The Korean Communications
in Statistics Vol. 7, No. 3, 2000
pp. 859-869

A Simulation Study for the Confidence Intervals of p by Using Average Coverage Probability

Daehak Kim[1)] and Hyeong Chul Jeong[2)]

Abstract

In this paper, various methods for finding confidence intervals for p of binomial parameter are reviewed. Also we introduce two bootstrap confidence intervals for p. We compare the performance of bootstrap methods with other methods in terms of average coverage probability by Monte Carlo simulation. Advantages of these bootstrap methods are discussed.

Keywords : Bootstrap, Confidence interval, Binomial distribution, Average coverage probability

1. 소개

의학이나 생물학 등 생명과학이나 사회학 등의 자료 중 많은 부분은 범주형으로 주어진다. 이 경우 고려되는 범주형 변수에 따라 개체들은 반응값이 성공 혹은 실패인 이항범주나 다항범주 등을 따르게 된다. 특히 반응값이 이항인 경우 모비율 p는 모집단에서 어떤 특정한 속성을 갖는 개체의 비율을 나타내는 모수라 할 수 있다. 이제, 크기가 n인 표본에서 특정한 속성을 갖는 개수를 X라 하고, $\hat{p}=X/n$을 표본비율이라 놓자. 본 연구에서는 모비율 p에 대한 신뢰구간을 구축하는 여러 방법들을 살펴보았다. 특히 붓스트랩을 이용하는 두 가지 신뢰구간을 소개하고 이들을 Woodroofe and Jhun(1989)의 평균포함확률측면(average coverage probability)에서 모의실험을 통하여 여러 방법들과 비교하였다.

약 200년 전부터 모비율 p의 신뢰구간 구축과 관련된 언급이 있었으며 최근에까지 Agresti and Coull(1998), Chen(1990) 그리고 Leemis and Trivedi(1996) 등에 의해 여러 성질들이 계속 연구되고 있다. Vollset(1993)은 모비율 p의 신뢰구간을 구축하는 다양한 방법 중 13가지를 비교하였으며, Leemis and Trivedi(1996)은 모비율의 신뢰구간을 구축함에 있어 정규근사방법과 포아송근사방법을 비교하여 표본비율이 낮은 경우 포아송 신뢰구간을 활용하는 것이 좋음을 보인바 있다. 또한 Chen(1990)은 베이지안 추정을 이용하여 최적의 신뢰구간을 구축하는 방법을 제공하기도 하였

1) Associate Professor, Department of Statistical Information, Catholic University of Daegu, Kyungbuk, 712-702, Korea E-Mail : dhkim@cuth.cataegu.ac.kr
2) Assistant Professor, Department of Computational Science and Statistics, Pyoungtaek University, Kyunggi, 450-701, Korea E-Mail : jhc@ptuniv.ac.kr

다. 특히 가장 최근의 연구로 Agresti and Coull(1998)을 들 수 있는데, 그들은 Clopper and Pearson(1934)의 "정확"신뢰구간보다 근사이론을 이용한 신뢰구간이 더 나을 수 있다는 것을 모의실험으로 보인바 있다. 한편, Efron(1979)의 붓스트랩을 이용하여 모비율 p의 신뢰구간을 구축할 수도 있다. 그런데, Singh(1981)에 의하면, 이항분포와 같이 격자분포를 따르는 분포에서는 붓스트랩분포와 실제분포와의 차이가 $O(1/\sqrt{n})$인데, 정규근사를 이용한 분포와 실제분포와의 차이 역시 $O(1/\sqrt{n})$으로 정규근사 방법과 붓스트랩 방법과의 차이가 없음을 보인바 있다. 하지만, Woodroofe and Jhun(1989)은 평균포함신뢰확률 측면에서 붓스트랩분포와 실제분포와의 차이를 실질적으로 $o(1/\sqrt{n})$까지 줄일 수 있음을 제시하였다. 이와 같은 사실에 따라 모비율 p의 신뢰구간 구축에 붓스트랩 방법을 사용하는 것도 의미가 있으리라 여겨진다. 2장에서는 모비율 p에 대한 신뢰구간을 구축하는 여러 방법들을 살펴보고, 붓스트랩을 이용한 두 가지 신뢰구간도 소개하였다. 3장에서는 평균포함확률 측면에서 모의실험을 통하여 신뢰구간을 구축하는 12가지 방법들을 비교하였다.

2. 모비율 p의 신뢰구간 구축

2.1 정확(exact) 방법

X가 모비율이 p인 이항분포를 따른다고 하자. 이제, p의 $100(1-\alpha)\%$ 정확신뢰구간을 구하기 위해서는 실현 값인 x에 대해, 유의수준 $\alpha/2$에서 귀무가설 $H_0: p=p_0$을 기각하지 않는 모든 p_0를 계산하면 된다. 즉, 주어진 x에 대해, 양측검정에서, 어떤 p_0를 사용하여야 귀무가설을 채택할 수 있는가하는 문제이다. 이때 얻어지는 모든 p_0 중 최소값이 정확신뢰구간의 하한, 최대값이 정확신뢰구간의 상한이 된다. 즉, 최소값 p_L과 최대값 p_U는

$$P(X \geq x | p=p_L) = \sum_{k=x}^{n} \binom{n}{k} p_L^k (1-p_L)^{n-k} \leq \alpha/2$$

$$P(X \leq x | p=p_U) = \sum_{k=0}^{x} \binom{n}{k} p_U^k (1-p_U)^{n-k} \leq \alpha/2$$

로 계산된다.

위의 정확신뢰구간을 구하기 위해서는 상당한 계산이 요구되지만 다행스럽게도 F 분포를 이용한 Blyth(1986)이나 Hald(1952)의 쉬운 계산방법이 존재한다. 또한 계산과정에 대해서는 Leemis and Trivedi(1996)을 참조할 수 있다. 정확신뢰구간에 대한 닫힌 형태는 다음과 같다.

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

여기서, $F_{a,b,c}$는 자유도 a,b를 따르는 F분포의 $100(1-c)\%$ 분위점이다. 이 신뢰구간을 Clopper-Pearson(1934)의 정확신뢰구간이라 한다. 이 신뢰구간은 가장 표준적인 정확한 방법임에도 불구하고, 이항분포가 지니는 이산성에 의해 신뢰구간을 넓게 추정하는 경향이 있다(Agresti and Coull, 1998).

2.2 포아송근사 방법

한편, 포아송분포의 모수에 대한 정확신뢰구간으로부터 모비율 p의 신뢰구간을 유도하는 방법을 살펴보자. 포아송분포의 모수를 μ라 했을 때, μ의 정확신뢰구간의 상한과 하한은 각각 다음을 만족하는 μ_U 와 μ_L이다.

$$P(X\geq x \mid \mu=\mu_L)=\sum_{k=x}^{n}\frac{(e^{-\mu_L}\mu_L^k)}{k!}\leq \alpha/2$$

$$P(X\leq x \mid \mu=\mu_U)=\sum_{k=0}^{x}\frac{(e^{-\mu_u}\mu_U^k)}{k!}\leq \alpha/2$$

여기서, 포아송분포의 누적확률인 $P(X\leq x|\mu)$는 자유도가 $v=2(1+x)$인 카이제곱분포 형태로 다음과 같이 표현된다.

$$\sum_{k=0}^{x}\frac{(e^{-\mu}\mu^k)}{k!}=P\{\chi_v^2>2\mu\}$$

위의 사실과, 이항분포의 포아송근사를 활용하여 $np\rightarrow\mu$ 로 놓으면, 다음과 같이 포아송분포의 정확확률로 모비율의 신뢰구간을 구할 수 있게된다. 즉

$$P(X\geq x \mid \mu=np_L)=\sum_{k=x}^{\infty}\frac{(e^{-np_L}(np_L)^k)}{k!}=1-\sum_{k=0}^{x-1}\frac{(e^{-np_L}(np_L)^k)}{k!}\leq \alpha/2$$

로 되고 포아송분포의 누적확률은 카이제곱분포 $P\{\chi_{2x}^2\geq 2np_L\}=1-\alpha/2$ 가 되어 신뢰구간의 하한 p_L 을 얻을 수 있으며, 비슷한 방법으로 상한 p_U 도 다음과 같이 계산된다.

$$p_L=\frac{1}{2n}\chi^2_{2x,\alpha/2}\quad,\quad p_U=\frac{1}{2n}\chi^2_{2(x+1),1-\alpha/2} \tag{2.2}$$

포아송 신뢰구간의 성질에 대해서는 Leemis and Trivedi(1996)에 의해 자세히 언급되었다. 특히 그들의 연구에서 표본의 크기와 표본비율이 주어질 때 정규근사를 사용하여야 하는지 포아송근사를 사용하여야 하는지의 경계점을 제시하기도 하였다.

2.3 정규근사에 의한 방법

이제 표본비율 $\hat{p}$ 의 분산을 $V(p)=\sqrt{p(1-p)/n}$ 로 놓으면, 확률변수 $(p-\hat{p})^2$는 정규근사에 의해 $(p-\hat{p})^2=cV(\tilde{p})$ 로 표현할 수 있다. 여기서, c 는 $\chi^2(1)$분포의 $100(1-\alpha)\%$ 분위점이며, $\tilde{p}$ 는 p에 대한 표현식이다. 이제, $\tilde{p}$에 적당한 값을 대입하여 위의 이차방정식을 풀면 두 값 $[p_L, p_U]$를 얻게된다. 잘 알려진 바대로, $\tilde{p}=\hat{p}$ 을 대입하여 방정식을 풀면, 모비율 p에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이 주어진다.

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

여기서, z_α 는 표준정규분포의 $100(1-\alpha)\%$ 분위점이다. 이 신뢰구간을 Wald 신뢰구간이라고 하며, 대부분의 통계학 책에서 소개하고 있다.

또한, $(p-\hat{p})^2 = cV(\tilde{p})$에서 $\tilde{p}=p$를 대입하여 이차방정식을 풀면 이는 주어진 귀무가설을 만족하는 신뢰영역(acceptance region)을 구하는 것과 같은 문제인데, 이에 대한 신뢰구간은 다음과 같이 주어진다.

$$\frac{X+z_{\alpha/2}^2/2 \pm z_{\alpha/2}\sqrt{X-X^2/n+z_{\alpha/2}^2/4}}{n+z_{\alpha/2}^2} \tag{2.4}$$

위의 신뢰구간을 Score 신뢰구간이라고 한다. Wald 신뢰구간은 분산 $V(p)$를 추정할 때 최대우도 추정량 $\hat{p}$을 이용하는 데 반해 Score 신뢰구간은 Wilson(1927)이 지적한 바처럼, χ^2분포의 일반화를 통해 p를 이용한다는 점에서 두 신뢰구간은 서로 다른 모습을 하고 있다. Blyth and Still(1983)은 Score 신뢰구간의 성질에 대해서 자세히 논하였다. 여기서, Score 신뢰구간의 중간값은, $\theta = z_{\alpha/2}/\sqrt{n}$ 일 때

$$\tilde{p}^* = \hat{p}\left(\frac{n}{n+z_\alpha^2}\right)+\frac{1}{2}\left(\frac{z_{\alpha/2}^2}{n+z_{\alpha/2}^2}\right)=\frac{X+\frac{1}{2}z_\alpha^2}{n+z_\alpha^2}=\hat{p}+\theta^2(1+\theta^2)^{-1}(\frac{1}{2}-\hat{p})$$

과 같이 주어지는데 이는 $\hat{p}$와 1/2의 가중 평균적인 의미를 지니고 있다. 또한 95% Score 신뢰구간은 $\hat{p}$을 2개의 성공과 2개의 실패를 더한 형태로 추정한 후 신뢰구간을 구하는 것과 비슷하며, 특히 $\hat{p}$의 추정을 $\tilde{p}^*$로 축소(shrink)시키는 의미를 지니고 있으며, 분산 $\tilde{p}^*(1-\tilde{p}^*)$도 축소되어 다음과 같이 추정된다.

$$\tilde{p}^*(1-\tilde{p}^*)-\frac{1}{4}\theta^2(1+\theta^2)^{-1}=(\hat{p}(1-\hat{p})+\frac{1}{4}\theta^2)(1+\theta^2)^{-2}$$

그러므로, $\sqrt{n}(\tilde{p}^*-p)[\tilde{p}^*(1-\tilde{p}^*)-\frac{1}{4}\theta^2/(1+\theta^2)]^{-1/2}$ 의 정규근사를 이용하여 식 (2.4)의 신뢰구간을 유도할 수 있다(Ghosh, 1979).

Agresti and Coull(1998)는 식 (2.4)의 신뢰구간이 정확신뢰구간보다 주어진 명목수준을 가장 근사하게 추정하며, 특히 중간값인 $\tilde{p}^*$로 p를 추정한 후 Wald 신뢰구간을 구축하는 것과 매우 유사함을 모의실험으로 보였다. 또한 Score 신뢰구간은 베이지안적 관점에서 유도한 신뢰구간과 비슷한 함을 논한 바 있다.

정규근사의 또 다른 방법으로 arcsin 변환을 이용하는 것이 있다. 이는 $\hat{p}$의 분산이 모비율 p의 함수이므로, 분산이 p에 의존하지 않는 함수를 유도하는 과정에서 arcsin 변환을 생각하게 되었다. 잘 알려진 바

$$\frac{2}{\sqrt{n}}\left[\arcsin(\hat{p}^{\frac{1}{2}})-\arcsin(p^{\frac{1}{2}})\right]\to N(0,1)$$

이므로 신뢰구간은 각각 다음과 같이 주어진다.

$$\sin^2[\arcsin(\hat{p}^{\frac{1}{2}})\pm\frac{Z_{\alpha/2}}{2\sqrt{n}}] \tag{2.5}$$

Chen(1990)은 arcsin 변환을 통하여, 정규분포로의 수렴속도가 빨라짐을 보인 바 있다.

2.4 연속성 수정방법

2.2절의 정규근사방법은 이산분포를 따르는 통계량을 연속확률분포인 정규분포로 근사함으로서 약간의 오차가 존재하리라고 예상할 수 있다. 그러므로 수정된 값을 대입시키는 연속성 수정(continuity correction)으로 오차를 보정하는 방법을 생각할 수 있다. 즉, 이항분포 X에 대해 $P(X=a)$의 정규근사는 $P(a-0.5\le Y\le a+0.5)$이다. 여기서, Y는 X와 평균과 분산이 같은 정규분포이다. 이에 따라, 식 (2.3)에 연속성 수정을 하면 다음과 같은 신뢰구간이 주어진다.

$$\frac{X}{n}\pm\left\{\frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\left(\frac{X}{n}\right)\left(1-\frac{x}{n}\right)}+\frac{1}{2n}\right\} \tag{2.6}$$

또한, Score 신뢰구간에 연속성 수정을 적용하면,

$$\frac{(X\pm0.5)+z_{\alpha/2}^2/2\pm z_{\alpha/2}\sqrt{(X\pm0.5)-(X^2\pm0.5)^2/n+z_{\alpha/2}^2/4}}{n+z_{\alpha/2}^2} \tag{2.7}$$

이다. 그런데, 평균이 np이고 분산이 $np(1-p)$인 정규분포 Y에 대해 비교적 큰 n에 대해

$$P(|X-np|\le z_\alpha\sqrt{np(1-p)})\approx 1-\alpha$$

이 성립하고 또한 $(X/n)(1-X/n)$는 $p(1-p)$로 확률수렴 하므로

$$P(|X-np|\le z_\alpha\sqrt{n\hat{p}(1-\hat{p})})\approx 1-\alpha$$

가 된다. 그러나 $\sqrt{\hat{p}(1-\hat{p})}$는 $\sqrt{p(1-p)}$를 상당히 하향추정하는 경향이 있다(Blyth and Still, 1983). 그러므로, 정규근사값 $z_{\alpha/2}$를 적당한 $\gamma z_{\alpha/2}$로 변환한 신뢰구간을 생각할 수 있다. 이에 대해, T-분포를 고려하여,

$$\frac{X}{n}\pm\left\{\frac{t(n-1,\alpha/2)}{\sqrt{n}}\sqrt{\left(\frac{X}{n}\right)\left(1-\frac{X}{n}\right)}+\frac{1}{2n}\right\} \tag{2.8}$$

로 수정할 수 있다. Blyth and Still(1983)은 $\gamma>1$ 을 만족하는 최적의 식을 다음과 같이 제시하였다.

$$\gamma^2=\frac{n}{n-z_{\alpha/2}^2-2z_{\alpha/2}/\sqrt{n}-1/n+O(1/n^3\sqrt{n}))}$$

$$\gamma\approx\sqrt{\frac{n}{n-z_{\alpha/2}^2-2z_{\alpha/2}/\sqrt{n}-1/n}}$$

이에 따라, Blyth and Still(1983)의 조정된 신뢰구간은 다음과 같이 주어진다.

$$\frac{X}{n}\pm\left\{\frac{z_{\alpha/2}}{\sqrt{n-z_{\alpha/2}^2-2z_{\alpha/2}/\sqrt{n}-1/n}}\sqrt{\left(\frac{X}{n}\right)\left(1-\frac{X}{n}\right)}+\frac{1}{2n}\right\} \tag{2.9}$$

2.5 $\hat{p}$ 을 조정하는 방법

Score 신뢰구간의 중간 추정값은 95%에서 $\tilde{p}=(X+z^2/2)/(n+z^2)\approx(X+2)/(n+4)$로 주어진다. 이는 모비율을 2개의 성공과 2개의 실패를 더한 값으로 이동(shift)하여 추정한다는 의미를 지닌다. Agresti and Coull(1998)은 $(X+2)/(n+4)$를 이용한 신뢰구간이 Score 신뢰구간과 매우 유사한 결과를 유도함을 모의실험으로 보였다. 그런데 비슷한 이동추정으로 베이지안 추정을 생각할 수 있다. 사전분포로 Beta (β,γ)를 이용하면, 모비율 p의 베이지안 추정량은 $(X+\beta)/(n+\beta+\gamma)$이다. 일반성을 잃지 않고, $\beta=\gamma$로 두면 $(X+\beta)/(n+2\beta)$이므로, 베이지안 추정량을 이용한 신뢰구간은 다음과 같다.

$$\left(\frac{X+\beta}{n+2\beta}\right)\pm\frac{z_{\alpha/2}}{\sqrt{n}}\left[\left(\frac{X+\beta}{n+2\beta}\right)\left(1-\frac{X+\beta}{n+2\beta}\right)\right]$$

Chen(1990)은 $\beta=kz_{\alpha/2}$에 대해서, 최적의 k를 제시하였는데, $\beta=z_{\alpha/2}^2/2$를 추천하고 있다. 그러므로 Chen(1990)의 결과를 이용한 신뢰구간은 다음과 같이 주어진다.

$$\left(\frac{X+z_{\alpha/2}^2/2}{n+z_{\alpha/2}^2}\right)\pm\frac{z_{\alpha/2}}{\sqrt{n}}\left[\left(\frac{X+z_{\alpha/2}^2/2}{n+z_{\alpha/2}^2}\right)\left(1-\frac{X+z_{\alpha/2}^2/2}{n+z_{\alpha/2}^2}\right)\right] \tag{2.10}$$

이 신뢰구간은 Score 신뢰구간의 중간값을 이용한 신뢰구간과 매우 유사함을 알 수 있다.

2.6 붓스트랩 방법

모비율 p에 대한 신뢰구간을 붓스트랩방법을 이용하여 추정할 수 있음은 앞에서도 언급하였다. 이제 붓스트랩방법을 이용한 모비율 p에 대한 신뢰구간을 살펴보자. 먼저 모비율 p의 붓스트랩 추정량을 $\hat{p}^*$ 라고 하면, B회의 붓스트랩 반복에 의해 얻은 붓스트랩분포로부터 모비율 p의 신뢰구간을 구할 수 있다. 즉, $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$[p_L, p_U]=\left[\hat{p}-q_{(1-\alpha/2)}\sqrt{\hat{p}(1-\hat{p})/n}\ ,\ \hat{p}-q_{(\alpha/2)}\sqrt{\hat{p}(1-\hat{p})/n}\ \right] \tag{2.11}$$

여기서, $q_{(1-\alpha/2)}$는 다음의 B회 붓스트랩 반복에 의해 얻은 Z^*의 $100(1-\alpha/2)\%$ 분위점이다.

$$Z^*(b)=\frac{\sqrt{n}(\hat{p}^*-\hat{p})}{\sqrt{\hat{p}^*(1-\hat{p}^*)}},\quad b=1,\cdots,B$$

또한, 식 (2.11)의 신뢰구간에서 최적의 β를 $z_{\alpha/2}^2/2$로 대입하는 대신에 붓스트랩분포로부터 $q_{(\alpha/2)}^2/2$ 를 입력하는 신뢰구간을 생각 할 수 있다. 이에 의하면 붓스트랩 신뢰구간은 다음과 같이 계산된다.

$$[p_L\ , p_U]=\left(\frac{X+q_{\alpha/2}^2/2}{n+q_{\alpha/2}^2}\right)\pm\frac{q_{\alpha/2}}{\sqrt{n}}\left[\left(\frac{X+q_{\alpha/2}^2/2}{n+q_{\alpha/2}^2}\right)\left(1-\frac{X+q_{\alpha/2}^2/2}{n+q_{\alpha/2}^2}\right)\right] \tag{2.12}$$

이는 정규분포를 이용하는 대신에 붓스트랩분포를 이용하여 이동(shift)한 신뢰구간을 의미한다. 지금까지 이항분포에서 붓스트랩을 잘 활용하지 않은 이유는 이항분포처럼 격자분포를 취하는 경우 정규근사와 붓스트랩 근사간에 큰 차이가 없기 때문이었다(Singh, 1981).

3. 모의실험 비교

본 절에서는 모의실험을 통하여 위에서 언급한 12가지 방법을 평균포함확률 측면에서 살펴보겠다. 먼저, 평균포함확률의 의미를 이항분포를 포함하는 격자분포의 경우에서 살펴보도록 하자. 예를 들어, $X_1, X_2, \cdots, X_n$ 을 평균이 θ 이고 분산이 σ^2인 격자분포에서 생성된 확률변수라 하자. 이제, $Z_n = \sqrt{n}(\overline{X_n} - \theta)/\sigma$ 의 분포함수를 $G_n(\omega, t) = P_\omega\{Z_n \le t\}$로 나타내자. 여기서, ω는 모수공간 Ω에서 정의되는 알려지지 않은 모수이다. 이제, $G_n(\omega, t)$은 다음과 같이 Edgeworth 연장된다.

$$G_n(\omega, t) = \Phi(t) + \frac{1}{\sqrt{n}}\varphi(t)\left[\frac{\rho}{6}(1-t^2) + R_n(\omega, t)\right] + o\left(\frac{1}{\sqrt{n}}\right)$$

여기서, Φ는 표준정규분포함수이고, φ는 표준정규확률밀도함수이며,

$$\rho = \rho(\omega) = \varphi'''(\omega)/\sigma^3 \ , R_n(\omega, t) = \frac{1}{\sigma}\left\{\frac{1}{2} - \langle n\theta + \sqrt{n}\sigma t\rangle\right\}$$

이다. 이제, $G_n^*(\omega, t) = \Phi(t) + \frac{\rho}{6\sqrt{n}}(1-t^2)\varphi(t)$라 하면,

$$\lim_n \int_\Omega \sqrt{n}\left[G_n(\omega, t) - G_n^*(\omega, t)\right]\xi(\omega)\, d(\omega) = 0$$

이 성립한다. 여기서, $G(\omega, t)$를 ω가 포함된 신뢰집합(confidence set)을 포함하는 확률로 보면, ω를 독립시행에 의해 확률밀도함수 ξ에서 발생했을 때, $\int_\Omega \mathrm{G_n}(\omega, \mathrm{t})\xi(\omega)\mathrm{d}\omega$를 평균포함확률(average coverage probability)이라 할 수 있다. Agresti and Coull(1998)은 이항분포에 대해서 실제 모수 p가 어떠한 값으로 주어지던, 평균적 수행능력을 평가하는 것에 의미를 두어 $\overline{C_n} = \int_0^1 C_n(p)\xi(p)dp$ 를 계산하였으며 이를 mean coverage probability라 칭하였다. 여기서, $C_n(p) = \sum_{k=0}^{n} I(k, p)\binom{n}{k}p^k(1-p)^{n-k}$ 이며, $I(k, p)$는 $X = k$에 대해 계산된 신뢰구간이 모수 p를 포함하면 1을 나타내는 지시함수이다.

우리는 여기서, 이항분포의 신뢰구간을 구축함에 있어서 평균포함확률 측면에서 붓스트랩방법이 가지는 이론적 특징을 소개하고자 한다. 모수 ω의 최대우도추정량을 $\hat{\omega}_n$이라 할 때 Z_n의 붓스트랩 분포함수를 $G_n(\hat{\omega}_n, t)$과 같이 표현하면, Woodroofe and Jhun(1989)에 의해 붓스트랩분포와 $G_n^*(\omega, t)$와 차의 기대값은 다음과 같다.

$$E_\omega\{\hat{G}_n(t) - G_n^*(\omega, t)\} = \frac{1}{\sigma\sqrt{n}}\varphi(t)e\left(\langle t\sqrt{n}\sigma\rangle, \frac{1}{2}\rho\sigma t\right) + o\left(\frac{1}{\sqrt{n}}\right)$$

이때 $e(m, r) = \sum_{k=1}^{\infty}\frac{1}{k\pi}\int_R \sin(2\pi kt)\frac{1}{r}\varphi\left(\frac{t-m}{r}\right)dt = \sum_{k=1}^{\infty}\frac{1}{k\pi}\sin(2\pi mk)\, e^{-2r^2\pi^2k^2}$로 실질적인 $e(m, r)$의 크기는 $\sup_m\left|e(m, \frac{1}{2})\right| < 0.01$, $\sup_m|e(m, 1)| < 10^{-9}$로 r 이 증가함에 따라 매우 작

아지는 함수이다. 그러므로 붓스트랩분포와 $G_n^*(\omega, t)$와 차이의 기대값은 실제적으로 $o(1/\sqrt{n})$에 근사하게 된다. 이는 평균포함확률을 계산하면, 붓스트랩근사가 정규근사에 비해 이론적 우월성을 격자분포인 경우에서도 그대로 지니고 있음을 나타내는 결과라 하겠다.

본 모의실험을 통해서는, 이항분포의 경우 기존의 여러 방법들의 평균적 수행능력을 평가하여 평균포함확률에서 각 방법들의 특징을 살펴보도록 하겠다. 특히 평균포함확률을 통해 붓스트랩 방법이 지니는 이론적 특징을 아울러 살펴보도록 하겠다.

평균포함확률을 계산하기 위하여, $\xi(p)$를 (1) 균등분포 (2) $p=0.5$를 중심으로 종형을 이루는 Beta(5, 5)분포를 고려하였다. 표본의 크기는 $n=10$, 20, 30, 50이며, 명목수준 $1-\alpha=0.9$, 0.95, 0.99를 상정하였다. 모의실험 절차를 요약하면 다음과 같다.

[단계1] $\xi(p)$로부터 모수 p_i를 발생한다.

[단계2] 모수 p_i를 따르는 베르누이 확률난수를 n개 발생한다.

[단계3] 주어진 난수로부터 각 방법들의 신뢰구간을 계산한 후 $I(p_i), L(p_i), LC(p_i)$를 계산한다. 여기서, $I(p_i)$는 신뢰구간이 모수를 포함하면 그 값이 1 인 지시함수며, $L(p_i)$는 신뢰구간의 길이, $LC(p_i)$는 신뢰구간의 중심값이다.

[단계4] 위의 [단계1]~[단계3]을 M회 반복한다.

[단계5] 다음의 통계량을 계산한다.

$$\overline{\beta}=\sum_{i=1}^{M} I(p_i)/M\ , \quad \overline{\text{length}}=\sum_{i=1}^{M} L(p_i)/M\ , \quad r=\#[LC(p_i)\geq p_i]/M$$

여기서, r은 주어진 명목수준에 대한 2차적인 통계량으로 사용되었다. 특별한 치우침이 없다면 r은 0.5에 근사하게 될 것이다.

사전분포가 균등분포일 때 모의실험 결과는 [표 1], Beta(5, 5)에 대해서는 [표 2]에 주어져 있다. 표현의 편리를 위하여 표에서의 각 방법들은 각각 식 (2.1)~식 (2.12)의 신뢰구간을 의미한다. 결과의 정도를 높이기 위하여 모의실험은 5000 회 반복되었고 붓스트랩 반복은 2000회를 고려하였다. 모의실험 결과를 요약하면 다음과 같다.

첫째, 정확신뢰구간과 포아송신뢰구간은 명목수준에 비해 과대추정 된다. 이는 표본의 크기가 작을수록 그 경향이 더 심하다. 정확신뢰구간은 그 계산의 어려움에 비해 과대추정 되므로 실제적인 응용문제에서의 활용이 떨어진다고 하겠다(Agresti and Coull, 1998). 포아송 근사에 의한 신뢰구간은 정확신뢰구간에 비해서 매우 크게 과대 추정되고 있다. 또한 포아송 근사방법은 r이 항상 0.5보다 큰데, 이는 신뢰구간의 상한이 약간 위로 추정되고 있음을 의미한다고 하겠다. 포아송 근사방법이 과대추정 되는 경향은 어느 정도 예상된 일인데 이에 대해 Leemis and Trivedi(1996)은 $n\geq 20, p\leq 0.05$ 이거나, $n\geq 100, np\leq 10$ 일 때 포아송 근사를 활용하는 것이 유리하다고 제시하였다. 그런데, 평균적 수행평가에서 포아송 근사를 활용하는 것은 문제가 있음을 알 수 있다.

둘째, Wald, Score, Arcsin 방법등 정규근사를 이용하는 대표적인 방법들을 비교하여보자. Wald 방법은 익히 알려진 바대로 상당히 하향추정하며, Score 방식은 매우 정확히 명목수준에 근사함을 볼 수 있다. 이는 Agresti and Coull(1998)의 $\overline{C_n}$의 결과와 상당히 유사하다. Score는 소표본인

경우에서도 매우 정확히 명목수준을 유지하고 있음을 보여주고 있다. 균등분포시 $n \le 50$에서, Arcsin 변환을 이용하여 신뢰구간을 구축하면, 주어진 명목수준보다 하향추정되고 있음을 볼 수 있다. 또한, Arcsin 변환은 붓스트랩방법보다 평균포함확률이 낮은 수준임을 볼 수 있다. 이는 균등분포시, Arcsin 변환을 이용한 정규근사가 붓스트랩보다 느림을 보여주는 결과라 할 수 있다.

[표 1] $\xi(p)$가 균등분포일 때 각 방법들의 평균포함확률

n	방법	0.90	길이	r	0.95	길이	r	0.99	길이	r
10	EXACT	.958	.466	.514	.982	.528	.512	.997	.637	.500
	POISSON	.989	.608	.568	.997	.680	.572	1.000	.795	.564
	WALD	.764	.357	.496	.796	.414	.500	.829	.512	.501
	SCORE	.903	.376	.499	.948	.440	.503	.988	.548	.499
	ARCSIN	.774	.353	.496	.809	.413	.499	.841	.517	.503
	Ad-WALD	.803	.430	.498	.823	.481	.502	.839	.567	.507
	Ad-SCORE	.947	.441	.480	.975	.499	.489	.994	.597	.495
	WALD-T	.817	.461	.501	.832	.527	.500	.845	.641	.503
	WALD-BS	.828	.502	.501	.823	.481	.502	.839	.567	.507
	SHIFT	.947	.437	.499	.984	.530	.502	.999	.700	.497
	BOOTSTRAP	.825	.650	.509	.848	.773	.507	.849	.844	.508
	SH-BOOT	.800	.525	.418	.834	.628	.403	.896	.744	.456
20	EXACT	.950	.325	.510	.981	.375	.506	.998	.468	.500
	POISSON	.988	.456	.558	.996	.520	.576	1.000	.631	.592
	WALD	.835	.276	.495	.883	.325	.495	.924	.414	.495
	SCORE	.901	.279	.495	.955	.330	.495	.992	.422	.497
	ARCSIN	.853	.274	.497	.901	.324	.498	.938	.415	.494
	Ad-WALD	.882	.318	.497	.909	.365	.496	.934	.450	.497
	Ad-SCORE	.939	.320	.481	.977	.368	.487	.996	.455	.495
	WALD-T	.892	.331	.497	.917	.384	.495	.938	.485	.498
	WALD-BS	.901	.343	.497	.925	.406	.496	.942	.536	.497
	SHIFT	.932	.306	.497	.979	.371	.497	.999	.499	.500
	BOOTSTRAP	.901	.421	.500	.932	.542	.500	.945	.725	.513
	SH-BOOT	.881	.405	.424	.916	.489	.405	.948	.646	.408
30	EXACT	.940	.263	.508	.974	.306	.511	.995	.386	.507
	POISSON	.982	.382	.555	.993	.438	.575	.999	.540	.601
	WALD	.872	.233	.498	.917	.275	.497	.959	.354	.501
	SCORE	.899	.233	.498	.951	.275	.500	.991	.355	.504
	ARCSIN	.888	.231	.501	.933	.274	.500	.969	.354	.501
	Ad-WALD	.912	.263	.497	.940	.304	.497	.965	.380	.499
	Ad-SCORE	.936	.262	.494	.971	.303	.494	.995	.381	.501
	WALD-T	.918	.270	.497	.945	.315	.499	.968	.402	.497
	WALD-BS	.922	.276	.498	.951	.326	.499	.971	.427	.498
	SHIFT	.924	.248	.499	.969	.300	.502	.997	.404	.505
	BOOTSTRAP	.918	.292	.501	.960	.413	.502	.978	.583	.510
	SH-BOOT	.902	.310	.422	.943	.400	.408	.972	.539	.408
50	EXACT	.933	.202	.495	.968	.236	.494	.993	.301	.493
	POISSON	.980	.302	.536	.992	.349	.555	1.000	.437	.593
	WALD	.867	.184	.495	.923	.218	.496	.970	.284	.494
	SCORE	.897	.183	.492	.950	.217	.493	.990	.282	.490
	ARCSIN	.886	.183	.493	.940	.217	.492	.984	.282	.492
	Ad-WALD	.907	.203	.496	.949	.237	.494	.979	.301	.495
	Ad-SCORE	.930	.202	.492	.966	.235	.490	.994	.300	.488
	WALD-T	.914	.206	.496	.953	.242	.494	.982	.312	.494
	WALD-BS	.918	.209	.496	.955	.247	.493	.984	.323	.493
	SHIFT	.915	.191	.492	.960	.230	.494	.996	.308	.491
	BOOTSTRAP	.923	.207	.487	.963	.263	.491	.991	.403	.504
	SH-BOOT	.905	.223	.430	.953	.286	.421	.985	.403	.417

※ 각 방법들은 제시된 식 (2.1) ~ 식 (2.12)의 신뢰구간을 각각 의미한다.

셋째, 4가지의 연속성 수정방법들의 평균포함확률을 살펴보자. Score 방법을 연속성 수정(Ad-SCORE)한 방법은 상당히 과대추정되는 경향을 나타내므로 Score 방법에 연속성 수정을 할 필요는 없어 보인다. 4가지 방법은 각각 식 (2.6)의 연속성 수정(Ad-WALD), T-분포의 연속성 수정(WALD-T), 식 (2.9)의 연속성 수정(WALD-BS) 그리고 Score 방법의 연속성 수정(Ad-SCORE) 순으로 평균포함확률이 큼을 볼 수 있다. 연속성 수정방법은, 표본의 크기가 작을 때 어느 정도 효율이 있는 것으로 알려져 있으나, 평균포함확률 측면에선 특별히 기준이 되는 표

본의 크기를 발견할 수가 없었다. 한 예로, 명목수준이 90%에서, 균등분포에서는 표본의 크기가 20이하일 때 연속성 수정방법이 어느 정도 잘 작동하나, Beta 분포에서는 과대추정하고 있음을 볼 수 있다. 결과적으로 연속성 수정은 표본의 크기가 크면 유용한 방법이 아님을 볼 수 있다.

[표 2] $\xi(p)$가 Beta(5, 5)일 때 각 방법들의 평균포함확률

n	방법	0.90	길이	r	0.95	길이	r	0.99	길이	r
10	EXACT	.956	.512	.505	.979	.577	.510	.997	.690	.508
	POISSON	.989	.692	.677	.998	.764	.682	1.000	.858	.664
	WALD	.837	.454	.498	.887	.534	.497	.939	.671	.498
	SCORE	.899	.423	.500	.953	.488	.504	.990	.593	.502
	ARCSIN	.871	.440	.503	.923	.514	.503	.963	.643	.505
	Ad-WALD	.898	.541	.498	.928	.612	.499	.957	.730	.502
	Ad-SCORE	.940	.497	.498	.972	.555	.501	.995	.649	.498
	WALD-T	.920	.585	.499	.945	.677	.498	.968	.820	.508
	WALD-BS	.937	.642	.499	.928	.612	.499	.957	.730	.502
	SHIFT	.940	.485	.501	.981	.582	.504	.999	.762	.507
	BOOTSTRAP	.958	.666	.505	.973	.848	.523	.975	.966	.519
	SH-BOOT	.916	.594	.396	.953	.738	.352	.971	.891	.397
20	EXACT	.941	.371	.493	.972	.427	.498	.997	.530	.500
	POISSON	.986	.532	.665	.997	.608	.696	1.000	.731	.725
	WALD	.874	.339	.497	.924	.404	.497	.968	.526	.497
	SCORE	.905	.323	.494	.952	.377	.497	.992	.472	.498
	ARCSIN	.895	.332	.496	.942	.392	.493	.983	.503	.497
	Ad-WALD	.920	.388	.497	.950	.452	.496	.978	.570	.496
	Ad-SCORE	.938	.366	.493	.969	.418	.496	.996	.509	.497
	WALD-T	.928	.405	.497	.958	.478	.496	.987	.621	.494
	WALD-BS	.938	.422	.497	.965	.508	.496	.992	.693	.497
	SHIFT	.925	.347	.493	.971	.416	.497	.998	.551	.498
	BOOTSTRAP	.944	.399	.501	.984	.490	.500	.997	.735	.527
	SH-BOOT	.932	.396	.460	.967	.473	.428	.992	.684	.390
30	EXACT	.939	.304	.492	.971	.353	.496	.993	.444	.499
	POISSON	.986	.441	.654	.996	.512	.687	1.000	.635	.730
	WALD	.892	.281	.497	.937	.334	.497	.976	.438	.498
	SCORE	.908	.271	.493	.957	.318	.498	.990	.403	.498
	ARCSIN	.904	.277	.494	.950	.327	.494	.985	.424	.497
	Ad-WALD	.926	.314	.497	.957	.367	.497	.982	.470	.497
	Ad-SCORE	.939	.301	.493	.971	.347	.498	.994	.430	.498
	WALD-T	.931	.323	.497	.963	.382	.497	.987	.500	.497
	WALD-BS	.937	.331	.497	.967	.396	.498	.990	.536	.498
	SHIFT	.925	.284	.493	.968	.340	.498	.995	.450	.498
	BOOTSTRAP	.928	.292	.494	.976	.368	.507	.998	.538	.520
	SH-BOOT	.926	.295	.474	.969	.368	.460	.993	.525	.440
50	EXACT	.922	.234	.514	.960	.274	.511	.993	.349	.514
	POISSON	.981	.343	.639	.995	.402	.670	1.000	.512	.737
	WALD	.889	.219	.516	.936	.261	.516	.982	.343	.516
	SCORE	.898	.214	.514	.945	.253	.512	.990	.325	.518
	ARCSIN	.895	.217	.511	.941	.258	.511	.988	.335	.512
	Ad-WALD	.913	.239	.516	.954	.281	.516	.986	.363	.516
	Ad-SCORE	.922	.233	.514	.960	.271	.512	.993	.342	.518
	WALD-T	.917	.243	.516	.958	.287	.516	.989	.376	.515
	WALD-BS	.921	.246	.516	.961	.293	.516	.991	.391	.515
	SHIFT	.907	.220	.514	.954	.263	.512	.994	.348	.518
	BOOTSTRAP	.913	.230	.514	.961	.276	.518	.996	.379	.531
	SH-BOOT	.914	.231	.504	.962	.278	.499	.994	.380	.484

※ 각 방법들은 제시된 식 (2.1) ~ 식 (2.12)의 신뢰구간을 각각 의미한다.

넷째, 식 (2.10)의 신뢰구간(SHIFT)은 Score 신뢰구간보다 항상 큰 명목수준을 유지하고 있으며, 표본의 크기가 커짐에 따라 명목수준에 근사하고 있음을 볼 수 있다.

다섯째, 붓스트랩 방법은 표본의 크기가 $n=10$ 처럼 작을 때 하향추정하는 경향이 있다. 이는 모비율이 0이나 1에 가까우면 그 경향이 더 심해진다. 즉, 표본의 크기가 작을 때 Wald 신뢰구간과 같이 하향추정하나 표본의 크기가 20이 되면서부터 빠른 속도로 명목수준에 근사함을 볼 수 있다. 우리는 여기서, Wald 방법은 표본의 크기가 커짐에도 불구하고 명목수준에 접근하는 속도가

붓스트랩에 비해 매우 느림을 볼 수 있다. 이는 Woodroofe and Jhun(1989)의 결과처럼 붓스트랩이 Wald의 정규근사보다 수렴속도가 평균포함확률 측면에서 빠름을 보여주는 결과라 하겠다. 식(2.12)방법의 붓스트랩 신뢰구간(SH-BOOT)은 기존의 붓스트랩이 지니는 약간의 과대추정을 보완해 준다고 하겠다. 균등분포의 경우 명목수준이 90%, 95%에서 SH-BOOT 방법의 평균포함확률이 상당히 명목수준에 근사함을 볼 수 있겠다.

사전분포가 Beta(0.5.,0.5)인 경우에도 모의실험의 결과 역시 이들과 비슷하나 명목수준에 접근하는 정도가 다소 떨어지며, 이 때는 붓스트랩 방법과 Wald 방법과 큰 차이가 나타나지 않음을 알 수 있었다. 결론적으로 p의 신뢰구간을 구축하는 방법에서는 Score 방법이 가장 정확하며, Score 방법이 지니는 복잡한 형태의 신뢰구간에 비해 붓스트랩 분포를 활용한 신뢰구간도 나름대로 장점을 지니고 있다고 할 수 있다.

참고문헌

[1] Agresti, A. and Coull, B.A.(1998) Approximate is better than "Exact" for interval estimation of binomial proportions, *The American Statistician*, Vol. 52, 119-126.

[2] Blyth, C.R.(1986) Approximate binomial confidence limits, *Journal of the American Statistical Association*, Vol. 81, 843-855;Corrigenda(1989), Vol. 84, 636.

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

[4] Chen, H.(1990) the accuracy of approximate intervals for a binomial parameter, *Journal of the American Statistical Association*, Vol. 85, 514-518.

[5] Clopper, C.J. and Pearson, E.S.(1934) The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika,* Vol. 26, 404-413.

[6] Efron, B.(1979) Bootstrap methods : another look at the jackknife. *The annals of Statistics*, Vol. 7, 1-26.

[7] Ghosh, B. K.(1979) A comparison of some approximate confidence intervals for the binomial parameter, *Journal of the American Statistical Association*, Vol. 74, 894-900.

[8] Hald, A.(1952) *Statistical Theory with Engineering Applications*, John Wiley, New York.

[9] Leemis, L.M. and Trivedi, K.S.(1996) A comparison of approximate interval estimators for the bernoulli parameter, *The American Statistician*, Vol. 50, 63-68.

[10] Singh, K.(1981) On the asymptotic accuracy of Efron's bootstrap, *The Annals of Statistics*, