

모비율 차이의 신뢰구간들에 대한 비교연구 *

정형철¹⁾ 전명식²⁾ 김대학³⁾

요약

본 논문에서는 두 모비율의 차에 대한 기존의 신뢰구간들을 소개하고 붓스트랩 신뢰구간도 제안하였다. 또한 모비율의 차에 대한 신뢰구간이 가지는 성질로서 근사신뢰구간의 하향추정의 문제와 정확신뢰구간의 상향추정의 문제점들을 확인하였고 평균포함확률, 구간기대폭 그리고 왜도성 측면에서 종합적인 비교를 하였다. 특히 모수에 대한 사전분포를 가정하여 여러 신뢰구간들이 지니는 특징도 살펴보았다. 기존의 신뢰구간들과 제안된 붓스트랩 신뢰구간은 소표본의 모의실험을 통하여 실제 포함확률의 평균을 기준으로 비교되었고 이항분포에서와 같이 정확신뢰구간이 지니는 보수성을 확인할 수 있었다. 신뢰구간의 평균포함확률의 등고선 그림도 소개 하였다.

주요용어: 모비율, 모비율의 차, 신뢰구간, 정확신뢰구간, 붓스트랩방법, 포함확률, 왜도

1. 소개

두 확률변수 X_1, X_2 가 각각 이항분포 $B(n_1, p_1), B(n_2, p_2)$ 를 따를 때 모비율의 차를 $d = p_1 - p_2$ 그리고 모비율의 합을 $\theta = p_1 + p_2$ 라 두면 X_1, X_2 의 결합확률밀도함수는

$$f_{p_1, d}(x_1, x_2) = \binom{n_1}{x_1} \binom{n_2}{x_2} p_1^{x_1} (1-p_1)^{n_1-x_1} (p_1-d)^{x_2} (1-p_1+d)^{n_2-x_2} \quad (1.1)$$

로 표현된다. 이제 모비율의 차이 d 에 대한 신뢰구간 구축에 대해 살펴보자. 각각의 표본비율을 $\hat{p}_1 = X_1/n_1, \hat{p}_2 = X_2/n_2$ 라 표기하면 정규근사를 활용한 d 의 신뢰구간으로 쉽게 고려되는 것은 Wald 신뢰구간으로

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \quad (1.2)$$

와 같이 표현된다. 여기서 z_{α} 는 표준정규분포의 $(1-\alpha)$ 백분위수를 나타낸다. Agresti와 Coull(1998)은 Wald 신뢰구간이 주어진 명목수준 $(1-\alpha)$ 를 크게 하향 추정하는 경향이 있음을 밝힌 바 있다. 이와 같은 하향추정의 문제는 이산형 확률분포를 연속형 확률분포로

* 이 연구는 2003년도 대구가톨릭대학교 교비연구비의 지원에 의한 것임

1) (450-701) 경기도 평택시 용이동 111, 평택대학교 정보통계학과, 조교수

E-mail : jhc@ptuniv.ac.kr

2) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과, 교수

E-mail : jhun@mail.korea.ac.kr

3) (712-702) 경북 경산시 하양읍 금락리 330, 대구가톨릭대학교 정보통계학과, 교수

E-mail : dhkim@cu.ac.kr

근사추정하는 것과 $\hat{d} = \hat{p}_1 - \hat{p}_2$ 의 분산추정에 있어 장애모수 p_1 (혹은 p_2)의 존재 때문이라 하겠다. Wald 신뢰구간 이외에도 근사이론을 활용하는 신뢰구간의 성질에 대한 많은 연구가 있는데 대표적으로 Beal(1987)과 Newcombe(1998b)를 들 수 있다. Beal(1987)은 정규 근사신뢰구간, Haldane신뢰구간 등 다섯가지 근사 신뢰구간에 대하여 주어진 임의의 모수 d 에서 포함확률을 모의실험을 통하여 비교하였다. Newcombe(1998b)는 다양한 표본의 크기(n_1, n_2)와 모수 d 가 주어질 때 포함확률의 평균을 모의실험을 통해 추정하였으며, 계산상의 복잡성을 이유로 스코어 신뢰구간을 추천하였다. 한편 Efron(1979)의 붓스트랩방법으로 d 의 신뢰구간을 구축할 수도 있다. Woodrooffe와 Jhun(1989)은 평균포함확률 측면에서는 붓스트랩분포와 실제분포와의 차이가 실제적으로 $o(1/\sqrt{n})$ 임을 보였으며 이는 붓스트랩 신뢰구간의 가능성을 보인것으로 여겨진다. 앞에서 언급한 근사신뢰구간들과 대별되는 신뢰구간으로 정확신뢰구간을 고려할 수 있다. 정확신뢰구간에 대해서는 Thomas와 Gart(1977)의 조건부 정확신뢰구간과 Santner와 Snell(1980)의 비조건부 정확신뢰구간을 들 수 있다. 이와 같은 정확신뢰구간은 확률변수의 이산성 때문에 주어진 명목수준을 과대추정하는 경향이 있음이 Agresti와 Coull(1989)에 의해 알려져 있다. 정확신뢰구간이 지니는 보수성을 제거하기 위하여 StatXact과 같은 패키지에서는 Berger와 Boos(1994)의 방법을 이용하여 수정된 정확신뢰구간을 제공한다. 기존의 연구결과를 살펴보면 신뢰구간 구축방법에 대한 제안과 기존의 방법들에 대해 주로 포함확률을 기준으로 부분적 비교연구가 시도되었다. 본 논문에서는 기존의 제안된 신뢰구간들을 전체적으로 취합하고 나아가 붓스트랩 방법도 추가하여 비교하였다. 또한 평가기준도 보완하였다. 표본의 크기가 고정된 상황에서 계산된 실제포함확률의 평균을 비교의 기준으로 사용하였으며, 신뢰구간의 기대폭과 왜도성까지 고려하는 종합적인 비교를 시도하였다.

2. 모비율의 차이 d 에 대한 신뢰구간

모비율의 차이에 대한 신뢰구간을 구축함에 있어 근사 신뢰구간과 정확 신뢰구간의 두 가지 접근방법이 가능하다. 본 절에서는 각 접근방법에 대한 내용을 기술하고자 한다.

2.1. 근사 신뢰구간

근사이론에 의한 d 의 신뢰구간을 살펴보자. $u = (1/n_1 + 1/n_2)/4$, $v = (1/n_1 - 1/n_2)/4$ 로 두고 $\hat{p}_1, \hat{p}_2, \hat{d}$ 그리고 $\hat{\theta}$ 를 각각 p_1, p_2, d 와 θ 의 최대우도추정량이라 하자. 이제 $\hat{d} = \hat{p}_1 - \hat{p}_2$ 의 분산은 다음 식 (2.1)과 같다.

$$V(\theta, d) = u\{(2 - \theta)\theta - d^2\} + 2v(1 - \theta)d. \quad (2.1)$$

중심극한정리에 의한 $100(1 - \alpha)\%$ 신뢰구간의 하한과 상한은 식 (2.2)에 주어진 이차방정식의 두 근에 해당된다.

$$(d - \hat{d})^2 = cV(\tilde{\theta}, \tilde{d}) \quad (2.2)$$

이때 c 는 $\chi_{(1)}^2$ 분포의 $(1 - \alpha)$ 분위점이며 $\tilde{\theta}$ 와 \tilde{d} 는 각각 θ, d 에 대한 표현식으로서 모수 θ 와 d 를 직접 포함하는 함수이거나 포함하지 않는 함수로 설명될 수 있다.

