

균형일원변량모형에서 분산성분비율의 새로운 추정량

이 장 태¹⁾

요 약

균형일원변량모형에서 분산성분비율의 점추정에 관한 문제가 고려되어진다. 분산성분비율에 대한 점추정량의 종류를 살펴보고 추정량의 평균자승오차(MSE)를 서로 비교하여 본다. 분산성분비율에 대한 새로운 추정량이 제안되며, 제안된 추정량을 사용하면 모의실험을 통하여 Das (1992)가 고려한 여러 가지 형태의 추정량들보다 급내상관계수 ρ 의 값이 대략 $0.2 < \rho < 0.7$ 인 경우에 MSE 효율성이 높아짐을 밝혔다.

1. 서론

일반적으로 반복수가 같은 균형일원변량모형은 다음 식(1.1)과 같이 나타낼 수 있다.

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J. \quad (1.1)$$

여기서 μ 는 미지의 모수, a_i 와 e_{ij} 는 서로 독립이며, 평균이 0이고 분산이 각각 σ_a^2 와 σ_e^2 인 정규 확률변수를 의미한다. 균형일원변량모형이 분석되어질 때 분산성분비율 $\theta = \sigma_a^2/\sigma_e^2$ 는 종종 관심의 대상이 되는 모수가 된다. 잘 알려져 있다시피 분산성분에 관한 신뢰한계를 구하려고 할 때, 신뢰한계에서 가장 빈번하게 나타나는 것이 모수 θ 이다. 또한 균형일원변량모형에서 σ_a^2 의 분산성분추정량이 음이 될 확률도 모수 θ 에 종속된다. 동물학을 연구하는 학자들은 이 비율을 유전력이라고도 한다.

분산성분비율의 추정에 관한 체계적인 연구결과는 Loh (1986)에 의해 처음으로 발표되었다. 그는 분산성분비율의 최우추정량(MLE)과 제한최우추정량(REMLE)보다 평균자승오차(MSE) 판정기준아래에서 효율성이 뛰어난 추정량 $\hat{\theta}_{opt}$ 를 제시하였다. 그 후로 Das, Meneghini 와 Giri (1990)는 $\hat{\theta}_{opt}$ 가 특정그룹변환에 있어서 동등인 구성원으로 구성된 특정 계급에 속하는 최적동등추정량(BEE)임을 밝혔다. 또한 Das (1992)는 θ 의 추정량이 MSE 효율성이 높아지게 되는 경우를 기존의 추정량인 경우와 음이 아닌 비해석적 추정량의 집합에 대하여 설명하였다. 하지만 지금까지 발표된 논문을 통하여 보았을 때 결론적으로 θ 의 추정량중 균형일원변량모형에서 임의의 θ 값에

1) (140-714) 서울시 용산구 한남동 산 8번지 단국대학교 전산통계학과 부교수.

대하여 다른 추정량에 비하여 월등히 MSE 효율성이 뛰어난 추정량은 존재하지 않았다.

따라서 본 논문에서는 θ 의 추정량들이 가져야 할 최적인 성질을 밝히는 방향에서 벗어나서 실제로 지금까지 나와 있는 여러 가지 추정량들보다 우수한 MSE를 갖는 추정량들을 제시하는 데 초점을 맞추기로 한다. 본 논문의 구성은 2절에서는 분산성분비율의 추정량으로서 사용되는 여러 가지 추정량들을 간략히 살펴보고, 3절에서는 새로운 분산성분비율의 추정량 $\hat{\theta}_{emp}$ 을 제시하고 이 추정량은 $\hat{\theta}_{opt}$ 보다 θ 가 적당한 크기의 값을 가질 때 항상 평균자승오차(MSE)가 작게 된다. 이렇게 제안된 $\hat{\theta}_{emp}$ 은 본질적으로 사전추측값을 고려한 MINQUE 타입의 추정량이라고 할 수 있다. 4절에서는 모의실험을 통하여 DAS(1992)가 고려한 여러 가지 추정량과 $\hat{\theta}_{emp}$ 을 MSE 판정아래에서 그 효율성을 알아본다. 또한 일원변량모형에 있어서 식(1.1)의 여러 가지 I 와 J 값에 대하여 $\hat{\theta}_{emp}$ 과 $\hat{\theta}_{opt}$ 의 MSE를 그래프로 비교하여 본다. 그리고 끝으로 5절에서는 결론을 제시한다.

