

독립표본에서 두 모비율의 차이에 대한 가중 POLYA 사후분포 신뢰구간*

이승천¹⁾

요약

모비율 차이의 구간 추정에서 표준으로 인식되고 있는 Wald 신뢰구간은 모비율 구간 추정과 마찬가지로 포함확률의 근사성에서 문제가 있다는 것이 알려져 있다. 이에 대한 대안으로 모비율 차이의 신뢰구간에 대한 많은 연구가 있어 왔으나 대부분의 신뢰구간은 매우 복잡한 과정을 통해 얻어지게 되어 있어 실용성에 대한 문제가 제기될 수 있다. 이와 비교하여 Agresti와 Caffo(2000)에 의해 제시된 신뢰구간은 매우 간편한 식에 의해 구할 수 있어 이해하기 쉽고 포함확률과 포함확률의 평균절대오차에 있어 다른 복잡한 신뢰구간과 필적할 수 있다. 그러나 Agresti-Caffo 신뢰구간은 포함확률이 명목 신뢰수준을 상회하는 보수적인 구간으로 알려져 있다. 본 논문에서는 이승천(2005)에서 이항비율의 신뢰구간을 구하기 위해 사용된 가중 Polya 사후분포를 이용하여 두 모비율 차이의 신뢰구간을 구하였다. 이렇게 구하여진 신뢰구간은 간편성은 물론 Agresti-Caffo 신뢰구간의 보수성을 개선하였다.

주요용어: 모비율 차이, 가중 Polya 사후분포, Wald 신뢰구간, Agresti-Caffo 신뢰구간, Hybrid score 신뢰구간, 근사 베이저안 신뢰구간

1. 서론

치료방법의 비교를 위한 임상실험과 같은 비교실험에서는 각 처리에서 치료율의 차이에 대해 관심을 갖게 되며, 이 경우 두 모비율 차이에 대한 구간추정이 중요한 통계적 분석방법이 된다. 일반적으로 모비율 차이의 신뢰구간을 구축함에 있어 근사적인 방법과 정확 신뢰구간을 구하는 두 가지 방법이 있다. 여기서 정확 신뢰구간은 모든 가능한 모수값에서 구한 포함확률의 최소값이 신뢰수준보다 큰 신뢰구간을 의미한다. 그러나 모비율 차이의 신뢰구간을 구할때 근간이 되는 이항분포는 이산형 확률분포로서 이산형 확률분포의 격자구조로 인하여 신뢰구간의 포함확률이 진동하는 것을 피할 수 없다. 그러므로 정확 신뢰구간과 같이 포함확률의 최소값이 신뢰수준보다 크려면 매우 보수적인 신뢰구간이 될 수 밖에 없어 문제가 될 수 있다. 일례로 Chan과 Zhang(1999)은 두 백신의 비교실험에 대한 예제에서 백신의 치료율이 각각 0.944 (17/18), 0.611 (11/18)으로 관찰되었을 때, 모비율 차이에 대한 추정값은 0.333이지만 Santner와 Snell(1980)의 95% 정확 신뢰구간은 (-0.019,

* 이 논문은 2006년도 한신대학교 학술연구비 지원에 의하여 연구되었음
1) (447-791) 경기도 오산시 양산동 411, 한신대학교 정보통계학과, 교수
E-mail: seung@hanshin.ac.kr

0.630)으로 추정되어 Santner-Snell 신뢰구간을 이용할 경우 두 백신의 치료율이 유의하게 다르지 않다는 결론을 얻을 수 밖에 없는 예제를 소개하였다.

이와 비교하여 Wald 신뢰구간과 같은 근사적인 방법은 포함확률이 평균적으로 신뢰수준에 근접한 신뢰구간을 구축하는 것으로 Brown 등(2001)는 $(1 - \alpha) \times 100\%$ 신뢰구간이라고 할 때 대부분의 사람들은 그 신뢰구간의 포함확률이 평균적으로 $(1 - \alpha) \times 100\%$ 에 가깝다고 인식한다는 점을 지적하였고, Agresti와 Coull (1998)도 이항비율의 신뢰구간에서 정확신뢰구간보다는 근사 신뢰구간이 상대적으로 효용성이 있다고 설명하고 있다. 앞의 예제에서 Wald 신뢰구간은 95% 유의수준에서 (0.049, 0.593)와 같이 구하여져 Santner-Snell 신뢰구간과는 다른 결론을 얻게 된다.