2.1.1. 정규근사에 기초한 신뢰구간

식 (1.2)의 Wald 신뢰구간은 식 (2.2)에 $\tilde{\theta} = \hat{\theta}$ 와 $\tilde{d} = \hat{d}$ 를 대입한 것이며 CL로 표기한다.

$$CL : \hat{d} \pm z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2) / n_2} \quad (2.3)$$

이에 대해 연속성을 고려하여 수정한 신뢰구간(CC)은 다음 식 (2.4)와 같다.

$$CC : \hat{d} \pm \{z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2) / n_2} + (1/2n_1 + 1/2n_2)\} \quad (2.4)$$

Anbar(1983)는 score신뢰구간의 개념을 도입하여 식 (2.2)에 $\tilde{\theta} = 2\hat{p}_1 - d$, $\tilde{d} = d$ 를 대입하는 방법을 제안하였다. 이러한 Anbar의 신뢰구간(AN)의 하한과 상한은 다음 식 (2.5)의 이차방정식의 근으로 표현된다.

$$AN : (d - \hat{d})^2 = cV(\theta_1^*, d), \quad \theta_1^* = 2\hat{p}_1 - d \quad (2.5)$$

Anbar(1983)는 식 (2.5)에서 유도한 신뢰구간을 Buhrman(1977)의 정확신뢰구간과 비교하여 점근효율(asymptotic efficiency)이 있음을 모의실험으로 보인바 있다. 그러나 p_1 과 d 대신에 p_2 와 d 즉 $\tilde{\theta} = 2\hat{p}_2 + d$, $\tilde{d} = d$ 를 대입하여 신뢰구간을 구하는 것과 차이가 발생하는 단점이 있다. 이에 대해서 Mee(1984)는 $\tilde{d} = d$, $\tilde{\theta} = \theta^*(d)$ 를 대입하는 방법을 제안하였다. 이때 $\theta^*(d)$ 는 d 가 알려졌다는 가정 하에서의 θ 의 최대우도추정량이다. Mee(1984)의 신뢰구간(ME)의 하한과 상한은 다음 식 (2.6)의 이차방정식의 근으로 유도된다.

$$ME : (d - \hat{d})^2 = cV(\theta^*(d), d), \quad \theta^*(d) = 2\hat{p}_1^* - d \quad (2.6)$$

여기서 \hat{p}_1^* 는 모수 d 가 알려져 있을 때 p_1 의 최대우도추정량이며 이를 구하는 방법은 Miettinen과 Nurminen(1985)에 자세히 제시되어 있다. 또한 Miettinen과 Nurminen(1985)은 식 (2.6)에서 \hat{d} 의 분산이 $(n_1 + n_2)/(n_1 + n_2 - 1)V(\theta^*(d), d)$ 라는 사실에 근거하여 식 (2.6)을 수정한 다음 식 (2.7)의 이차방정식의 근으로 신뢰구간(MN)의 하한과 상한을 제시하였다.

$$MN : (d - \hat{d})^2 = c^*V(\theta^*(d), d), \quad c^* = c(n_1 + n_2)/(n_1 + n_2 - 1), \theta^*(d) = 2\hat{p}_1^* - d \quad (2.7)$$

식 (2.7)에 의해 주어지는 신뢰구간은 표본의 크기가 작을 때 $c^* > c$ 의 경향이 심해지므로 Mee(1984)의 신뢰구간보다 폭이 더 넓게 형성된다. Beal(1987)은 Mee(1984)나 Miettinen과 Nurminen(1985)의 신뢰구간을 구하기 위해서는 \hat{p}_1^* 을 추정하는 것과 더불어 반복작업이 필요하다고 하였는데 그는 임의의 초기값 d_0 를 선정하여 \hat{p}_1^* 을 추정한 후 각각의 신뢰구간을 구하는 방법에 의해 새로운 d_1 를 유도하여 $|d_1 - d_0|/|d_1| \leq 10^{-4}$ 이 될 때까지 반복작업을 하기를 권하고 있다. Gart와 Nam(1990)의 신뢰구간(GN)은 Mee(1984)의 신뢰구간에 있어 $V(\theta^*(d), d)$ 를 추정하는데 있어 왜도(skewness)까지 고려한 신뢰구간의 하한과 상한으로 다음의 식 (2.8)의 근에 근거한 결과를 제시하였다.

$$GN : (d - \hat{d})^2 = c^*V(\theta^*(d), d) \quad (2.8)$$

여기서 $c^* = (c^{1/2} + r(c - 1)/6)^2$ 이고 $r = V(\theta^*(d), d)^{3/2} \{ \hat{p}_1^* \hat{q}_1^* (\hat{q}_1^* - \hat{p}_1^*) / n_1^2 - \hat{p}_2^* \hat{q}_2^* (\hat{q}_2^* - \hat{p}_2^*) / n_2^2 \}$ 이다. 이러한 신뢰구간은 모수 d 가 1 혹은 -1에 가까워도 큰 영향을 받지 않고 안정

적으로 신뢰구간을 구축해주는 장점이 있으나 계산이 복잡하다는 단점을 지니고 있다고 하겠다. Miettinen과 Nurminen(1985)의 profile 신뢰구간은 우도함수비 $\chi_{(\alpha)}^2 = 2\{\log L(d, \theta^*(d)) - \log L(\hat{d}, \hat{\theta})\}$ 에 기초한 카이제곱분포로부터 유도되는데 profile 신뢰구간(PR)의 하한과 상한은 다음의 식을 만족하는 두 근으로 주어진다.

$$\begin{aligned} \text{PR} : & \log \left[\frac{(\theta^* + d)/2}{x_1/n_1} \right]^{x_1} + \log \left[\frac{(\theta^* - d)/2}{x_2/n_2} \right]^{x_2} \\ & + \log \left[\frac{1 - (\theta^* + d)/2}{(n_1 - x_1)/n_1} \right]^{n_1 - x_1} + \log \left[\frac{1 - (\theta^* - d)/2}{(n_2 - x_2)/n_2} \right]^{n_2 - x_2} = -\frac{c}{2} \end{aligned} \quad (2.9)$$

여기서 셀의 값이 0인 항은 계산에서 제외한다. Newcombe(1998b)는 일변량 스코어 신뢰구간에 기초하여 d 에 대한 신뢰구간(SO)으로 다음 식 (2.10)을 제시하였다.

$$\text{SO} : \hat{d} \pm z_{\alpha/2} \sqrt{l_1(1-l_1)/x_1 + u_2(1-u_2)/n_2} \quad (2.10)$$

여기서 l_1, u_1 은 $(p_1 - \hat{p}_1)^2 = cp_1(1-p_1)/n_1$ 의 일변량 스코어 신뢰구간의 하한과 상한값이며 l_2, u_2 는 $(p_2 - \hat{p}_2)^2 = cp_2(1-p_2)/n_2$ 의 일변량 스코어 신뢰구간의 하한과 상한값이다. 또한 일변량 스코어 신뢰구간에 대해 각각 연속성 수정을 통해 SO 신뢰구간에 대한 연속성 수정 신뢰구간(SC)을 식 (2.11)과 같이 구할 수 있다.

$$\text{SC} : \hat{d} \pm z_{\alpha/2} \sqrt{l_1^*(1-l_1^*)/x_1 + u_2^*(1-u_2^*)/n_2} \quad (2.11)$$

여기서 l_i^* 와 u_i^* 는 $|p_i - \hat{p}_i| - 1/(2n_i) = c^{1/2} \sqrt{p_i(1-p_i)/n_i}$, $i = 1, 2$ 의 근임을 Blyth와 Still(1983)이 밝힌바 있다.

2.1.2. θ 의 사후평균을 이용하는 신뢰구간

임의의 $a \geq -1$ 에 대해,

$$\hat{\theta}(a) = \frac{n_1}{n_1 + 2(a+1)} \hat{p}_1 + \frac{a+1}{n_1 + 2(a+1)} + \frac{n_2}{n_2 + 2(a+1)} \hat{p}_2 + \frac{a+1}{n_2 + 2(a+1)}$$

라 두자. $\hat{\theta}(a)$ 는 $(p_1 q_1 p_2 q_2)^a$ 에 비례하는 (p_1, p_2) 의 사전확률분포가 주어졌을 때 θ 의 사후평균에 해당한다. 적당한 a 에 대해 $\hat{\theta} = \hat{\theta}(a)$, $\tilde{d} = d$ 를 대입하여 이차방정식을 풀면 신뢰구간을 얻을 수 있다. $a = -1$ 이면 $\hat{\theta}(-1) = \hat{\theta}$ 이며 $a = \infty$ 이면 $\hat{\theta} = 1$ 이 된다. 또한 $n_1 = n_2$ 일 때 $a = -1$ 일때의 신뢰구간이 다른 a 값에 대한 신뢰구간에 비해 항상 작은 신뢰구간을 형성한다고 알려져 있다. Beal(1987)은 $a = -1$ 일 때의 신뢰구간을 Haldane(1945) 신뢰구간(HA), $a = -1/2$ 의 신뢰구간을 Jeffreys-Perks 신뢰구간(JP)이라 불렀다. 즉 Haldane 신뢰구간과 Jeffreys-Perks 신뢰구간의 하한과 상한은 각각 다음의 이차방정식 (2.12)와 (2.13)의 근으로 유도된다.

$$\text{HA} : (d - \hat{d})^2 = cV(\hat{\theta}(-1), d) \quad (2.12)$$

$$\text{JP} : (d - \hat{d})^2 = cV(\hat{\theta}(-1/2), d) \quad (2.13)$$

2.1.3. 붓스트랩방법에 의한 신뢰구간

통계량 $T = \frac{(\hat{d}-d)}{\sqrt{V(\hat{d},\hat{\theta})}}$ 에 근거하여 d 의 붓스트랩 신뢰구간을 구하는 과정은 다음과 같다.

