한 국 통 계 학 회 논 문 집
2011, 18권, 5호, 615–623
DOI: http://dx.doi.org/10.5351/CKSS.2011.18.5.615

이항 신뢰구간에서 극단값의 영향

류제복[1,a]

[a]청주대학교 통계학전공

요 약

이항비율에 대한 구간추정에 다양한 신뢰구간들이 사용된다. 그러나 대부분의 신뢰구간들은 모비율 p가 0이나 1에 근사할 때 포함확률이 신뢰수준(또는 명목수준, $1-\alpha$)을 크게 벗어난다. 이는 극단적인 관찰값의 영향 때문이다. Vollset (1993), Agresti와 Coull (1998), Newcombe (1998), Brown 등 (2001) 등은 극단값의 조정을 통해서 이러한 문제를 해결하는 방법들을 제시하였다. 본 연구에서는 극단값들이 이항비율에 대한 신뢰구간에 어느 정도 영향을 미치는 지를 6개의 신뢰구간들에 대해서 수치적으로 비교해 보았다.

주요용어: 이항비율, 신뢰구간, 극단값, 포함확률, 기대폭.

1. 서론

이항비율에 대한 Wald신뢰구간, $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$은 점근적 정규이론에 바탕을 둔 대표적인 신뢰구간이다. 그러나 Wald신뢰구간은 관찰값이 $X = 0$(또는 n)일 때 모비율 p에 대한 신뢰구간을 제공하지 못한다. 그리고 p가 0이나 1의 근방에 있을 때 포함확률이 아주 낮고 p가 1/2 근방에서는 신뢰구간의 길이가 길다. 이와 같이 Wald신뢰구간은 신뢰구간의 평가에 사용되는 (평균)포함확률이나 신뢰구간의 (평균)기대폭이 다른 신뢰구간들에 비해 바람직하지 않기 때문에 이항비율에 대한 신뢰구간으로 적절치 않다. 이러한 문제를 보완해주기 위해서 다양한 신뢰구간들이 제시되고, Wald신뢰구간을 다른 신뢰구간으로 대체해야 한다는 주장이 지배적이다 (Ghosh, 1979; Blyth와 Still, 1983; Vollset, 1993; Newcombe, 1998; Agresti와 Coull, 1998; Brown 등, 2001, 2002; 등).

Wilson (1927)은 귀무가설($H_0 : p = p_0$)하에서 정규근사이론을 사용한 score검정에 기반한 신뢰구간을 제시하였고, Clopper와 Pearson (1934)은 정규근사이론을 사용하지 않고 양측에 동일 확률을 부여한 이항양측검정에 기초한 정확신뢰구간을 제안하였다. 또한 정확신뢰구간에서 관측결과에 1/2의 확률을 부여한 Mid-p신뢰구간도 종종 사용된다. 한편 Agresti와 Coull (1998)은 95% Wald신뢰구간에서 "두 번의 성공과 두 번의 실패를 더한" 수정된 Wald신뢰구간을 제시(이를 Agresti-Coull신뢰구간이라 부름)하여 실무에서 간편하게 사용할 수 있게 하였다. 또한 Borkowf (2006)는 Agresti-Coull신뢰구간과 유사하게 Wald신뢰구간을 변형한 신뢰구간을 제시하였다. 그리고 무정보사전분포인 Beta(1/2, 1/2)를 사용한 Jefffereys사전신뢰구간, Arcsine변환에 의한 신뢰구간, 우도비 검정에 기초한 우도비신뢰구간, 로짓신뢰구간, 연속성 수정을 한 신뢰구간 등 많은 신뢰구간들이 사용되고 있다.