모비율 및 모비율 차이의 구간추정에서는 정규근사에 의한 Wald 신뢰구간이 표준으로 인식되고 있지만 최근의 연구 결과에 따르면 Wald 신뢰구간은 근사성에 있어 많은 문제가 있다. 즉, 모비율에 대한 Wald 신뢰구간은 표본크기가 큰 경우 유용한 것으로 알려져 있기 때문에 여러 기초 통계학 책에서는 Wald 신뢰구간을 사용할 수 있는 표본크기 n 에 대해 $np, n(1 - p) \geq 5$ (또는 10), 또는 $np(1 - p) \geq 5$ (또는 10) 등의 조건들이 제시되고 있으나 Brown 등(2001)에 의하면 이러한 조건들이 Wald 구간의 유용성을 보장하여 주지는 못한다고 한다. 즉, 표본크기가 매우 큰 경우라고 할지라도 Wald 신뢰구간의 포함확률은 신뢰수준과 상당한 차이가 있을 수 있다. 이러한 Wald 신뢰구간의 오류에 대해 Ghosh(1973), Blyth와 Still(1983), Vollet(1993), Newcombe(1998a), Agresti와 Coull(1998), Brown 등(2001, 2002) 등 여러 문헌에서 언급되고 있다.

모비율 차이의 구간추정에 있어서도 사정은 비슷하여 Newcombe(1998b), Agresti와 Caffo (2000) 등에서 Wald 신뢰구간의 근사성에 대한 문제점을 지적하고 있다. 이러한 Wald 신뢰구간의 문제점으로 인하여 Wilson(1927), Anbar(1983), Mee(1984), Agresti와 Coull(1998), Newcombe(1998a, 1998b), Agresti와 Caffo(2000), 정형철 등(2003), 이승천(2005) 등 많은 학자들에 의해 모비율 및 모비율 차이에 대한 새로운 신뢰구간이 발표되었다.

Brown 등(2001)는 모비율에 대한 신뢰구간에 대해 표본크기가 작은 경우($n < 40$)에는 Wilson(1927)의 score 신뢰구간과 Jefferys 사전분포에 의한 Bayesian 신뢰구간을 추천하였고, 표본크기가 큰 경우 ($n \geq 40$)는 Agresti와 Coull(1998)에 의한 신뢰구간을 추천하였다. 표본크기가 큰 경우 세 신뢰구간은 거의 비슷한 성질을 갖고 있지만 Agresti-Coull 신뢰구간이 추천된 이유는 Agresti-Coull 신뢰구간은 Wald 신뢰구간과 비슷한 식에 의해 구할 수 있는 계산의 간편성 때문이라고 한다. 한편 Newcombe(1998b)은 11개의 모비율 차이에 대한 신뢰구간의 성능을 비교하여 성능과 계산의 간편성을 이유로 Hybrid score 신뢰구간을 추천하였다.

모든 학자들이 동의하는 것은 아니지만 모비율 및 모비율 차이의 구간추정에서 오류가 많은 Wald 신뢰구간이 현재까지도 많이 사용되고 있는 이유를 생각하면 신뢰구간의 간편성이 매우 중요하다고 판단된다. 또 Brown 등(2001)와 그 논문의 대부분 토론자들도 계산의 간편성에 대한 중요성에 대해 언급한 바가 있다. 이러한 관점에서 모비율 차이의 구간추정에서 Agresti와 Caffo(2001)의 신뢰구간은 Wald 신뢰구간과 같은 계산의 간편성을 갖고 있으며 평균 포함확률의 근사성도 비교적 양호한 실용적인 신뢰구간으로 생각된다. 그

러나 Agresti-Caffo 신뢰구간은 평균 포함확률이 신뢰수준을 상회하는 보수적인 신뢰구간으로 알려져 있다.

본 논문에서는 이승천(2005)에서 모비율의 신뢰구간을 구하기 위하여 유도된 가중 Polya 사후분포를 이용하여 Agresti-Caffo 신뢰구간과 같은 간편성을 갖는 모비율 차에 대한 근사 신뢰구간을 구하였다. 여기서 구하여진 신뢰구간의 평균 포함확률은 명목 신뢰수준에 매우 근접하여 Agresti-Caffo 신뢰구간의 보수성을 개선하였다. 또 모든 표본크기에서 무리없이 사용될 수 있다는 것을 추천된 몇 개의 모비율 차이 신뢰구간과 비교하여 설명하기로 한다.

2. 가중 Polya 사후분포 신뢰구간과 비교 신뢰구간

Polya 사후분포는 주어진 데이터에 의한 미관측값들에 대한 예측 결합분포(predictive joint distribution)로 Feller(1968)에 잘 기술되어 있다. 예측분포로서의 타당성 등 보다 자세한 내용은 Ghosh와 Meeden(1998)를 참조할 수 있다. 또 가중 Polya 사후분포에 대해서는 Meeden(1999) 및 이승천(2005)을 참조할 수 있다.