[단계1] 모수적 붓스트랩 표본 X_1^*, X_2^* 를 각각 $B(n_1, \hat{p}_1), B(n_2, \hat{p}_2)$ 로부터 표본추출한다. 붓스트랩 표본비율을 $\hat{p}_1^* = X_1^*/n_1, \hat{p}_2^* = X_2^*/n_2$ 라 놓자. 또한 붓스트랩 표본비율의 차를 $\hat{d}^* = \hat{p}_1^* - \hat{p}_2^*$ 그리고 붓스트랩 표본비율의 합을 $\hat{\theta}^* = \hat{p}_1^* + \hat{p}_2^*$ 라 놓자.

[단계2] $T^* = \frac{(\hat{d}^*-d)}{\sqrt{V(\hat{d}^*,\hat{\theta}^*)}}$ 를 계산한다. 이때 $V(\hat{d}^*, \hat{\theta}^*) = \hat{p}_1^*(1 - \hat{p}_1^*)/n_1 + \hat{p}_2^*(1 - \hat{p}_2^*)/n_2$ 이다.

[단계3] [단계1]과 [단계2]를 B회 반복하여 식 (2.14)의 $100(1-\alpha)\%$ 붓스트랩 신뢰구간(BT)의 하한과 상한을 각각

$$BT : \hat{d} - Q_{(1-\alpha/2)}V(\hat{d}, \hat{\theta}), \quad \hat{d} - Q_{(\alpha/2)}V(\hat{d}, \hat{\theta}) \tag{2.14}$$

으로 구한다. 여기서 $Q_{(\alpha)}$ 는 T^* 의 $100\alpha\%$ 백분위점이다.

붓스트랩 신뢰구간은 식 (2.3)의 신뢰구간에서 정규분포를 이용한 대신에 붓스트랩 분포를 이용한다는 점에 차이가 있다고 하겠다. 또한 Woodroofe와 Jhun(1989)에 의하면 기저분포가 격자분포일 때 평균포함확률 측면에서 붓스트랩 분포와 실제분포와의 차이가 실제적으로 $o(1/\sqrt{n})$ 와 근사함을 보였으므로 본 연구와 같은 이항분포에서도 붓스트랩 신뢰구간이 Wald 신뢰구간이 지니는 하향추정 문제를 다소 극복해 주리라 예상할 수 있다.

2.1.4. 정확신뢰구간

근사신뢰구간에 대한 대안으로 정확신뢰구간을 생각할 수 있다. 정확신뢰구간을 구축하는데 있어서 가장 큰 문제점은 식 (1.1)에서 장애모수 p_1 의 제거에 있다. 여기에 있어서 중요한 두 가지 통계적 방법으로 조건부정확신뢰구간과 비조건부 정확신뢰구간이 존재한다. 조건부정확신뢰구간은 성공의 합 $m = x_1 + x_2$ 을 고정시켜 충분통계량으로 p_1 을 제거하는 방법이며 Barnard(1945)의 비조건부 정확신뢰구간은 m 을 확률변수로 고려하여 장애모수의 주어진 범위에서 통계량의 최대값으로 장애모수를 제거하는 방법이다.

2.1.5. 비조건 정확신뢰구간

모든 가능한 표본 $\mathbf{y} = (y_1, y_2) \in n_1 \times n_2$ 에 대해 표본비율의 차를 $D(\mathbf{y}) = y_1/n_1 - y_2/n_2$ 라 두면 주어진 표본 $\mathbf{x} = (x_1, x_2)$ 에 대한 상한의 표본 U 와 하한의 표본 L 은 다음과 같다.

$$U = \{(y_1, y_2) | D(\mathbf{y}) \geq D(\mathbf{x})\}, \quad L = \{(y_1, y_2) | D(\mathbf{y}) \leq D(\mathbf{x})\}$$

이제 모수 d 의 $100(1-\alpha)\%$ 정확신뢰구간을 구하려면, 관찰값 $D(\mathbf{x})$ 에 대해 유의수준 $\alpha/2$ 에서 귀무가설 $H_0 : d = d_0$ 를 기각하지 않는 모든 d_0 를 계산해야 한다. 이는 주어진 $D(\mathbf{x})$ 에 대해 양측검정에서 어떤 d_0 를 사용하여야 귀무가설을 채택할 수 있는가 하는 문제이다. 이

때 얻어지는 모든 d_0 중 최소값 d_L 을 정확신뢰구간의 하한, 최대값 d_U 를 정확신뢰구간의 상한이라 한다. 즉 최대값과 최소값은 다음과 같다.

$$Pr(U|d = d_L) = Q_{p_1, d}(\mathbf{x}) \leq a/2, \quad Pr(L|d = d_U) = P_{p_1, d}(\mathbf{x}) \leq a/2 \quad (2.15)$$

여기서 $Q_{p_1, d}(\mathbf{x}) = \sum_{D(y) \geq D(x)} f_{p_1, d}(y)$ 이고 $P_{p_1, d}(\mathbf{x}) = \sum_{D(y) \leq D(x)} f_{p_1, d}(y)$ 이다. 그러나 식 (2.15)에서 미지의 장애모수(nuisance parameter) p_1 (혹은 p_2)때문에 정확신뢰구간을 계산할 수 없다. 이러한 문제를 피하기 위해서 Santner와 Snell(1980)은 주어진 임의의 모수 d 에 대해 장애모수의 전구간에서 $Q_{p_1, d}(\mathbf{x})$ 와 $P_{p_1, d}(\mathbf{x})$ 의 최대값을 선택하는 방법을 생각하였다. 장애모수 p_1 의 범위는 d 의 선택에 따라 $I(d) = \{p_1 : \max(0, d) \leq p_1 \leq \min(1, 1+d)\}$ 와 같이 결정된다. $I(d)$ 에서 최대값을 각각 다음과 같이 놓으면

$$Q_d(\mathbf{x}) = \sup\{Q_{p_1, d}(\mathbf{x}) : p_1 \in I(d)\}, \quad P_d(\mathbf{x}) = \sup\{P_{p_1, d}(\mathbf{x}) : p_1 \in I(d)\} \quad (2.16)$$

정확신뢰구간의 상한과 하한은 식 (2.16)이 주어진 유의수준과 일치하는 d_L, d_U 가 된다. 즉 비조건부 정확신뢰구간의 하한과 상한은 다음 식 (2.17)을 만족한다.