이론적이든 실무적이든 신뢰구간을 사용하고자 하는 사람들은 항상 자신들의 연구에 어떤 신뢰구간을 사용하는 것이 적절한지에 대한 의문을 갖고 있다. 이러한 질문의 답을 얻기 위해서 많은 연구자들이 신뢰구간들에 대한 비교연구를 하였다. 신뢰구간들을 비교하기 위해서 (평균)포함확률, 신뢰구간의 (평균)기대폭, 오차 등 다양한 척도와 그림 등이 사용되고 있다. Schader과 Schmid (1990), Vollset

이 논문은 2010년도에 청주대학교 산업과학연구소가 지원한 특별연구과제에 의해서 연구되었음.
[1] (360-764) 충북 청주시 상당구 대성로 298, 청주대학교 자연과학부 통계학전공, 교수. E-mail: jbryu@cju.ac.kr

(1993), Newcombe (1998) 등은 Wilson신뢰구간이 적절하다고 주장하고 있다. 그러나 임상분야 등에서는 안전성을 확보하기 위해서 정확신뢰구간이, 그리고 사용의 간편함을 위해서는 Agresti-Coull신뢰구간이 추천된다. 한편 Brown 등 (2001)은 표본이 작을 때($n \leq 40$)는 Wilson과 Jeffreys사전신뢰구간을, 표본이 큰 경우($n > 40$)에는 Wilson, Jeffreys, 그리고 Agresti-Coull신뢰구간을 구미에 맞게 사용할 것을 제안하고 있다. 이와 같이 상황에 따라 선호되는 신뢰구간이 다를 수 있다.

Brown 등 (2001, 그림 5, 그림 11), Agresti와 Coull (1998, 그림 4, 그림 5), Vollset (1993, 그림 1–3), 그리고 류제복과 이승주 (2006, 그림 4.1과 그림 4.2)에서와 같이 p가 0이나 1근처에 있을 때 정확신뢰구간과 Agresti-Coull신뢰구간은 포함확률이 명목수준을 크게 초과하고 그 밖의 다른 신뢰구간들은 지나치게 작은 경향이 있다. 이런 현상은 표본이 작은 경우 더 두드러진다. 그 이유 중의 하나는 $X = 0$(또는 n)과 같이 극단값에서의 신뢰구간이 포함확률이나 신뢰구간의 길이에 영향을 주기 때문이다. 따라서 이런 문제를 해결하기 위해서 Vollset (1993), Agresti와 Coull (1998), Newcombe (1998), Brown 등 (2001), 그리고 류제복 (2010) 등은 극단값에서 수정한 신뢰구간의 사용을 제시하였다. 특히 표본이 작은 경우 극단값에서의 신뢰구간이 포함확률이나 신뢰구간의 길이에 미치는 영향이 크게 되므로 적절한 극단값 조정이 필요하다. 따라서 극단값에서의 조정을 하기 전에 이항비율에 대한 신뢰구간에서 표본에 따른 극단값의 영향력이 어느 정도인지를 살펴보고자 한다.

본 논문의 구성은 2장에서는 이항비율에 대한 신뢰구간 추정에 널리 사용되는 방법들을 소개하고 3장에서는 신뢰구간의 평가기준들을 다룬다. 그리고 4장에서는 극단값에서의 신뢰구간이 신뢰구간의 포함확률과 기대폭에 미치는 영향을 살펴본다. 이를 위해서 2장에서 다룬 6개 신뢰구간들에 대해서 표본크기와 모비율의 변화에 따라 극단값에서 영향력이 어느 정도인지를 수치적으로 비교해본다.

2. 신뢰구간

이항비율에 대한 Wald신뢰구간의 대안으로 제시되고 있는 신뢰구간들은 매우 다양하다. 여기서는 이론적이고 실무적으로 널리 사용되는 6개의 신뢰구간들을 살펴본다.

2.1. 정확신뢰구간

Clopper와 Pearson (1934)은 정규근사이론을 사용하지 않고 선형보간법(interpolation)을 사용해서 이항 비율에 대한 신뢰구간을 제시하였는데, 이를 정확신뢰구간이라 한다. 이 신뢰구간은 모든 모수값에서 포함확률이 명목수준 $(1-\alpha)$를 초과해서 보수적인 신뢰구간으로 간주된다. $x = 1, 2, \ldots, n-1$일 때, 정확신뢰구간은 다음과 같다.

$$\left[1 + \frac{n-x+1}{x F_{2x,2(n-x+1),1-\alpha/2}}\right]^{-1} < P < \left[1 + \frac{n-x}{(x+1)F_{2(x+1),2(n-x),\alpha/2}}\right]^{-1}. \tag{2.1}$$