2. 여러 가지 추정량들의 모양

수식의 간편한 표기를 위하여 v_1 과 v_2 를 각각 $v_1 = I-1, v_2 = I(J-1)$ 이라고 하자. 그러면 식(1.1)에 대한 모형제곱합 SSA, 오차 제곱합 SSE, 평균 모형제곱합 MS_A 와 평균 오차 제곱합 MS_E 는 다음과 같이 정의된다.

$$SSA = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2, \quad SSE = J \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2.$$

$$MS_A = SSA / v_1, \quad MS_E = SSE / v_2.$$

이 경우 우리는 통계량 $\bar{y}_{..}$, SSA, SSE는 식(1.1)에 대한 완전충분통계량(Graybill 1976, 정리 15.2.4)임을 알 수 있으며, 따라서 이 논문에 언급되는 θ 의 추정량은 SSA와 SSE의 함수에 국한하기로 한다. 또한 θ 의 값이 항상 0이상의 값을 취하므로 θ 의 추정량으로써 항상 음이 아닌 값을 제공하는 추정량인 경우에만 논의한다. 이상을 토대로 하여 기존에 나와 있는 θ 의 추정량은 다음과 같은 세 가지 형태로 대별될 수 있다.

$$\begin{aligned} \text{타입 A: } & \hat{\theta} = aG, \\ \text{타입 B: } & \hat{\theta} = (bG - c)^+, \\ \text{타입 C: } & \hat{\theta} = f(G). \end{aligned}$$

여기서 a, b, c 는 음이 아닌 상수 값이며 $f(\cdot)$ 는 임의의 음이 아닌 함수이다. 또한 $(\cdot)^+$ 은 양수 부분이며 G 는 MS_A / MS_E 를 의미한다. 그리고 평균제곱합의 비율 G 는 $(1+J\theta)F$ 와 같은 분포를 가지며 여기서 F 는 자유도가 각각 v_1, v_2 인 F 분포를 따르는 확률변수이다.

앞으로 논문의 전개를 위하여 θ 의 추정량을 각각 타입-A,B,C라고 부르기로 한다. 타입-A 추정량중에서 관심의 대상이 되는 것은 MINQUE 타입의 추정량이다. MINQUE 타입의 추정량은 분자는 σ_e^2 의 MINQUE 추정량, 분모는 σ_e^2 의 분산분석추정량 MS_E 를 사용하면 타입 A와 같이 구할 수 있다. 물론 이 경우에 σ_e^2 의 추정량도 MINQUE 추정량으로 대체하면 MINQUE 추정량을 구할 수 있지만, 타입-A 추정량에 관한 비교분석을 하는 경우에 계산이 복잡하며 또한 모의실험을 통한 효율성은 분산분석추정량을 사용한 경우와 거의 비슷하게 나타났기 때문에 (Cornley, 1987), MINQUE 타입의 추정량을 이 논문에서 사용하였다. 다음은 타입-A중에서 특별히 관심을 갖는 a 값들이다.

$$a_1 = \frac{1}{J}, \quad a_2 = \frac{J}{1+J^2}, \quad a_3 = \frac{1}{1+J}, \quad a_4 = \frac{I-1}{I} \frac{J}{(1+J)^2}. \quad (2.1)$$

표기의 간편함을 위하여 $\hat{\theta}(a_i) = a_i G, i=1,2,3,4$ 로 정의하자. 그러면 $\hat{\theta}(a_1)$ 는 CMINQUE (Chaubey, 1983), $\hat{\theta}(a_2)$ 는 MB-MINQE (Hartung, 1981), $\hat{\theta}(a_3)$ 는 AUE (J.N.K.Rao, 1973), $\hat{\theta}(a_4)$ 는 MINQE (P.S.R.S.Rao와 Chaubey, 1978) 타입 추정량이 된다. 다음 보조정리 1은 Loh (1986)의 논문에 나와 있는 정리를 약간 수정한 형태이며 타입-A 추정량들의 평균자승오차 (MSE)를 비교하는 데 유용하게 사용된다.

보조정리 1. 만일 $v_2 > 4$ 이고 임의의 $\theta \geq 0$ 에 대하여 위험함수 $R(\delta, \theta) = E(\delta G - \theta)^2$ 는 $\delta^* < \delta < \infty$ 에 있어서 δ 에 대한 단조증가함수이며, δ 의 값이 $\delta = \delta^*$ 인 경우에 최소값을 갖는다. 이 경우 δ^* 는 다음과 같이 구해진다.

$$\delta^* = K\theta / (1+J\theta), \quad K = v_1(v_2 - 4)(v_2(v_1 + 2))^{-1} < 1. \quad (2.2)$$

보조정리 1을 이용하면 다음과 같은 보조정리 2와 3을 쉽게 구할 수 있다.

보조정리 2. 만일 $v_2 > 4$ 이고 δ^* 를 (2.2)식과 같이 정의하면, $\hat{\theta}_k = \delta_k G, k=1,2$ 인 경우에 있어서 $w_i = |\delta^* - \delta_i|, i=1,2$ 에 대하여, 만일 $w_1 < w_2$ 이면 임의의 $\theta \geq 0$ 에 대하여 다음이 성립한다.

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$$

보조정리 3. 만일 $v_2 > 4$ 이고 a_i 와 δ^* 를 식(2.1)과 식(2.2)로 정의하면, $\delta^* < a_i, i=1,2,3,4$ 와 임의의 $\theta \geq 0$ 에 대하여 다음이 성립한다.

$$MSE(\hat{\theta}(a_4)) < MSE(\hat{\theta}(a_3)) < MSE(\hat{\theta}(a_2)) < MSE(\hat{\theta}(a_1)).$$

증명: $\delta^* < a_i$, $i = 1, 2, 3, 4$ 인 경우에 간단한 계산을 통해서 $a_4 < a_3 < a_2 < a_1$ 임을 확인할 수 있다.

타입-B 추정량의 대표적인 것은 분산분석추정량(ANOVA), 최소분산불편추정량(MVUE), 최우추정량(MLE), 제한최우추정량(REMLE)등이 있으며 구체적인 모양은 다음과 같다.

$$\begin{aligned}\hat{\theta}_{\text{ANOVA}} &= \frac{1}{J} (G-1)^+, & \hat{\theta}_{\text{MVUE}} &= \left(\frac{J-I-2}{J(J-1)} G - \frac{1}{J} \right)^+, \\ \hat{\theta}_{\text{MLE}} &= \frac{1}{J} \left(\frac{I-1}{I} G - 1 \right)^+, & \hat{\theta}_{\text{REMLE}} &= \frac{1}{J} (G-1)^+.\end{aligned}\quad (2.3)$$

여기서 원래 분산분석추정량(ANOVA)이나 최소분산불편추정량(MVUE)은 θ 의 추정량으로서 모두 음수의 값을 가질 수 있으므로 항상 0이상의 값을 갖도록 정의하였다. 또 다른 타입-B 추정량 중에는 opt 추정량과 mode 추정량이 있는 데 다음과 같이 각각 정의된다.