왜도가 심한 연속형 분포의 평균에 대한 구간추정에서 표본크기가 작은 경우 대부분 신뢰구간의 포함확률은 명목 신뢰수준보다 상당히 낮은 값을 갖는 것이 일반적이다. Meeden(1999)는 가중 Polya 사후분포를 이용하여 신뢰구간을 구하므로써 포함확률의 하향성에 대한 문제를 성공적으로 해결하였다. 이항확률의 구간추정 문제도 이와 크게 다르지 않다. 대부분의 신뢰구간은 표본크기가 작고 모비율 p 가 0 또는 1에 가까운 즉, 왜도가 심한 경우 문제가 발생하게 된다.

이승천(2005)은 $X \sim B(n, p)$ 일 때 가중치 $\xi (> 0)$ 를 갖는 가중 Polya 사후분포에서 모비율에 대한 사후 평균과 분산을 다음과 같이 유도하였고

$$E(p | X) = \frac{X + \xi}{n + 2\xi} \tag{2.1}$$

$$\text{Var}(p | X) = \frac{1}{n + 2\xi + 1} \frac{X + \xi}{n + 2\xi} \left(1 - \frac{X + \xi}{n + 2\xi} \right) \tag{2.2}$$

사후평균의 정규 근사성을 이용하여 다음과 같은 모비율의 근사 신뢰구간을 구하였다.

$$\tilde{p} \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n} + 1}} \tag{2.3}$$

여기서 $\tilde{n} = n + z_{\alpha/2}^2, \tilde{p} = (X + z_{\alpha/2}^2/2)/\tilde{n}$ 이다. \tilde{p} 는 가중치 ξ 의 값이 $z_{\alpha/2}^2/2$ 일 때 사후평균인데 이와 같이 가중치를 정한 이유는 Wilson(1927)의 Score 신뢰구간과 밀접한 관계가 있다. 또 $z_{0.025}^2 = 1.96^2 \approx 4$ 으로 Agresti와 Coull(1998)의 4 개의 인위적인 추가 관측 갯수와도 연관이 있다고 하겠다.

서로 독립인 두 확률변수 X_1, X_2 가 각각 이항분포 $B(n_1, p_1), B(n_2, p_2)$ 를 따를 때 $\theta = p_1 - p_2$ 에 대한 가중 Polya 사후분포에 의한 $(1 - \alpha) \times 100\%$ 신뢰구간은 $\tilde{n}_i = n_i + z_{\alpha/2}^2/2, \tilde{p}_i =$

$(X_i + z_{\alpha/2}^2/4)/\tilde{n}_i, i = 1, 2$ 라고 할 때

$$CI_{WP} : \tilde{p}_1 - \tilde{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1+1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2+1}} \quad (2.4)$$

와 같이 정의하며 이를 가중 Polya 사후분포 신뢰구간이라고 하기로 한다. 이는 각 이항분포에서 가중치를 $\xi = z_{\alpha/2}^2/4$ 으로 부여하여 신뢰구간을 구한 것이다.

인위적인 2개의 성공과 2개의 실패를 추가하여 구한 Agresti-Caffo의 $(1 - \alpha) \times 100\%$ 신뢰구간은

$$CI_{AC} : \bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{\bar{n}_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{\bar{n}_2}} \quad (2.5)$$

와 같다. 단 $\bar{n}_i = n_i + 2, \bar{p}_i = (X_i + 1)/\bar{n}_i, i = 1, 2$ 이다. 즉, Agresti와 Coull(1998)에서와 같이 인위적인 2개의 성공과 2개의 실패를 추가하여 $\hat{p}_i = X_i/n_i$ 대신 \bar{p}_i 에 의해 모비율을 추정하고 정규근사에 의해 신뢰구간을 구한 것으로 Wald 신뢰구간과 같은 형태를 취하고 있다. 그러나 베이저안의 관점에서 볼 때 (2.5)는 분산을 과대 추정하여 신뢰구간을 구한 것으로 Agresti-Caffo 신뢰구간의 보수성을 시사하고 있다. 즉, (2.1)과 (2.2)는 이항분포의 공액분포인 Beta(ξ, ξ) 사전분포로부터 얻어진 사후 평균과 분산과 같은데 \bar{p}_i 는 $\xi = 1$ 인 경우이므로 사후분산은 $\bar{p}_i(1-\bar{p}_i)/(\bar{n}_i+1)$ 과 같이 구하여 진다. 그러므로 베이저안 관점에서는 근사 신뢰구간이

$$CI_{BA} : \bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{\bar{n}_1+1} + \frac{\bar{p}_2(1-\bar{p}_2)}{\bar{n}_2+1}} \quad (2.6)$$

와 같이 구하여지며 이때 모든 X_1 과 X_2 의 관찰값에서 $CI_{BA} \subset CI_{AC}$ 가 성립된다.