$x = 0$일 때 하한은 0이고 $x = n$일 때 상한은 1이며 $F_{a,b,\alpha}$는 자유도가 a, b인 F분포의 $(1-\alpha)$분위수이다. 또한 정확신뢰구간의 하한과 상한은 각각 $B(\alpha/2; x, n-x+1)$와 $B(1-\alpha/2; x+1, n-x)$로 나타낼 수 있다. 여기서 $B(\alpha; a, b)$는 모수가 a와 b인 베타분포의 α분위수이다.

2.2. Mid-p신뢰구간

정확신뢰구간의 보수성을 보완하기 위해서 관측 결과에 1/2의 확률을 부여한 식 (2.2)로부터 Mid-p신뢰구간을 얻는다. Mid-p신뢰구간은 명목수준에 근사한 신뢰구간으로 경계값에 이를 사용하면 정확신뢰구간을 사용한 경우보다 변동이 적다. 류제복과 이승주 (2006)는 표본이 크고 p가 작은 경우

Mid-p신뢰구간의 사용이 적절함을 보여주고 있다.

$$\frac{1}{2}\binom{n}{x}p^x(1-p)^{n-x} + \sum_{k=x+1}^{n}\binom{n}{k}p^k(1-p)^{n-k} = \frac{\alpha}{2},$$
$$\frac{1}{2}\binom{n}{x}p^x(1-p)^{n-x} + \sum_{k=0}^{x-1}\binom{n}{k}p^k(1-p)^{n-k} = \frac{\alpha}{2}. \tag{2.2}$$

Mid-p신뢰구간의 상한은 $1/2\, B(1-\alpha/2; x, n-x+1) + 1/2B(1-\alpha/2; x+1, n-x)$이고, 하한은 대칭이다.

2.3. Wilson신뢰구간

Wilson신뢰구간은 정규근사이론에 근거해서 귀무가설 $H_0: p = p_0$에 대한 score검정을 역 변환한 $|\hat{p} - p_0|/\sqrt{p_0(1-p_0)/n} \le z_{\alpha/2}$를 p_0에 대해 풀어서 얻는다. 즉,

$$\left(\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} \pm z_{\frac{\alpha}{2}}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z_{\frac{\alpha}{2}}^2}{4n}\right]\Big/ n}\right)\Big/\left(1 + \frac{z_{\frac{\alpha}{2}}^2}{n}\right). \tag{2.3}$$

Wilson신뢰구간은 Wald신뢰구간의 단점을 보완해주고 정확신뢰구간보다 보수성이 덜하면서 포함확률이 명목수준에 근사하기 때문에 Schader과 Schmid (1990), Vollset (1993), Agresti와 Coull (1998), Newcombe (1998), Brown 등 (2001, 2002) 등 많은 이들이 추천하고 있다.

2.4. Agresti-Coull(AC)신뢰구간

Agresti와 Coull (1998)은 Wald신뢰구간에 두 번의 성공과 두 번의 실패를 추가한 신뢰구간을 제안하였다. 이는 Wald신뢰구간에서 n과 $\hat{p}$ 대신에 n^*와 $\hat{p}^*$를 대입한 것이다. 즉,

$$\hat{p}^* \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}^*(1-\hat{p}^*)}{n^*}}, \tag{2.4}$$

여기서 총 시행수 $n^* = n+4$이고 $\hat{p}^* = (X+2)/(n+4)$이다.

AC신뢰구간은 Wald신뢰구간보다 포함확률이 크게 개선되며, 평균포함확률이 Wilson신뢰구간보다 크고 p가 0이나 1근처에 있을 때에도 포함확률이 심각하게 낮지 않다. 약간 보수적이고 신뢰구간의 기대폭이 크지만 그리 우려할만하지 않다. 그리고 작은 표본에서도 사용이 가능하다. 다만, 95%신뢰수준에서만 위 원리를 적용할 수 있다. 이 신뢰구간의 최대의 장점은 간편함에 있다.

