한 국 통 계 학 회 논 문 집
2009, 16권, 5호, 731–743

이항신뢰구간에 대한 소고

류제복[1,a]

[a]청주대학교 생명·유전·통계학부

요 약

이항비율에 대한 구간추정의 문제는 오래전부터 많이 다루어져 왔다. 본 논문에서는 주요 신뢰구간들의 특성을 비교하고 신뢰구간의 평가기준인 포함확률과 신뢰구간의 길이에 대해 이제까지 다루어져온 문제들을 종합 정리해 보았다. 실제로 이항신뢰구간 문제를 다룰 때 고려해야 할 3가지 추가 사항들을 살펴보고, 이항비율 추정에 늘 문제가 되는 낮은 이항비율에 대한 향후 논의 사항들을 제시하였다.

주요용어: 이항비율, 신뢰구간, 포함확률, 신뢰구간길이.

1. 서론

이항비율의 신뢰구간 추정에 대한 연구는 상당히 많이 이루어지고 있다. 다양한 방법들이 제시되고 방법들 간의 비교를 위한 적절한 기준의 설정과 이의 적합성에 대한 논의가 계속되고 있다. 하지만 어느 누구도 최적의 신뢰구간을 제시하지 못하고 심지어 연구자들 간에도 의견의 차이를 보이고 있다. 다만 한 가지 공통점은 지금까지 널리 사용되고 있는 Wald신뢰구간이 더 이상 사용되어서는 곤란하다는 것이다 (Ghosh, 1979; Blyth와 Still, 1983; Vollset, 1993; Newcombe, 1998; Agresti와 Coull, 1998; Brown 등, 2001, 2002, 2003). 심지어 기존 교과서에서 Wald신뢰구간을 대체해야 한다고 강력히 주장하기도 한다. 응용분야에서는 검정보다는 구간추정이 조사결과를 보다 효율적으로 사용하는 측면이 강해서 응용통계분야에서는 신뢰구간이 보다 널리 사용되고 있다. 특히 통계학을 전공하지 않은 사람들이 가설검정에서 얻게 되는 확률값인 p-값을 이해하기가 쉽지 않고 이를 조사 자료와 연관시켜서 설명하기는 더 어렵다. 가설검정에서 1종오류인 α가 의미하는 바는 귀무가설이 옳음에도 불구하고 귀무가설을 기각할 확률이다. 따라서 귀무가설의 채택역을 역변환해서 얻은 신뢰구간이 모수를 포함할 확률 $1-\alpha$를 신뢰수준(confidence level)이라고 한다(검정이론에서 일양 최량기각력 또는 최량기각역 등을 이용하면 최적의 채택역을 설정할 수 있고 이를 바탕으로 신뢰구간의 선정이 가능하다. 이때 Blyth와 Still (1983)이 제시한 기준들을 만족하는 신뢰구간을 선정하면 바람직한 신뢰구간을 얻을 수 있다).

신뢰수준, 포함확률과 신뢰계수를 혼용해서 사용하고 있는 경우가 있는데, 이들을 정확히 정의하면 모든 모수값에서 신뢰구간이 모수를 포함할 확률을 신뢰수준이라 하고. 특정 모수값에서 구한 신뢰구간이 그 모수를 포함할 확률을 실제포함확률(actual coverage probability)이라 하는데, 이는 신뢰수준보다 크거나 작을 수 있다. 그리고 통계이론에서는 포함확률 중에서 가장 작은 포함확률을 신뢰계수(confidence coefficient)라 정의해 주고 있으나 정작 실무자들은 이를 평균적 개념으로 해석하고 있다 (Agresti와 Coull, 1998). 정확신뢰구간에서는 신뢰계수가 신뢰수준보다 약간 크다 (Angus와 Schafer, 1984).

이 논문은 2008년도 청주대학교가 지원하는 연구년으로 연구되었음.
[1] (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 교수. E-mail: jbryu@cju.ac.kr

과거에는 신뢰구간에 대한 평가 시 중요한 기준으로 사용한 포함확률이 $1-\alpha$를 넘어야만 그 신뢰구간이 가치를 인정받았다. 그러나 이러한 기준은 안전성을 지나치게 추구하여 포함확률이 설정한 명목수준(또는 신뢰수준)보다 커서 보수적(conservative)이라 비현실적이라는 의견이다. 따라서 최근에는 가급적 명목수준에 근사한 신뢰구간을 설정하는 것에 초점을 맞추고 있다.

이항비율에 대한 신뢰구간추정에서 늘 문제가 되는 것은 이항분포가 이산분포이기 때문에 연속성 가정이 어렵고, 특히 표본 규모가 작은 경우 정규성 이론의 적용이 곤란하다는 것이다. 그럼에도 불구하고 많은 신뢰구간들은 점근적 이론들에 기초를 두고 있다. 또한 이항분포에서 모비율 p가 0이나 1 근방에 있는 경우 신뢰구간 사용에 어려움이 많다. 그 이유는 이항자료에서 표본 규모가 크지 않을 경우 p에 대한 신뢰구간의 하한이나 상한이 0이나 1을 초과하는 경우가 발생하기 때문이다.

대표적인 Wald신뢰구간이 편향(bias)된 이유는 신뢰구간의 중심(center)이 치우쳐있기 때문인데(음으로 편향), Wilson이나 Agresti-Coull 신뢰구간은 중심이 1/2쪽으로 이동하여 편향이 어느 정도 제거되어 결과적으로 포함확률이 Wald보다 크게 된다. 또한 포함확률의 변동이 매우 심한 것은 이항분포의 이산성과 왜도(skewness)인데, 연속성 수정은 이산성을 어느 정도 완해해 주는 효과가 있어서 포함확률이 대체적으로 개선된다. 한편 Brown 등 (2002, 2003)은 이러한 편향과 진동의 정확한 평가를 위해 신뢰구간들의 포함확률과 기대길이에 대한 Edgeworth 전개를 사용하였다.

본 논문에서는 이항비율에 대한 신뢰구간 추정에서 지금까지 논의되고 있는 사항들을 상세히 검토해서 이 분야의 연구자들과 실무자들이 사용할 신뢰구간을 선정하는데 도움을 주고자 한다. 논문의 구성은 2장은 일반적으로 널리 사용되고 있는 신뢰구간들을 소개하고 이들의 특성을 비교하였다. 3장과 4장에서는 신뢰구간들의 평가에 가장 널리 사용되고 있는 포함확률과 신뢰구간의 길이에 대한 사항들을 다루며, 5장에서는 이항비율의 신뢰구간추정과 관련되는 추가 문제들을 다루었다. 끝으로 6장에서는 이항신뢰구간 추정에서 늘 문제가 되는 것은 낮은 이항비율의 경우에 검토할 몇 가지 사항들을 살펴보았다.