$$\hat{\theta}_{\text{opt}} = \frac{1}{J} \left(\frac{v_1(v_2-4)}{v_2(v_1+2)} G - 1 \right)^+, \quad \hat{\theta}_{\text{mode}} = \frac{1}{J} \left(\frac{v_1(v_2-2)}{v_2(v_1+2)} G - 1 \right)^+.\quad (2.4)$$

식(2.4)에 정의된 $\hat{\theta}_{\text{opt}}$ 는 Loh(1986)에 의해 제안된 추정량으로써 이 추정량은 $J^{-1}((1-\delta)G-1)^+$, $-\infty < \delta < \delta^*$, $\delta^* = v_1(v_2-4)/(v_2(v_1+2))$ 와 같은 추정량중에서 가장 작은 MSE를 갖는다. 그리고 $\hat{\theta}_{\text{mode}}$ 는 Klotz et al.(1969)에 의해 제안된 베이지안 모드추정량이다. 타입-B 추정량들에 대한 MSE의 비교는 부분적으로 알려져 있으며 다음은 REML, ML, mode, opt 추정량에 대한 MSE의 비교이다 (Loh,1986).

$$\text{MSE}(\hat{\theta}_{\text{opt}}) < \text{MSE}(\hat{\theta}_{\text{mode}}) < \text{MSE}(\hat{\theta}_{\text{MLE}}) < \text{MSE}(\hat{\theta}_{\text{REMLE}}).$$

타입-B 추정량과 비슷한 분류로 Das(1992)는 θ 의 추정량으로서 $a + bG$, $a \geq 0$, $b \geq 0$ 와 같은 집합을 고려하여 이 집합에 속하는 추정량이 최소분산불편추정량(MVUE)보다 작은 MSE를 갖기 위한 조건을 서술하였다. 관심 있는 독자는 Das(1992)의 논문을 참조하기 바란다. 타입-C 추정량들은 일반적으로 매우 복잡한 모양을 가지고 있음에도 불구하고 타입-A나 타입-B 추정량보다 MSE측면에서 좋은 추정량은 아직 제안되지 못하고 있다. 타입-C 추정량들의 여러가지 모양은 Loh(1986)이나 Das(1992)의 논문에서 찾아 볼 수 있다.

3. 제안된 추정량

이 절에서 제안되는 추정량은 근본적으로 타입-A 추정량에 속한다. 하지만 타입-A,B,C 추정량에 대한 관계를 수학적으로 분명하게 밝히는 것은 매우 어려운 일이라고 생각되어지나 타입-A 추정량의 MSE 하한은 G 의 계수가 식(2.2)에 정의된 δ^* 일때 구하여 지는데, 만일 G 의 계수가

δ^* 가 되면 우리는 다음과 같은 바람직한 결과들을 알 수 있다. 첫째로 Das(1992)의 논문에 의하면 θ 의 추정량의 집합으로 $\{\hat{\theta} | \hat{\theta} = a + bG, a \geq 0, b \geq 0\}$ 와 같은 집합과 타입-A 추정량의 집합간의 MSE 효율성을 논의하였는데, 임의의 θ 에 대하여 $b = \delta^*$ 인 모든 b 값에 대하여 항상 b 를 이용한 타입-A 추정량의 MSE 효율성이 높다는 사실을 밝혔다. 둘째로 δ^* 를 이용하여 구하여진 추정량 $\hat{\theta}_{\delta^*}$ 은 $\{\hat{\theta} | \hat{\theta} = aG - 1/J, a > 0\}$ 와 같은 모양의 모든 추정량보다 항상 MSE 효율성이 높다. 이 집합에 속하는 추정량들은 2절에서 소개한 타입-B 추정량중에서 음의 값을 허용하면 전부 이 집합에 속하게 되며 또한 θ 의 추정량으로 흔히 사용하는 분산분석추정량이나 최소분산 불편추정량도 이 집합에 속하게 된다. 이 사실은 다음과 같이 증명되는 데, $\hat{\theta}_{\delta^*}$ 를 $\hat{\theta}_{\delta^*} = \delta^*G$ 로 두면 G 는 $(1 + J\theta)F$, $F \sim F(v_1, v_2)$ 임으로 이 사실과 F 분포의 1,2차 적률을 이용하면 $MSE(\hat{\theta}_{\delta^*}) = 2\theta^2 \frac{(v_1 + v_2 - 2)}{(v_1 + 2)(v_2 - 2)}$ 임을 알 수 있다. 또한 $E(aG - 1/J - \theta)^2$ 를 최소로 하는 a 값은 K/J , $K = v_1(v_2 - 4)(v_2(v_1 + 2))^{-1} < 1$, 임을 알 수 있다. 따라서 $a = K/J$ 를 이용하여 MSE를 계산하면 $2(\theta + 1/J)^2 \frac{(v_1 + v_2 - 2)}{(v_1 + 2)(v_2 - 2)}$ 와 같이 됨을 확인할 수 있으며, 따라서 이 사실은 항상 참임을 알 수 있다.