2.5. Borkowf신뢰구간

Borkowf (2006)는 AC신뢰구간을 구할 때와는 달리 신뢰하한을 구할 때 관측된 이항자료에 하나의 가상의 실패를 더해주고, 신뢰상한을 구할 때는 관측된 이항 자료에 하나의 가상의 성공을 더해주는 방법을 사용하였다. Borkowf신뢰구간의 하한(LB)과 상한(UB)은 각각 다음 식으로 구할 수 있다.

$$\mathrm{LB}(x, n, \alpha) = p_l - z_{\frac{\alpha}{2}}\sqrt{\frac{p_l q_l}{n}}, \qquad \mathrm{UB}(x, n, \alpha) = p_u + z_{\frac{\alpha}{2}}\sqrt{\frac{p_u q_u}{n}}. \tag{2.5}$$

여기서 하한과 상한을 구할 때는 각각 $p_l = (X+0)/(n+1)$, $p_u = (X+1)/(n+1)$을 이용하고 $q_l = 1 - p_l$이고 $q_u = 1 - p_u$이다. 95%신뢰수준에서 $X = 0$일 때의 Borkowf신뢰구간의 상한은 3의 법칙(rule of three)에 의한 신뢰상한인 $3/n$과 거의 유사하게 된다 (Jovanovic와 Levy, 1997).

2.6. Jeffreys사전신뢰구간

이항비율에 대한 신뢰구간 추정에 모비율 p에 대한 사전정보를 활용한 베이지안 신뢰구간도 널리 사용된다. 사전분포로 무정보적 사전분포(noninformative prior)를 사용할 것인가 정보적 사전분포(informative prior)를 사용할 것인가는 많은 논란이 있지만 빈도학파들의 주장을 최대한 수용하는 무정보적 사전분포의 사용이 보편적이다. 대표적인 Jeffreys사전신뢰구간은 무정보사전분포로 Beta(1/2, 1/2)를 사용한다. Brown 등 (2001, 2002)은 표본이 작을 경우 Jeffreys사전신뢰구간의 사용을 추천하고 류제복과 이승주 (2006)도 p가 작고 표본이 큰 경우 이 신뢰구간의 사용이 적절함을 보여주고 있다. $100(1-\alpha)\%$ 동일 꼬리 Jeffereys사전신뢰구간은 다음과 같다.

$$L_J(x) \le p \le U_J(x), \tag{2.6}$$

여기서 $L_J(x) = B(\alpha/2; x + 1/2, n - x + 1/2)$, $U_J(x) = B(1 - \alpha/2; x + 1/2, n - x + 1/2)$이고 $L_J(0) = 0$, $U_J(n) = 1$이다.

3. 신뢰구간의 평가기준

이항비율에 대한 신뢰구간을 추정하는 다양한 방법들이 많은 연구자들에 의해서 제시되었다. 그러나 이러한 신뢰구간들을 사용하기에 앞서 그 신뢰구간의 적절성을 평가해야 한다. 신뢰구간의 적절성을 평가하는 기준은 여러 가지가 있으나 포함확률과 신뢰구간에 대한 기대폭이 널리 사용된다. 본 논문에서도 이들 두 가지의 평가기준을 사용해서 극단값이 포함확률과 기대폭에 미치는 영향을 살펴보고자 한다.

3.1. 포함확률

포함확률(coverage probability)은 주어진 모수값에서의 신뢰구간이 그 모수값을 포함할 확률로 다음과 같이 정의된다.

$$C(n, p) = \sum_{x=0}^{n} I(x, p)\binom{n}{x}p^x(1-p)^{n-x}, \tag{3.1}$$

여기서 $I(x, p)$는 $X = x$일 때의 신뢰구간이 모수 p를 포함하고 있으면 1이고 그렇지 않으면 0이다. 포함확률은 n과 p의 함수이고 모든 X값에 대해 계산된다.

3.2. 기대폭

신뢰구간의 길이도 포함확률과 마찬가지로 모수값에 의존한다. 이항비율에 대한 구간추정에서 n과 p가 주어졌을 때 신뢰구간의 기대폭은 다음과 같다.

$$E(n, p) = \sum_{x=0}^{n}(U(x) - L(x))\binom{n}{x}p^x(1-p)^{n-x}, \tag{3.2}$$

여기서 $U(x)$와 $L(x)$는 주어진 p값에서 $X = x$일 때 신뢰구간의 상한과 하한 값이다.

