서 각각의 다른 ν값이 주어졌을 때 $\delta\sqrt{n}$ = 0.1(0.2)0.5(0.5)2.0, 5.0($\delta = \triangle/\sigma$)인 경우를 택했다. 여기서 사용되는 난수는 미국 Purdue 대학교 전자계산소에 있는 RVP 부프로그램을 이용해서 얻었으며 시뮬레이션의 횟수는 모든 경우 각 1,000번씩 했다. 또 여기서 사용된 상수 d는 Gupta, Panchapakesan 그리고 Sohn(1985)의 논문으로부터 나온 값을 이용했다. 모든 계산결과는 미국 Purdue 대학교에 있는 컴퓨터 CDC6600을 이용해서 얻어졌다. 이 계산에서 얻어진 P(CS), E(| I |)과 EFF(R)들의 값들은 표－1과 표－2에 수록되어 있다. 이 표들로부터 다음과 같은 사실들을 알 수가 있다.

1. P* = 0.90과 P* = 0.95일 때의 효율을 비교해 보면 어떤 배열이든 거의 모든 k, $\delta\sqrt{n}$과 ν의 값에서 P* = 0.90일 때의 효율이 P* = 0.95일 때의 효율보다 높은 것을 알 수 있다.

따라서 P^*-조건을 강하게 하는 것이 반드시 절차 R의 효율을 높이는 것은 아니다.

2. $P^* = 0.90$과 $P^* = 0.95$에서 효율의 차이는 $\delta\sqrt{n}$에 관계없이 등간격배열상에서의 차이가 스리피지배열에서의 차이보다 일반적으로 크다.
3. 스리피지배열에서나 등간격배열에서 $\delta\sqrt{n}$의 값이 0.5이상이면 P(CS)의 추정치가 $P^* = 0.90$인 조건하에서 대부분 0.95이상이 되어 있어 굳이 효율이 떨어지는 $P^* = 0.95$인 강한 조건은 필요가 없다. 각 모집단으로부터 얻어지는 표본의 크기 n이 25 이상이 되면 모집단의 수 k에 상관없이 $P^* = 0.90$에서 P(CS)의 추정치들은 0.95 이상이 됨을 보여주고 있다.
4. 스리피지배열상에서 $\delta\sqrt{n}$이 충분히 크더라도 절차 R이 뽑는 부분집합 I의 기대치 E(|I|)의 값이 상당히 크다. 이는 절차 R의 취약점으로 여겨진다. 따라서 모집단의 수가 크면 절차 R을 사용할 것인가는 생각해 볼 필요가 있다. 왜냐하면 E(|I|)가 커질 때는 원래의 목적인 최고의 정규모집단을 찾아내는데 어려움이 있을 뿐만 아니라 절차 R 자체가 무의미해지기 때문이다.

4. 절차 R에 대한 오염된 데이타의 영향

일반적으로 데이타가 오염이 되는 경우, 분포뿐만 아니라 평균과 분산이 전혀 다른 경우이나 여기서는 원래의 모집단과 같은 평균을 가지나 분포가 다르거나 분산이 다른 모집단으로부터 오염(contamination)되었을 경우 절차 R의 성취도를 시뮬레이션을 통해 알아 보고자 한다.

평균이 μ_i이고 미지의 분산이 σ_0^2인 정규확률밀도함수를 $f_1(x : \mu_i, \sigma_0^2)$이라 하자. 정규모집단 Π_i로부터 얻은 n개의 독립인 확률표본 X_{ij}중 $100\epsilon\%$는 평균은 μ_i이나 분산이 τ^2인 확률밀도함수 $f_2(x : \mu_i, \tau^2)$를 가지는 모집단으로부터 오염된 것이라 하자. 그러면 X_{ij}, $j = 1, 2, \cdots, n$에 대한 확률밀도함수 $f(x : \mu_i, \sigma^2)$는

$$f(x : \mu_i, \sigma^2) = (1-\epsilon)f_1(x : \mu_i, \sigma_0^2) + \epsilon f_2(x : \mu_i, \tau^2)$$

가 된다. 이 때 우리는 이 오염현상을 모르고 있다고 가정한다. 이러한 경우 ϵ이 1로 가까이 갈수록 즉 $f(x : \mu_i, \sigma^2)$가 원래의 $f_1(x_i : \mu_i, \sigma_0^2)$으로부터 멀어질수록 절차 R의 성취도에 대한 오염현상의 영향력은 일반적으로 커지게 된다. Gupta와 Singh(1980)은 $f_2(x_i : \mu_i, \tau^2)$가 정규확률밀도함수이며 σ_0^2과 τ^2이 기지인 경우 절차 R을 중앙값선택절차론과 비교하였다. 그 결과 $\tau^2 = 9.1\sigma_0^2$인 경우 절차 R이 P^*-조건을 만족치 못함을 보여주었다.

여기서 우리가 고려하는 경우는 $f_2(x : \mu_i, \tau^2)$가 우선 정규확률밀도함수인 경우 $\tau^2 = 4\sigma_0^2$과 $\tau^2 = 9\sigma_0^2$인 경우와 $f_2(x : \mu_i, \tau^2)$가 logistic확률밀도함수인 경우 $\tau^2 = \sigma_0^2$과 $\tau^2 = 4\sigma_0^2$인 경우에 대해 절차 R의 성취도를 보고자 하였다. 즉, $f_2(x : \mu_i, \tau^2)$가 정규확률밀도함수인 경우는 분산이 원래의 정규모집단보다 훨씬 큰 모집단으로부터의 오염이 된 경우이다. 또 logistic분포

로부터 오염이 된 경우는, 비록 두 모집단의 분산이 같더라도 정규분포보다 더 두터운 꼬리 부분을 갖게되거나 분산이 더 큰 logistic분포로부터 오염이 되는 경우이다. 여기서 고려되는 평균값 μ_i들의 배치는 스리피지배열과 등간격배열 두가지 경우이다.

시뮬레이션 및 결과

모평균 μ_i 들의 배치에 상관없이 $P^* = 0.90$, 0.95와 $k = 3$인 경우 $\nu = 15$, 30, 60, $k = 5$인 경우 $\nu = 30$, 60, 120, $k = 8$인 경우 $\nu = 48$, 120을 택했다. 또 각각의 k와 ν에 대해 오염율 ϵ의 값은 다음과 같다.

k	3			5			8	
ν	15	30	60	30	60	120	48	120
ϵ	0.167	0.182	0.143	0.143	0.154	0.120	0.143	.0.188
	0.333	0.273	0.238	0.286	0.231	0.200	0.286	0.250

스리피지배열에서는 $\delta\sqrt{n} = 0.1$, 0.5, 1.0인 경우가, 등간격배열하에서는 $\delta\sqrt{n} = 0.1$, 0.3, 0.5인 경우가 고려되었다. 이 시뮬레이션에서 사용되는 난수는 혼합된 적합난수발생방법(mixed-congruential generation method)인

$$x_i \equiv (\alpha x_{i-1} + c)\,(\mathrm{mod}\,m),\ i = 1,\ \cdots$$

을 사용하되 $m = 2^{59} - 1$을 이용해서 얻었다. 시뮬레이션 횟수는 모든 경우 각 1,000번씩 했으며 표본에서 σ^2의 합동추정량 S^2을 계산하고 Gupta, Panchapakesan 그리고 Sohn(1985)의 논문으로부터 필요한 상수 $d \equiv d(k, P^*, \nu)$를 사용하였다. 위의 모든 계산은 경북대학교 전자계산소에 있는 컴퓨터 Cyber 170/835를 이용했다. 여기서 얻어지는 값은 P(CS)의 추정치와 표준오차, $E(|I|)$의 추정치 및 표준오차 그리고 앞 절에서 정의된 절차 R의 효율 EFF(R)이며 얻어진 모든 결과는 표-3에서 표-10에 수록되어 있다.