$$EX : Q_{d_L}(\mathbf{x}) = a/2, \quad P_{d_U}(\mathbf{x}) = a/2 \quad (2.17)$$

이경우에 $\mathbf{x} = (n_1, 0)$ 일 때 $d_U = +1$ 이 되고 $\mathbf{x} = (0, n_2)$ 일 때 $d_L = -1$ 이다. 또한 $Q_d(\mathbf{x})$ 는 d 에 대한 단조증가함수이며 $P_d(\mathbf{x})$ 는 d 에 대한 단조감소함수이다. 여기서 $D(y)$ 의 이산성과 $p_1 \in I(d)$ 의 구간에서 최대값을 취하는 것에 의해 정확신뢰구간이 지나치게 넓게 추정됨을 짐작할 수 있다. 그러므로 EX 신뢰구간은 보수적인 신뢰구간이라고 할 수 있다. StatXact(1997)은 d 의 정확신뢰구간을 구하기 위해 Berger와 Boos(1994)의 검정의 방법을 신뢰구간 문제에 적용하여 정확신뢰구간보다 덜 보수적인 신뢰구간을 구하고 있다. 이 방법의 핵심적인 사항은 $I(d)$ 의 구간을 재조정하는 것이다. 먼저 $I(d)$ 의 구간을 조정하기 위해 이항분포의 모비율 p_1, p_2 각각에 대해 $100(1-\beta)\%$ 의 정확신뢰구간을 구축한다. p_1 에 대한 정확신뢰구간을 $A_1 = [l_1, u_1]$ 그리고 p_2 에 대한 정확신뢰구간을 $A_2 = [l_2, u_2]$ 라고 하면 $A_1 \times A_2$ 의 모든 사건에 대해 위에서 언급한 Santner와 Snell(1980)의 방법을 다시 적용하는 것이다. 즉 p_1 에 대한 조정된 구간은 다음과 같다.

$$I_\beta(d) = \{p_1 : \max(l_1, l_2 + d) \leq p_1 \leq \min(u_1, u_2 + d)\}$$

여기서 $I_\beta(d) \subseteq I(d)$ 이다. 이제 주어진 구간에서

$$Q_d(\mathbf{x}, \beta) = \sup\{Q_{p_1, d}(\mathbf{x}) : p_1 \in I_\beta(d)\}, \quad P_d(\mathbf{x}, \beta) = \sup\{P_{p_1, d}(\mathbf{x}) : p_1 \in I_\beta(d)\}$$

라 두고 조정된 정확신뢰구간의 하한과 상한은 신뢰구간을 구축하는데 사용된 β 만큼의 유의수준을 조정한 다음의 식 (2.18)를 만족하는 d_L, d_U 로 결정된다.

$$BB : Q_{d_L}(\mathbf{x}) = \alpha/2 - \beta, \quad P_{d_U}(\mathbf{x}) = \alpha/2 - \beta \quad (2.18)$$

유의수준 조정에 있어서 중요한 역할을 하는 것은 β 의 값임을 알 수 있다. 만일 β 가 0이면 이는 전혀 조정하지 않은 Santner와 Snell(1980)의 정확신뢰구간과 같으며 β 를 늘리면 구간

을 줄이는 대신 조정된 유의수준을 작게하므로써 상대적으로 낮은 유의수준에서 보상을 받게되는 현상이 발생한다. 그러므로 최적의 β 선정은 어려운 일이라 할 수 있다. StatXact에서는 $\beta=0.001$ 을 사용하는데 본 연구에 의하면 두 집단의 표본의 크기가 비슷하거나 두 모확률의 차이가 크지 않은 경우에는 오히려 조정된 방법이 더 보수적으로 행동함을 알 수 있었다. 이는 조정된 방법의 유의수준이 조정하지 않은 유의수준보다 작은 값을 취하기 때문이다.

한편 정확신뢰구간을 계산하기 위해서는 많은 시간이 소유되는 단점이 있는데, 본 연구에서는 $p_1 \in I(d)$ 의 구간에서 최대값을 찾는 알고리즘으로 golden section search방법과 가상수정위치방법을 사용하였다. 특히 $n_1 = n_2$ 인 상황에서는 Buhrman(1977)의 방법에 따라 이항분포의 정확신뢰구간을 구축하는 Clopper-Pearson 방법을 이용하여 식 (2.17)을 빠르고 간편하게 구할 수 있다.

2.1.6. 조건부 정확신뢰구간(conditional exact confidence interval)

Thomas와 Gart(1977)는 두 집단의 성공의 주변합이 고정된 가정하에서 조건부 정확신뢰구간을 제시하였다. $m = x_1 + x_2$ 가 주어질 때 x_1 의 조건부분포는 다음과 같다.

$$f(x_1|m; \psi) = \frac{\binom{n_1}{x_1} \binom{n_2}{m-x_1} \psi^{x_1}}{\sum_{i=0}^m \binom{n_1}{i} \binom{n_2}{m-i} \psi^i}, \quad x_1 = 0, 1, \dots, m \tag{2.19}$$

여기서 $\psi = [p_1(1 - p_2)]/[p_2(1 - p_1)]$ 인 오즈비를 나타낸다. 또한 방정식 $\sum_{i=0}^{x_1} f(i|m; \psi_U) = \alpha/2$ 와 $\sum_{i=x_1}^m f(i|m; \psi_L) = \alpha/2$ 를 만족하는 ψ_U 와 ψ_L 을 오즈비 ψ 의 $100(1 - \alpha)\%$ 정확신뢰구간이라 두면 Thomas와 Gart(1977)의 d 에 대한 조건부정확신뢰구간의 하한과 상한은 다음 식 (2.20)에 주어진 d_L, d_U 와 같다.

$$TG : d_L = p_1^L - p_2^U, \quad d_U = p_1^U - p_2^L \tag{2.20}$$

이때 (p_1^U, p_2^L) 과 (p_1^L, p_2^U) 은 다음의 이차방정식에서 각각 유도된다.

$$\psi_U = \frac{p_1^U(1 - p_2^L)}{(1 - p_1^U)p_2^L}, \quad n_1p_1^U + n_2p_2^L = m ; \quad \psi_L = \frac{p_1^L(1 - p_2^U)}{(1 - p_1^L)p_2^U}, \quad n_1p_1^L + n_2p_2^U = m$$

이와 같은 조건부정확신뢰구간에 대해 Santner와 Snell(1980)은 다음 식 (2.21)과 같은 수정된 조건부신뢰구간을 제시하였다.

$$SS : d_L = (\psi_L^{1/2} - 1)/(\psi_L^{1/2} + 1), \quad d_U = (\psi_U^{1/2} - 1)/(\psi_U^{1/2} + 1), \tag{2.21}$$

여기서 $0 \leq \psi_L \leq 1, 1 \leq \psi_U \leq \infty$ 이고 오즈비가 각각의 범위를 벗어나면 신뢰구간의 상한과 하한은 0이다. Santner와 Snell(1980)의 신뢰구간은 오즈비의 정확한 신뢰구간을 구할 수 있으면 쉽게 활용할 수 있는 장점이 있으며 $n_1 = n_2 = m$ 인 경우에는 TG 신뢰구간과 이론상 일치하게 된다. 그런데 오즈비의 범위에 따라 TG 신뢰구간과 큰 차이점을 두고 있다. 즉 Santner와 Snell(1980)의 조건부 신뢰구간은 $\psi_L < 1$ 혹은 $\psi_U < 1$ 인 경우에는 극단적인 추정을 하기 때문에 실제적인 활용에는 너무 보수적인 경향을 띤다고 하겠다.

3. 비교방법

3.1. 포함확률

모든 가능한 표본 $X_1 = x_1, X_2 = x_2$ 에서 신뢰구간을 구축하면 주어진 모수 d 를 포함할 실제포함확률(actual coverage probability)을 계산할 수 있다. 주어진 모비율 p_1, p_2 에서 실제포함확률은 다음과 같다.

$$ACP(\alpha) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} I(x_1, x_2, 1 - \alpha) \times f_{p_1, d}(x_1, x_2). \quad (3.1)$$

이때 $I(x_1, x_2, 1 - \alpha)$ 는 $X_1 = x_1, X_2 = x_2$ 에서 $100(1 - \alpha)\%$ 신뢰구간이 주어진 모수 d 를 포함하면 1, 그렇지 않으면 0이 되는 지시함수이다. 또한 임의의 모비율 p_1, p_2 에 대한 평균적 수행능력을 평가하기 위해 모비율 p_1, p_2 에 사전분포 $g(p_1, p_2)$ 를 가정하여

$$\overline{ACP}(\alpha) = \int_{\Omega} ACP(\alpha)g(p_1, p_2)dp_1dp_2$$

를 추정할 수 있다. 우리는 이를 평균포함확률(average coverage probability)이라 할 수 있다. 이는 모비율의 차 d 가 어떠한 값으로 주어지던 포함확률에 대한 평균적 능력을 평가하는 지표라고 하겠다. Agresti와 Coull(1998)은 이항분포 신뢰구간의 비교에 있어 위와 같은 개념의 평균포함확률을 활용한 바 있다. 또한 얼마나 명목신뢰수준이 실제포함확률이 떨어져 있는가를 측정하기 위해 다음과 같은 $RMSE$ 를 계산하였다.

$$RMSE = \sqrt{\int_{\Omega} (ACP(\alpha) - \alpha)^2 dp_1dp_2}.$$

3.2. 신뢰구간의 기대 폭

신뢰구간을 비교함에 있어서 동일한 유의수준에서는 신뢰구간의 상한과 하한의 차이인 신뢰구간의 폭이 좁을수록 바람직하다고 할 수 있다. 본 연구에서는 신뢰구간의 기대 폭(expected widths)을 다음 식 (3.2)와 같이 계산하였다.

$$E_d(w) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} W(x_1, x_2, 1 - \alpha) \times f_{p_1, d}(x_1, x_2) \quad (3.2)$$

여기서 $W(x_1, x_2, 1 - \alpha)$ 는 $X_1 = x_1, X_2 = x_2$ 일 때 $100(1 - \alpha)\%$ 신뢰구간의 폭을 나타낸다.