2. 신뢰구간들과 특성

이항비율의 구간추정에 사용되는 신뢰구간들은 매우 다양하고, 각 방법들이 나름대로의 특성이 있다. 여기서는 이론적으로나 실무적으로 널리 사용되고 있는 신뢰구간들을 중심으로 다룬다.

2.1. Wald신뢰구간

이항비율에 대한 구간추정에서 대표적인 신뢰구간으로 불리는 Wald신뢰구간은 이항비율에 대한 점근적 정규분포이론에 근거해서 유도된다. 이 신뢰구간의 중심은 모비율에 대한 최우추정치이고 상한과 하한의 계산에 추정치의 추정표준오차가 사용된다.

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{2.1}$$

여기서 $z_{\alpha/2}$는 표준정규분포의 $1-\alpha/2$분위수이다.

대부분의 통계학 교과서에서 이항모비율의 신뢰구간 추정에 Wald신뢰구간이 사용되는데, 이를 사용하기 위한 전제 조건들이 있다. Leemis와 Trivedi (1996) 그리고 Brown 등 (2001)이 추려낸 조건들은 다음과 같다.

1) $np,\ n(1-p) \geq 5$ (or 10)

2) $n\hat{p},\ n(1-\hat{p}) \geq 5$ (or 10)

3) $np(1-p) \geq 5$ (or 10)

4) $n > 9\max\{(1-p)/p,\ p/(1-p)\}$

5) $p \pm 2\sqrt{p(1-p)/n} \in (0,1)$

6) $\hat{p} \pm 3\sqrt{\hat{p}(1-\hat{p})/n} \notin (0,1)$

7) n이 아주 큰 경우

8) p가 아주 작지 않으면 $n \geq 50$

그러나 이러한 조건들이 충족된다 하여도 Wald신뢰구간의 사용이 바람직하지 않다는 것은 1장에서 언급한 바와 같이 많은 논문들에서 입증해주고 있다. 특히 Brown 등 (2001, 2002, 2003)은 그 원인을 상세히 다루었으며 대체 방법의 사용이 반드시 이루어져야 한다고 주장하고 있다.

2.2. Wilson신뢰구간

Wald신뢰구간 (2.1)은 추정치의 추정표준오차를 사용하지만, Wilson (1927)의 신뢰구간 (2.2)는 정규근사이론에 근거해서 귀무가설, $H_0 : p = p_0$에 대한 score검정을 역 변환한 $|\hat{p} - p_0|/\sqrt{p_0(1-p_0)/n} \leq z_{\alpha/2}$를 p_0에 관해 풀어서 얻는다.

$$\left(\hat{p} + \frac{z^2_{\frac{\alpha}{2}}}{2n} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p}) + z^2_{\frac{\alpha}{2}}/4n}{n}}\right) \Bigg/ \left(1 + \frac{z^2_{\frac{\alpha}{2}}}{n}\right). \tag{2.2}$$

Wald신뢰구간의 포함확률이 아주 빈약한 것은 신뢰구간의 중심점이 1/2에서 떨어져 편향이 되었기 때문인데, Wilson신뢰구간은 Wald신뢰구간의 중심 $\hat{p}$를 1/2쪽으로 이동시켜서 편향을 줄여주고 있다. Agresti와 Coull (1998)은 Wilson신뢰구간의 중심이 $\hat{p}$와 1/2의 가중평균이고, $z_{\alpha/2}$의 계수부분도 표본비율의 분산과 $p = 1/2$일 때의 표본비율의 분산의 가중평균 형태로 표본이 증가하게 되면 자료에 의존도가 커지게 됨을 보여주었다.

Wilson신뢰구간은 Wald신뢰구간의 단점을 보완해주고 정확신뢰구간보다 보수성이 덜하면서 포함확률이 명목수준에 근사하기 때문에 Schader과 Schmid (1990), Vollset (1993), Newcombe (1998), Santner (1998) Agresti와 Coull (1998), Brown 등 (2001, 2002) 외에 많은 이들이 추천하고 있다. 또한 Brown 등 (2002)은 약간의 조건만 갖추면 포함확률이 적어도 명목수준을 유지하면서 신뢰구간의 길이를 최소로 하는 신뢰구간이 Wilson신뢰구간임을 증명하였다.

2.3. Agresti-Coull신뢰구간

Agresti와 Coull (1998)은 Wilson신뢰구간의 신뢰수준을 95%로 두면 신뢰구간의 중심이 $(X + z^2/2)/(n+z^2) \approx (X+2)/(n+4)$가 되므로 "두 번의 성공과 두 번의 실패를 더한" Wald신뢰구간과 같아서 이를 수정된 Wald신뢰구간이라 하였다. 이후 수정된 Wald신뢰구간을 Agreti-Coull(AC)신뢰구간이라 부른다. 이 신뢰구간의 총 시행 수는 $\tilde{n} = n+4$이고 이항비율 p에 대한 점추정치는 $\tilde{p} = (X+2)/(n+4)$이다. 또한 중심은 사전분포가 $B(2,2)$인 Bayes 추정량과 같다. 여기서 X는 이항모집단에서의 확률변수로 성공의 수이다.

Wald신뢰구간에 n과 $\hat{p}$대신에 $\tilde{n}$와 $\tilde{p}$를 대입하면 AC신뢰구간이 된다. 즉,

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

AC신뢰구간은 Wald신뢰구간보다 포함확률이 크게 개선되며, 평균포함확률이 Wilson신뢰구간보다 크고 p가 0이나 1근처에 있을 때에도 포함확률이 심각하게 낮지 않다. 이 신뢰구간의 장점은 간편함에 있다. 비록 약간 보수적이고 신뢰구간의 기대 길이가 크지만 그리 우려할 만하지 않고 적은 표본에서도 사용이 가능하다. Agreti와 Coull이 1998년 논문을 발표한 이후 거의 모든 논문에서 이 신뢰구간의 사용을 추천하고 있다. Borkowf (2006)는 Agresti-Coull의 방법을 약간 수정해서 신뢰구간의 하한을 구할 때는 실패를 하나 더해주고 상한을 구할 때는 성공을 하나 더해주는 방법을 제시해주고 있으나 간편성이나 보수성 등을 고려하면 오히려 AC신뢰구간이 좋다 할 수 있다. 다만, 이는 95%신뢰수준에서만 그 원리를 사용할 수 있다. 일찍이 Wilson (1927)은 이 축소 추정치를 표본비율의 대체 추정치로 언급하였으며 2σ인 경우 $(X+2)/(n+4)$ 사용을 제안하였다.