결과적으로 (2.4)와 (2.6)는 각각 Beta($z_{\alpha/2}^2/4, z_{\alpha/2}^2/4$)와 Beta(1, 1)의 사전분포를 이용한 근사 베이저안 신뢰구간이라고 할 수 있고 $z_{0.025}^2/4 \approx 1$ 이므로 95% 신뢰수준에서 두 신뢰구간은 매우 유사한 성질을 갖게 될 것이다. 그러나 99% 신뢰수준에서는 표본크기가 작을 경우 차이가 있을 수 있겠다. 또 표본크기가 작은 경우 Agresti-Caffo 신뢰구간은 두 개의 신뢰구간과 비교하여 포함확률에 있어 뚜렷한 차이를 나타낼 것으로 예상할 수 있다.

Newcombe(1998b)에 의해 추천된 Hybrid score 신뢰구간은 l_i, v_i 를 각각 p_i 의 이차방정식 $z_{\alpha/2}^2 p_i(1-p_i) = n_i(\hat{p}_i - p_i)^2$ 의 하한근과 상한근이라고 할 때

$$CI_{HS} : \left[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{v_2(1-v_2)}{n_2}}, \right. \\ \left. (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\frac{l_2(1-l_2)}{n_1} + \frac{v_1(1-v_1)}{n_2}} \right] \quad (2.7)$$

과 같이 표현된다. CI_{HS} 는 p_1 과 p_2 가 모두 0 또는 1에 가까울 경우 보수적인 신뢰구간이 되지만 Agresti-Caffo 신뢰구간보다는 덜 보수적이며 평균 포함확률도 Agresti-Caffo 신뢰구간

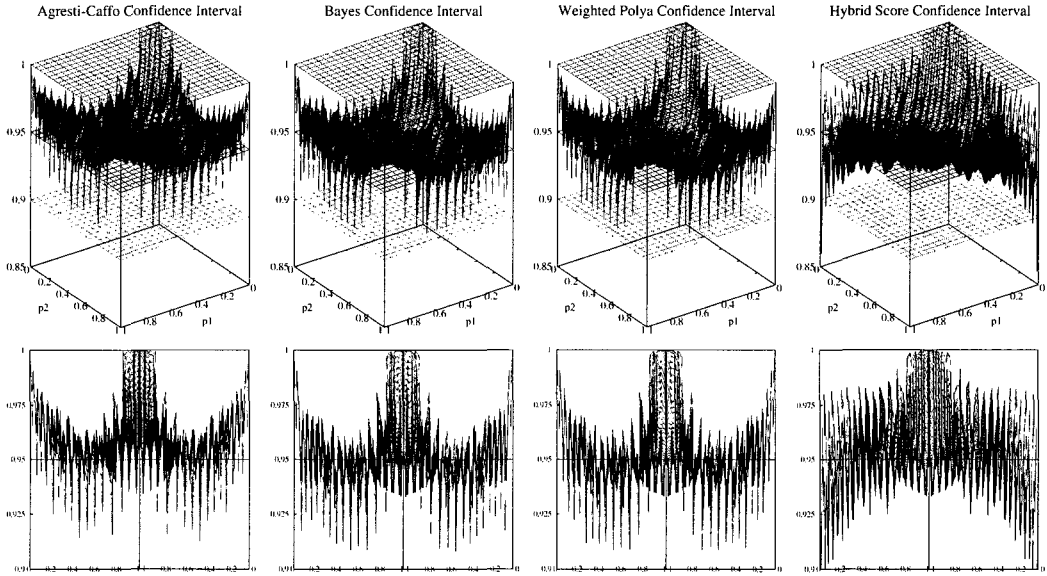


그림 2.1: $n_1 = 15, n_2 = 15$ 에서 95% Agresti-Caffo, 근사 베이지안, 가중 Polya 사후분포, Hybrid Score 신뢰구간의 포함확률 3차원 그림 및 투시도

보다 명목 신뢰수준에 더 근접한 것으로 알려져 있다. 그러므로 CI_{WP} 의 비교 신뢰구간으로 매우 적절할 것으로 판단된다.

모비율 차이의 네 신뢰구간에 대해 95% 신뢰수준에서 구한 전형적인 포함확률은 그림 2.1와 같다. 예상과 같이 근사 베이지안 신뢰구간과 가중 Polya 사후분포 신뢰구간의 포함확률은 매우 유사한 형태를 갖고 있으며 Agresti-Caffo 신뢰구간도 유사하지만 약간 보수적인 경향을 보이고 있다. 한편 Hybrid Score 신뢰구간과는 다른 세 개의 신뢰구간과는 차이를 보이고 있다. 즉, (p_1, p_2) 의 값이 $(0,0)$, $(1,1)$ 에 가까울 때 네개의 신뢰구간은 극히 보수적인 경향을 띄고 있고 $(0,1)$ 또는 $(1,0)$ 에 가까울 때 Agresti-Caffo, 근사 베이지안, 가중 Polya 사후분포에 의한 신뢰구간은 대체로 보수적인 경향을 갖는다. 그러나 이 점들에서 Hybrid Score 신뢰구간은 포함확률의 진동이 매우 심한 것을 볼 수 있다.