추정된 신뢰구간의 포함확률은 신뢰수준에 근접할수록 바람직하다. 그러나 많은 신뢰구간들은 일부 구간에서 포함확률이 지나치게 크거나 작다. 통상적으로 포함확률은 신뢰구간의 길이에 의존하는데, 신뢰구간의 길이가 길면 포함확률이 커지게 되고 반대로 신뢰구간이 짧으면 포함확률이 작아지는 경향이 있다. 포함확률이 너무 작은 신뢰구간은 추정하고자 하는 모수를 포함할 확률이 낮아 구간추정

표 1: 95%신뢰수준에서 $X = 0$일 때 각 방법들에 따른 신뢰상한

신뢰구간	$n = 5$	$n = 10$	$n = 30$	$n = 100$
Exact	0.5218	0.3085	0.1157	0.0362
Mid-p	0.4507	0.2589	0.0950	0.0295
Wilson	0.4345	0.2775	0.1135	0.0370
AC	0.4938	0.3262	0.1379	0.0456
Borkowf	0.4933	0.2691	0.0955	0.0293
Jeffreys	0.3794	0.2172	0.0797	0.0247

에 적절하지 않다. 또한 신뢰구간이 너무 길면 포함확률은 커지지만 추정의 의미가 없게 된다. 따라서 포함확률이 신뢰수준에 근사하면서 신뢰구간의 폭이 짧은 신뢰구간이 바람직하다.

4. 극단값이 신뢰구간의 평가기준에 미치는 영향

이항비율에 대한 신뢰구간들의 기대폭은 모비율 p가 0.5에서 멀리 떨어질수록 작아진다. 그러나 기대폭은 신뢰구간들에 따라 다르고, p가 0이나 1에 근접할 때 포함확률에 영향을 준다. 이항 모비율 p가 0이나 1에 가까우면 관측값이 극단값($X = 0$ 또는 n)을 가질 가능성이 커지고, 결국에는 극단값이 신뢰구간의 포함확률과 기대폭에 영향을 주게 된다. 본 장에서는 극단값이 신뢰구간의 평가기준인 포함확률과 기대폭에 어떠한 영향을 주는 지를 2장에서 다룬 6가지의 신뢰구간을 대상으로 살펴본다.

표 1은 관측값이 $X = 0$인 경우 6개 신뢰구간의 상한을 나타내고 있다. $n = 5$인 경우 신뢰구간의 상한은 정확신뢰구간이 가장 크고 다음으로는 AC신뢰구간 그리고 Borkowf신뢰구간 순이다. $n = 10$과 $n = 30$인 경우는 AC신뢰구간의 상한이 가장 크고 다음으로 정확신뢰구간, Wilson신뢰구간의 순이다. $n = 100$인 경우도 AC의 상한이 가장 크지만 Wilson신뢰구간과 정확신뢰구간의 차이는 거의 없어진다. 그러나 Jeffreys사전신뢰구간의 상한은 모든 n에서 가장 작다. 표본이 작은 경우 정확신뢰구간의 상한이 크지만 표본이 증가하면 AC신뢰구간이 크고 정확신뢰구간과 Wilson신뢰구간의 차이는 작아진다.

신뢰구간의 상한은 대체적으로 신뢰구간의 포함확률과 관련이 높다. AC신뢰구간이 p가 0이나 1에 가까운 곳에서의 포함확률이 지나치게 큰데 이는 표 1에서와 같이 $X = 0$인 극단값에서의 신뢰상한이 크기 때문이다. 정확신뢰구간과 Wilson신뢰구간의 경우도 유사하다.

$X = 0$일 때 6개 신뢰구간의 길이는 바로 상한값이 되며(AC나 Jeffrey의 경우는 상한 값 보다 약간 작음) 표본이 증가할수록 신뢰구간의 길이는 작아진다.