3.3. 신뢰구간의 왜도성

모수를 포함하지 않은 신뢰구간이 모수의 위나 아래에 균일하게 분포한다면 신뢰구간이 상당히 안정적이라고 할 수 있다. 이제 $R_1 = P[d_L \geq d]$, $R_2 = P[d_U \leq d]$ 라 두면 R_1 은 구축된 신뢰구간이 주어진 모수 d 보다 위에 구축될 확률, R_2 는 아래에 구축될 확률을 나타

표 4.1: 16가지 신뢰구간의 모의실험결과
(신뢰수준= 95%, $n_1 = n_2$, 기대폭= 평균기대폭, $\overline{SM}=SM$ 의 평균)

n1(n2)	10(10)			20(20)			50(50)		
	방법	ACP(RMSE)	기대폭	SM	ACP(RMSE)	기대폭	SM	ACP(RMSE)	기대폭
CL	.894(.086)	.653	.063	.925(.042)	.484	.032	.941(.015)	.313	.016
CC	.960(.039)	.853	.035	.967(.027)	.581	.026	.967(.019)	.353	.020
AN	.898(.079)	.596	.070	.923(.044)	.457	.044	.939(.022)	.305	.025
ME	.953(.015)	.669	.017	.948(.010)	.485	.011	.949(.006)	.313	.006
MN	.959(.017)	.686	.017	.951(.009)	.491	.010	.950(.006)	.315	.006
GN	.953(.016)	.683	.015	.949(.010)	.491	.008	.949(.006)	.309	.004
PR	.935(.024)	.657	.022	.942(.013)	.483	.011	.941(.049)	.317	.011
SO	.953(.018)	.656	.029	.952(.013)	.482	.021	.952(.007)	.312	.013
SC	.979(.031)	.759	.032	.974(.026)	.542	.027	.968(.019)	.339	.019
HA	.928(.056)	.626	.044	.942(.025)	.471	.024	.947(.008)	.309	.013
JP	.950(.016)	.649	.027	.949(.010)	.478	.018	.950(.006)	.311	.011
BT	.938(.072)	.709	.050	.969(.048)	.535	.032	.973(.030)	.326	.012
EX	.983(.035)	.823	.033	.979(.031)	.587	.029	.975(.028)	.369	.025
BB	.984(.036)	.826	.034	.980(.032)	.583	.030	.974(.026)	.359	.024
TG	.869(.151)	.624	.124	.893(.121)	.463	.087	.905(.103)	.299	.060
SS	.991(.041)	1.005	.041	.990(.041)	.816	.040	.990(.041)	.650	.040

낸다. 그러므로 R_1 과 R_2 에 대한 추정량으로 각각 다음을 생각하자.

$$R_1 = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} I_L(d_L \geq d, 1 - \alpha) \times f_{p_1, d}(x_1, x_2)$$

$$R_2 = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} I_U(d_U \geq d, 1 - \alpha) \times f_{p_1, d}(x_1, x_2).$$

여기서 $I_L(d_L \geq d, 1 - \alpha)$ 는 $X_1 = x_1, X_2 = x_2$ 에서 $100(1 - \alpha)\%$ 신뢰구간의 상한이 d 보다 위에 구축되면 1, 그렇지 않으면 0이 되는 지시함수이다. 우리는 $R_1 + R_2$ 의 확률이 주어진 유의수준 α 와 일치하면서 동시에 R_1, R_2 가 각각 $\alpha/2$ 에 근사하기를 기대한다. 그러나 Gart와 Nam(1990)에 의하면 확률이 양극단에서는 어느 한쪽의 R_i 값이 지나치게 커지는 경향이 있음을 지적하였다. 그러므로 우리는 다음의 SM (skewness measure) 통계량을 측정하여 신뢰구간의 성질을 파악하였다.

$$SM = |\alpha/2 - R_1| + |\alpha/2 - R_2| \tag{3.3}$$

SM 이 0에 가까운 신뢰구간이 좋은 신뢰구간이라 할 수 있다.

4. 모의실험 결과

본 절에서는 모비율의 차이에 대한 앞에서 언급된 16가지 방법에 대한 성능 및 특징을 모의실험을 통해 비교하고자 한다.

4.1. 평균포함확률

표4.1과 표4.2는 각각 표본수가 같을때와 다를때의 95% 신뢰수준에 대한 평균포함확률(\overline{ACP})과 $RMSE$ 를 보여주고 있다. 평균포함확률을 계산하기 위해서는 주어진 p_1, p_2 에서 실제포함확률(ACP)을 먼저 계산하여야 한다. 예를 들어 $n_1(n_2)$ 가 10(10)인 경우 주어진 p_1, p_2 에서 $x_1 = 0$ 부터 $x_1 = n_1 (= 10)$ 까지 가능한 11개와 마찬가지로 x_2 에 대해서도 11개의 값을 고려한 총 121(11×11)개의 신뢰구간을 통해 실제포함확률 식 (3.1)을 계산한다. 그리고 이를 가능한 모든 모수공간 p_1, p_2 에서 반복하여 실제포함확률의 평균으로 평균포함확률을 구축하였다. 우리는 여기서 p_1, p_2 에 구간 [0,1] 사이의 균일분포를 가정하여 모수 p_1, p_2 에 대해 0.01의 간격으로 격자점을 형성하여 실제포함확률을 계산하였다. 그러므로 모든 가능한 모수 p_1, p_2 에 대해 식 (3.1)의 실제포함확률은 총 10,201개의 격자점에서 유도되어진다. 여기서 모수 d 는 0을 중심으로 하는 삼각형분포를 따른다. 붓스트랩 신뢰구간(BT)을 구축하기 위해서는 B=1000회의 반복을 실시하였다.

표4.1과 표4.2의 \overline{ACP} 를 통해 Wald 유형의 신뢰구간은 주어진 명목수준을 하향추정하는 반면 정확신뢰구간은 상향추정을 하는 보수적인 신뢰구간임을 볼 수 있다. 정확신뢰구간의 수정방법인 BB 방법 역시 $\beta = 0.001$ 수준에서 정확신뢰구간과 큰 변화가 없음을 알 수 있다. 그리고 조건부 정확신뢰구간인 SS 방법은 과대추정을 TG 방법은 과소추정을 하고 있다. 이에 반해 Mee의 방법(ME), Skewness 조정방법(GN), Miettinen과 Nurminen 방법(MN), Jeffrey-Perk 방법(JP) 그리고 Score 방법(SO) 등 근사이론을 이용한 방법의 신뢰수준이 명목수준에 근사함을 볼 수 있다. $RMSE$ 값에 의하면 $n_1 = n_2$ 인 경우 대체적으로 ME, GN, MN과 JP 방법이 유리하며, $n_1 \neq n_2$ 인 경우에 SO 방법이 상대적으로 유리함을 알 수 있다. Skewness 조정방법은 ME 방법에 비해 상대적으로 많은 계산이 요구됨에도 불구하고 $RMSE$ 측면에서는 다소 효율성이 떨어진다고 볼 수 있다.