3절에서는 신뢰구간의 성능을 평가할 때 많이 이용되는 평균 포함확률, 포함확률의 평균 절대오차 및 신뢰구간의 평균적인 기대길이 등을 조사하여 네 신뢰구간의 성능을 비교하여 보기로 한다.

3. 모비율 차이 신뢰구간들의 성능 비교

표본크기 n_1 과 n_2 가 클 경우 ($n_1, n_2 \geq 40$) Agresti-Caffo, 근사 베이지안, 가중 Polya 사후분포 신뢰구간은 매우 유사하여 성능에서 유의한 차이가 나지 않는다. 또 Hybrid Score 신뢰구간도 평균 포함확률 및 기대길이 등에서 다른 세 개의 신뢰구간과 뚜렷한 차이를 보

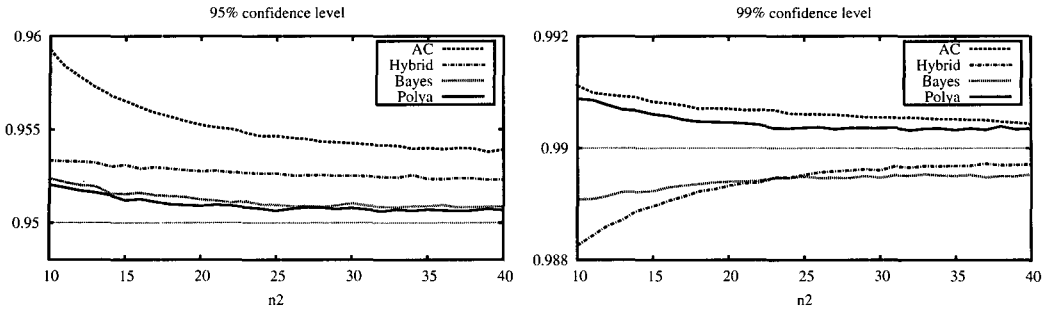


그림 3.1: $n_1 = 30$ 과 $n_2 = 10, \dots, 40$ 에서 Agresti-Caffo, 근사 베이저안, Hybrid Score, 가중 Polya 사후분포 신뢰구간들의 평균 포함확률

이지 않는다. 그러므로 표본크기가 큰 경우에는 계산의 간편성을 고려할 때 세 개의 신뢰구간이 Hybrid Score 신뢰구간보다 선호된다고 하겠다. 그러나 표본크기가 작은 경우에는 네 신뢰구간은 평균 포함확률 등에서 차이를 보이고 있으므로 여기에서는 표본크기가 작을 때 신뢰구간의 성능을 비교하기로 한다.

3.1. 평균 포함확률

네 신뢰구간은 모두 근사 신뢰구간이므로 포함확률이 명목 신뢰수준에 얼마나 근접한지를 비교하기 위해 각 표본크기에서 평균 포함확률을 구하였다. 즉, $C(p_1, p_2; n_1, n_2)$ 를 $B(n_1, p_1)$ 와 $B(n_2, p_2)$ 에서 구한 θ 신뢰구간의 포함확률이라고 할 때, 평균 포함확률을 다음과 같은 식에 의해 구한다.

$$\bar{C}(n_1, n_2) = \iint C(p_1, p_2; n_1, n_2) dp_1 dp_2 \quad (3.1)$$

(3.1)는 모비율에 대한 무정보 사전분포인 균일분포를 가정하여 구한 평균으로 이를 구하기 위해 0.005 간격의 격자점에서 신뢰구간들의 포함확률을 구하였다. 그림 3.1는 $n_1 = 30, n_2 = 10, \dots, 40$ 에서 95%와 99% 신뢰수준에서 구한 각 신뢰구간의 평균 포함확률을 n_2 의 크기에 따라 나타낸 것으로 여러 n_1 의 값에서 모두 비슷한 결과를 얻을 수 있었다.

95% 신뢰수준에서는 네 개의 신뢰구간이 모두 보수적인 것으로 나타났으나 가중 Polya 사후분포 신뢰구간이 평균적으로 신뢰수준에 가장 근접하였고, 근사 베이저안 구간도 이와 유의한 차이를 보이지 않고 있다. 그러나 Hybrid Score 신뢰구간과 Agresti-Caffo 신뢰구간의 평균 포함확률은 신뢰수준을 많이 상회하고 있으며 특히 Agresti-Caffo 신뢰구간의 평균 포함확률은 신뢰수준과 유의한 차이가 있는 것으로 볼 수 있다. 한편 99% 신뢰수준에서는 Agresti-Caffo와 Hybrid Score 신뢰구간의 평균 포함확률과 명목 신뢰수준과의 차이가 $n_2 = 10$ 에서 각각 최대 0.001과 0.0018 정도의 차이를 보이고 있어 네 신뢰구간이 모두 근사성에 있어서는 큰 문제가 없다고 하겠다. 그러나 이 경우도 가중 Polya 사후분포에 의한 근사 신뢰구간이 근사성에 있어 가장 좋다고 할 수 있다.