$p = 0.5$를 기준으로 포함확률이 대칭이 되므로 낮은 p값에서는 극단값으로 $X = 0$인 경우를 고려한다. $n = 10, 20, 30$이고 $p = 0.01, 0.05, 0.1, 0.2$일 때 포함확률에서 극단값인 $X = 0$이 차지하는 비율이 표 2에 있다. n이 작고 p가 0에 가까우면 극단값이 포함확률에 영향을 준다는 것이 여러 논문에서 지적되고 있지만 이를 표 2에서 수치적으로 확인할 수 있다. 류제복 (2010)은 Wald신뢰구간에서 극단값을 수정할 경우 포함확률과 기대폭이 크게 개선됨을 보여주고 있다. 그러나 n이 크면 극단값의 영향력은 예상한 바와 같이 작아지므로 극단값의 조정이 굳이 필요하지 않다. 따라서 이항비율에 대한 신뢰구간에서 n이 작고 p가 0에 가까운 경우 극단값의 조정이 필요하다.

표 3은 신뢰구간의 기대폭에서 극단값이 차지하는 비율을 나타내고 있다. 표 2의 포함확률에서는 방법들 간에 차이가 작은 반면에 표 3의 기대폭은 p가 증가하면 방법들 간에 차이가 생긴다. AC신뢰구간이 기대폭에서 극단값의 영향이 가장 크고 Jeffreys사전신뢰구간이 가장 작지만 n이 증가하면 그 차이는 미미해 진다. 표 1에서 신뢰상한이 AC신뢰구간이 가장 큰데, 이것이 신뢰구간의 기대폭에 영향을 주고 있음을 알 수 있다. 포함확률과 신뢰구간의 기대폭에서 극단값의 영향은 n과 p에 매우 민감하다.

이항분포는 이산형이기 때문에 포함확률과 신뢰구간의 기대폭은 이항모수 n과 p에 따라 변동이 크

표 2: n과 p에 따른 포함확률에서 극단값($X = 0$)이 차지하는 비율

n	신뢰구간	포함확률			
		$p = 0.01$	$p = 0.05$	$p = 0.1$	$p = 0.2$
10	Exact	0.9083	0.6057	0.3532	0.1081
	Mid-p	0.9083	0.6057	0.3532	0.1110
	Wilson	1.0000	0.6552	0.3750	0.1110
	AC	0.9083	0.6057	0.3750	0.1110
	Borkowf	0.9044	0.5994	0.3492	0.1081
	Jeffreys	1.0000	0.6057	0.3532	0.1110
20	Exact	0.8319	0.3643	0.1230	0.0000
	Mid-p	0.8319	0.3643	0.1230	0.0000
	Wilson	0.8319	0.3878	0.1271	0.0000
	AC	0.8319	0.3643	0.1271	0.0000
	Borkowf	0.8179	0.3594	0.1219	0.0000
	Jeffreys	0.8319	0.3643	0.1271	0.0000
30	Exact	0.7422	0.2180	0.0427	0.0000
	Mid-p	0.7674	0.2180	0.0000	0.0000
	Wilson	0.7674	0.2285	0.0435	0.0000
	AC	0.7422	0.2180	0.0435	0.0000
	Borkowf	0.7397	0.2153	0.0000	0.0000
	Jeffreys	0.7674	0.2180	0.0000	0.0000

표 3: n과 p에 따른 기대폭에서 극단값($X = 0$)이 차지하는 비율

n	신뢰구간	기대폭			
		$p = 0.01$	$p = 0.05$	$p = 0.1$	$p = 0.2$
10	Exact	0.8673	0.4985	0.2541	0.0656
	Mid-p	0.8589	0.4799	0.2397	0.0607
	Wilson	0.8708	0.5075	0.2620	0.0688
	AC	0.8764	0.5226	0.2766	0.0754
	Borkowf	0.8564	0.4706	0.2288	0.0550
	Jeffreys	0.8459	0.4533	0.2199	0.0542
20	Exact	0.7494	0.2537	0.0701	0.0053
	Mid-p	0.7347	0.2381	0.0644	0.0048
	Wilson	0.7575	0.2640	0.0744	0.0057
	AC	0.7675	0.2777	0.0808	0.0065
	Borkowf	0.7327	0.2307	0.0598	0.0042
	Jeffreys	0.7130	0.2173	0.0571	0.0041
30	Exact	0.6494	0.1337	0.0208	0.0005
	Mid-p	0.6310	0.1235	0.0188	0.0004
	Wilson	0.6604	0.1412	0.0224	0.0005
	AC	0.6730	0.1506	0.0247	0.0006
	Borkowf	0.6290	0.1179	0.0171	0.0004
	Jeffreys	0.6048	0.1106	0.0164	0.0004