2.4. 정확신뢰구간

Clopper와 Pearson (1934)은 $n = 20$이고 $p = .45$인 이항자료에서 95%, 99%신뢰구간의 하한과 상한을 선형보간법(interpolation)을 사용해서 누적확률이 정확히 .025와 .005가 되는 x(또는 x/n)를 얻었다. 정확신뢰구간이라고도 불리는 이 신뢰구간은 정규근사이론을 사용하지 않고 이항양측검정에 기초해서 식 (2.4)를 풀어 직접 신뢰 하한과 상한을 얻는다.

$$\sum_{k=x}^{n}\binom{n}{k}p^k(1-p)^{n-k} = \frac{\alpha}{2}, \qquad \sum_{k=0}^{x}\binom{n}{k}p^k(1-p)^{n-k} = \frac{\alpha}{2}. \tag{2.4}$$

이 신뢰구간은 모든 모수 값에서 포함확률이 명목수준 $(1-\alpha)$를 초과해서 보수적인 신뢰구간으로 간주된다.

$x = 1, 2, \ldots, n-1$일 때, 정확신뢰구간은 다음과 같다.

$$\left(1 + \frac{n-x+1}{xF_{2x,2(n-x+1),1-\frac{\alpha}{2}}}\right)^{-1} < P < \left(1 + \frac{n-x}{(x+1)F_{2(x+1),2(n-x),\frac{\alpha}{2}}}\right)^{-1} \tag{2.5}$$

$x = 0$일 때 하한은 0이고 $x = n$일 때 상한은 1이며 $F_{a,b,c}$는 자유도가 a, b인 F분포의 $(1-c)$분위수이다. 또한 이 신뢰구간의 하한은 모수가 x와 $n-x+1$인 베타분포의 $\alpha/2$분위수이고, 상한은 모수가 $x+1$과 $n-x$인 베타분포의 $(1-\alpha/2)$분위수와 같다. 정확신뢰구간의 단점으로 계산이 복잡한 것이 문제가 되었으나 최근에 들어서는 범용통계소프트웨어에서도 쉽게 계산된다.

임상시험 등의 보건의료 분야에서는 다소 보수적인 신뢰구간을 선호하는 경향이 있어서 이 방법이 널리 사용된다. 또한 다른 방법과의 비교를 위해서도 정확신뢰구간이 사용되며, Vollset (1993)은 신뢰구간의 포함확률이 p가 0이나 1에 가까운 경우 아주 작거나 크게 되는 문제를 해결하기 위해서 경계값을 정확신뢰구간으로 대체하고 포함확률의 개선 상태를 비교하였다.

2.5. Mid-p신뢰구간

정확신뢰구간은 최소포함확률이 항상 명목수준을 유지해서 매우 보수적인 신뢰구간이다. 정확신뢰구간을 얻는 식 (2.2)에서 관측 결과에 1/2의 확률을 부여한 식 (2.6)으로부터 Mid-p신뢰구간을 얻는다. 이 신뢰구간은 정확신뢰구간이 갖는 보수성을 완화시켜준다.

$$\begin{aligned}\frac{1}{2}\binom{n}{x}p^x(1-p)^{n-x} + \sum_{k=x+1}^{n}\binom{n}{k}p^k(1-p)^{n-k} &= \frac{\alpha}{2},\\ \frac{1}{2}\binom{n}{x}p^x(1-p)^{n-x} + \sum_{k=0}^{x-1}\binom{n}{k}p^k(1-p)^{n-k} &= \frac{\alpha}{2}\end{aligned} \tag{2.6}$$

이 신뢰구간의 상한은 $(1/2)B(1-\alpha/2; x, n-x+1) + (1/2)B(1-\alpha/2; x+1, n-x)$이고 하한은 대칭이다. Berry와 Armitage (1995)는 Mid-p신뢰구간에 대한 특성들을 상세히 설명해 주고 있다. 또한 Brown 등 (2001)은 Mid-p와 정확신뢰구간을 Jeffreys신뢰구간과 연계해서 보여주고 있다.

Vollset (1993)과 Newcombe (1998)는 명목수준에 근사한 신뢰구간을 원할 경우 Mid-p 신뢰구간의 사용을 추천하고 있으며, 류제복과 이승주 (2006)는 표본규모가 크고 p가 작은 경우 Mid-p신뢰구간의 사용이 적절함을 보여주고 있다. 또한 경계 값에 Mid-p신뢰구간을 사용하면 정확신뢰구간을 사용한 경우보다 변동이 적다.

2.6. Jeffreys prior 신뢰구간

앞에서 소개한 신뢰구간들은 이항비율에 대한 사전적 지식을 고려하지 않았다. 그러나 이항비율 p에 대한 사전정보를 활용할 수 있다면 매우 유용하다. 특히 모비율 p가 0이나 1에 근사한 경우 사전분포의 사용이 필요하다(자세한 것은 5장에서 다룸). Jeffreys 사전 신뢰구간은 객관적이고 대표적인 무정보사전분포인 Beta$(1/2, 1/2)$를 사용한 것으로 $100(1-\alpha)\%$ 동일 꼬리 Jeffereys 신뢰구간은 다음과 같다.

$$L_J(x) \le p \le U_J(x) \tag{2.7}$$

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

Brown 등 (2001, 2002)은 표본이 적을 경우 Jeffreys prior 신뢰구간의 사용을 추전하고 류제복과 이승주 (2006)도 p가 작고 표본이 큰 경우 이 신뢰구간의 사용이 적절함을 보여주고 있다. 한편 Jovanovic과 Levy (1997), Winkler 등 (2002) 그리고 Tuyl 등 (2008)은 p가 작은 경우 이에 대한 사전정보인 베타 분포족을 이용한 베이지안 방법의 사용을 제안하고 있다.

2.7. 그 밖의 신뢰구간

앞에서 소개한 신뢰구간들 이외에도 종종 사용되는 몇 가지 신뢰구간을 추가로 살펴본다.

연속성 수정에 의한 신뢰구간; 이항분포는 이산분포이므로 포함확률이 n과 p의 변화에 매우 민감하다. 따라서 기존의 신뢰구간에 간단히 연속성 수정을 하면 포함확률과 신뢰구간의 기대폭을 개선할 수 있다. Blyth와 Still (1983)은 연속성 수정이 필요함을 주장하고 Schader과 Schmid (1990), Vollset (1993) 그리고 Casella (Brown 등 (2001)의 토론자, pp. 120–122) 등도 연속성 수정을 한 Wilson신뢰구간을 강력히 추천하지만, Agresti와 Coull (1998) 그리고 Brown 등 (2001)은 연속성 수정을 한 Wilson신뢰구간이 너무 보수적이라서 사용에 부정적인 의견을 제시하고 있다. 이와 같이 연속성 수정에 대해서 각기 다른 의견이 있는데, 이는 연속성 수정이 모든 신뢰구간에서 긍정적인 효과를 주는 것은 아니고, 이를 사용할 관점에 따라 다르기 때문이다. 즉, 포함확률의 개선을 바라는 경우와 사용의 간편성을 주로 하는 경우에 따라 다르므로 사용자들이 여러 신뢰구간에 연속성 수정을 적용해서 그 결과를 보고 판단해야 할 것이다.

포아송근사에 의한 신뢰구간; 일반적으로 n이 크고 p가 작은 경우 이항비율의 신뢰구간 추정에 포아송근사를 사용하는 것이 바람직하다. Leemis와 Trivedi (1996)는 이항비율에 대한 신뢰구간 추정에 정규근사와 포아송근사를 비교하고 정규근사와 포아송근사의 사용에 대한 가이드 라인을 제시하였다. Brown 등 (2001)은 Jeffreys prior신뢰구간과 Wilson신뢰구간이 모수가 경계근처에 있을 때 포함확률이 급격한 하향 값을 가지므로 이에 대한 수정을 위해서 X가 0이나 n에 근사한 값일 때 포아송근사의 사용

을 제안하고 있다. 그러나 Newcombe (1998)가 지적하는 바와 같이 포함확률이 불필요하게 커지는 문제가 생긴다.

그밖에 이항분포에 대한 분산의 안정화를 위해서 Arcsine 변환에 의한 신뢰구간, 우도비 검정을 역변환해서 얻는 신뢰구간 그리고 로그오즈, $\lambda = \log(p/(1-p))$에 대한 Wald형태의 신뢰구간을 역변환해서 얻는 로짓신뢰구간이 사용되기도 한다.

3. 포함확률(Coverage probability)

2장에서 이항비율에 대한 다양한 신뢰구간 방법들이 소개되었다. 그러나 어떤 신뢰구간을 사용할지를 정할 때 선정 기준이 필요한데, 그 첫 번째 기준으로 포함확률이 사용된다. 포함확률은 특정 모수가 신뢰구간에 포함될 확률로 다음과 같이 정의 된다.

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

여기서 $I(k,p)$는 신뢰구간이 $X = k$일 때 p를 포함하고 있으면 1이고 그렇지 않으면 0이다. 포함확률은 이항모수 n과 p의 함수로 이들 값에 따라 변한다. 그런데 이항분포는 이산형이기 때문에 n과 p의 조그만 변화에도 포함확률의 변동이 매우 크다. 특히 n과 p가 작은 경우 그 정도가 심하지만, n이 크고 p가 0.5에 가까운 경우도 이러한 현상이 생긴다.

포함확률의 이러한 변동 원인을 규명하기 위해서 Ghosh (1979)는 점근적 전개방법을 이용해서 포함확률을 순위 $n^{-1/2}$, n^{-1} 그리고 오차항 $O(n^{-3/2})$으로 구분하고 순위 $n^{-1/2}$까지는 포함확률이 항상 $1-\alpha$보다 크다고 하였다. 한편 Brown 등 (2002)은 2항 Edgeworth 전개를 이용해서 포함확률을 다음과 같은 5개 항으로 분해하고 각 항의 성질과 포함확률에의 영향을 상세히 보여주었다.

$$C_n(p) = (1-\alpha) + \text{"}O\left(n^{-\frac{1}{2}}\right)\text{ 진동"} + \text{"}O\left(n^{-1}\right)\text{ 고정"} + \text{"}O\left(n^{-1}\right)\text{ 진동"} + O\left(n^{-\frac{3}{2}}\right) \tag{3.2}$$

그런데 우측 두 번째 항인 "$O(n^{-1/2})$진동"은 이산성에 의한 반올림 오차(rounding error)와 왜도에 의한 오차로 구성되는데, Ghosh (1979)는 반올림 오차에 의한 항을 누락했기 때문에 포함확률이 순위 $n^{-1/2}$까지는 항상 $1-\alpha$보다 크다는 오류를 범하였다. 물론 양측 신뢰구간에서는 반올림오차는 커지고 왜도에 의한 오차는 $O(n^{-1})$로 줄어든다. Brown 등 (2002)은 포함확률에서 핵심인 $O(n^{-1})$항을 고정부분과 진동부분으로 나누고, 비중이 크고 편향을 일으키는 고정 부분의 비교를 통해서 여러 신뢰구간들의 포함확률을 비교하였다.

이항 비율에 대한 신뢰구간 문제를 다룰 때 거의 대부분의 논문들은 비교 대상인 신뢰구간의 포함확률에 대한 그래프나 수치적 계산 결과를 바탕으로 비교하였는데, Edgeworth 전개를 이용한 Brown 등 (2002, 2003)의 해부학적 연구는 포함확률이 갖고 있는 근본적인 문제를 다루었다는 점에서 매우 의의가 크다. 그렇지만, 실무자들이 사용할 신뢰구간을 결정할 때는 여전히 어려움이 있다. 이유는 상황에 따라 신뢰수준과 표본규모, 그리고 모수값에 대한 정보의 유무 등이 다르고, 이론적 타당성과 함께 계산의 편리성이나 분야별로 요구되는 안정성에 차이가 있기 때문이다. 앞서 언급한 바와 같이 포함확률은 표본규모와 모수 값에 의존하기 때문에 모수 구간 전체에 대해 일정 수준의 포함확률을 평가 기준으로 삼는 것이 바람직하다.