표-3에서부터 표-6에 의한 스리피지배열상에서는 다음과 같은 사실들을 찾아 볼 수가 있다.

1. 전체적으로 우선 P^*-조건을 위배하는 경우가 $\delta\sqrt{n} = 0.1$인 경우 상당히 빈번하게 나타나는 반면 $\delta\sqrt{n}$이 0.5보다 큰 경우에는 P^*-조건이 위배되지는 않고 있다. 그러나 데이타가 오염된 경우 $\delta\sqrt{n}$이 0.5 이상이면 P(CS)는 오염되지 않았을 때보다 낮게 나타난다.
2. 오염되는 표본이 서로 다른 분포에서 오더라도 분산이 같으면 절차 R에 미치는 영향은 거의 대등하여 P(CS)나 $E(|I|)$에 큰 차이가 없다. 단 logistic분포의 경우에서 P^*-조건을 위배하는 사례가 분산이 다른 정규분포일 때보다 조금 더 빈번히 나타난다.
3. 절차 R의 효율은 모든 경우를 통해서 $P^* = 0.90$일 때가 $P^* = 0.95$일 때보다 낫다. 이런

현상은 앞 절에서나 다음에 보게 될 등간격배열에서와 동일하다.

4. k가 커지면 E(|I|)의 증가에 의해 절차 R의 효율이 거의 모든 경우에서 낮아진다. 이는 다음에서 보게 될 등간격배열에서의 현상과 동일한 것이다. 이 현상은 데이타가 오염이 되지 않았을 경우에도 절차 R의 취약점으로 나타나고 있기도 하다.

표-7부터 표-10까지로부터 등간격배열상에서 다음을 알 수가 있다.

1. 분산이 다른 정규분포로부터 오염되었을 때 분산이 커질수록 E(|I|)가 커지는 경향이 있어 절차론 R의 효율은 떨어진다. 그러나 거의 모든 경우에 P*-조건은 만족함을 알 수 있다. 이는 절차 R이 다른 정규분포에 의한 어느 정도의 오염에 비교적 강건함을 알 수 있다.
2. 만약 꼬리부분이 정규분포보다 두터운 logistic분포로부터 오염이 되었을 시 이 logistic분포의 분산이 동일하면 그 오염효과가 그다지 나쁘게 나타나지 않음을 볼 수가 있다. 그러나 분산이 커지면 1에서와 마찬가지로 절차 R의 효율이 대체로 감소되는 현상이 있으며 P*-조건은 거의 대부분 만족이 된다. 또한 E(|I|)의 최대값은 kP*를 넘지 않고 있다. 스리피지배열에서는 분산이 다른 정규분포에 의한 오염시 절차 R의 효율이 logistic분포에 의한 오염시 보다 약간 나은 경향이 있으며 등간격배열에서는 그 반대의 현상을 보이고 있다.
3. 어떤 경우든 k가 커지면 P(CS)가 낮아지고 E(|I|)가 커져 효율이 떨어짐을 보여 주고 있다. 이는 앞 절에서 보았듯이 절차 R에서 사용되는 상수 d가 너무 조심스러웁게 계산되어 있기 때문이다.
4. 전체적으로 P*-조건은 거의 위배되지 않고 있다. 따라서 절차 R은 P*-조건에 관해서는 오염현상에 대해 강건함을 보여준다. 단지 R의 효율만이 어느 정도 감소된다.
5. 오염의 증가에 따라 P(CS)가 낮아지고 E(|I|)가 커지는 현상은 분산이 큰 분포로부터 오염이 되는 경우와 $\delta\sqrt{n}$ 이 커질수록 P*의 값이 0.95보다 0.90인 경우가 더 분명하게 나타난다. 또 이 현상은 E(|I|)에서가 P(CS)보다 더 분명히 나타난다.
6. 효율면에서는 모든 경우에 P* = 0.90일 때가 P* = 0.95일 때보다 높은 현상이 나타나고 있다. 이는 앞 절에서 본 바와 같이 지나치게 P*-조건을 강하게 할 필요가 없음을 의미한다.

5. 결　　론

Gupta형 부분집합선택절차 R에서 사용되는 상수 d는 대체로 조심스럽게 고안되어 있으므로 만약 데이타의 오염현상이 없다고 생각되면 P*-조건을 지나치게 강하게 할 필요가 없음을 알 수가 있다. 스리피지배열에서나 등간격배열에서 $\delta\sqrt{n} = 0.5$ 이상되면 P(CS)의 추정치가 비록 P* = 0.90이라도 거의 대부분 0.95 이상이 되고 있으며 P* = 0.95이면 0.97 이

상이 됨을 알 수가 있다. 따라서 k에 관계없이 각 모집단으로부터 얻어지는 표본의 크기가 25 이상 되면 $P^* = 0.90$에서 P(CS)의 참값은 0.95 이상이 될뿐만 아니라 절차 R의 효율이 좋다고 할 수 있다. 이는 가설검정시 유의수준을 높게 잡을수록 검출력의 저하를 가져옴으로써 유의수준을 0.05가 아닌 0.1로 해도 좋은 결과를 얻을 수 있는 현상과 유사하다고 할 수 있다.

그러나 데이타가 오염이 되어 있을 경우 P^*-조건이 만족되지 않을 경우가 있다. 하지만 스리피지배열에서도 $\delta\sqrt{n}$이 어느 정도 크다면 P^*-조건을 위배하지도 않을 뿐만 아니라 P(CS)의 추정치도 충분히 크기 때문에 절차 R은 오염에 대해서도 비교적 강건함을 보여주고 있다. 더우기 Gupta와 Singh(1980) 또는 Sohn(1985)의 논문에서 보았듯이 중앙값선택절차론은 이러한 오염현상에 절차 R보다 상당히 강건함을 보여주고 있지만 주어진 정보를 충분히 활용하지 못하는 점과 같은 값의 P(CS)를 얻기 위해 필요한 표본의 크기가 절차 R보다 상당히 크다는 것을 감안한다면 표본의 획득에 따른 비용이 비싸게 되는 경우 절차 R의 사용이 상당히 매력적이라 할 수 있다. 단 모집단의 수가 아주 큰 경우 E(|I|)의 크기가 매우 커짐으로 2단계추출법(two-stage selection procedures)의 사용이 고려되어야 한다.