다. 따라서 모비율의 모든 구간에 대해 포함확률과 신뢰구간의 기대폭을 평균한 평균포함확률과 평균기대폭이 신뢰구간의 평가기준으로 또한 널리 사용된다. 포함확률과 마찬가지로 평균포함확률도 표 4에 의하면 $X = x$가 차지하는 비율은 거의 비슷하다. 즉, 6개 신뢰구간들은 n의 크기에 관계없이 평균포함확률에서 극단값이 차지하는 비율이 거의 동일하고, n이 증가하면 극단값의 비중은 줄어든다. 하지만 신뢰구간의 평균기대폭은 표 3에서와 같이 방법들 간에 차이가 있다. 즉, 평균기대폭에서 극단값의 영향이 AC신뢰구간이 가장 큰 반면에 Jeffreys사전신뢰구간이 가장 작음을 알 수 있다. 이는

표 4: 평균포함확률과 평균기대폭에서 극단값인 $X = 0$과 n이 차지하는 비율

n	신뢰구간	평균포함확률		평균기대폭	
		$X = 0$	$X = n$	$X = 0$	$X = n$
5	Exact	0.1665	0.1660	0.1285	0.1281
	Mid-p	0.1660	0.1655	0.1220	0.1216
	Wilson	0.1686	0.1686	0.1299	0.1296
	AC	0.1696	0.1691	0.1388	0.1384
	Borkowf	0.1653	0.1648	0.1134	0.1131
	Jeffreys	0.1644	0.1638	0.1130	0.1126
10	Exact	0.0911	0.0906	0.0553	0.0550
	Mid-p	0.0907	0.0902	0.0512	0.0509
	Wilson	0.0924	0.0924	0.0581	0.0578
	AC	0.0932	0.0927	0.0641	0.0637
	Borkowf	0.0897	0.0892	0.0457	0.0454
	Jeffreys	0.0892	0.0887	0.0458	0.0455
20	Exact	0.0480	0.0480	0.0220	0.0218
	Mid-p	0.0480	0.0480	0.0199	0.0197
	Wilson	0.0490	0.0490	0.0237	0.0235
	AC	0.0492	0.0490	0.0270	0.0267
	Borkowf	0.0470	0.0465	0.0178	0.0176
	Jeffreys	0.0466	0.0461	0.0173	0.0171
30	Exact	0.0327	0.0322	0.0126	0.0124
	Mid-p	0.0324	0.0319	0.0112	0.0110
	Wilson	0.0333	0.0328	0.0136	0.0134
	AC	0.0328	0.0322	0.0190	0.0187
	Borkowf	0.0320	0.0314	0.0101	0.0099
	Jeffreys	0.0316	0.0311	0.0096	0.0095

Jeffreys사전신뢰구간의 평균기대폭이 가장 짧고 AC신뢰구간이 항상 길다는 것을 이론적으로 입증한 Brown 등 (2002)과도 연관된다.