그림4.1은 $0 \leq p_1 \leq 1, 0 \leq p_2 \leq 1$ 인 10,201개의 격자점에서 95% 신뢰수준에 대한 실제포함확률(ACP)의 등고선 그림이다. 본 연구에서는 16가지 방법중 대표적인 4가지 방법(CL, ME, BT, EX)의 등고선 그림만 제시하였다. 이 등고선 그림은 $p_1 - p_2 = k$ (기울기 45도의 직선)에서 대체적으로 같은 수준의 포함확률을, $p_1 + p_2 = k$ 에서 실제포함확률이 다양한 변화를 보이고 있다. 그러므로 우리는 포함확률의 다양한 변화를 파악하기 위해 $p_1 + p_2 = 1$ 에서 포함확률을 측정하였다. 그림4.2는 95% 신뢰수준에 대한 $p_1 + p_2 = 1$ 이고 $n_1(n_2) = 20$ 일때의 16가지 방법의 실제포함확률을 나타낸다. 그림4.2를 통하여 각 방법들의 특징을 살펴보자. 우선 정규근사에 의한 Wald 방법(CL)은 톱날모양을 하면서 하향추정을 하고 있다. 이와 비슷한 하향추정의 경향을 보이는 것은 카이제곱분포에 의한 근사방법인 Anbar 방법(AN)이다. AN 방법은 모비율이 0 이나 1에 가까울수록 Wald 방법이 지니는 급격한 하향추정문제는 어느 정도 해결하고 있음을 볼 수 있다. Wald 방법(CL)에 비해 정

표 4.2: 16가지 신뢰구간의 모의실험결과
(신뢰수준= 95%, $n_1 \neq n_2$, 기대폭= 평균기대폭, $\overline{SM}=SM$ 의 평균)

방법	10(20)			10(30)			20(50)		
	$ACP(RMSE)$	기대폭	SM	$ACP(RMSE)$	기대폭	SM	$ACP(RMSE)$	기대폭	SM
CL	.902(.070)	.574	.057	.897(.072)	.540	.064	.925(.042)	.484	.032
CC	.960(.031)	.716	.030	.955(.030)	.667	.032	.967(.027)	.581	.026
AN	.897(.076)	.540	.065	.891(.080)	.516	.070	.923(.044)	.457	.044
ME	.953(.012)	.583	.018	.955(.011)	.548	.019	.948(.010)	.485	.011
MN	.957(.013)	.593	.017	.958(.012)	.554	.018	.951(.009)	.491	.010
GN	.952(.029)	.561	.014	.955(.024)	.531	.012	.949(.010)	.491	.008
PR	.939(.018)	.575	.017	.939(.027)	.542	.018	.942(.013)	.483	.011
SO	.953(.013)	.574	.026	.953(.010)	.538	.026	.952(.013)	.482	.021
SC	.978(.029)	.657	.030	.977(.028)	.615	.029	.974(.026)	.542	.027
HA	.939(.030)	.557	.030	.941(.026)	.527	.026	.942(.025)	.471	.024
JP	.951(.013)	.573	.021	.952(.014)	.542	.019	.949(.010)	.478	.018
BT	.959(.056)	.667	.047	.962(.053)	.667	.042	.961(.031)	.454	.025
EX	.980(.033)	.710	.030	.979(.032)	.672	.029	.974(.029)	.480	.024
BB	.981(.033)	.702	.031	.980(.032)	.652	.030	.974(.027)	.465	.024
TG	.886(.134)	.545	.103	.899(.121)	.522	.090	.893(.121)	.463	.087
SS	.990(.041)	.919	.040	.990(.041)	.888	.040	.990(.041)	.816	.040

확신뢰구간인 EX 나 EX 방법의 보정방법인 BB, SS, CC 그리고 SC 방법은 상향추정하고 있음을 볼 수 있다. 즉 BB 방법은 EX에 비해 계산상 많은 시간이 소요됨에도 불구하고 포함확률에서는 큰 차이가 없음을 알 수 있다. 이를 통해보면 연속성 수정방법은 정규근사의 하향추정을 상당히 보정하여 주나 과도하게 보정하고 있음을 알 수 있다. 정확신뢰구간 중 TG 방법은 하향추정을 SS 방법은 상향추정을 하는데 이는 조건부신뢰구간으로 m 이 고정된 상황에서 사용하는 신뢰구간이기에 우리의 모의실험 상황에는 적합하지 못하다고 할 수 있으나 대체적으로 문제점을 지니고 있음을 볼 수 있다.

ME 방법, MN 방법, JP 방법 그리고 HA 방법은 포함확률에 있어서 비슷한 형태를 보이며 상당히 안정적임을 볼 수 있다. 그러나 θ 의 사후분포의 평균을 이용하는 베이지안 방법인 JP 와 HA 방법은 표본의 크기가 10(10)등 표본의 크기가 작을 때는 모수가 0이나 1 근처에서 급격한 하향추정의 경향이 있으나 20(20)에서는 상당부분 ME 방법과 포함확률이 유사함을 볼 수 있었다. 또한 ME 방법의 보정방법인 GN 방법과 PR 방법은 그 형태가 유사함을 볼 수 있다. 그러나 포함확률측면에서 GN 방법은 ME 방법에 비해 큰 보완이 되지 못하고 있으며 PR 방법은 약간의 하향추정 경향이 있음을 볼 수 있다. 이제 붓스트랩 신뢰구간의 특징을 살펴보자. 평균포함확률 측면에서 붓스트랩 방법이 항상 잘 작용한다고 할 수는 없으며 오히려 다소 과대 추정하는 경향이 있음을 발견할 수 있다. 그 이유는 붓스트랩 신뢰구간은 $p_1(p_2)$ 가 0이나 1에 근접해 있는 경우와 표본의 크기가 작은 경우 자료의 희소성 때문에 신뢰구간 구축이 어려우며, 구축한다고 하더라도 신뢰구간의 폭이 너무 넓어지게 되는 단점이 있기 때문이다. 그러므로 자료가 희소한 경우에는 붓스트랩 방법은 치명

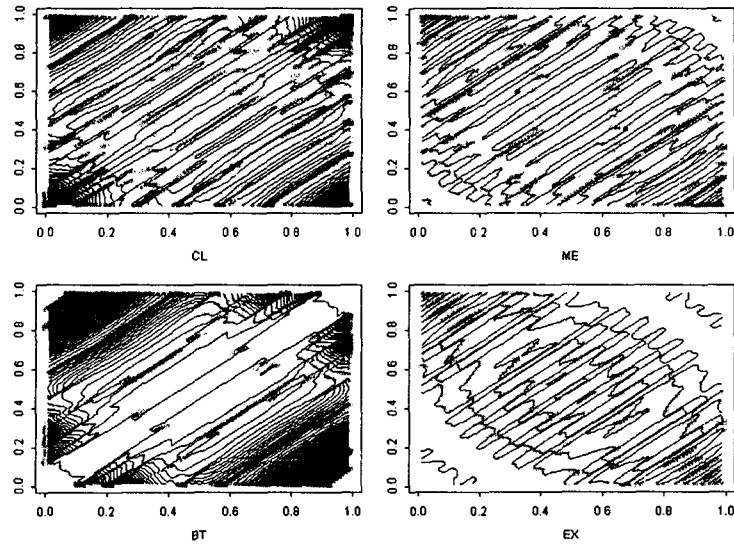


그림 4.1: 4가지 방법의 실제포함확률(ACP)등고선 그림
(가로축 p_1 , 세로축 p_2 , $n_1 = n_2 = 10$)

적인 한계성을 지니고 있다고 할 수 있다. 그러므로 표4.1과 표4.2의 평균포함확률에서 보듯 붓스트랩 신뢰구간은 다소 과대추정하는 경향이 있음을 볼 수 있었다. 하지만 붓스트랩 방법은 정규근사를 이용한 통계량을 추측통계량으로 사용하여 신뢰구간을 구축한다는 점에서 정규근사 방법이 지니는 태생적인 약점을 충분히 상쇄해준다고 할 수 있겠다.