우리가 알고자 하는 모수값이 어느 위치에 있는지를 모르므로 모든 모수 구간에 대해서 일정 수준의 포함확률을 유지하는 것이 필요하다. 대부분의 신뢰구간의 포함확률은 모비율 p가 0.5근방에 있는 경우 명목수준에 근사하게 된다. 그러나 p가 0.5로부터 멀어지게 되면 명목수준에 훨씬 못 미치거나

크게 상회하는 등 그 차이가 크다. Stein (1985)과 Woodroofe (1986)는 모비율에 대한 사전분포를 이용한 평균포함확률 (3.3)을 신뢰구간의 평가 기준으로 사용하였다. 즉, 모비율의 모든 범위에 대해 포함확률의 기댓값을 구하는 것이다.

$$\bar{C}_n(p) = \int_0^1 C_n(p)f(p)dp. \tag{3.3}$$

비록 평균포함확률이 명목수준에 근사하지만, 여러 모수 값에서 포함확률이 명목수준에 훨씬 못 미친다면 그 신뢰구간을 사용할 수 있는가? 이점을 보완하기 위해서 최소포함확률의 개념을 사용하게 되었다 (Newcombe, 1998; Brown 등, 2001; Reiczigel, 2003; Borkowf, 2006). 최소포함확률은 최소한의 안정성을 보장하기 위한 수단으로 사용되나 포함확률이 특별히 작은 경우로 인해 신뢰구간의 평가가 정확해지지 않을 수 있다. 이런 단점을 보완하기 위해서 Agresti와 Coull (1998) 그리고 Reiczigel (2003)은 포함확률이 일정한 범위에 속하는 비율을 이용하여 신뢰구간을 평가하였다. 즉, 각 신뢰구간들에 대해 포함확률이 명목수준의 일정 범위에 있는(또는 벗어나는) 모수 공간의 비율을 사용한다. 한편 Newcombe (1998)는 신뢰구간의 위치를 하나의 기준으로 사용하였다. 즉, 모비율이 $[0,1]$사이에서 대칭이면 왼쪽과 오른쪽의 비포함확률도 대칭이지만 모비율이 $[0,0.5]$사이에 있을 때는 왼쪽과 오른쪽 비포함확률이 다를 수 있으므로 이들을 위치에 대한 지표로 사용하였다.

연속분포의 경우와 달리 이산인 이항분포에서의 포함확률은 모수값에 따라 다르기 때문에 평균포함확률과 최소포함확률(신뢰계수)은 모수 공간을 일정 수로 나눈 상태에서 모의실험으로 구한다. 평균포함확률이나 최소포함확률을 보다 정확하게 구하기 위해서는 모수 공간을 가급적 많은 구간으로 쪼개어 계산을 하게 되므로 많은 계산이 필요하게 된다. 이에 Wang (2007, 2009)은 몇 가지 조건하에서 관측값에 따른 신뢰구간의 상한과 하한 값들을 이용해서 최소포함확률과 평균포함확률을 정확히 구하는 과정을 제시하였다. 이 방법은 기존의 모의실험에 의한 방법보다 계산과정을 대폭 줄여준다.

모수 전체에 대한 평균포함확률은 모수에 대한 사전 정보가 전혀 없는 경우 사용이 적절하지만, 과거 경험이나 자료로부터 모수에 대한 정보가 있을 때 이를 활용하면 신뢰구간 선정에 매우 유용하다. 류제복과 이승주 (2006)는 사전정보($p \le 0.1$)가 있는 국내기생충조사 자료에서 이항 비율의 추정에 적절한 신뢰구간의 방법들을 비교하고 있다. 특히 모비율이 0이나 1근방에 있을 때 신뢰구간의 포함확률에 많은 차이가 있고 이들이 평균포함확률에 크게 영향을 주므로 사전 정보를 충분히 반영하는 것이 바람직하다. 무정보사전분포를 활용하는 베이지안 신뢰구간 방법은 사전정보의 오류를 최소화하고 객관적인 입장을 취하는 장점이 있지만 활용할 수 있는 정보를 충분히 사용하지 못해서 추정의 효율을 떨어뜨리는 문제도 안고 있다.

만약 사용하고자 하는 신뢰구간의 포함확률이 명목수준(신뢰수준)과 비교해서 작으면 추정의 신뢰성이 낮아 그 신뢰구간을 사용하기가 곤란한 반면에 지나치게 포함확률이 크게 되면 불필요한 낭비가 된다. 따라서 어떤 수준이 적절한가? 초기에는 포함확률이 신뢰수준$(1-\alpha)$을 초과하는 것이 바람직하다는 의견이 지배적이었다. 아직도 Corcoran과 Mehta (Brown 등 (2001) 논문의 토론자)는 검정통계량의 이산성으로 인해 요구되는 신뢰수준을 얻기 어려우므로 과소 포함확률보다는 과대 포함확률이 보다 안전하고, 요구수준보다 더 신뢰할 수 있는 신뢰구간을 제시해 주므로 손해 볼 일이 없다며 포함확률의 보수성을 옹호하고 있다. 반면에 Brown 등 (2001)은 과대 포함확률이 과소포함확률만큼 나쁘다는 주장을 하고 있는데, 최근의 전반적이 추세는 포함확률이 명목수준을 초과하거나 못 미치는 경우보다는 최대한 명목수준에 근사한 신뢰구간을 요구하고 있다. 이에 대한 Newcombe (1998)의 "최소포함확률은 명목수준에 약간 못 미치고 평균포함확률은 명목수준을 약간 초과하지만 거의 같게 되는 신뢰구간이 바람직하다"는 의견이 적절한 표현이라 생각된다.

