응용통계연구 제 19권 2호, 2006년, pp. 217-230

낮은 이항 비율에 대한 신뢰구간

류제복[1)] 이승주[2)]

요 약

본 연구에서는 낮은 이항비율에 관한 구간추정을 위해서 어떤 신뢰구간이 바람직한지를 살펴보았다. 실제적으로 희귀질병, 특정 산업재해율, 그리고 기생충에 관한 실태조사를 위해서 대규모 표본조사가 실시된다. 표본 규모가 크고, $0 < p \leq 0.1$인 상황에서 모비율 p의 추정에 바람직한 신뢰구간을 살펴보았다. 위의 조건에서 6가지의 신뢰구간들에 대해 평균포함확률과 평균제곱오차의 제곱근, 그리고 평균기대폭을 사용한 결과 Mid-p 신뢰구간이 가장 바람직하고 다음으로 AC, score와 Jeffrey 신뢰구간들이 적절한 것으로 밝혀졌다.

주요용어: 이항비율, 신뢰구간, 평균포함확률, 평균제곱오차의 제곱근, 평균기대폭

1. 머리말

성공의 확률, 제품의 불량률, 희귀질병 발생률, 특수 산업재해율, 그리고 특정 기생충의 감염률 등과 같은 이항비율에 관한 구간추정 문제는 생활 속에서 자주 발생한다. 이항비율의 추론에 사용되는 신뢰구간은 일반 통계학 교과서에서 Wald 신뢰구간이 가장 보편적으로 사용되고 있다. 역사적으로 가장 오래된 Wald 신뢰구간은 모비율에 대한 MLE를 이용하고 추정량의 정규근사를 이용한 것으로 사용하기에 간편하다. 그러나 Wald 신뢰구간을 사용하는 데는 $np \geq 5$(또는 10)나 $np(1-p) \geq 5$(또는 10)와 같은 조건이 만족되어야 한다. 이항비율 p가 0 또는 1에 근사하고 위의 조건이 충족되지 않은 경우 Wald 신뢰구간의 사용에는 문제가 있다. 특히 p가 0 또는 1에 가까운 경우 신뢰구간의 하한이 음수가 되거나 상한이 1을 초과하는 경우가 발생하여 신뢰구간이 성립하지 않게 된다.

Wald 신뢰구간 사용의 적합성에 대한 연구는 Vollset(1993), Agresti와 Coull(1998), Newcombe(1998), Brown, et al.(2001, 2002), Kott, et al.(2001), 이승천(2005) 등 많은 학자들에 의해서 이루어졌는데, 대부분의 연구에서 Wald 신뢰구간의 사용이 적절치 않다는 것을 보여주고 있다. 이들은 Wald 신뢰구간 대신에 다른 신뢰구간의 사용을 주장하고 있으며, 아울러 기존 통계학 교과서에서 전통적으로 사용되고 있던 Wald 신뢰구간 이외에 보다 적

1) (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 교수
E-mail: jbryu@cju.ac.kr
2) (교신저자) (360-764) 충북 청주시 상당구 내덕동 36번지, 청주대학교 생명·유전·통계학부, 부교수
E-mail: access@cju.ac.kr

절한 신뢰구간의 사용을 권장하고 있다(예를 들어, Wilson의 score 신뢰구간 또는 Clopper-Pearson의 정확신뢰구간). 한편 Ghosh는 Wald 신뢰구간을 새로운 신뢰구간으로 대체해야 한다는 의견에는 동의하지만, 새로운 신뢰구간들을 유도하는 것이 쉽지 않음을 피력하고 있다(Brown, et al.(2001), 125쪽).

신뢰구간은 표본크기 n과 모비율 p에 의존하므로 표본크기와 모비율에 따라서 선호되는 신뢰구간은 달라진다. 많은 선행 연구에서도 표본크기 n과 모비율 p에 따라 선호되는 신뢰구간이 다름을 보여주고 있다. 특히, p가 0 또는 1에 가깝고 표본크기가 작은 경우는 적절한 신뢰구간 선정에 어려움이 있다. 본 연구진들은 2004년도에 실시한 2만 3천여 명에 대한 장내 기생충 실태조사 자료로부터 기생충 감염률을 추정하는 과정에서 신뢰구간의 하한이 음수 값을 갖는 경우를 경험하였다. 산업재해율에 대한 추정도 유사한데, 감염률이나 산재율은 매우 낮았다. 이때의 표본크기는 기존 연구에서 다룬 규모보다 훨씬 큰 경우이다. 따라서 표본규모가 큰 대규모 조사에서 p가 0에 가까운 경우 어떤 신뢰구간의 사용이 적절한지를 살펴보고자 한다.