$n_1(n_2) = 50$ 일 때 $p_1 + p_2 = 1$ 에서 실제포함확률은 제시하지 않았으나 실험결과 실제 포함확률의 변동 폭이 급격히 줄어들음을 볼 수 있었다. 표본의 크기가 50(50)일 때 EX, BB, CC, SC 그리고 BT 방법의 실제포함확률은 p_1 이 0 이나 1 근처를 제외하곤 거의 비슷한 형태를 띠고 있었다. 또한 ME, MN, JP, HA 그리고 SO 방법이 비슷하며 GN, PR 방법이 서로 유사한 형태임을 볼 수 있었다. 그러므로 표본의 크기가 어느 정도 커지면 반복계산에 의존하는 방법보다 상대적으로 계산이 간단한 JP나 HA 방법 그리고 SO방법을 사용하는 것을 추천할 수 있겠다.

4.2. 신뢰구간의 기대폭

좋은 신뢰구간을 선택하는 하나의 방법으로 주어진 신뢰수준에서 신뢰구간이 너무 넓지 않기를 바랄 것이다. 그러므로 4.1절의 포함확률의 그림을 통한 좋은 신뢰구간을 선택하는 기준의 하나로 신뢰구간의 기대폭을 생각할 수 있다. 앞의 표4.1과 표4.2는 $0 < p_1 < 1, 0 < p_2 < 1$ 인 상황에서 다양한 표본의 크기에 대한 신뢰구간의 평균 기대폭을 보여주고 있다. 여기서 신뢰구간의 평균 기대폭이 가장 넓은 것은 SS 방법이다. 이는 조건부 신뢰구

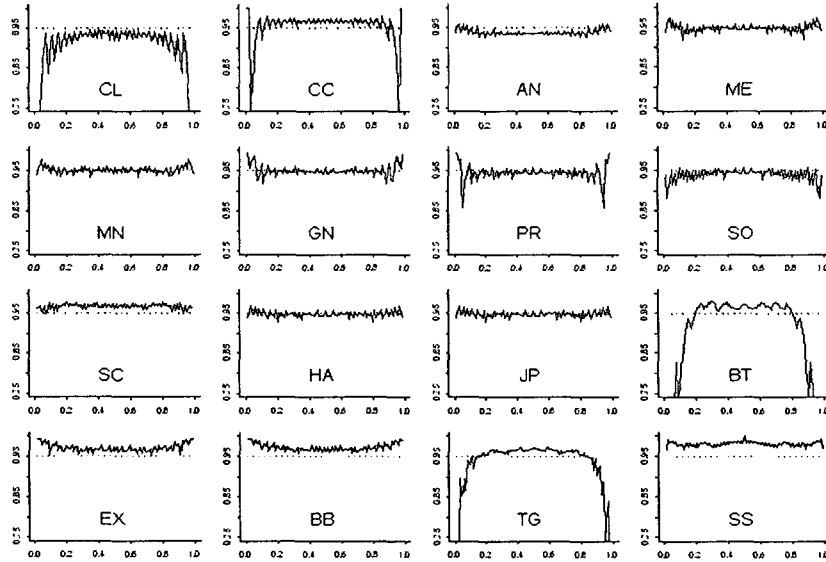


그림 4.2: 16가지 방법의 실제포함확률(*ACP*)
 $(p_1 + p_2 = 1, n_1 = n_2 = 20)$

간이므로 m 을 고정시키지 않은 우리의 모의실험 상황에 영향을 받을 수도 있으나 또한 TG 신뢰구간은 오즈비에 크기에 영향을 받기 때문에 이를 수정한 SS 신뢰구간은 소표본의 아주 좋은 조건($\psi_L < 1, \psi > 1$ 인 경우)이외에서는 적용하기가 어렵다는 것을 알 수 있다. 반대의 현상은 TG 신뢰구간에서 나타난다.

정확신뢰구간의 수정방법인 BB방법은 신뢰구간의 기대폭에서 상당한 효과가 있음을 볼 수 있다. 신뢰구간의 기대폭은 표본의 크기가 10(10)인 경우를 제외하고 정확신뢰구간(EX)보다 기대폭이 작음을 볼 수 있다. 이는 두 방법이 거의 같은 수준의 포함확률을 유지하나 신뢰구간의 폭에 있어서는 BB 방법이 상대적으로 좁다는 것을 의미한다. 그리고 두 집단의 표본의 크기가 차이가 날수록 BB 방법의 신뢰구간의 기대폭이 정확신뢰구간의 기대폭보다 더 좁아진다. 이와 같은 현상은 $x_1(x_2)$ 이 0(n_2)에 근접할수록 심해진다. 예를 들어 $n_1(n_2)$ 가 10(50)일 때 $x_1(x_2)$ 가 1(1)인 경우 정확신뢰구간(EX)은 (-0.2672, 0.4195)이나 수정된 정확신뢰구간(BB)은 (-0.1309, 0.4182)로 하한의 신뢰구간에 상당한 수정이 가해진다. 여기서 정확신뢰구간은 $\hat{d} = 0.08$ 을 중심으로 하지만 BB 신뢰구간은 \hat{d} 을 중심으로 하지 않는다는 점이다. 이는 $I_\beta(d)$ 의 선택에 영향을 받는데 BB 신뢰구간은 표본의 크기가 작은 집단의 확률에 가중치를 더 두는 방법임을 알 수 있다.

명목신뢰수준에 근접한 ME, GN, MN, JP 그리고 SO 신뢰구간 중 기대폭이 가장 짧은 신뢰구간은 GN 신뢰구간이다. 그런데 GN 신뢰구간은 표본의 크기가 커질수록 모비율 p_1, p_2 가 0이나 1 근처에서 급격한 하향추정을 보이는 경향이 있다. 그러므로 급격한 하향추

정이 없는 신뢰구간 중 기대폭이 짧은 신뢰구간을 선정한다면 ME, JP, HA 그리고 SO 신뢰구간을 들 수 있다. HA 신뢰구간은 표본의 크기가 다를 경우에 가장 짧은 기대폭을 유지하나 이는 표본의 크기가 다를 때 하향추정하는 경향에 따르기 때문이다. 그러므로 우리는 표본의 크기가 다른 경우나 같은 경우에도 잘 작동하는 JP 와 SO 신뢰구간을 신뢰구간의 폭 기준에서 좋은 신뢰구간으로 추천하고자 한다.

4.3. 왜도성

표4.1과 표4.2는 앞에서 언급한 SM 의 평균을 보여주고 있다. 우리는 SM 이 작은 방법은 모수 d 를 포함하지 않을 확률이 α 에 근사하며 각각 오분류된 신뢰구간의 하한과 상한이 $\alpha/2$ 에 균등하게 분배되어 있음을 의미한다고 하겠다. SM 관점에서 가장 좋은 성질을 보여주는 신뢰구간은 Gart와 Nam(1990)의 GN 신뢰구간이다. GN 신뢰구간은 대부분의 표본에서 가장 작은 SM 을 유지하고 있다. 이는 GN 신뢰구간이 지니는 성질을 잘 반영해 주는 결과라 하겠다. 그런데 SM 이 큰 EX, BB, CL, CC, SC 그리고 BT 등의 신뢰구간은 그 자체가 왜도성을 지니고 있어서가 아니라 주어진 명목수준을 과대추정하거나 과소추정하는 것에 기인한다고 할 수 있다. 실제적으로 대부분의 신뢰구간은 양극단의 표본에서만 GN 신뢰구간에 비해 다소 어느 한쪽 방향으로 치우치는 경향을 보이고 있으므로 SM 의 평균에는 그 반영비율이 전체적으로 미비하다고 할 수 있다. 결론적으로 SM 기준에서는 두 집단의 x_1, x_2 가 크게 차이가 나는 경우 GN 신뢰구간이 비교적 우수한 신뢰구간이라고 하겠다.

5. 토의와 결론

모비율의 신뢰구간구축에서 Vollset(1993), Agresti와 Coull(1998), Newcombe(1998a) 등은 Wald 방법이 지니는 문제점을 지적하였고 그 대안으로 Score 방법들이 주로 추천되어 왔다. 본 연구와 같은 모비율의 차 $p_1 - p_2$ 에 대한 신뢰구간의 연구로 Miettinen과 Nurminen(1985), Beal(1987), Gart와 Nam(1990) 그리고 Newcombe(1998b)등을 들 수 있다. Beal(1987)은 복잡하며 보다 정교한 신뢰구간을 원한다면 ME나 MN 신뢰구간을, 보다 빠르고 간단한 계산을 원한다면 JP나 HA 신뢰구간을 추천하였다. Newcombe(1998b)는 PR 신뢰구간과 SO 신뢰구간을 추천하였다.