표－1 스리피지배열에서의 효율(단, 여기서 $\triangle/\sigma=\delta$이며 $\nu=k(n-1)$)

주어진 P*, $\delta\sqrt{n}$과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(｜I｜) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k	2						5						10					
P*	0.90			0.95			0.90			0.95			0.90			0.95		
$\delta\sqrt{n}$ ν	16	24	60	16	24	60	20	45	120	20	45	120	30	60	120	30	60	120
0.1	.907	.919	.918	.948	.963	.953	.902	.908	.905	.947	.960	.960	.904	.921	.905	.951	.968	.957
	(.009)	(.009)	(.009)	(.009)	(.006)	(.007)	(.009)	(.009)	(.009)	(.007)	(.006)	(.006)	(.009)	(.009)	(.009)	(.007)	(.006)	(.006)
	1.786	1.805	1.791	1.883	1.897	1.891	4.498	4.465	4.507	4.737	4.724	4.748	8.989	8.993	8.984	9.481	9.471	9.495
	(.013)	(.013)	(.013)	(.010)	(.010)	(.010)	(.027)	(.027)	(.026)	(.020)	(.019)	(.018)	(.045)	(.045)	(.043)	(.043)	(.031)	(.030)
	1.016	1.018	1.025	1.007	1.015	1.008	1.003	1.017	1.004	1.000	1.016	1.011	1.007	1.024	1.007	1.003	1.022	1.008
0.5	.963	.939	.960	.986	.973	.983	.949	.952	.960	.981	.980	.984	.959	.960	.954	.980	.983	.979
	(.006)	(.008)	(.006)	(.004)	(.005)	(.004)	(.007)	(.007)	(.006)	(.004)	(.004)	(.004)	(.006)	(.006)	(.007)	(.004)	(.004)	(.005)
	1.796	1.781	1.767	1.888	1.879	1.876	4.435	4.498	4.452	4.720	4.738	4.712	8.914	8.992	8.862	9.433	9.433	9.430
	(.013)	(.013)	(.013)	(.010)	(.010)	(.010)	(.027)	(.026)	(.028)	(.019)	(.020)	(.020)	(.048)	(.047)	(.046)	(.034)	(.035)	(.032)
	1.033	1.054	1.087	1.044	1.036	1.048	1.070	1.058	1.078	1.039	1.034	1.044	1.076	1.076	1.077	1.039	1.041	1.038
1.0	.968	.974	.968	.984	.994	.984	.978	.977	.984	.992	.993	.995	.976	.986	.985	.989	.991	.995
	(.006)	(.005)	(.006)	(.004)	(.002)	(.004)	(.005)	(.005)	(.004)	(.003)	(.003)	(.002)	(.005)	(.004)	(.004)	(.003)	(.003)	(.002)
	1.686	1.701	1.666	1.813	1.830	1.805	4.341	4.242	4.329	4.610	4.564	4.628	8.677	8.741	8.670	9.305	9.295	9.276
	(.015)	(.014)	(.015)	(.012)	(.012)	(.013)	(.030)	(.032)	(.030)	(.025)	(.027)	(.023)	(.053)	(.053)	(.052)	(.039)	(.039)	(.039)
	1.148	1.145	1.162	1.085	1.086	1.090	1.126	1.152	1.137	1.076	1.088	1.075	1.125	1.128	1.136	1.063	1.066	1.073
2.0	.998	.996	.997	.999	.998	.999	1.000	.997	.999	1.000	1.000	.999	.998	.999	.997	.999	.999	.999
	(.001)	(.002)	(.002)	(.001)	(.001)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.001)	(.001)	(.001)	(.002)	(.001)	(.001)	(.001)
	1.428	1.473	1.439	1.591	1.628	1.592	3.627	3.618	3.589	4.099	4.109	4.044	7.575	7.515	7.496	8.490	8.428	8.382
	(.016)	(.016)	(.016)	(.016)	(.015)	(.016)	(.041)	(.041)	(.042)	(.036)	(.035)	(.037)	(.077)	(.075)	(.072)	(.064)	(.064)	(.062)
	1.398	1.352	1.386	1.256	1.226	1.255	1.379	1.378	1.392	1.220	1.217	1.235	1.317	1.329	1.330	1.177	1.185	1.192

표-2 등간격배열에서의 효율(단, 여기서 $\triangle/\sigma = \delta$이며, $\nu = k(n-1)$)

주어진 P^*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k	2						5						10					
P^*	0.90			0.95			0.90			0.95			0.90			0.95		
$\delta\sqrt{n}$ \ ν	16	24	60	16	24	60	20	45	120	20	45	120	30	60	120	30	60	120
0.1	.909	.917	.912	.961	.946	.965	.927	.937	.929	.971	.968	.972	.947	.950	.951	.972	.974	.985
	(.009)	(.009)	(.009)	(.006)	(.007)	(.006)	(.008)	(.008)	(.008)	(.005)	(.006)	(.005)	(.007)	(.007)	(.007)	(.005)	(.005)	(.004)
	1.798	1.808	1.811	1.910	1.876	1.911	4.478	4.503	4.478	4.734	4.747	4.744	8.740	8.719	8.681	9.374	9.354	9.390
	(.013)	(.012)	(.012)	(.009)	(.010)	(.009)	(.027)	(.025)	(.026)	(.019)	(.019)	(.019)	(.050)	(.050)	(.047)	(.036)	(.037)	(.032)
	1.011	1.014	1.007	1.006	1.009	1.010	1.036	1.040	1.037	1.026	1.020	1.024	1.084	1.090	1.095	1.037	1.041	1.049
0.5	.941	.940	.954	.973	.973	.977	.967	.978	.985	.988	.990	.994	.992	.992	.991	.996	.991	1.000
	(.007)	(.008)	(.007)	(.005)	(.005)	(.005)	(.006)	(.005)	(.004)	(.003)	(.003)	(.002)	(.003)	(.003)	(.003)	(.002)	(.003)	(.000)
	1.778	1.749	1.769	1.870	1.875	1.869	4.023	3.985	3.993	4.415	4.354	4.359	5.684	5.632	5.489	6.549	6.379	6.497
	(.013)	(.014)	(.013)	(.011)	(.010)	(.011)	(.033)	(.032)	(.033)	(.028)	(.028)	(.027)	(.055)	(.054)	(.058)	(.059)	(.056)	(.052)
	1.058	1.075	1.079	1.041	1.038	1.045	1.202	1.227	1.233	1.119	1.137	1.140	1.745	1.761	1.805	1.521	1.554	1.539
1.0	.977	.976	.974	.990	.993	.993	.990	.994	.993	.999	.998	.998	.996	.999	.993	.999	1.000	1.000
	(.005)	(.005)	(.005)	(.003)	(.003)	(.003)	(.003)	(.002)	(.003)	(.001)	(.001)	(.001)	(.002)	(.001)	(.003)	(.001)	(.000)	(.000)
	1.704	1.694	1.700	1.823	1.830	1.789	3.015	2.697	2.933	3.458	3.388	3.324	3.392	3.327	3.398	3.896	3.712	3.735
	(.014)	(.015)	(.014)	(.012)	(.012)	(.013)	(.034)	(.033)	(.033)	(.034)	(.033)	(.033)	(.036)	(.036)	(.035)	(.037)	(.037)	(.037)
	1.147	1.152	1.146	1.086	1.085	1.110	1.642	1.675	1.693	1.444	1.473	1.501	2.936	3.003	2.922	2.564	2.694	2.677
2.0	.998	.996	.998	.997	.996	1.000	.999	1.000	1.000	1.000	1.000	1.000	.999	1.000	1.000	1.000	1.000	1.000
	(.001)	(.002)	(.001)	(.002)	(.002)	(.000)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)
	1.460	1.456	1.440	1.627	1.595	1.585	1.869	1.859	1.858	2.125	2.082	2.056	2.040	2.012	1.996	2.250	2.210	2.201
	(.016)	(.016)	(.016)	(.015)	(.016)	(.016)	(.021)	(.021)	(.021)	(.023)	(.022)	(.022)	(.023)	(.021)	(.022)	(.022)	(.022)	(.022)
	1.367	1.368	1.386	1.226	1.249	1.262	2.673	2.690	2.691	2.353	2.402	2.432	4.897	4.370	5.010	4.444	4.525	4.543

표－3 스리피지배열하에서 분산이 $4\delta^2$인 정규분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma = \delta$이며 $\nu = k(n-1)$이다.)

주어진 P^*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

	k	3						5					
	ν	15		30		60		30		60		120	
P^*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.895	.915	.914	.909	.919	.924	.910	.914	.905	.903	.902	.907
		(.010)	(.009)	(.009)	(.009)	(.009)	(.008)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)
		2.673	2.691	2.699	2.718	2.707	2.690	4.517)	4.519	4.461	4.471	4.487	4.516
		(.018)	(.018)	(.017)	(.016)	(.017)	(.017)	(.026)	(.025)	(.026)	(.026)	(.026)	(.024)
		1.004	1.020	1.016	1.003	1.018	1.030	1.007	1.011	1.014	1.010	1.005	1.004
	0.5	.943	.951	.943	.943	.935	.968	.951	.935	.940	.947	.956	.944
		(.007)	(.007)	(.007)	(.007)	(.008)	(.006)	(.007)	(.008)	(.008)	(.007)	(.006)	(.007)
		2.678	2.689	2.669	2.668	2.653	2.699	4.477	4.422	4.463	4.395	4.456	4.476
		(.018)	(.017)	(.018)	(.019)	(.018)	(.017)	(.026)	(.028)	(.027)	(.027)	(.025)	(.027)
		1.056	1.061	1.060	1.060	1.057	1.076	1.062	1.057	1.053	1.077	1.073	1.055
	1.0	.976	.969	.971	.979	.979	.977	.979	.963	.976	.962	.982	.967
		(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.006)	(.005)	(.006)	(.004)	(.006)
		2.585	2.613	2.615	2.622	2.577	2.601	4.364	4.401	4.316	4.329	4.383	4.313
		(.020)	(.020)	(.019)	(.019)	(.020)	(.020)	(.030)	(.028)	(.030)	(.030)	(.028)	(.031)
		1.133	1.113	1.114	1.120	1.140	1.127	1.121	1.094	1.131	1.111	1.120	1.121
0.95	0.1	.963	.956	.953	.960	.959	.955	.956	.965	.967	.958	.954	.950
		(.006)	(.006)	(.007)	(.006)	(.006)	(.007)	(.006)	(.006)	(.006)	(.006)	(.007)	(.007)
		2.880	2.847	2.846	2.841	2.844	2.828	4.767	4.786	4.779	4.775	4.764	4.756
		(.011)	(.013)	(.013)	(.013)	(.013)	(.013)	(.018)	(.016)	(.017)	(.017)	(.017)	(.018)
		1.003	1.007	1.005	1.014	1.012	1.013	1.003	1.008	1.012	1.003	1.001	.999
	0.5	.979	.972	.976	.977	.987	0.976	.984	.974	.962	.974	.978	.973
		(.005)	(.005)	(.005)	(.005)	(.004)	(.005)	(.004)	(.005)	(.006)	(.005)	(.005)	(.005)
		2.840	2.822	2.831	2.852	2.830	2.838	4.752	4.727	4.706	4.694	4.728	4.741
		(.013)	(.014)	(.014)	(.012)	(.014)	(.013)	(.019)	(.020)	(.020)	(.020)	(.020)	(.019)
		1.034	1.033	1.034	1.028	1.046	1.032	1.035	1.030	1.022	1.037	1.036	1.026
	1.0	.991	.990	.985	.992	.988	.990	.991	.982	.994	.990	.990	.993
		(.003)	(.003)	(.004)	(.003)	(.003)	(.003)	(.003)	(.004)	(.002)	(.003)	(.003)	(.003)
		2.766	2.795	2.784	2.797	2.763	2.790	4.648	4.635	4.679	4.678	4.651	4.648
		(.016)	(.015)	(.015)	(.015)	(.016)	(.015)	(.021)	(.023)	(.021)	(.020)	(.022)	(.021)
		1.075	1.063	1.061	1.064	1.073	1.065	1.066	1.059	1.062	1.058	1.064	1.068

표－4 스리피지배열하에서 분산이 $9\delta^2$인 정규분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma=\delta$이며 $\nu=k(n-1)$)

주어진 P*, $\delta\sqrt{n}$과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.915	.898	.902	.901	.905	.908	.913	.914	.906	.908	.910	.895
		(.009)	(.010)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.010)
		2.731	2.688	2.694	2.684	2.690	2.700	4.535	4.518	4.511	4.490	4.522	4.454
		(.017)	(.017)	(.017)	(.018)	(.018)	(.017)	(.024)	(.024)	(.026)	(.025)	(.024)	(.026)
		1.005	1.002	1.004	1.007	1.009	1.009	1.007	1.012	1.004	1.011	1.006	1.005
	0.5	.945	.937	.936	.942	.936	.943	.954	.947	.943	.950	.950	.936
		(.007)	(.008)	(.008)	(.007)	(.008)	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)	(.008)
		2.696	2.698	2.678	2.715	2.659	2.704	4.507	4.472	4.428	4.459	4.496	4.481
		(.017)	(.018)	(.018)	(.017)	(.019)	(.018)	(.025)	(.025)	(.028)	(.026)	(.025)	(.026)
		1.052	1.042	1.049	1.041	1.056	1.046	1.058	1.059	1.065	1.065	1.056	1.044
	1.0	.974	.957	.970	.955	.966	.955	.979	.952	.969	.960	.972	.962
		(.005)	(.006)	(.005)	(.007)	(.006)	(.007)	(.005)	(.007)	(.005)	(.006)	(.005)	(.006)
		2.652	2.670	2.625	2.642	2.634	2.624	4.420	4.447	4.394	4.380	4.389	4.474
		(.018)	(.018)	(.019)	(.019)	(.019)	(.019)	(.027)	(.027)	(.028)	(.027)	(.028)	(.025)
		1.102	1.075	1.109	1.084	1.100	1.092	1.107	1.070	1.103	1.096	1.107	1.075
0.95	0.1	.963	.971	.961	.962	.946	.955	.965	.960	.964	.943	.950	.969
		(.006)	(.005)	(.006)	(.006)	(.007)	(.007)	(.006)	(.006)	(.006)	(.007)	(.007)	(.005)
		2.874	2.883	2.860	2.862	2.843	2.843	4.804	4.789	4.781	4.748	4.725	4.779
		(.012)	(.011)	(.012)	(.012)	(.013)	(.013)	(.016)	(.016)	(.017)	(.018)	(.019)	(.016)
		1.005	1.010	1.008	1.008	.998	1.008	1.004	1.002	1.008	.993	1.005	1.014
	0.5	.977	.977	.971	.968	.971	.976	.976	.979	.977	.970	.965	.978
		(.005)	(.005)	(.005)	(.006)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.006)	(.005)
		2.881	2.858	2.853	2.839	2.840	2.863	4.773	4.799	4.744	4.750	4.707	4.749
		(.012)	(.012)	(.012)	(.013)	(.013)	(.012)	(.018)	(.016)	(.018)	(.018)	(.020)	(.017)
		1.017	1.026	1.021	1.023	1.026	1.023	1.022	1.020	1.030	1.021	1.025	1.030
	1.0	.986	.987	.992	.988	.982	.984	.993	.987	.993	.979	.988	.987
		(.004)	(.004)	(.003)	(.003)	(.004)	(.004)	(.003)	(.004)	(.003)	(.005)	(.004)	(.004)
		2.812	2.826	2.823	2.832	2.791	2.798	4.716	4.722	4.715	4.708	4.667	4.709
		(.014)	(.014)	(.014)	(.013)	(.015)	(.015)	(.019)	(.019)	(.020)	(.020)	(.021)	(.019)
		1.052	1.048	1.054	1.047	1.056	1.056	1.053	1.045	1.053	1.040	1.056	1.048