하지만 δ^* 는 실제로 θ 의 함수이기 때문에 우리는 활용할 수가 없으며 따라서 δ^* 를 어떤 값 δ 로 대체한 추정량 $\hat{\theta} = \delta G$ 는 $\delta^* = \delta$ 가 되는 θ 의 값 θ_0 에서는 MSE가 가장 작아지나 θ 의 실제 값이 θ_0 에서 멀어지면 멀어질수록 MSE 효율성이 떨어진다. 이런 이유로 θ 를 확률변수로 간주하고 한 가지 가능한 판단기준으로서 $E(\delta^* - \delta)^2$ 를 최소화하는 상수 δ 를 선택하여 보자. 이 경우 δ 의 값은 $\delta = E(\delta^*)$ 가 된다는 것을 우리는 알고 있으므로, δ^* 의 기대값을 구하기 위하여 θ 에 대한 정보가 전혀 없으므로 급내상관계수 ρ 가 $(0, 1)$ 에서 일양분포를 한다고 가정하면, 따라서

$$\begin{aligned} E(\delta^*) &= E\left(\frac{K\rho}{1+\rho(J-1)}\right) = K \int_0^1 \frac{\rho}{1+\rho(J-1)} d\rho \\ &= \frac{K}{(J-1)^2} \int_0^{J-1} \frac{t}{1+t} dt = K \frac{((J-1) - \log(J))}{(J-1)^2}. \end{aligned}$$

그러므로 δ^* 의 추정값으로 다음 식(3.1)을 얻을 수 있다.

$$\hat{\delta}^* = K \frac{((J-1) - \log(J))}{(J-1)^2}. \quad K = v_1(v_2 - 4)(v_2(v_1 + 2))^{-1} < 1. \quad (3.1)$$

식(3.1)을 이용한 추정량 $\hat{\theta}_{emp}$ 의 MSE 효율성은 4절에서 논의된다. 한편 식(3.1)은 MINQUE 이론에 근거하면 결론적으로 사전추측값이 r 인 MINQUE 타입 추정량이라고 할 수 있다. 따라서 이 절에서 서술된 방향은 결국 사전추측값 r 의 값을 잘 설정함으로써 사전추측값이 1인 MINQUE 타입 추정량보다 평균적으로 MSE 효율성을 높일 수 있다는 사실로 설명될 수 있다.