2장에서는 Wald 신뢰구간 이외에 대표적으로 많이 사용되고 있는 Clopper-Pearson의 정확(exact) 신뢰구간, Mid-p 신뢰구간, Wilson의 score 신뢰구간, Agresti-Coull 신뢰구간(조정된 Wald 신뢰구간이라고도 함), 그리고 Jeffreys prior 신뢰구간 등에 대해서 다루었다. 3장에서는 신뢰구간 선정 시 주요 기준이 되는 평균포함확률(mean coverage probability)과 포함확률이 신뢰수준으로부터 얼마나 떨어져 있는지를 나타내는 평균제곱오차의 제곱근(root mean square error), 그리고 평균기대폭(mean expected width)에 관해서 살펴보았다. 4장에서는 3장의 기준들을 사용해서 2장에서 다룬 6가지 신뢰구간들을 표본규모가 크고 p가 0에 가까운 경우에 비교하여 가장 바람직한 신뢰구간을 찾아보았다. 그리고 5장에서는 연구결과를 요약, 정리하였다.

2. 신뢰구간

이항비율에 대한 신뢰구간들 중에서 Wald 신뢰구간의 대안으로 제시되는 대표적인 신뢰구간들을 살펴본다. 물론 연구자들에 의해 선호되는 신뢰구간은 다를 수 있고, 연속성수정기법들도 좋은 대안으로 제시되고 있으나 여기서는 Vollset(1993), Agresti와 Coull(1998), Newcombe(1998), Brown, et al.(2001, 2002) 등이 사용한 신뢰구간들로 제한한다.

2.1. Wald 신뢰구간

이항비율에 대한 신뢰구간으로 널리 사용되고 있는 Wald 신뢰구간은 모비율 p에 대한 최우추정량 $\hat{p}$의 점근적 정규성에 근거한다. $100(1-\alpha)\%$ Wald 신뢰구간은 다음과 같다.

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

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

2.2. Clopper-Pearson의 정확 신뢰구간

Clopper-Pearson(1934)이 제안한 신뢰구간으로 정확(exact) 신뢰구간이라고 한다. Wald 신뢰구간이 점근적 정규성을 가정해서 얻는 반면에 정확신뢰구간은 이러한 가정의 한계를 피하고(정규성 가정은 p가 1/2에 근사하거나 n이 큰 경우), $H_0 : p = p_0$에 대한 이항양측검정에 기초해서 구한 것으로 다음과 같은 등식의 해가 신뢰한계점이 된다.

$$\sum_{k=x}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2, \quad \sum_{k=0}^{x} \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2. \tag{2.2}$$

단, $x = 0$일 때 하한은 0이고, $x = n$일 때 상한이 1이므로 이 경우를 제외한다. 이러한 구간추정은 모든 가능한 모수 값에 대해 포함확률이 최소 $(1-\alpha)$가 된다.

$x = 1, 2, \cdots, n-1$일 때, Clopper-Pearson의 정확신뢰구간은 다음과 같다.

$$\left[1 + \frac{n-x+1}{xF_{2x,2(n-x+1),1-\alpha/2}}\right]^{-1} < p < \left[1 + \frac{n-x}{(x+1)F_{2(x+1),2(n-x),\alpha/2}}\right]^{-1}. \tag{2.3}$$

여기서, $F_{a,b,c}$는 자유도가 a, b인 F 분포의 $(1-c)$ 분위수이다. 정확신뢰구간의 하한은 모수가 x와 $n-x+1$인 베타분포의 $\alpha/2$ 분위수이고, 상한은 모수가 $x+1$와 $n-x$인 베타분포의 $(1-\alpha/2)$ 분위수와 같다. 따라서 이러한 신뢰구간은 기존의 통계소프트웨어로 쉽게 계산할 수 있다(SAS, Minitab, S-plus 등).

2.3. Mid-p 신뢰구간

정확신뢰구간을 유도하는 식(2.2)에서 관측된 결과의 확률에 1/2를 부여한 식(2.4)로부터 신뢰구간을 유도한다. 이 신뢰구간을 Mid-p 신뢰구간이라 한다.

$$\begin{aligned} \frac{1}{2}\binom{n}{x} p^x (1-p)^{n-x} + \sum_{k=x+1}^{n} \binom{n}{k} p^k (1-p)^{n-k} &= \alpha/2, \\ \frac{1}{2}\binom{n}{x} p^x (1-p)^{n-x} + \sum_{k=0}^{x-1} \binom{n}{k} p^k (1-p)^{n-k} &= \alpha/2. \end{aligned} \tag{2.4}$$

Berry와 Armitage(1995)는 검정에서 Mid-p를 사용하는 이유와 구간추정에서 Mid-p 개념을 신뢰구간으로 확대하고 이항분포, 포아송분포, 그리고 2×2 분할표에서 Mid-p 신뢰구간 문제를 상세히 다루고 있다.

2.4. Wilson의 score 신뢰구간

Wald 신뢰구간은 추정치의 표준오차를 사용하는데, Wilson은 표준오차 대신에 $H_0 : p = p_0$에 대한 점근적 정규검정을 기초로 한 식(2.5)를 p_0에 대해 풀어서 신뢰구간 식(2.6)을 얻었다.

$$|\hat{p} - p_0| / \sqrt{p_0(1-p_0)/n} \le z_{\alpha/2}. \tag{2.5}$$

$$\left(\hat{p}+\frac{z_{\alpha/2}^2}{2n}\pm z_{\alpha/2}\sqrt{[\hat{p}(1-\hat{p})+z_{\alpha/2}^2/4n]/n}\right)/(1+z_{\alpha/2}^2/n). \tag{2.6}$$

이 신뢰구간은 Wilson(1927)에 의해 처음으로 제시되었고 모비율 p에 대한 score 검정으로부터 얻어져서 간단히 score 신뢰구간이라고 한다.