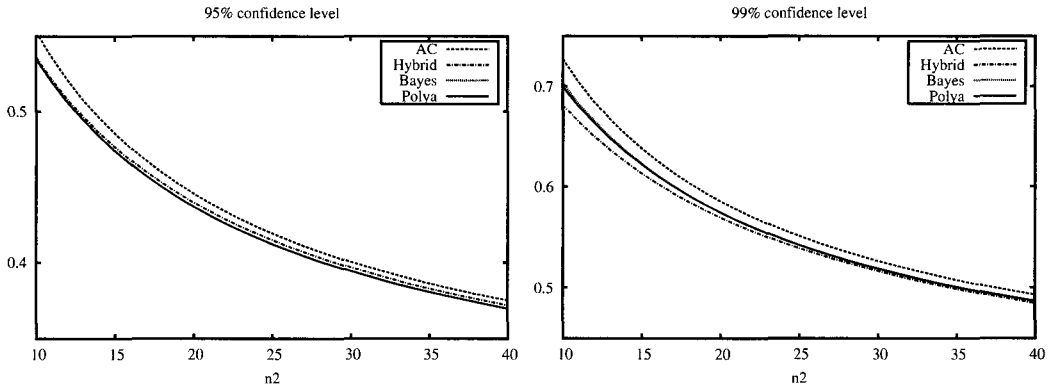


그림 3.2: $n_1 = 30$ 과 $n_2 = 10, \dots, 40$ 에서 Agresti-Caffo, 근사 베이지안, Hybrid Score, 가중 Polya 사후분포 신뢰구간의 평균 기대길이

Agresti와 Caffo(2000)는 Hybrid Score 신뢰구간이 보수적인 신뢰구간이라고 언급하였으나 이는 95% 신뢰수준에 해당되는 것으로 99% 신뢰수준에서는 오히려 평균 포함확률이 신뢰수준을 하회하는 것을 확인할 수 있었다.

3.2. 평균 기대길이

일반적으로 같은 신뢰수준에서는 신뢰구간의 길이가 짧을 수록 보다 많은 정보를 제공하므로 구간의 길이가 짧은 신뢰구간을 선호하게 된다. 이와 같은 점에서 네 신뢰구간의 평균 기대길이를 비교하였다. 즉, 관찰값 x_1, x_2 에서 구한 신뢰구간의 하한과 상한을 각각 $L(x_1, x_2; n_1, n_2), U(x_1, x_2; n_1, n_2)$ 이라고 할 때 p_1, p_2 에서의 기대길이는

$$E_{p_1, p_2; n_1, n_2}(CI) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} (U(x_1, x_2; n_1, n_2) - L(x_1, x_2; n_1, n_2)) \times \binom{n_1}{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1} \binom{n_2}{x_2} p_2^{x_2} (1 - p_2)^{n_2 - x_2}$$

와 같으며 모비율이 균일분포를 따른다고 가정할 때 평균 기대길이 (mean of expected length)는

$$AEL(n_1, n_2) = \iint E_{p_1, p_2; n_1, n_2}(CI) dp_1 dp_2$$

와 같이 정의한다.

그림 3.2은 $n_1 = 30, n_2 = 10, \dots, 40$ 에서 구한 평균 기대길이를 n_2 의 값에 따라 나타낸 것으로 95% 신뢰수준에서는 가중 Polya 신뢰구간과 베이지안 근사 신뢰구간이 Agresti-Caffo와 Hybrid Score 신뢰구간과 비교하여 모든 n_2 에서 일률적으로 작은 평균 기대길이를 갖고 있다. 95% 신뢰수준에서 Agresti-Caffo와 Hybrid Score는 보수적인 신뢰구간이었

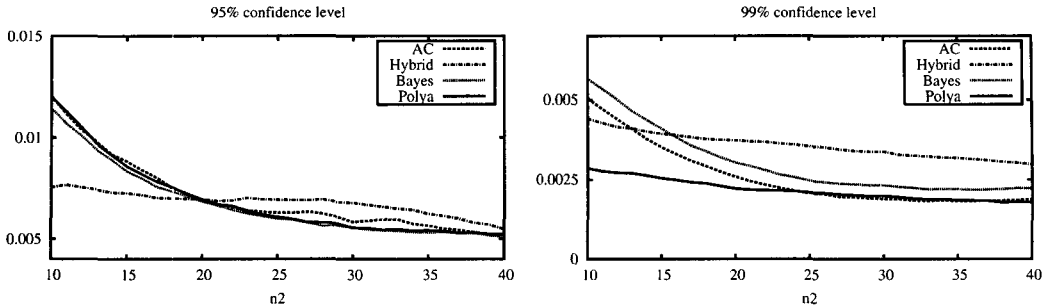


그림 3.3: $n_1 = 30$ 과 $n_2 = 10, \dots, 40$ 에서 Agresti-Caffo, Bayes, Hybrid Score, 가중 Polya posterior 신뢰구간 포함확률의 평균절대오차