4. 모의실험

이 절에서는 3절에서 제안된 $\hat{\theta}_{emp}$ 의 MSE 효율성이 뛰어난을 밝히기 위하여 Das(1992)가 제안한 6개의 추정량과 비교하여 보기로 한다. Das(1992)는 그의 논문에서 $I=5, J=5$ 인 경우와 $I=4, J=7$ 인 두 가지 경우에 모의실험을 하였는데, 본 논문에서도 같은 두 가지 경우를 사용하였다. 표(4-1)과 (4-2)는 2가지 경우에 대한 Das(1992)의 6개의 추정량, $\hat{\theta}_{emp}$ 에 대한 $MSE(\hat{\theta})/MSE(\hat{\theta}_{UNB})$ 의 값이며, 여기서 $\hat{\theta}_{UNB}$ 는 식(2.3)에 소개된 $\hat{\theta}_{MVUE}$ 추정량에서 음의 값을 취할 수 있도록 허용한 원래의 θ 의 최소분산불편추정량이다. 또한 $\hat{\theta}_0$ 는 Loh의 opt 추정량이며, 나머지 추정량에 대한 자세한 설명은 Das(1992)의 논문을 참고하면 된다.

모의실험은 통계패키지 SAS를 이용하였는데 통계량 G 의 값을 구하기 위하여 난수 RANGAM을 이용하였고, 모든 MSE값은 10000번의 반복을 이용하여 구하였다. 또한 θ 의 값은 급내상관계수 $\rho = \theta/(1+\theta)$ 를 이용하여 ρ 의 값이 0부터 0.9까지 0.1간격으로 선택되어지는 것을 택하였다. 모의실험 결과 Das(1992)의 논문에 제시된 6개의 추정량에 대한 추정된 MSE값은 Das(1992)의 결과와 비교하여 소숫점 2자리에서 약간의 차이가 생기는 경우가 있었으나 전체적으로 거의 비슷하였다. <표 4.1>과 <표 4.2>에 표시된 *는 7개의 추정량중에서 가장 MSE 효율성이 높은 추정량이다.

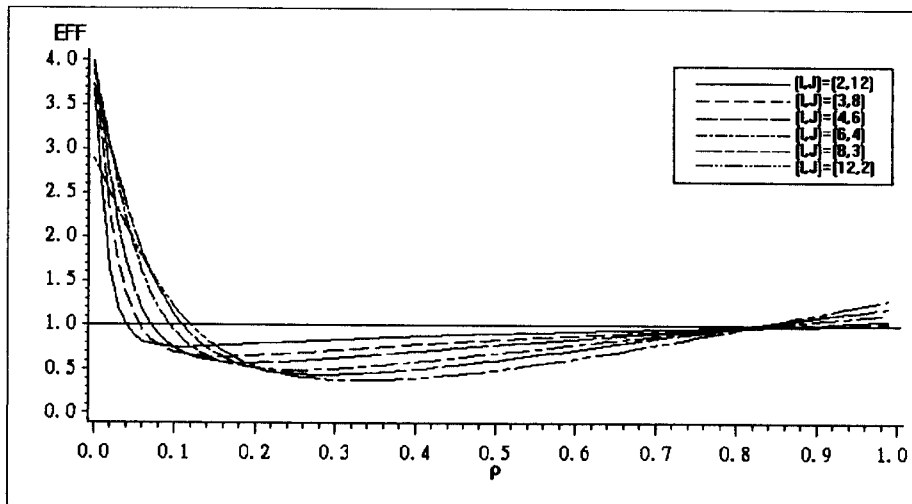
<표 4.1> $MSE(\hat{\theta})/MSE(\hat{\theta}_{MVUE}) : I=5, J=5$ 인 경우.

ρ	$\hat{\theta}_0$	$\hat{\theta}_{10}$	$\hat{\theta}_{11}$	$\hat{\theta}^{**}$	$\hat{\theta}_0^*$	$\hat{\theta}_1^*$	$\hat{\theta}_{emp}$
0.0	0.12	0.08	0.09	0.06*	0.08	0.07	0.48
0.1	0.25	0.22	0.23	0.21	0.08*	0.08*	0.21
0.2	0.40	0.38	0.39	0.38	0.24*	0.25	0.21*
0.3	0.49	0.48	0.48	0.48	0.40	0.42	0.28*
0.4	0.55	0.53	0.53	0.54	0.53	0.55	0.35*
0.5	0.57	0.57	0.54	0.57	0.67	0.71	0.42*
0.6	0.58	0.58	0.57	0.58	0.73	0.76	0.48*
0.7	0.59	0.58	0.58	0.59	0.83	0.86	0.53*
0.8	0.59	0.58	0.58	0.59	0.90	0.94	0.57*
0.9	0.59*	0.59*	0.59*	0.59*	0.96	0.99	0.61