2.5. Agresti-Coull 신뢰구간

Agresti와 Coull(1998)은 Wald 신뢰구간에 두 번의 성공과 두 번의 실패를 추가하여 신뢰구간을 구하고 이를 조정된 Wald 신뢰구간이라 하였다. 그 후 조정된 Wald 신뢰구간을 Agresti-Coull(AC) 신뢰구간이라 한다. 총 시행수는 $\tilde{n}=n+4$이고 이항비율에 대한 점추정치는 $\tilde{p}=(X+2)/(n+4)$이다. 식(2.1)에 n과 p 대신에 $\tilde{n}$와 $\tilde{p}$를 대입한 것이 AC 신뢰구간이다. 즉,

$$\tilde{p}\pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}. \tag{2.7}$$

score 신뢰구간에서 $z^2=1.96^2\approx 4$로 두면 이는 Agresti-Coull 신뢰구간과 같게 된다.

2.6. Jeffreys prior 신뢰구간

이항비율 p에 대한 신뢰구간 추정에 p에 대한 사전분포로 $B(1/2,1/2)$인 베타분포를 이용한다. $100(1-\alpha)\%$ 꼬리가 같은 Jeffreys prior 신뢰구간은 다음과 같다.

$$L_J(x)\le p\le U_J(x). \tag{2.8}$$

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

3. 평가기준

이항비율 p에 대한 신뢰구간은 2장에서 다룬 것 이외에도 다양한 방법들이 소개되고 있다. Vollset(1993)과 Newcombe(1998)은 각각 13가지와 7가지 신뢰구간들을 비교하였다. Agresti와 Coull(1998), Santner(1998), Brown, et al.(2001, 2002), 그리고 Reiczigel(2003) 등도 여러 신뢰구간들을 비교하여 조건에 따라 적절한 신뢰구간의 사용을 제안하고 있다. 이항비율에 대한 신뢰구간 평가에 사용한 기준들은 일부 차이는 있으나 대부분이 유사하며, 대표적인 것으로 다음 3가지 평가기준들이 주로 사용된다.

3.1. 평균포함확률

고정된 모수값에 대한 신뢰구간의 포함확률(coverage probability)은 그 구간이 고정된 모수값을 포함할 확률이다. 이는 모수값에 따라 변한다. 통계적으로 신뢰계수는 모든 가능한 모수값들에 대한 포함확률의 최소값으로 정의된다.

하나의 신뢰구간에 대해서 고정된 p값에서의 포함확률은 다음과 같다.

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

여기서 $I(k,p)$는 그 신뢰구간이 $X=k$일 때 p를 포함하고 있으면 1이고 그렇지 않으면 0이다. 포함확률을 p가 취하는 모든 가능한 값에 대해 평균한 값인 평균포함확률(mean coverage probability; MCP)을 하나의 평가기준으로 사용하기도 한다.

$$\bar{C}_n = \int_0^1 C_n(p)g(p)dp. \tag{3.2}$$

3.2. 평균제곱오차의 제곱근

포함확률이 신뢰수준으로부터 얼마나 떨어져 있는가를 나타내는 척도로 평균제곱오차의 제곱근(root mean square error; RMSE)을 사용한다.

$$RMSE = \sqrt{\int_0^1 (C_n(p)-0.95)^2 dp}. \tag{3.3}$$

여기서 신뢰수준은 $1-\alpha = 0.95$이다. 점근적 정규근사를 이용한 경우는 신뢰구간이 추정량을 중심으로 대칭이 된다. 그러나 신뢰 하한이 음수가 되거나 반대로 상한이 1을 넘는 경우 대칭성이 유지되지 않는다. 만약 모수 p가 $[0,1]$상에서 대칭분포라면, 왼쪽과 오른쪽의 비포함확률은 같게 되지만, 모수 p의 분포가 $[0,0.5]$로 제한되면, 왼쪽의 비포함확률인 원점비포함확률(distal non-coverage probability)과 오른쪽의 비포함확률인 중심비포함확률(mesial non-coverage probability)을 각각 계산해야 하는 것을 Newcombe(1998)은 주장하였다. 즉, 신뢰구간의 위치가 신뢰구간의 기대폭 만큼 중요하다는 의미이다.

3.3. 평균기대폭

신뢰구간의 평가에 평균기대폭(mean expected width; MEW)은 포함확률을 보완해주는 하나의 기준으로 널리 사용된다. 기대폭은 신뢰구간의 상한과 하한의 차이이다. 주어진 n과 p에 대한 기대폭은 다음과 같이 계산된다.

$$E_{(n,p)}(w) = \sum_{x=1}^{n} w_x \binom{n}{x} p^x (1-p)^{n-x}. \tag{3.4}$$

여기서 $w_x = U(x,n) - L(x,n)$이다. 그리고 기대폭의 평균은 다음과 같다.

$$\bar{E}_{(n,p)}(w) = \int_0^1 E_{(n,p)}(w)dp. \tag{3.5}$$

가능한 한 기대폭이 작으면서 요구되는 포함확률을 만족시키는 신뢰구간이 바람직한 신뢰구간이 된다.