으므로 신뢰구간의 폭이 필요 이상으로 넓다는 것을 알 수 있다. 한편 99% 신뢰수준에서는 Hybrid Score 신뢰구간의 평균 기대길이가 가장 짧은 것을 볼 수 있으나 99% 신뢰수준에서 Hybrid Score 신뢰구간의 평균 포함확률이 신뢰수준보다 작은 값을 갖고 있으므로 신뢰구간의 길이가 필요보다 약간 짧은 것으로 생각할 수 있다.

가중 Polya 신뢰구간과 베이지안 근사신뢰구간은 95%와 99%에서 모두 거의 같은 평균 기대길이를 갖고 있지만 특히 99% 신뢰수준에서 평균 포함확률이 차이를 보이는 이유는 두 신뢰구간의 중심위치가 다르기 때문이다. 즉, 가중 Polya 신뢰구간은 $\bar{p}_1 - \bar{p}$ 를 중심으로 신뢰구간이 형성되어 있으나 베이지안 근사 신뢰구간은 $\bar{p}_1 - \bar{p}_2$ 를 중심으로 구성되어 있어 평균 포함확률에서 차이를 보이고 있다. 가중 Polya 신뢰구간의 평균 포함확률이 신뢰수준에 보다 근접해 있으므로 빈도학파적 입장에서는 모비율 차이의 신뢰구간을 구하기 위한 사전분포는 Beta(1, 1)보다는 Beta($z_{\alpha/2}^2/4, z_{\alpha/2}^2/4$)이 더 적절하다고 판단된다.

3.3. 평균 절대오차

그림 2.1에서 볼 수 있듯이 네 개 신뢰구간의 포함확률은 신뢰수준을 전후하여 심한 진동을 하고 있는데 진동의 폭이 좁을 수록 선호되는 신뢰구간이라고 하겠다. 포함확률의 진동 정도를 측정하기 위해 여기에서는 포함확률의 평균 절대오차를 구하였다. 즉,

$$\iint |C(p_1, p_2; n_1, n_2) - (1 - \alpha)| dp_1 dp_2$$

를 구하여 이를 그림 3.3으로 나타내었다.

95% 신뢰수준에서 $n_2 < 20$ 인 경우 Hybrid Score 신뢰구간이 가장 작은 평균 절대오차를 갖고 있고, $n_2 \geq 20$ 에서는 가중 Polya 또는 베이지안 근사 신뢰구간이 비슷한 정도로 작은 평균 절대오차의 값을 갖는다. 또 99% 신뢰수준에서는 가중 Polya 신뢰구간이 평균 절대오차에 있어 가장 선호된다고 하겠다. 그러나 평균 절대오차의 최대 차이는 95% 신뢰수준과 99% 신뢰수준에서 각각 0.004와 0.0025 정도이므로 유의하게 차이가 있다고 하기는

표 4.1: 95% 신뢰수준에서 가중 Polya 신뢰구간(WP)와 Jefferys-Perk(JP) 신뢰구간의 평균 포함확률, 평균 기대길이, 평균제곱오차

(n_1, n_2)	(10,10)		(10,20)		(10,30)		(20,20)		(40,40)	
	WP	JP	WP	JP	WP	JP	WP	JP	WP	JP
평균 포함확률	0.951	0.949	0.951	0.950	0.952	0.952	0.951	0.949	0.951	0.949
평균 기대길이	0.647	0.645	0.568	0.569	0.535	0.539	0.476	0.475	0.344	0.343
평균 제곱오차	0.012	0.012	0.011	0.010	0.012	0.010	0.007	0.007	0.005	0.004

어렵다고 생각된다. 그러므로 각 신뢰구간의 포함확률은 신뢰수준을 중심으로 거의 비슷한 정도의 진동을 갖고 있다고 할 수 있다.

4. 결론

모비율 차이의 구간추정에서 Agresti-Caffo 신뢰구간은 추정식의 간편성으로 인하여 선호되고 있으나 구간의 보수성으로 인하여 문제가 제기될 수도 있다. 또 비교적 간편한 식에 의해 구할 수 있다는 이유로 Newcombe(1998b)이 추천한 Hybrid score 신뢰구간은 Agresti-Caffo 신뢰구간과 비교하여 보수성은 개선되었지만 잘 알려진 식에 의한 신뢰구간이 아니라는 점에서 실용성에서 다소 떨어진다고 볼 수 있다.