4. 신뢰구간의 길이

신뢰구간 평가의 또 다른 기준으로 신뢰구간의 길이가 사용된다. 3장에서 다룬 포함확률은 특정 모수가 신뢰구간에 포함될 확률을 나타내는데, 일반적으로 신뢰구간의 길이가 길으면 포함확률은 커지게 된다. 그러나 포함확률은 커지지만 신뢰구간의 길이가 지나치게 길면 그 신뢰구간은 추정의 의미가 없게 된다. 따라서 연구자들은 가급적 요구 수준의 포함확률을 유지하면서 신뢰구간의 길이가 짧은 신뢰구간을 선호하게 된다. 그런데, 신뢰구간도 모수 값에 따라 결정되므로 신뢰구간의 기대길이는 표본 크기 n과 모비율 p가 주어졌을 때 $E(n,p) = \sum_{x=0}^{n}(U(x) - L(x))\binom{n}{x}p^x(1-p)^{n-x}$가 된다. 하지만 기대길이도 모비율 p에 따라 달라지므로 포함확률의 경우와 마찬가지로 모든 모수 공간에서의 평균 기대길이 (4.1)을 신뢰구간 선정의 기준으로 사용한다.

$$\bar{E}(n,p) = \int_0^1 E(n,p)dp. \tag{4.1}$$

$100(1-\alpha)\%$신뢰구간을 구할 때 양쪽 꼬리부분에 동일하게 $\alpha/2$의 확률을 부여하는 경우와 그렇지 않은 경우 포함확률이 같다 해도 신뢰구간의 길이는 달라진다. 물론 정규근사를 이용하는 경우는 대칭이므로 양쪽에 동일 확률을 부여한다. 2.4절에서 다룬 정확신뢰구간도 양쪽 꼬리 부분에 동일 확률을 부여해서 신뢰구간을 얻으므로 신뢰구간의 길이가 길다. 그래서 포함확률이 지나치게 커지는 문제가 발생한다. 이런 점을 보완하기 위해서 Stern (1954)은 가장 보편적인(popular) X을 포함하는 신뢰구간을 구하고 다음으로 보편적인 X을 포함하는 신뢰구간을 구하는 과정을 반복해서 총 확률이 $1-\alpha$보다 같거나 크게 될 때까지 신뢰구간을 구축하였다. 또한 Crow (1956)는 하나의 신뢰구간 구축에 사용되는 가능한 모든 신뢰구간의 길이의 합이 최소가 되는 신뢰구간 족을 제시하였다. 그러나 Blyth와 Still (1983)은 신뢰구간을 구축할 때는 일정 조건을 만족해야 한다며 4가지의 조건들을 제시하고, Crow (1956)의 신뢰구간은 자신들의 조건에 위배된다고 지적하였다. 하지만 Reiczigel (2003)은 Blyth와 Still (1983)의 방법이 너무 복잡해서 적절치 않으므로 신뢰구간의 기대 길이가 작고 평균포함확률에 대해 자동적으로 조정되는 Stern (1956)의 방법이 적절하다고 주장하였다. 요구되는 포함확률을 유지하면서 기대 길이가 짧은 신뢰구간을 구하는 과정은 매우 복잡한데, Casella (1986)는 동일 신뢰수준 하에서 신뢰구간의 길이를 항상 짧게 해주는 알고리즘을 제시하였다. Ghosh (1979)는 |포함확률 − 명목수준| = $|\alpha(p,n) - \alpha|$이 작은 경우에 두 신뢰구간 길이를 비교하는데 확률적 개념을 사용하였다.

Brown 등 (2002)에 의하면 신뢰구간의 기대 길이에 대한 2항 Edgeworth 전개는 포함확률의 경우와 달리 순위 $n^{-1/2}$, $n^{-3/2}$ 그리고 오차항 $O(n^{-2})$으로 구성되고, $n^{-1/2}$항의 계수는 모든 신뢰구간에서 같고 $n^{-3/2}$항의 계수만 다르므로 이것이 각 신뢰구간의 길이에 영향을 준다. 그리고 평균 기대 길이는 Wald신뢰구간과 Wilson신뢰구간이 순위 $n^{-3/2}$까지 같으며, Wald, Wilson, AC, Jeffreys, 우도비신뢰구간들을 비교할 때 Jeffreys신뢰구간의 평균 기대 길이가 항상 짧고 AC신뢰구간이 항상 길다는 것을 이론적으로 입증하였다.

포함확률과 함께 신뢰구간의 길이가 중요한 평가 기준이지만, 표본크기가 작을 경우에는 기대 길이의 차이는 어느 정도 분명하지만 표본크기가 증가하게 되면 평균 기대 길이의 차이는 미미해져서 이를 평가기준으로 사용하기에 적절치 않다.

3장과 4장에서 다룬 2가지의 신뢰구간의 평가 기준 이외에도 포함확률의 변동을 측정하는 평균제곱오차의 제곱근이나 포함확률이 요구수준(명목수준)을 얼마나 잘 충족시키고 있는 가를 나타내는 절대오차(또는 상대오차)를 사용하기도 한다. 또한 Ghosh (1979)는 바람직한 신뢰구간은 잘못된 모수값을 포함할 확률이 작아야 하므로 Neyman shortneess를 사용하였고, Schader와 Schmid (1990)는 이들을 그림으로 그려서 신뢰구간들을 비교하였다.

5. 추가 고려사항

이항비율에 대한 구간추정에서 현실적으로 고려해야 할 추가적인 사항들을 살펴본다.

5.1. 종속이항자료의 경우

앞에서 다룬 신뢰구간은 자료들이 독립인 이항자료의 경우이다. 그러나 자료들이 독립이 아니고 종속인 경우는 이 성질을 구간추정에 반영해 주어야 한다. Miao와 Gastwirth (2004)는 주경찰관을 살해한 혐의로 구속 기소된 Moultrie V. Martin이 흑인 3명과 백인 15명으로 구성된 대배심원에서 흑인의 대표성이 낮다고 이의를 제기한 문제를 종속이항비율에 대한 신뢰구간 추정에 적용하였다. 문제가 제기된 South Carolina의 대배심원은 매년 새롭게 선정된 12명(자격이 갖추어져 있는 목록에서 지역의 인구구성 비율에 맞게 랜덤 추출)과 전년도에 새롭게 선출된 12명 중에서 6명을 랜덤하게 추출하여 총 18명으로 구성된다. 18명의 대배심원 중에서 6명이 전년에 이어 연임되므로 구성된 대배심원의 인종구성은 통계적으로 독립이 아니다.