4. 신뢰구간들의 비교

Vollset(1993)은 13개 방법들에 대해 표본크기가 $n = 10$, $n = 100$, $n = 1,000$일 경우 포함확률을 구하고 그 패턴이 거의 동일함을 보여주고 있다. 그러나 포함확률의 진동(fluctuation)은 p가 0이나 1에 가까울 때를 제외하고는 표본이 증가할수록 줄었다. Brown, et al.(2001, 2002)도 표본크기에 따라 포함확률의 변동이 심함을 밝혀냈는데, p가 1/2에서 멀어질수록 그 진동 폭은 컸다. 그러나 표본이 증가하게 되면 이러한 포함확률의 진동 폭은 역시 줄어들게 됨을 보여주고 있다.

Agresti와 Coull(1998)은 표본크기 n이 100이하인 경우 평균포함확률만을 사용하여 신뢰구간들을 비교하였다. 그러나 표4.1은 표본이 증가함에 따라서 평균포함확률은 신뢰구간에 따라 차이는 있지만 대체적으로 신뢰수준에 근사하게 됨을 보여주고 있다. 따라서 표본이 큰 경우 평균포함확률만으로는 신뢰구간들 간에 큰 차이를 발견할 수 없다. 본 연구에서는 표본이 큰 경우 6가지 신뢰구간들에 대해 3장에 다룬 평가기준들인 평균포함확률, 평균제곱오차의 제곱근, 그리고 평균기대폭을 적용해서 비교하였다. 평균포함확률을 계산하기 위해서 밀도함수 $g(p)$가 균등분포인 경우를 가정하였고 S-plus를 사용하였다.

표 4.1: $0 < p \leq 1$이고 $n = 10,\ 100,\ 500,\ 1,000$일 때 p에 대한 95% 신뢰구간의 평균포함확률, 평균제곱오차의 제곱근과 평균기대폭

Method	$n = 10$			$n = 100$		
	MCP	RMSE	MEW	MCP	RMSE	MEW
Exact	0.98365	0.03625	0.50842	0.96447	0.01934	0.16143
Score	0.95402	0.02377	0.43540	0.95104	0.01276	0.15227
Wald	0.76929	0.28632	0.40343	0.92228	0.09312	0.15183
AC	0.96374	0.02367	0.45650	0.95528	0.01474	0.15431
Mid-P	0.96830	0.02581	0.46066	0.95298	0.01015	0.15307
Jeffrey	0.95315	0.02613	0.43251	0.94985	0.01370	0.15171

Method	$n = 500$			$n = 1,000$		
	MCP	RMSE	MEW	MCP	RMSE	MEW
Exact	0.95715	0.01312	0.07062	0.95517	0.01167	0.04960
Score	0.95028	0.01043	0.06869	0.95012	0.01001	0.04862
Wald	0.94301	0.04190	0.06867	0.94617	0.02974	0.04862
AC	0.95175	0.01118	0.06893	0.95098	0.01041	0.04871
Mid-P	0.95072	0.00485	0.06876	0.95039	0.00358	0.04865
Jeffrey	0.94983	0.01063	0.06863	0.94984	0.01011	0.04860

표4.1에서 평균포함확률을 기준으로 보면 Wald의 신뢰구간이 가장 나쁘며, 표본 수가 증가해도 다른 신뢰구간들에 비해서 신뢰수준 0.95에서 가장 멀리 떨어져 있다. 또한 정확신뢰구간도 다른 방법들에 비해 상대적으로 좋은 방법이라 할 수 없다. 그러나 표본 수가 커짐에 따라서 Jeffrey, score, Mid-p와 AC 신뢰구간들은 근소한 차이를 보이고 있으므로 소수점 2자리까지 고려한다면 평균포함확률이 95% 신뢰수준에 근사하여 Wald와 정확신뢰구간보다 더 좋은 방법이라 할 수 있다. 한편 포함확률의 변동 정도를 나타내는 평균제곱오차의 제곱근의 관점에서는 Wald 방법이 가장 크고 Mid-p가 가장 작으므로 Mid-p를 선택하는 것이 가장 바람직하다. 반면에 평균기대폭은 정확신뢰구간이 가장 크지만, 표본 수가 증가하면 평균기대폭의 차이는 거의 없게 된다. 따라서 표4.1에 의하면 표본 수가 큰 경우 어떤 신뢰구간을 사용할 지의 판단 기준은 평균포함확률과 평균제곱오차의 제곱근을 사용하는 것이 바람직하다.