표-5 스리피지배열하에서 분산이 σ^2인 logistic분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma = \delta$이며 $\nu = k(n-1)$이다.)

주어진 P^*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P^*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.231	0.154	0.231	0.120	0.200
0.90	0.1	.922	.921	.911	.909	.910	.922	.892	0.908	.921	.907	.913	.915
		(.008)	(.009)	(.009)	(.009)	(.009)	(.008)	(.010)	(.009)	(.009)	(.009)	(.009)	(.009)
		2.720	2.690	2.688	2.699	2.698	2.726	4.500	4.496	4.490	4.509	4.493	4.523
		(.017)	(.018)	(.018)	(.017)	(.017)	(.017)	(.026)	(.027)	(.025)	(.026)	(.026)	(.025)
		1.017	1.027	1.017	1.010	1.012	1.015	.991	1.010	1.026	1.006	1.016	1.011
	0.5	.938	.961	.940	.962	.944	.950	.957	.951	.952	.961	.945	.960
		(.008)	(.006)	(.008)	(.006)	(.007)	(.007)	(.006)	(.007)	(.007)	(.006)	(.007)	(.006)
		2.645	2.648	2.680	2.670	2.656	2.651	4.480	4.458	4.506	4.460	4.430	4.439
		(.019)	(.019)	(.018)	(.018)	(.018)	(.018)	(.027)	(.027)	(.025)	(.026)	(.027)	(.027)
		1.064	1.089	1.052	1.081	1.066	1.075	1.068	1.067	1.056	1.077	1.067	1.081
	1.0	.983	.983	.975	.976	.981	.981	.979	.978	.986	.981	.982	.977
		(.004)	(.004)	(.005)	(.005)	(.004)	(.004)	(.005)	(.005)	(.004)	(.004)	(.004)	(.005)
		2.612	2.586	2.565	2.552	2.605	2.567	4.270	4.317	4.354	4.329	4.251	4.302
		(.019)	(.020)	(.021)	(.020)	(.020)	(.020)	(.032)	(.030)	(.030)	(.030)	(.031)	(.031)
		1.129	1.140	1.140	1.147	1.130	1.146	1.146	1.133	1.132	1.133	1.155	1.136
0.95	0.1	.959	.954	.960	.955	.966	.951	.968	.952	.963	.946	.957	.966
		(.006)	(.007)	(.006)	(.007)	(.006)	(.007)	(.006)	(.007)	(.006)	(.007)	(.006)	(.006)
		2.855	2.846	2.853	2.837	2.868	2.856	4.733	4.748	4.758	4.743	4.770	4.739
		(.012)	(.013)	(.013)	(.013)	(.012)	(.013)	(.020)	(.019)	(.017)	(.018)	(.018)	(.019)
		1.008	1.006	1.009	1.010	1.010	.999	1.023	1.003	1.012	.997	1.003	1.019
	0.5	.984	.977	.981	.979	.988	.984	.976	.985	.982	.966	.978	.976
		(.004)	(.005)	(.004)	(.005)	(.003)	(.004)	(.005)	(.004)	(.004)	(.006)	(.005)	(.005)
		2.822	2.814	2.831	2.844	2.821	2.844	4.684	4.712	4.716	4.705	4.714	4.682
		(.014)	(.014)	(.013)	(.014)	(.014)	(.013)	(.021)	(.020)	(.019)	(.021)	(.021)	(.020)
		1.046	1.042	1.040	1.033	1.051	1.038	1.042	1.045	1.041	1.027	1.037	1.042
	1.0	.988	.994	.993	.993	.987	.993	.991	.990	.989	.987	.993	.991
		(.003)	(.002)	(.003)	(.003)	(.004)	(.003)	(.003)	(.003)	(.003)	(.004)	(.003)	(.003)
		2.729	2.759	2.782	2.711	2.762	2.746	4.567	4.626	4.629	4.599	4.620	4.575
		(.017)	(.016)	(.016)	(.017)	(.016)	(.016)	(.025)	(.023)	(.023)	(.024)	(.022)	(.025)
		1.086	1.081	1.071	1.099	1.072	1.085	1.085	1.070	1.068	1.073	1.075	1.083

표－6 스리피지배열하에서 분산이 $4\delta^2$인 logistic분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma = \delta$이며 $\nu = k(n-1)$)