이와 비교하여 근사 베이지안 신뢰구간과 가중 Polya 신뢰구간은 Agresti-Caffo와 같이 Wald 신뢰구간과 거의 같은 식에 의해 구할 수 있으며 평균 포함확률 및 기대길이에서 Hybrid score 신뢰구간보다 선호되어 모비율 차이의 구간추정에서 매우 유용하게 사용될 수 있다. 특히 표 4.1에 의하면 가중 Polya 신뢰구간은 다소 복잡하지만 Beal(1987)에 의해 추천된 Jefferys-Perk 신뢰구간과 비교하여도 평균 포함확률 등 여러 기준에 의한 비교에서 손색이 없어 현실적으로 유용하게 사용될 수 있을 것이다.

참고문헌

이승천 (2005). 이항비율의 가중 Polya posterior 구간추정, <응용통계연구> **18**, 607-615.
 정형철,전명식,김대학 (2003). 모비율 차이의 신뢰구간들에 대한 비교연구, <응용통계연구> **16**, 377-393.
 Agresti, A. and Coull, B. A. (1998). Approximation is better than “exact” for interval estimation of binomial proportions, *The American Statistician*, **52**, 119-126.
 Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures, *The American Statistician*, **54**, 280-288.
 Anbar, D. (1983). On estimating the difference between two probabilities with special reference to clinical trials, *Biometrics*, **39**, 257-262.
 Beal, S. L. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples, *Biometrics*, **43**, 941-950.

- Blyth, C. R. and Still, H. A. (1983) Binomial confidence intervals, *Journal of the American Statistical Association*, **78**, 108–116.
- Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion, *Statistical Science*, **16**, 101–133.
- Brown, L. D., Cai, T. T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions, *The Annals of Statistics*, **30**, 160–201.
- Chan, I. S. F. and Zhang, Z. (1999). Test-based exact confidence intervals for the difference of two binomial proportions, *Biometrics*, **55**, 1202–1209.
- Feller, W. (1968). *An introduction of probability theory and its applications*, volume I, Wiley, New York.
- Ghosh, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter, *Journal of the American Statistical Association*, **74**, 894–900.
- Ghosh, M. and Meeden, G. D. (1998) *Bayesian methods for finite population sampling*, Chapman & Hall, London.
- Mee, R. (1984). Confidence bounds for the difference between two probabilities, *Biometrics*, **40**, 1175–1176.
- Meeden, G. D. (1999). Interval estimators for the population mean for skewed distributions with a small sample size, *Journal of Applied Statistics*, **26**, 81–96.
- Newcombe, R. (1998a) Two-sided confidence intervals for the single proportion: Comparison of seven methods, *Statistics in Medicine*, **17**, 857–872.
- Newcombe, R. (1998a) Interval estimation for the difference between independent proportions: Comparison of eleven methods, *Statistics in Medicine*, **17**, 873–890.
- 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, *Statistics in Medicine*, **17**, 873–890.
- Vollet, S. E. (1993). Confidence intervals for a binomial proportion, *Statistics in Medicine*, **12**, 809–824.
- Wilson, E. B. (1927). Probable inference, the law of succession and statistical inference, *Journal of the American Statistical Association*, **22**, 209–212.

[2005년 7월 접수, 2005년 12월 채택]

The Weighted Polya Posterior Confidence Interval For the Difference Between Two Independent Proportions*

Seung-Chun Lee¹⁾

ABSTRACT

The Wald confidence interval has been considered as a standard method for the difference of proportions. However, the erratic behavior of the coverage probability of the Wald confidence interval is recognized in various literatures. Various alternatives have been proposed. Among them, Agresti-Caffo confidence interval has gained the reputation because of its simplicity and fairly good performance in terms of coverage probability. It is known however, that the Agresti-Caffo confidence interval is conservative. In this note, a confidence interval is developed using the weighted Polya posterior which was employed to obtain a confidence interval for the binomial proportion in Lee(2005). The resulting confidence interval is simple and effective in various respects such as the closeness of the average coverage probability to the nominal confidence level, the average expected length and the mean absolute error of the coverage probability. Practically it can be used for the interval estimation of the difference of proportions for any sample sizes and parameter values.

Keywords: difference of proportions, weighted Polya posterior, Wald confidence interval, Agresti-Caffo confidence interval, hybrid score confidence interval, approximated Bayesian confidence interval

* This Work was Supported by Hanshin University Research Grant in 2006.

1) Professor, Dept. of Statistics, Hanshin University, 411 Yangsan-Dong, Osan, Kyunggi-Do, 447-791
E-mail: seung@hanshin.ac.kr