본 연구에서의 결과는 기존의 연구의 결과들과 대부분이 유사하지만 몇 가지 점에서는 차이를 보이고 있다. Vollset(1993)은 정확과 Mid-p 그리고 score 신뢰구간들을, Newcombe(1998)은 score 신뢰구간을, 그리고 Agresti와 Coull(1998)은 score와 자신들이 제안한 방법인 AC 신뢰구간들을 추천하고 있다. 또한 Brown, et al.(2001, 2002)은 표본이 작을 때($n \le 40$)는 score와 Jeffrey 신뢰구간을, 표본이 클 때는 score, AC, 그리고 Jeffrey 신뢰구간이 좋지만, 실용적인 면에서 AC 신뢰구간의 사용을 제시하고 있다. 그러나 표4.1에서 평균포함확률을 기준으로 할 때는 표본의 크기가 큰 경우 score, Jeffrey, Mid-p와 AC 신뢰구간들이 바람직하고 평균제곱오차의 제곱근을 기준으로 할 때는 Mid-p 신뢰구간이 바람직하다. 전체적으로 볼 때, score, Jeffrey, Mid-p와 AC신뢰구간들이 이항비율에 대한 구간추정에 바람직한 방법이라 할 수 있지만 평균포함확률과 평균제곱오차의 제곱근을 동시에 고려하면 Mid-p가 바람직한 방법이라 할 수 있다.

선행 연구들은 대부분이 신뢰구간을 선택함에 있어서 표본의 크기를 변화시키면서 평가기준들을 비율의 전 구간에 대해 계산하였는데, 모비율이 0에 가까운 경우 정확한 비교가 곤란하다. 더욱이 실제 조사에서 모비율 p가 0에 가까운 경우에 위와 같이 p의 모든 범위에 대해 적절한 신뢰구간을 찾는 것은 큰 의미가 없다. 왜냐하면, 본 연구진들이 사용한 기생충 실태조사 자료(표본규모는 2만 3천여 명)에 의하면 특정 기생충의 감염률은 시간이 지남에 따라 줄어들고 있으며, 감염률이 $p = 0.1$ 보다도 훨씬 작으므로 이러한 범위에서 어떤 신뢰구간을 사용하는 것이 바람직한 가를 결정해야 한다. 또한 특정 산업재해률에 관한 경우도 동일하다. 따라서 $0 < p \le 0.1$과 같이 p가 작고 표본규모가 큰 경우 본 연구에서 다룬 6개의 신뢰구간들에 대해 3가지의 평가기준을 적용할 때 어떤 신뢰구간의 사용이 적절한 지를 살펴보고자 한다.

표 4.2: $0 < p \leq 0.1$이고 $n = 1000, 5000, 10000$일 때 p에 대한 95% 신뢰구간의 평균포함확률, 평균제곱오차의 제곱근과 평균기대폭

Method	$n = 1,000$			$n = 5,000$			$n = 10,000$		
	MCP	RMSE	MEW	MCP	RMSE	MEW	MCP	RMSE	MEW
Exact	0.9591	0.0325	0.0263	0.9539	0.0308	0.0115	0.9526	0.0305	0.0081
Score	0.9500	0.0307	0.0255	0.9492	0.0303	0.0113	0.9492	0.0302	0.0080
Wald	0.9335	0.0697	0.0252	0.9455	0.0377	0.0113	0.9472	0.0326	0.0080
AC	0.9536	0.0313	0.0258	0.9504	0.0304	0.0114	0.9498	0.0302	0.0080
Mid-P	0.9516	0.0072	0.0254	0.9504	0.0043	0.0113	0.9502	0.0033	0.0080
Jeffrey	0.9490	0.0308	0.0253	0.9492	0.0303	0.0113	0.9491	0.0302	0.0080

표4.2는 $0 < p \leq 0.1$이고 $n = 1,000$, $n = 5,000$, $n = 10,000$이며 밀도함수 $g(p)$가 균등분포인 경우 평균포함확률, 평균제곱오차의 제곱근, 그리고 평균기대폭들을 보여주고 있다. 표4.2는 표4.1과 다른 결과를 나타내고 있다. 평균포함확률을 기준으로 할 때 $n = 1,000$인 경우는 표4.1과 같이 score와 Jeffrey 신뢰구간들이 좋은 방법이나 표본크기가 커지면 오히려 AC와 Mid-p 신뢰구간들이 바람직한 방법으로 평가된다. 그러나 소수점 2자리까지만 고려하면 평균포함확률에는 차이가 없는 것을 알 수 있다. 한편 평균제곱오차의 제곱근은 표본 크기에 관계없이 Mid-p가 항상 작다. 표본규모가 크고 p가 작은 경우 신뢰구간은 Mid-p가 가장 바람직하고 다음으로 AC, score 그리고 Jeffrey가 거의 비슷하다. 그림4.1과 그림4.2는 $0 < p \leq 0.1$이고 표본수가 각각 $n = 1,000$, $n = 5,000$일 때의 포함확률을 보여주고 있다. 이들 두 그림에서 p가 0에 가까울수록 포함확률이 0.95에서 벗어나지만 Mid-p, score와 Jeffrey가 다른 방법에 비해 안정적이다.