주어진 P^*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P^*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.231	0.154	0.231	0.120	0.200
0.90	0.1	.910	.907	.914	.899	.901	.886	.898	.927	.918	.915	.919	.913
		(.009)	(.009)	(.009)	(.010)	(.009)	(.010)	(.010)	(.008)	(.009)	(.009)	(.009)	(.009)
		2.705	2.701	2.672	2.662	2.700	2.662	4.485	4.502	4.497	4.477	4.553	4.531
		(.017)	(.017)	(.018)	(.018)	(.017)	(.018)	(.026)	(.024)	(.026)	(.025)	(.025)	(.024)
		1.009	1.007	1.026	1.013	1.001	.998	1.001	1.030	1.021	1.022	1.014	1.008
	0.5	.952	.961	.944	.941	.940	.935	.952	.944	.947	.943	.945	.955
		(.007)	(.006)	(.007)	(.007)	(.008)	(.008)	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)
		2.701	2.666	2.682	2.676	2.660	2.665	4.489	4.459	4.480	4.435	4.453	4.457
		(.018)	(.018)	(.018)	(.018)	(.018)	(.019)	(.026)	(.027)	(.026)	(.027)	(.027)	(.026)
		1.057	1.081	1.056	1.055	1.060	1.053	1.060	1.059	1.057	1.063	1.061	1.071
	1.0	.974	.977	.968	.964	.971	.977	.980	.978	.954	.966	.977	.973
		(.005)	(.005)	(.006)	(.006)	(.005)	(.005)	(.004)	(.005)	(.005)	(.006)	(.005)	(.005)
		2.572	2.689	2.620	2.619	2.572	2.631	4.387	4.367	4.362	4.369	4.384	4.356
		(.020)	(.018)	(.019)	(.019)	(.020)	(.019)	(.029)	(.030)	(.029)	(.028)	(.027)	(.029)
		1.136	1.090	1.108	1.104	1.133	1.114	1.117	1.120	1.116	1.106	1.114	1.117
0.95	0.1	.971	.951	.955	.953	.957	.963	.953	.940	.956	.956	.967	.954
		(.005)	(.007)	(.007)	(.007)	(.006)	(.006)	(.007)	(.008)	(.006)	(.006)	(.006)	(.007)
		2.856	2.840	2.835	2.852	2.857	2.872	4.728	4.747	4.734	4.778	4.784	4.729
		(.013)	(.013)	(.013)	(.012)	(.012)	(.012)	(.019)	(.019)	(.019)	(.016)	(.017)	(.019)
		1.020	1.005	1.011	1.006	1.005	1.006	1.008	.990	1.010	1.000	1.011	1.009
	0.5	.975	.979	.978	.972	.977	.969	.983	.977	.978	.977	.977	.965
		(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.004)	(.005)	(.005)	(.005)	(.005)	(.006)
		2.835	2.868	2.848	2.803	2.849	2.810	4.737	4.718	4.728	4.758	4.728	4.731
		(.013)	(.012)	(.013)	(.013)	(.013)	(.014)	(.018)	(.019)	(.019)	(.018)	(.017)	(.018)
		1.032	1.024	1.030	1.040	1.029	1.035	1.038	1.035	1.034	1.027	1.033	1.020
	1.0	.992	.982	.990	.990	.987	.993	.991	.990	.992	.984	.989	.979
		(.003)	(.004)	(.003)	(.003)	(.004)	(.003)	(.003)	(.003)	(.003)	(.004)	(.003)	(.005)
		2.810	2.793	2.746	2.805	2.768	2.776	4.654	4.635	4.656	4.680	4.659	4.637
		(.014)	(.015)	(.016)	(.015)	(.016)	(.015)	(.022)	(.022)	(.022)	(.020)	(.020)	(.023)
		1.059	1.055	1.082	1.059	1.070	1.073	1.065	1.068	1.065	1.051	1.061	1.056

표－7 등간격배열하에서 분산이 $4\delta^2$인 정규분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma=\delta$이며 $\nu=k(n-1)$이다.)

주어진 P*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.914	.926	.912	.907	.905	.910	.921	.924	.924	.926	.928	.918
		(.009)	(.008)	(.009)	(.009)	(.009)	(.009)	(.009)	(.008)	(.008)	(.008)	(.008)	(.009)
		2.693	2.700	2.709	2.684)	2.690	2.700	4.488	4.491	4.473	4.498	4.472	4.482
		(.018)	(.017)	(.017)	(.018)	(.018)	(.018)	(.026)	(.026)	(.026)	(.025)	(.026)	(.026)
		1.018	1.029	1.010	1.014	1.009	1.011	1.026	1.029	1.033	1.029	1.038	1.024
	0.3	.943	.924	.926	.936	.924	.944	.956	.961	.956	.954	.946	.961
		(.007)	(.008)	(.008)	(.008)	(.008)	(.007)	(.006)	(.006)	(.006)	(.007)	(.007)	(.006)
		2.679	2.686	2.664	2.663	2.631	2.679	4.361	4.459	4.369	4.376	4.327	4.408
		(.018)	(.018)	(.018)	(.018)	(.020)	(.018)	(.028)	(.027)	(.028)	(.027)	(.028)	(.027)
		1.056	1.032	1.043	1.054	1.054	1.057	1.096	1.078	1.094	1.090	1.093	1.090
	0.5	.960	.958	.950	.954	.961	.955	.969	.966	.971	.973	.976	.972
		(.006)	(.006)	(.007)	(.007)	(.006)	(.007)	(.005)	(.006)	(.005)	(.005)	(.005)	(.005)
		2.634	2.677	2.619	2.621	2.609	2.662	4.146	4.181	4.140	4.157	4.057	4.164
		(.019)	(.018)	(.020)	(.019)	(.019)	(.018)	(.032)	(.032)	(.031)	(.031)	(.033)	(.031)
		1.093	1.074	1.088	1.092	1.105	1.076	1.169	1.155	1.173	1.170	1.203	1.167
0.95	0.1	.961	.961	.964	.965	.965	.957	.955	.962	.965	.961	.961	.953
		(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.007)	(.006)	(.006)	(.006)	(.006)	(.007)
		2.856	2.859	2.862	2.874	2.855	2.854	4.739	4.746	4.733	4.741	4.698	4.733
		(.013)	(.012)	(.012)	(.012)	(.012)	(.012)	(.019)	(.019)	(.018)	(.018)	(.020)	(.019)
		1.009	1.008	1.010	1.007	1.014	1.006	1.008	1.013	1.019	1.013	1.023	1.007
	0.3	.977	.967	.976	.975	.979	.971	.985	.979	.985	.977	.977	.981
		(.005)	(.006)	(.005)	(.005)	(.005)	(.005)	(.004)	(.005)	(.004)	(.005)	(.005)	(.004)
		2.837	2.827	2.813	2.843	2.809	2.844	4.641	4.673	4.648	4.679	4.647	4.681
		(.013)	(.014)	(.014)	(0.13)	(.014)	(.013)	(.023)	(.020)	(.021)	(.020)	(.022)	(.020)
		1.033	1.026	1.041	1.029	1.046	1.024	1.061	1.048	1.060	1.044	1.051	1.048
	0.5	.986	.972	.982	.983	.987	.979	.988	.989	.990	.991	.979	.986
		(.004)	(.005)	(.004)	(.004)	(.004)	(.005)	(.003)	(.003)	(.003)	(.003)	(.005)	(.004)
		2.831	2.806	2.794	2.814	2.805	2.801	4.452	4.523	4.473	4.507	4.429	4.474
		(.014)	(.014)	(.015)	(.014)	(.014)	(.015)	(.027)	(.024)	(.026)	(.025)	(.026)	(.025)
		1.042	1.038	1.054	1.048	1.056	1.049	1.110	1.093	1.107	1.099	1.105	1.102

표－8 등간격배열하에서 분산이 $9\delta^2$인 정규분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\Delta/\sigma=\delta$이며 $\nu=k(n-1)$이다)