<표 4.2> $MSE(\hat{\theta}) / MSE(\hat{\theta}_{MVUE}) : I=4, J=7$ 인 경우.

ρ	$\hat{\theta}_0$	$\hat{\theta}_{10}$	$\hat{\theta}_{11}$	$\hat{\theta}^{**}$	$\hat{\theta}_0^*$	$\hat{\theta}_1^*$	$\hat{\theta}_{emp}$
0.0	0.09	0.06	0.07	0.05	0.02*	0.65	0.41
0.1	0.26	0.23	0.24	0.22	0.14*	0.31	0.18
0.2	0.40	0.37	0.38	0.37	0.34	0.31	0.24*
0.3	0.47	0.44	0.45	0.45	0.49	0.34	0.31*
0.4	0.51	0.48	0.49	0.49	0.61	0.38	0.37*
0.5	0.53	0.50	0.50	0.50	0.69	0.42*	0.42*
0.6	0.54	0.52	0.51	0.52	0.80	0.47	0.46*
0.7	0.54	0.53	0.52	0.53	0.85	0.48*	0.50
0.8	0.54	0.54	0.53	0.53	0.92	0.51*	0.53
0.9	0.55	0.54	0.53	0.54	0.92	0.52*	0.55

<표 4.1>과 <표 4.2>를 통해서 알 수 있듯이 I 와 J 의 두 가지 조합과 급내상관계수 ρ 의 값이 대략 $0.2 < \rho < 0.7$ 인 경우에 있어서 $\hat{\theta}_{emp}$ 는 Das(1992)의 6개의 추정량보다 효율성이 높다. 따라서 $\hat{\theta}_{emp}$ 는 θ 의 값이 아주 크거나 작은 경우를 제외하고 MSE 효율성이 높다고 할 수 있다. 이러한 경향을 좀 더 자세히 관찰하기 위하여 다양한 실험계획에 대하여 $\hat{\theta}_{emp}$ 의 $\hat{\theta}_{opt}$ 에 대한 MSE 효율성을 알아보면 <그림 4.1>과 같다. <그림 4.1>이 작성된 과정은 $MSE(\hat{\theta}_{opt})$ 를 구하기 위하여 필요한 수치적분계산은 10점-가우스 구적법을 이용하였고, 플롯은 SAS/GRAPH를 이용하였다.



<그림 4.1> $MSE(\hat{\theta}_{emp}) / MSE(\hat{\theta}_{opt}) : N=24$ 인 경우.

<그림 4.1>은 총관측치 $N=24$ 인 경우에 두 추정량의 MSE 비율 $MSE(\hat{\theta}_{emp})/MSE(\hat{\theta}_{opt})$ 을 보여준다. $N=24$ 인 경우에는 가능한 I 와 J 의 두가지 조합은 모두 6가지인데 6가지 조합 모두에 대하여 두 추정량의 MSE 비율은 정도의 차이는 있으나 전체적으로 $0.1 < \rho < 0.8$ 인 경우에 항상 $\hat{\theta}_{emp}$ 의 효율성이 높다. 이러한 경향은 실제로 N 의 크기가 변화하여도 비슷한 양상을 가지게 된다. 따라서 대략적으로 $\hat{\theta}_{emp}$ 는 $0.1 < \rho < 0.8$ 인 경우에 MSE 판정기준아래에서 $\hat{\theta}_{opt}$ 추정량보다 효율성이 높아진다고 할 수 있다.