한편 그림4.3은 $0 < p \leq 0.1$이고 $n = 1,000$, $n = 5,000$, $n = 10,000$일 때 포함확률에 대한 상자그림인데, $n = 1,000$일 때 Wald는 구간의 폭이 너무 크고, 정확신뢰구간과 AC 신뢰구간은 0.95의 위에 위치하고 있어 보수적이다. 반면에 score, Mid-p, 그리고 Jeffrey 신뢰구간들은 0.95를 중심으로 분포하고 있다. n이 증가할수록 전체적으로 상자그림의 구간 폭이 줄어들고 있으나 Mid-p, score, Jeffrey 그리고 AC 신뢰구간들이 선호된다. 현실적으로 $0 < p \leq 0.1$이고 표본 규모가 큰 경우 평가기준들을 모두 만족시키는 신뢰구간을 찾는데 어려움이 있으므로, 기존 연구에서와 같이 상대적으로 평가기준들을 충실히 만족하는 신뢰구간을 사용하는 것이 바람직하다.

5. 토의 및 결론

이항비율에 대한 구간추정은 오래전부터 다루어져 온 주요한 내용이다. 일반적으로 Wald 신뢰구간이 널리 사용되고 있는데, 이 방법의 사용이 적절치 않다는 주장은 많은 연구자들에 의해서 제기되었고 또한 설득력을 더해가고 있다. 본 연구에서는 발생 빈도가 낮은 산재율이나, 희귀 질병의 발생률, 그리고 특정 기생충의 감염률과 같은 낮은 이항비

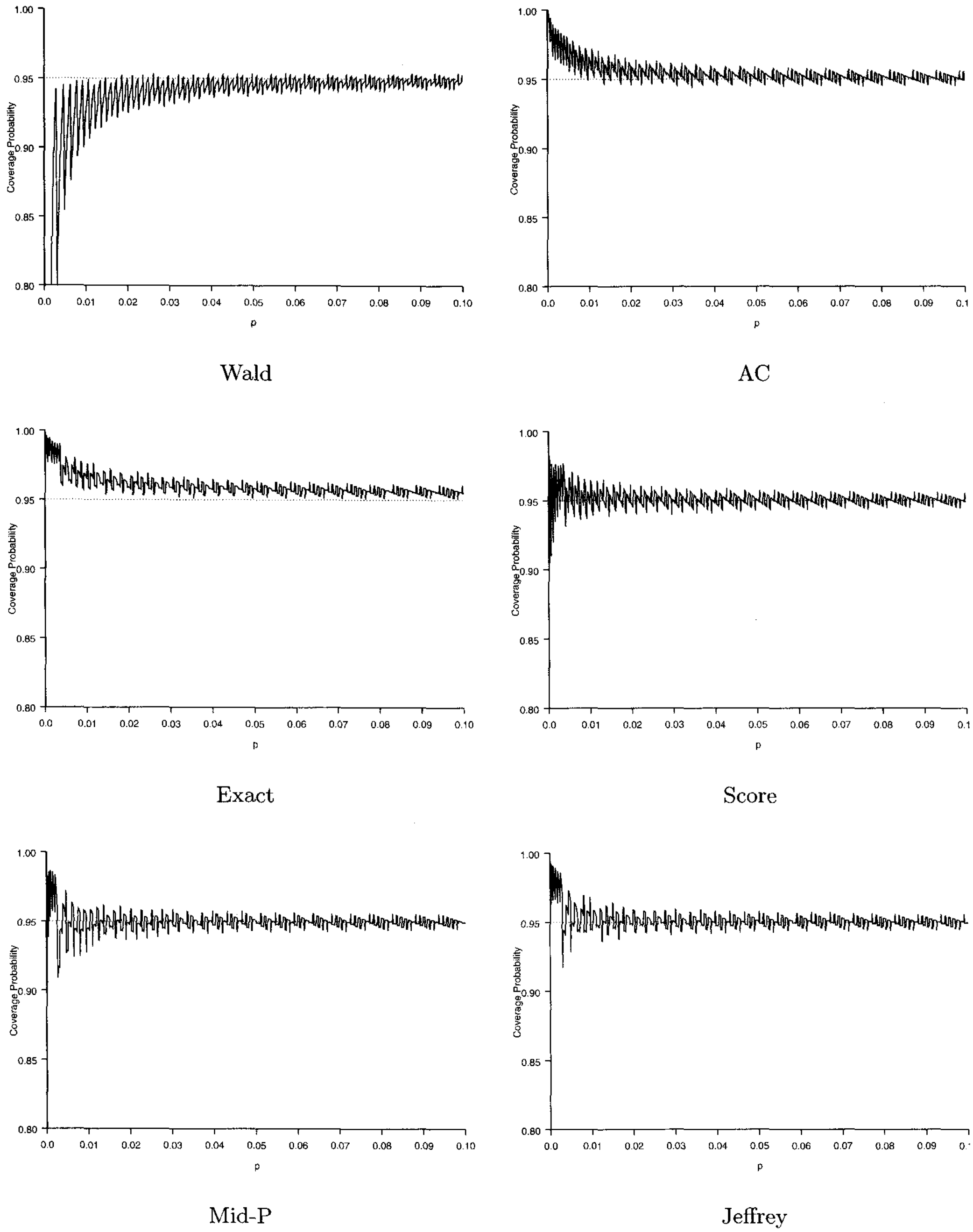
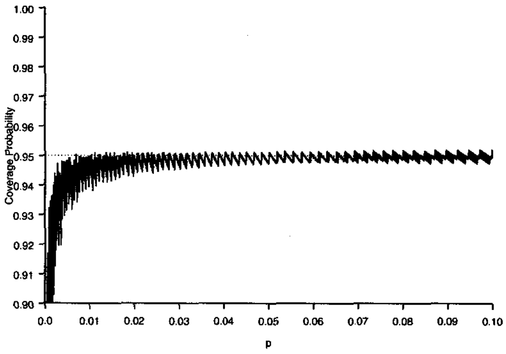
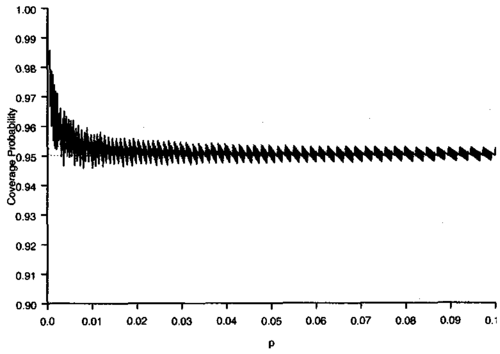


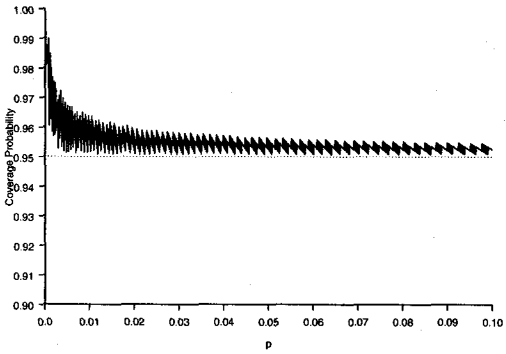
그림 4.1: $0 < p \leq 0.1$이고 $n = 1,000$인 경우 p에 대한 95% 신뢰구간의 포함확률



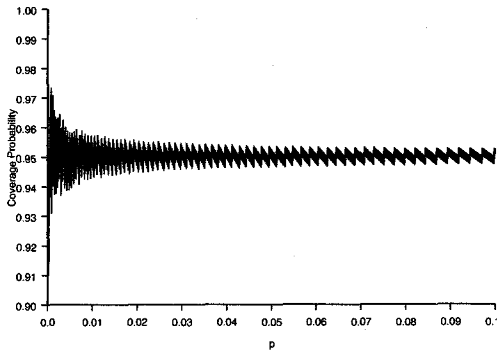
Wald



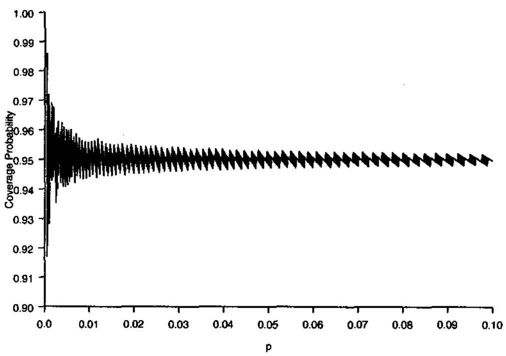
AC



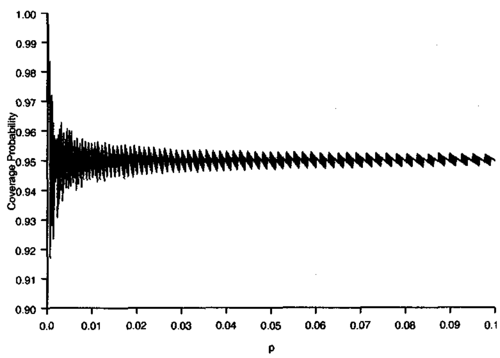
Exact



Score

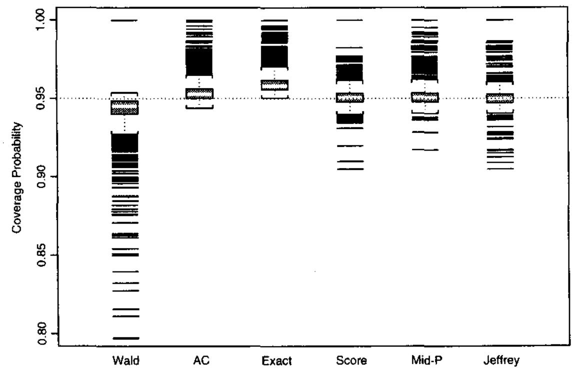


Mid-P

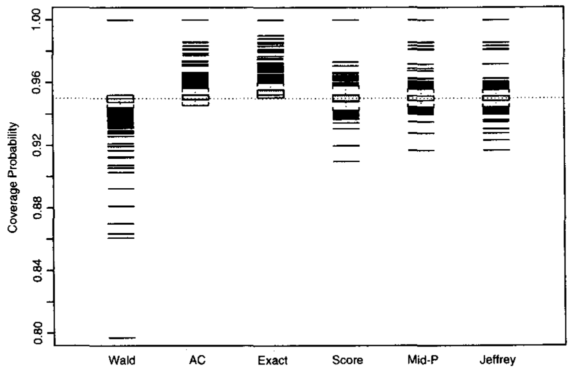


Jeffrey

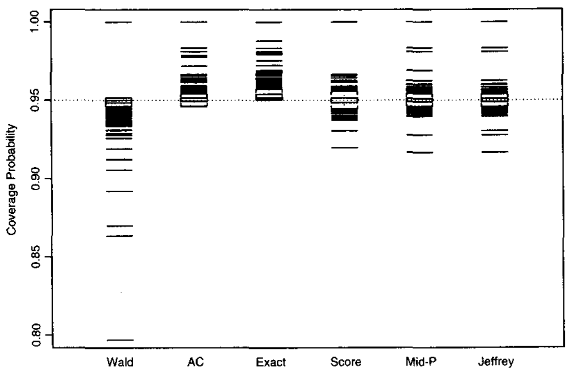
그림 4.2: $0 < p \leq 0.1$이고 $n = 5,000$인 경우 p에 대한 95% 신뢰구간의 포함확률



$n = 1,000$



$n = 5,000$



$n = 10,000$

그림 4.3: $0 < p \leq 0.1$이고 $n = 1,000$, $n = 5,000$, $n = 10,000$인 경우 포함확률의 상자그림

율에 대한 구간추정에 있어서 어떠한 신뢰구간이 적절할 지를 살펴보았다. 많은 연구자들이 주로 사용하였던 6가지 방법인 Wald, score, exact, AC, Mid-p, 그리고 Jeffrey 신뢰구간들을 비교하였다. 그리고 신뢰구간의 판단기준으로 평균포함확률, 평균제곱오차의 제곱근, 그리고 평균기대폭을 사용하였다.

이항비율 p의 전 구간에 대해서 신뢰구간들을 비교한 결과(참고: 표4.1)는 기존의 연구들과 큰 차이가 없었으나 평균포함확률을 기준으로 할 때는 표본의 크기가 큰 경우 score, Jeffrey, Mid-p와 AC 신뢰구간들이 바람직하고 평균제곱오차의 제곱근을 기준으로 할 때는 Mid-p 신뢰구간이 바람직함을 알 수 있었다. 또한 평균포함확률과 평균제곱오차의 제곱근을 동시에 고려한다면 Mid-p가 바람직한 방법이라 할 수 있다.

본 연구진들이 수행한 기생충실태조사는 표본 규모는 크지만 기생충의 감염율은 아주 낮아 이러한 경우 어떤 신뢰구간추정법이 적절한 지가 관심거리이다. 따라서 $0 < p \leq 0.1$이고 표본의 규모가 큰 $n = 1,000$, $n = 5,000$, $n = 10,000$인 경우를 살펴보았다. 비교 결과는 p의 전 구간에서와는 달리 평균포함확률을 기준으로 할 때는 표본크기가 커지면 오히려 AC와 Mid-p가 바람직한 방법으로 평가된다. 그러나 평균제곱오차의 제곱근은 표본 크기에 관계없이 Mid-p가 변함없이 작게 된다. 그리고 그림4.1~그림4.3에 의하면 score, Mid-p, Jeffrey 그리고 AC 신뢰구간이 바람직하다. 이상으로 볼 때 표본규모가 크고 p가 작은 경우 구간추정은 Mid-p 신뢰구간을 사용하는 것이 가장 바람직하고 다음으로 AC, score와 Jeffrey 방법의 사용을 권장한다. 한편 이항비율에 대한 구간추정에 Blyth와 Still(1983), 그리고 Leemis와 Trivedi(1996)들이 사용한 정규근사나 Poisson 근사 등과 같은 근사적 방법의 적용도 의미가 있을 것이다.

참고문헌

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.

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.

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.

Kott, P. S., Anderson, 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.

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.

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.

[2005년 9월 접수, 2006년 1월 채택]

Confidence Intervals for a Low Binomial Proportion

Jea-Bok Ryu[1)] Seung-Joo Lee[2)]

Abstract

We discuss proper confidence intervals for interval estimation of a low binomial proportion. A large sample surveys are practically executed to find rates of rare diseases, specified industrial disaster, and parasitic infection. Under the conditions of $0 < p \leq 0.1$ and large n, we compared 6 confidence intervals with mean coverage probability, root mean square error and mean expected widths to search a good one for interval estimation of population proportion p. As a result of comparisons, Mid-p confidence interval is best and AC, score and Jeffreys confidence intervals are next.

Keywords: Binomial proportion, Confidence interval, Mean coverage probability, Root mean square error, Mean expected width

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

2) (Corresponding author) Associate Professor, Division of Life Science·Genetic Engineering·Statistics Cheongju University, 36, Naedok-Dong, Sangdang-Gu, Cheongju, Chungbuk, 360-764, Korea.
E-mail: access@cju.ac.kr