주어진 P*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.920	.890	.908	.912	.906	.911	.938	.917	.937	.919	.924	.924
		(.009)	(.010)	(.009)	(.009)	(.009)	(.009)	(.008)	(.009)	(.008)	(.009)	(.008)	(.008)
		2.731	2.687	2.703	2.712	2.700	2.700	4.458	4.481	4.525	4.469	4.511	4.488
		(.016)	(.018)	(.017)	(.017)	(.018)	(.017)	(.026)	(.025)	(.024)	(.026)	(.025)	(.026)
		1.011	0.994	1.008	1.009	1.007	1.012	1.052	1.023	1.035	1.028	1.024	1.029
	0.3	.925	.930	.942	.927	.936	.930	.956	.955	.955	.952	.957	.941
		(.008)	(.008)	(.007)	(.008)	(.008)	(.008)	(.006)	(.007)	(.007)	(.007)	(.006)	(.007)
		2.682	2.719	2.688	2.682	2.690	2.700	4.431	4.477	4.396	4.463	4.461	4.394
		(.018)	(.017)	(.018)	(.017)	(.017)	(.017)	(.027)	(.025)	(.027)	(.026)	(.026)	(.027)
		1.035	1.026	1.051	1.037	1.044	1.033	1.079	1.067	1.086	1.067	1.073	1.071
	0.5	.955	.944	.939	.949	.968	.944	.976	.965	.969	.968	.966	.956
		(.007)	(.007)	(.008)	(.007)	(.006)	(.007)	(.005)	(.006)	(.005)	(.006)	(.006)	(.006)
		2.665	2.679	2.609	2.658	2.689	2.650	4.237	4.301	4.238	4.310	4.214	4.295
		(.018)	(.018)	(.019)	(.018)	(.017)	(.018)	(.031)	(.030)	(.028)	(.030)	(.031)	(.028)
		1.075	1.057	1.080	1.071	1.080	1.069	1.152	1.122	1.143	1.123	1.146	1.113
0.95	0.1	.973	.957	.956	.964	.958	.965	.960	.964	.961	.962	.964	.970
		(.005)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.005)
		2.900	2.857	2.841	2.863	2.861	2.865	4.783	4.740	4.762	4.756	4.775	4.744
		(.010)	(.012)	(.013)	(.012)	(.012)	(.012)	(.016)	(.018)	(.018)	(.018)	(.018)	(.018)
		1.007	1.005	1.010	1.010	1.005	1.010	1.004	1.017	1.009	1.011	1.009	1.022
	0.3	.979	.977	.971	.976	.981	.973	.990	.979	.978	.979	.985	.975
		(.005)	(.005)	(.005)	(.005)	(.004)	(.005)	(.003)	(.005)	(.005)	(.005)	(.004)	(.005)
		2.854	2.880	2.851	2.871	2.828	2.872	4.725	4.722	4.680	4.700	4.657	4.670
		(.013)	(.011)	(.013)	(.012)	(.014)	(.012)	(.019)	(.019)	(.021)	(.020)	(.021)	(.021)
		1.029	1.018	1.022	1.020	1.041	1.016	1.048	1.037	1.045	1.041	1.058	1.044
	0.5	.987	.981	.981	.966	.976	.970	.989	.984	.985	.991	.992	.984
		(.004)	(.004)	(.004)	(.006)	(.005)	(.005)	(.003)	(.004)	(.004)	(.003)	(.003)	(.004)
		2.843	2.847	2.838	2.829	2.822	2.816	4.583	4.688	4.569	4.630	4.583	4.625
		(.013)	(.013)	(.013)	(.013)	(.014)	(.014)	(.024)	(.021)	(.023)	(.021)	(.023)	(.021)
		1.042	1.034	1.037	1.024	1.038	1.033	1.079	1.049	1.078	1.070	1.082	1.064

표－9 등간격배열하에서 분산이 δ^2인 logistic분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R에 대한 효율. (단, 여기서 $\triangle/\sigma=\delta$이며 $\nu=k(n-1)$이다)

주어진 P*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.918	.912	.908	.912	.914	.916	.928	.932	.917	.934	.924	.921
		(.009)	(.009)	(.009)	(.009)	(.009)	(.009)	(.008)	(.008)	(.009)	(.008)	(.008)	(.009)
		2.702	2.703	2.711	2.693	2.678	2.694	4.458	4.509	4.412	4.491	4.488	4.455
		(.018)	(.017)	(.017)	(.018)	(.018)	(.017)	(.028)	(.025)	(.027)	(.025)	(.025)	(.027)
		1.019	1.012	1.005	1.016	1.024	1.020	1.041	1.033	1.039	1.040	1.029	1.034
	0.3	.946	.944	.944	.954	.938	.959	.951	.962	.963	.958	.957	.954
		(.007)	(.007)	(.007)	(.007)	(.008)	(.006)	(.007)	(.006)	(.006)	(.006)	(.006)	(.007)
		2.660	2.653	2.627	2.674	2.669	2.692	4.296	4.276	4.268	4.299	4.288	4.319
		(.019)	(.019)	(.019)	(.018)	(.018)	(.017)	(.030)	(.030)	(.030)	(.029)	(.029)	(.029)
		1.067	1.067	1.078	1.070	1.054	1.069	1.107	1.125	1.128	1.114	1.116	1.104
	0.5	.955	.963	.952	.968	.963	.965	.982	.980	.981	.985	.973	.964
		(.007)	(.006)	(.007)	(.006)	(.006)	(.006)	(.004)	(.004)	(.004)	(.004)	(.005)	(.006)
		2.603	2.601	2.588	2.584	2.607	2.587	3.982	3.927	3.960	4.079	3.999	3.951
		(.020)	(.020)	(.019)	(.019)	(.019)	(.020)	(.033)	(.035)	(.034)	(.031)	(.033)	(.035)
		1.101	1.111	1.104	1.124	1.108	1.119	1.233	1.248	1.239	1.207	1.217	1.220
0.95	0.1	.961	.968	.952	.952	.956	.957	.971	.969	.975	.965	.954	.969
		(.006)	(.006)	(.007)	(.007)	(.006)	(.006)	(.005)	(.005)	(.005)	(.006)	(.007)	(.005)
		2.862	2.863	2.847	2.825	2.847	2.831	4.764	4.748	4.741	4.719	4.716	4.740
		(.012)	(.012)	(.013)	(.014)	(.013)	(.013)	(.018)	(.018)	(.018)	(.020)	(.019)	(.018)
		1.007	1.014	1.003	1.011	1.007	1.014	1.019	1.020	1.028	1.022	1.011	1.022
	0.3	.972	.974	.976	.975	.979	.974	.988	.980	.996	.985	.985	.982
		(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.003)	(.004)	(.002)	(.004)	(.004)	(.004)
		2.843	2.792	2.829	2.819	2.839	2.810	4.612	4.610	4.626	4.596	4.574	4.612
		(.013)	(.015)	(.013)	(.014)	(.013)	(.014)	(.023)	(.023)	(.022)	(.023)	(.024)	(.022)
		1.026	1.047	1.035	1.038	1.035	1.040	1.071	1.063	1.077	1.072	1.077	1.065
	0.5	.980	.980	.981	.987	.979	.978	.994	.993	.989	.993	.995	.991
		(.004)	(.004)	(.004)	(.004)	(.005)	(.005)	(.002)	(.003)	(.003)	(.003)	(.002)	(.003)
		2.774	2.771	2.775	2.788	2.761	2.816	4.375	4.401	4.399	4.350	4.370	4.380
		(.016)	(.015)	(.016)	(.015)	(.016)	(.014)	(.028)	(.027)	(.028)	(.028)	(.027)	(.026)
		1.060	1.061	1.061	1.062	1.064	1.033	1.136	1.128	1.124	1.141	1.138	1.131