이에 본 연구에서는 이들의 연구에 기초하여 여러 신뢰구간들의 성질을 평균포함확률, 구간기대폭, 왜도성 측면에서 비교하였다. 특히 모수에 대한 사전분포를 가정하여 여러 신뢰구간들이 지니는 특징을 살펴보았다. 본 연구를 통해 이항분포에서처럼 정확신뢰구간이 지니는 보수성을 확인 할 수 있었고 Wald 신뢰구간(CL)이 지니는 하향추정의 문제점도 확인 할 수 있었다. 그리고 Berger와 Boos(1993)의 정확신뢰구간의 수정방법(BB) 역시 보수성에 대한 큰 개선이 없음을 발견할 수 있었다. 물론 이산형 자료에서 표본의 크기가 작은 경우 정확추론은 매우 중요한 역할을 한다는 점을 간과할 수 없다. 그런데 표본의 크기가 적당히 크면 근사이론에 기초한 통계적 추론이 경우에 따라 정확하게 작용한다는 점이다. Agresti와 Coull(1998)의 결과에서 단일모수 p 에 대한 신뢰구간으로 가장 좋은 방법으로 알려진 Score 방법과 유사한 방법은 $p_1 - p_2$ 의 신뢰구간에서는 Mee(1984)의 신뢰구간(ME)

이라 할 수 있다. GN, PR 그리고 MN 신뢰구간은 ME 신뢰구간에서 파생된 변형 Mee 방법이라 할 수 있는데 이들 방법은 상당한 반복계산을 요구한다고 하겠다. 이에 반해 상대적으로 간단한 계산방법인 JP와 HA 신뢰구간은 사전확률분포를 이용하여 $p_1 + p_2$ 의 사후평균을 신뢰구간 구축에 활용 하는 베이지안 신뢰구간이라고 할 수 있다. 이들 방법은 정확신뢰구간(EX, BB)이나 정규근사에 의존하는 방법(CL, AN)에 비해 상당히 정확하며 Mee 방법들과 유사한 결과를 얻는다는 장점을 지니고 있다고 할 수 있다. 또한 Woodroofe와 Jhun(1989)의 결과에 의해 평균포함확률 측면에서 붓스트랩 신뢰구간이 이산형 분포에서도 CL 방법에 대한 대안으로 활용될 수 있다는 점을 주목하여야 한다.

본 연구를 통해 다음의 몇가지 결과와 제안을 제시하고자 한다. 첫째, 정규근사를 이용하는 Wald 방법(CL)이나 카이제곱분포에 근사하는 Anbar 방법(AN)은 포함확률을 하향 추정하는 경향이 있다. 둘째, 정확방법(EX), 정확방법의 수정방법(BB), 연속형 수정방법(CC) 그리고 Score 수정방법(SC)등은 상향추정한다. 그러므로 특별히 연속형 수정을 가할 필요는 없다고 할 수 있다. 특히 조건부 정확신뢰 구간은 해당 조건이 벗어나면 심하게 과대추정 혹은 과소추정하는 경향이 있다. 셋째, 비교적 정확하고 간단한 계산에 의해 표본의 크기가 서로 다른 경우에도 효과적인 결과를 얻을 수 있는 방법으로 베이지안 방법인 JP 나 HA 방법 그리고 SO 방법을 사용하길 권한다. 특히 JP 방법이 HA 방법보다 우수하다고 할 수 있다. 넷째, 표본의 비 차이가 극단적인 경우 GN 신뢰구간이 계산의 어려움은 있으나 비교적 정확하다고 할 수 있다. 다섯째, 붓스트랩 방법은 Mee 방법(ME, SK, PR, MN)들이나 베이지안 방법(JP, HA)들이 지니는 이론적 어려움을 피하여 주며 Wald 방법(CL)의 훌륭한 대안이 될 수 있다는 점이다. 특히 $p_1(p_2)$ 이 0.1 이상일때 붓스트랩 신뢰구간은 정확신뢰구간의 특징을 보인다고 하겠다.

참고문헌

- [1] Agresti A. and Coull B. (1998) Approximate is better than "Exact" for interval estimation of binomial proportions. *The American Statistician*, 52, 119-126
- [2] Anbar, D.(1983) On estimating the difference between two probabilities with special reference to clinical trials. *Biometrics*, 39, 257-262.
- [3] Barnard G.A. (1945) A new test for 2x2 tables, *Nature*, 156-177.
- [4] Beal S.L.(1987) Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples, *Biometrics*, 43, 941-950
- [5] Berger R. and Boos D. (1994) p values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89, 1012-1016.
- [6] Blyth C. and Still H. (1983) Binomial Confidence Intervals. *Journal of the American Statistical Association*. 78. 108-116.

- [7] Buhrman, J.M (1977) Tests and confidence intervals for the difference and ratio of two probabilities. *Biometrika*, 64, 160-162.
- [8] Efron, B.(1979) Bootstrap Methods: Another look at the Jackknife. *The annals of Statistics*, 7, 1-26
- [9] Gart, J.J and Nam, J.M (1990) Approximate interval estimation of the difference in binomial parameters: correction for skewness and extension to multiple tables, *Biometrics*, 46, 637-643.
- [10] Haldane, J.B.S. (1945) On a method of estimating frequencies. *Biometrika*, 33, 222-225.
- [11] Mee, R.(1984) Confidence bounds for the difference between two probabilities. *Biometrics*, 40, 1175-1176.
- [12] Miettinen, O. and Nurminen, M.(1985) Comparative analysis of two rates. *Statistics in Medicine*, 4, 213-226
- [13] Newcombe(1998a) Two-sided confidence intervals for the single proportion : Comparison of seven methods, *Statistics in Medicine*, 17, 857-872
- [14] Newcombe(1998b) Interval estimation for the difference between independent proportions :Comparison of eleven methods, *Statistics in Medicine*, 17, 873-890
- [15] Santner, T.J. and Snell, M.K(1980) Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 contingency tables, *Journal of the American Statistical Association*, 75, 386-394.
- [16] StatXact (1997) *A Statistical Package for Exact Nonparametric Inference*, Cytel Software Corporation, 137 Erie Street, Cambridge, MA.
- [17] Thomas D.G. and Gart J.J. (1977) A Table of Exact Confidence Limits for Differences and Ratios of Two Proportions and Their Odds Ratios, *Journal of the American Statistical Association*, 72, 73-76
- [18] Vollset, S.E. (1993) Confidence intervals for a binomial proportion, *Statistics in Medicine*, 12, 809-824
- [19] Woodroffe, M.and Jhun, M.(1989) Singh's theorem in the lattice case. *Statistics and Probability Letters*, 7, 201-205.

[2003년 2월 접수, 2003년 6월 채택]

A Comparison of Confidence Intervals for the Difference of Proportions*

Hyeong Chul Jeong ¹⁾ Myoungshic Jhun ²⁾ Daehak Kim ³⁾

ABSTRACT

Several confidence interval estimates for the difference of two binomial proportions were introduced. Bootstrap confidence interval is also suggested. We examined the over estimation property of approximate intervals and under estimation trend of exact intervals for the difference of proportions. We compared these confidence intervals based on the average coverage probability, expected width and skewness measure. Particularly actual coverage probability were calculated by using the prior distribution of parameters. Monte Carlo simulation for small sample size is conducted. Some interesting contour plots of average coverage probability and marginal plots for several interval estimates are presented.

Keywords: Population proportion; difference of proportions; confidence interval; exact confidence interval; bootstrap method; coverage probability; skewness

* This research was support by Catholic University of Daegu research grant in 2003.

1) Assistant Professor, Department of information Statistics, Pyongtaek University.

E-mail: jhc@ptuniv.ac.kr

2) Professor, Department of Statistics, Korea University.

E-mail: jhun@mail.korea.ac.kr

3) Professor, Department of Statistical Information, Catholic University of Daegu.

E-mail: dhkim@cu.ac.kr