5. 토의 및 결론

이항비율에 대한 신뢰구간의 적절성을 평가하기 위해서 포함확률과 신뢰구간의 기대폭이 널리 사용된다. 포함확률이 신뢰수준에 근사하고 신뢰구간의 기대폭이 가능한 한 작은 것이 선호되는데, Wald신뢰구간은 포함확률이 지나치게 작고 정확신뢰구간은 기대폭이 지나치게 크다. 이를 보완해주는 다양한 신뢰구간들이 제시되고 상황에 따른 방법들이 추천된다. 하지만 대부분의 신뢰구간들은 이항 모비율 p가 0이나 1에 근사한 경우 포함확률이 신뢰수준을 크게 벗어난다. 정확신뢰구간, Mid-p신뢰구간, AC신뢰구간 등은 신뢰수준보다 크고 Wald신뢰구간은 작으며 Wilson신뢰구간이나 Jeffreys사전신뢰구간은 변동이 크다. 특히 Wald신뢰구간은 극단값에서 신뢰구간을 제공하지 못하므로 다른 방법으로 대체해야 된다는 의견이 일반적이다. 류제복 (2010)은 Wald신뢰구간에서 극단값을 대체하였을 경우 포함확률이 크게 개선됨을 보여주고 있다.

이러한 극단값에서의 영향력을 조정해주기 위해서 Vollset (1993), Agresti와 Coull (1998), Newcombe (1998), Brown 등 (2001), 그리고 류제복 (2010) 등은 이항비율에 대한 구간추정에서 극단값에서의 대체를 제안하고 있다. 따라서 본 연구에서는 이항비율에 대한 신뢰구간 추정에서 극단값의 영향력을 대표적인 6개의 신뢰구간에 대해서 수치적으로 비교하였다. 비교 결과 n과 p가 작은 경우 6개 신뢰구간 모두에서 극단값에 민감함을 확인 할 수 있다. 포함확률에서는 신뢰구간들 간에 차이가 작은

반면에 신뢰구간의 기대폭은 차이가 있음을 알 수 있다. 특히 AC신뢰구간이 기대폭에서 극단값의 영향이 가장 크고 Jeffreys사전신뢰구간이 가장 작다. 한편 평균포함확률은 극단값에서 모든 $X = x$가 차지하는 비율은 거의 같으나 평균기대폭은 차이가 있음을 확인할 수 있다.

참고 문헌

류제복 (2010). Wald 신뢰구간에서 극단값 조정의 효과, <산업과학연구>, **28**, 29–34.

류제복, 이승주 (2006). 낮은 이항 비율에 대한 신뢰구간, <응용통계연구>, **19**, 217–230.

Agresti, A. and Coull, B. A. (1998). Approximate is better than “Exact” for interval estimation of Binomial proportions, *The American Statistician*, **52**, 119–126.

Blyth, C. R. and Still, H. A. (1983). Binomial confidence intervals, *Journal of the American Statistical Association*, **78**, 108–116.

Borkowf, C. B. (2006). Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failure or success, *Statistics in Medicine*, **25**, 3679–3695.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion), *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.

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, **26**, 403–413.

Ghosh, B. K.(1979). A comparison of some approximate confidence intervals for the Binomial parameter, *Journal of the American Statistical Association*, **74**, 894–900.

Jovanovic, B. D. and Levy, P. S. (1997). A look at the Rule of three, *The American Statistician*, **51**, 137–139.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods, *Statistics in Medicine*, **17**, 857–872.

Schader, M. and Schmid, F. (1990). Charting small sample characteristics of asymptotic confidence intervals for the binomial parameter, *Statistical Papers*, **31**, 251–264.

Vollset, 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.

2011년 7월 접수; 2011년 8월 채택

The Influence of Extreme Value in Binomial Confidence Interval

Jea-Bok Ryu[1,a]

[a]Division of Natural Science, Cheongju University

Abstract

Several methods are used in interval estimation for binomial proportion; however the coverage probabilities of most confidence intervals depart from the confidence level when the binomial population proportion closes to 0 or 1 due to the extreme value. Vollset (1993), Agresti and Coull (1998), Newcombe (1998), and Brown *et al.* (2001) suggested methods to adjust the extreme value. This paper discusses the influence of extreme value in a binomial confidence interval through the numerical comparison of 6 confidence intervals.

Keywords: Binomial proportion, confidence interval, extreme value, coverage probability, expected width.

This work was supported by the research grant of Cheongju University in 2010.

[1] Professor, Division of Natural Science, Cheongju University, 298 Daeseongro, Sangdang-gu, Cheongju, Chungbuk 360-764, Republic of Korea. E-mail: jbryu@cju.ac.kr