5. 결론

이 논문에서는 일원변량모형에 있어서 분산성분비율 θ 에 대한 새로운 추정량을 고려하였다. 제안된 추정량은 모수 θ 에 대한 정보가 전혀 없을 때 급내상관계수 대략 ρ 의 값이 2 $0.2 < \rho < 0.7$ 인 경우에 기존의 다른 추정량들보다 효율성이 뛰어나다. 하지만 $\rho > 0.7$ 인 경우와 $\rho < 0.2$ 인 경우에도 95%정도의 신뢰도에 대하여 구한 모수 θ 의 신뢰구간의 길이가 짧은 경우에2는 좀 더 추정량의 정밀도를 높일 수 있다. 예를 들어 ρ 의 신뢰구간의 하한을 $\hat{\rho}_L$, 신뢰구간의 상한을 $\hat{\rho}_U$ 라고 했을 때 $|\hat{\rho}_U - \hat{\rho}_L| < 0.3$ 정도가 되면, $\hat{\rho}_L$ 과 $\hat{\rho}_U$ 에 대응되는 $\hat{\theta}_L$ 과 $\hat{\theta}_U$ 를 이용하여 $\hat{\theta} = (\hat{\theta}_U + \hat{\theta}_L)/2$ 를 구하고, 3절의 δ^* 값을 $\delta^* = K\hat{\theta}/(1+J\hat{\theta})$ 로 추정하면 좀 더 효율성을 높일 수 있다. 이와 같은 결과는 논문에서는 제시되지는 않았지만, 신뢰구간의 길이가 매우 짧으면 짧을수록 신뢰구간의 하한및 상한의 평균값이 θ 의 점추정량으로 바람직하다는 모의실험의 결과에 따른 것이다. 하지만 이런 경우는 모의실험의 경우에서도 알 수 있드시 흔히 발생하는 일은 아니다.

참고문헌

- [1] Chaubey, Y.P. (1983). A Non Negative Estimator of Variance Component Closest to MINQUE, *Sankhya*, Vol. 45, 201-211.
- [2] Conerly, M.D. and Webster, J.T. (1987). MINQE for the One-Way Classification, *Technometrics*, Vol. 29(2), 229-236.
- [3] Das, K. (1992). Improved Estimation of the Ratio of Variance Components for a Balanced One-Way Random Effects Model, *Statistics & Probability Letters*, Vol. 13, 99-108.
- [4] Das, K., Q. Meneghini and N. Giri (1990). Inadmissibility of an Estimator for the Ratio of Variance Components, *Statistics & Probability Letters*, Vol. 10, 151-157.
- [5] Graybill, F.A. (1976). *Theory and Applications of the Linear Models*, Duxbury Press, North Scituate, Massachusetts.

- [6] Hartung, J. (1981). Nonnegative Minimum Biased Invariant Estimation in Variance Components Models, *Annals of Statistics*, Vol. 9, 278-292.
- [7] Horn, S.D. and Horn, R.A. (1975). Comparison of Estimators of Heteroscedastic Variances in Linear Models, *Journal of the American Statistical Association*, Vol. 70, 872-879.
- [8] Loh, W.Y. (1986). Improved Estimators for Ratios of Variance Components, *Journal of the American Statistical Association*, Vol. 81, 699-702.
- [9] Klotz, J.H., R.C. Milton and S. Zacks (1969). Mean Square Efficiency of Estimators of Variance Components, *Journal of the American Statistical Association*, Vol. 64, 1383-1402.
- [10] Rao, J.N.K. (1973). On the Estimation of heteroscedastic Variances, *Biometrics*, Vol. 29, 11-24.
- [11] Rao, P.S.R.S. and Chaubey, Y.P. (1978). Three Modifications of the principle of MINQUE, *Commun. Statist.- Theor. Meth.*, Vol. A7(8), 767-778.