표－10 등간격배열하에서 분산이 $4\sigma^2$인 logistic분포로부터 표본이 100ϵ%오염이 된 경우의 절차 R 에 대한 효율. (단, 여기서 $\triangle/\sigma = \delta$이며 $\nu = k(n-1)$이다)

주어진 P*, $\delta\sqrt{n}$ 과 ν에 대해 첫째 줄은 P(CS) 및 그 오차, 둘째 줄은 E(| I |) 및 그 오차, 세째 줄은 EFF(R)을 나타낸다.

k		3						5					
	ν	15		30		60		30		60		120	
P*	$\delta\sqrt{n}$ ϵ	0.167	0.333	0.182	0.273	0.143	0.238	0.143	0.286	0.154	0.231	0.120	0.200
0.90	0.1	.916	.905	.906	.915	.926	.912	.922	.924	.934	.923	.934	.926
		(.009)	(.009)	(.009)	(.009)	(.008)	(.009)	(.008)	(.008)	(.008)	(.008)	(.008)	(.008)
		2.694	2.695	2.691	2.682	2.713	2.700	4.476	4.512	4.510	4.500	4.522	4.495
		(.018)	(.018)	(.017)	(.018)	(.017)	(.017)	(.027)	(.025)	(.025)	(.026)	(.025)	(.026)
		1.020	1.007	1.010	1.023	1.024	1.013	1.030	1.024	1.035	1.026	1.033	1.030
	0.3	.932	.930	.930	.957	.925	.935	.958	.954	.962	.960	.958	.967
		(.008)	(.008)	(.008)	(.006)	(.008)	(.008)	(.006)	(.007)	(.006)	(.006)	(.006)	(.006)
		2.656	2.684	2.660	2.708	2.661	2.692	4.360	4.375	4.398	4.418	4.359	4.363
		(.018)	(.018)	(.018)	(.017)	(.018)	(.018)	(.028)	(.029)	(.027)	(.027)	(.028)	(.028)
		1.053	1.039	1.049	1.060	1.043	1.042	1.099	1.090	1.094	1.086	1.099	1.108
	0.5	.956	.961	.970	.951	.951	.949	.977	.965	.976	.977	.980	.974
		(.006)	(.006)	(.005)	(.007)	(.007)	(.007)	(.005)	(.006)	(.005)	(.005)	(.004)	(.005)
		2.635	2.649	2.642	2.630	2.602	2.612	4.141	4.166	4.150	4.172	4.059	4.157
		(.019)	(.019)	(.018)	(.019)	(.020)	(.020)	(.032)	(.032)	(.031)	(.031)	(.034)	(.030)
		1.088	1.088	1.101	1.085	1.096	1.090	1.180	1.158	1.176	1.171	1.207	1.172
0.95	0.1	.965	.967	.965	.961	.964	.964	.968	.968	.961	.953	.972	.963
		(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.007)	(.005)	(.006)
		2.854	2.861	2.862	2.862	2.841	2.853	4.771	4.772	4.735	4.745	4.750	4.734
		(.013)	(.014)	(.012)	(.012)	(.013)	(.013)	(.018)	(.017)	(.019)	(.019)	(.018)	(.019)
		1.014	1.026	1.012	1.007	1.018	1.014	1.014	1.014	1.015	1.004	1.022	1.017
	0.3	.976	.976	.975	.981	.977	.976	.975	.979	.981	.982	.979	.975
		(.005)	(.005)	(.005)	(.004)	(.005)	(.005)	(.005)	(.005)	(.004)	(.004)	(.005)	(.005)
		2.864	2.845	2.822	2.829	2.825	2.833	4.652	4.682	4.658	4.676	4.615	4.682
		(.012)	(.013)	(.014)	(.013)	(.013)	(.013)	(.022)	(.020)	(.022)	(.020)	(.023)	(.020)
		1.022	1.029	1.036	1.040	1.038	1.034	1.048	1.045	1.053	1.050	1.061	1.041
	0.5	.990	.979	.986	.977	.988	.980	.993	.990	.992	.988	.997	.990
		(.003)	(.005)	(.004)	(.005)	(.003)	(.004)	(.003)	(.003)	(.003)	(.003)	(.002)	(.003)
		2.823	2.835	2.797	2.809	2.804	2.797	4.443	4.582	4.480	4.485	4.452	4.521
		(.014)	(.013)	(.014)	(.014)	(.014)	(.014)	(.027)	(.022)	(.026)	(.025)	(.026)	(.025)
		1.052	1.036	1.058	1.043	1.057	1.051	1.117	1.080	1.107	1.101	1.120	1.095

참 고 문 헌

(1) Gupta, S. S., "On a decision rule for a problem in ranking means", Ph. D. Thesis(Mimeo. Series No. 150), Inst. of Statis., Univ. of North Carolina, Chapel Hill, North Carolina, U. S. A. (1956).

(2) Gupta, S. S. and Hsu, J. C., "On the performance of some subset selection procedures", Communications in Statistics-Simulation and Computation B7 ; 561 – 591(1978)

(3) Bjornstad, J. F., "Comparison of three minimax subset selection precedures", Technometrics 22 ; 617 – 620(1980)

(4) Gupta, S. S. and Singh, A. K., "On rules based on sample medians for selection of the largest location parameter", Communications in Statistics-Theory and Methods A9 ; 1277 – 1298(1980)

(5) Bickel, P. J. and Yahav, J. A., "Asymptotic theory of selection procedures and optimality of Gupta's rules", Statistics and Probability : Essays in Honor of C. R. Rao(G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) ; 109 – 124(1982)

(6) Sohn, J., "Multiple decision procedures for Tukey's generalized lambda distributions" Technical Report No. 85 – 20, Department of Statistics, Purdue University, West Lafayette, Indiana 47907, U. S. A. (1985)

(7) Gupta, S. S., Panchapakesan, S. and Sohn, J., "On the distribution of the Studentized Maximum of Equally Correlated Normal Random Variables", Communications in Statistics-Simulation and Computation 14 ; 103 – 135(1985)

Operating Characteristics of a Subset Selection Procedure for Selecting the Best Normal Population with Common Unknown Variance

Joong K. Sohn* Shanti S. Gupta**

Abstract

The subset selection approach introduced by Gupta plays an important role in the multiple decision procedures. For the normal means problem with common unknown variance, some operating characteristics of the selection procedure have been investigated via Monte Carlo simulation. Also some properties including efficiencies of the selection procedure are examined when the data are contaminated.

* Dept. of Statistics, Kyungpook National University, Sankyuk-dong, Buk-gu, Taegu.

** Dept. of Statistics, Purdue University, West Lafayette, Indiana 47907, U. S. A