독립인 확률변수 X_i를 i번째 해의 흑인 대배심원 수이고 p를 그 지역에서 대배심원의 자격을 갖춘 흑인의 비율이라 하면, $S_N = \sum_{i=1}^{n} X_i$은 이항분포를 한다. 그러나 Moultrie의 예에서 매년 대배심원이 중복해서 선출되므로 n해 동안 흑인 대배심원의 총수 S_N은 더 이상 이항분포를 하지 않는다. 결과적으로 종속을 무시한 경우, 1972–1977년 6년 동안에 연속성 수정된 p값은 0.059인 반면에 종속을 고려한 경우의 p값은 0.129로 큰 차이가 난다. 독립이 아닌 자료에 대해서 기존의 신뢰구간을 사용하면 포함확률이 명목수준에 못 미치게 된다. 따라서 Miao와 Gastwirth (2004)는 종속구조를 고려한 변형된 신뢰구간을 제시하고 실제 종속된 자료들에 대해 적용하고 비교하였다. 즉, 독립인 경우 S_N의 분산 $V(S_N) = Np(1-p)$ 대신에 종속자료에 대한 S_N의 분산은 $V(S_N) = N^{*}p(1-p)$을 사용하였다. 여기서 N은 독립이 경우이고 N^{*}는 종속을 고려한 경우의 대배심원 총 수이다. 그러나 신뢰구간들에 대한 비교 결과는 종전의 연구와 같이 (Brown 등, 2001) 종속을 고려해서 수정한 Wilson, A.C. 그리고 Jeffreys 신뢰구간을 추천하고 있다. 하지만 종속의 정도가 큰 경우에는 변형된 신뢰구간의 포함확률이 지나치게 커지는 문제가 있다.

5.2. 복합표본설계에서 신뢰구간

이항비율에 대한 신뢰구간의 문제는 대부분이 단순확률추출의 경우이다. 특히 각 시행의 확률이 동일한 경우를 가정하므로 편의상 복원추출의 경우이다. 그러나 5.1절에서 다룬 바와 같이 독립이 아닌 경우에는 분산공식에서 공분산을 고려해야 하고 결과적으로 이를 반영한 신뢰구간을 사용해야 한다.

마찬가지로 단순확률추출에 의한 분산을 V_{srs}이라하고 복합표본설계에 의한 분산을 V_{com}이라 하면 설계효과(design effect)는 Deff $= V_{com}/V_{srs}$가 된다. 따라서 복합표본설계에 의한 신뢰구간은 설계효과를 이용한 유효표본(efficient sample)인 $n^{*} = n \times$ Deff를 사용한다. 이때 n은 단순확률추출을 사용하였을 때의 주어진 허용오차를 만족시키는 표본크기이다. 신뢰구간의 형식은 그대로 단순확률추출의 형태를 유지하고 표본 n 대신에 n^{*}를 사용하여 간편하게 구할 수 있다. Kott 등 (2001)은 Wilson신뢰구간에 이 방법을 적용하였고, Gray 등 (2004)도 복합표본설계에서 유효 표본에 대해 언급하고 있으며 추가로 포함확률의 특성을 비교하기 위해서 bootstrap분석을 사용하였다.

5.3. Bayesian 접근

이항분포에서 모비율에 대한 신뢰구간추정에서는 모비율이 미지이고 사전에 이에 대한 가정이 불가한 경우를 다루고 있다. 그러나 신뢰구간의 평가에 사용되는 평균포함확률이나 평균기대길이 등을

계산할 때 미지 모수 p의 분포를 가정한다. Agresti와 Coull (1998)은 세 개의 무정보 사전베타분포를 사용한 결과 그 특성이 비슷하여 p의 사전분포로 균등분포를 사용하였으며, 다른 연구에서도 같은 가정을 하고 있다. 그러나 미지 모수에 대한 사전정보를 이용할 수 있다면 신뢰구간추정에 매우 도움이 된다.

모수에 대한 사전분포로 가정의 위험을 줄여주고 객관성을 높일 수 있는 공액사전분포인 베타분포가 가장 널리 사용되는데, 이의 밀도함수는 $\pi(p) = p^{a-1}(1-p)^{b-1}/B(a,b)$이고 평균은 $a/(a+b)$가 된다. 이러한 조건에서 n번 시행하여 x번 성공하면 p에 대한 사후분포는 Beta$(a+x,\ b+n-x)$가 되므로 이들 신뢰구간 추정에 이용한다. 이승천 (2005)은 이항비율에 대한 구간추정에 가중 Polya 사후 근사신뢰구간을 이용하여 표본이 작을 경우 AC신뢰구간의 보수성을 줄여 주었다.

특히, 이항비율 p가 0에 가깝다는 것이 알려져 있는 경우에는 사전분포의 활용이 용이한데, Jovanovic과 Levy (1997)는 p의 사전분포로 Beta$(1,b)$, $b > 1$가 적절하다고 주장하였다. 사전정보의 사용에 대해서 Winkler 등 (2002)은 $X = 0$일 때(관측값이 없을 때) 3의 법칙(신뢰구간의 상한으로 $3/n$을 사용)의 결과와 무정보사전분포 그리고 정보적사전분포를 사용한 결과를 Hanley와 Lippman-Hand (1983)의 예를 들어 비교하였다. 또한 그들은 무정보사전분포가 빈도학파들이 선호하는 보다 객관적인 사전분포이지만 어떤 무정보사전분포를 사용하는 가에 따라 결과가 달라지므로, 적절한 정보적 사전분포를 사용할 수 있다면 보다 유용한 결과를 얻을 수 있음을 주장하였다. 그리고 $X = 0$일 때 사후분포는 a의 선택에 매우 민감하므로 Jovanovic과 Levy (1997)가 제시한 바와 같이 사전분포를 Beta$(1,b)$로 제한하는 것은 너무 임의적이고 과도하므로 최소의 노력으로 풍부한 분포족을 제공해 주는 $a > 0, b > 0$인 Beta(a,b)분포 족의 사용을 제안하였다. Tuyl 등 (2008)도 이항모수에 대한 추정에 두 개의 모수가 동일한 Beta(a,a), $a = 0, 1/2, 1$인 무정보사전분포와 모수가 다른 Beta(a,b) 사전분포를 비교하였다. $X = 0$인 경우 일반적으로 널리 사용되는 Jeffrey 사전분포보다는 Bayes-Laplce 사전분포인 Beta$(1,1)$가 바람직함을 주장하였다. 한편 모수값이 작을 때는 Jovanovic과 Levy (1997)와 같이 Beta$(1,b)$, $b > 1$사전분포의 사용에 동의하지만, Winkler 등 (2002)의 주장과는 달리 $a < 1$을 때는 너무 정보적이라는 것을 예를 통해 보여주었다.

6. 토의

본 논문에서는 이항신뢰구간의 추정과 관련해서 널리 사용되고 있는 신뢰구간 방법들의 특성을 살펴보고 중요한 평가 기준에 대한 기존의 연구 결과를 종합 분석해 보았다. 신뢰구간을 평가하는 주요 기준들의 성질들을 상세히 분석한 Brown 등 (2002, 2003)의 연구 결과는 향후 신뢰구간의 선정에 반영되어야 한다. 그리고 제시된 다양한 대안들에 대해서도 역시 많은 추가적인 논의가 필요하다. 5장에서 다룬 신뢰구간 추정 시에 고려해야 할 3가지 사항들을 살펴보았는데, 실제 문제에서 신뢰구간을 이용할 때 이들을 사전에 충분히 점검해야 할 것이다.

이항신뢰구간 추정에서 늘 문제가 되는 것은 이항비율이 작은 경우이다. 이와 관련하여 향후 논의가 필요한 몇 가지 사항들을 언급하자 한다.

Vollset (1993), Agresti와 Coull (1998), Brown 등 (2001) 그리고 류제복과 이승주 (2006) 등 많은 연구자들에 의하면 모비율 p가 0이나 1근처에 있을 때 정확신뢰구간과 AC신뢰구간은 포함확률이 지나치게 크고, 다른 신뢰구간들은 지나치게 작다. 이들이 평균포함확률이나 평균기대길이에 영향을 주어 신뢰구간의 선정에 영향을 미친다. 따라서 이런 문제를 해결하기 위해서 관측이 $X = 0$(또는 n)이거나 그 근방의 값에 적절한 신뢰구간을 대체하여 극단적인 포함확률을 조정해 주기도 하였는데, 이것이 실제로 포함확률이나 기대길이에 어느 정도 영향을 주는 지를 비교해 보아야 한다. 그리고 모비율에 관한 사전적 지식이 있을 때 어떤 대체 방법이 적절한지를 검토해 보는 것이 필요하다.

모비율이 낮다는 것은 모비율에 대한 사전적 지식이 있다는 의미이므로 5.3절에서 다룬 베이지안적 접근이 필요하다. 어떤 사전분포의 사용이 신뢰구간 추정에 바람직하며, 특히 관측값이 없는 경우의 신뢰구간 추정문제를 고려해야 한다.

희귀질병에 대한 병리 검사결과 양성반응률, 새로운 치료약에 대한 부작용 발생률, 비행기의 추락확률, 제품의 불량률 등과 같이 발생 빈도가 매우 낮은 이항 모집단에서 관심 특성이 관측되지 않았을 때, $X = 0$, 미래 관측 자료에 대한 추정의 문제도 다루어져야 한다.

참고 문헌

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

이승천 (2005). 이항비율의 가중 Polya Posterior 구간추정, <응용통계연구>, **18**, 607–615.

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

Angus, J. E. and Schafer, R. E. (1984). Improved confidence statements for the Binomial parameter, *The American Statistician*, **38**, 189–191.

Berry, G. and Armitage, P. (1995). Mid-P confidence intervals: A brief review, *The Statistician*, **44**, 417–423.

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.

Brown, L. D., Cai, T. T. and DasGupta, A. (2003). Interval estimation in exponential families, *Statistica Sinica*, **13**, 19–49.

Casella, G. (1986). Refining binomial confidence intervals, *The Canadian Journal of Statistics*, **14**, 113–129.

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.

Crow, E. L. (1956). Confidence intervals for a proportion, *Biometrika*, **43**, 423–435.

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

Gray, A., Haslett, S. and Kuzmicich, G. (2004). Confidence intervals for proportions estimated from complex sample designs, *Journal of Official Statistics*, **20**, 705–723.

Hanley, J. A. and Lippman-Hand, A. (1983). If nothing goes wrong, Is everything all right?, *Journal of the American Medical Association*, **249**, 1743–1745.

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

Kott, P. S., Andersson, P. G. and Nerman, O. (2001). Two-sided coverage intervals for small proportions based on survey data, *Federal Committee on Statistical Methodology Research Conference*, 2001.

Leemis, L. M. and Trivedi, K. S. (1996). A comparison of approximate interval estimators for the bernoulli parameters, *The American Statistician*, **50**, 63–68.

Miao, W. and Gastwirth, J. L. (2004). The effect of dependence on confidence intervals for a population proportion, *The American Statistician*, **58**, 124–130.

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

Reiczigel, J. (2003). Confidence intervals for a binomial parameter: Some new considerations, *Statistics in Medicine*, **22**, 611–621.

Santner, T. J. (1998). Teaching large-sample Binomial confidence intervals, *Teaching Statistics*, **20**, 20–23.

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

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution, *Sequential Methods in Statistics*, **16**, 485–514.

Sterne, T. E. (1954). Some remarks on confidence or fiducial limits, *Biometrika*, **41**, 275–278.

Tuyl, R., Gerlach, R. and Mengersen, K. (2008). A comparison of Bayes-Laplace, Jeffreys, and other priors: The case of zero events, *The American Statistician*, **62**, 40–44.

Vollset, S. E. (1993). Confidence intervals for a binomial proportion, *Statistics in Medicine*, **12**, 809–824.

Wang, H. (2007). Exact confidence coefficients of confidence intervals for a binomial proportion, *Statistica Sinica*, **17**, 361–368.

Wang, H. (2009). Exact average coverage probabilities and confidence coefficients of confidence intervals for discrete distributions, *Statistic and Computing*, **19**, 139–148.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference, *Journal of the American Statistical Association*, **22**, 209–212.

Winkler, R. L., Smith, J. E. and Fryback, D. G. (2002). The role of informative priors in zero-numerator problems: Being conservative versus being candid, *The American Statistician*, **56**, 1–4.

Woodroofe, M. (1986). Very weak expansions for sequential confidence levels, *The Annals of Statistics*, **14**, 1049–1067.

2009년 7월 접수; 2009년 7월 채택

A Short Consideration of Binomial Confidence Interval

Jea-Bok Ryu[1,a]

[a]Division of Life Science·Genetic Engineering·Statistics, Cheongju University

Abstract

The interval estimation for binomial proportion has been treated practically as well as theoretically for a long time. In this paper we compared the properties of major confidence intervals and summarized current issues for coverage probability and interval length which are the criteria of evaluation for confidence interval. Additionally, we examined the three topics which were considered in using the binomial confidence interval in the field. And finally we discussed the future studies for a low binomial proportion.

Keywords: Binomial proportion, confidence interval, coverage probability, interval length.

This paper was researched by 2008 research year program of Cheongju University.

[1] Professor, Division of Life Science·Genetic Engineering·Statistics, Cheongju University, 36 Naedok-Dong, Sangdang-Gu, Cheongju, Chungbuk 360-764, Republic of Korea. E-mail: jbryu@cju.ac.kr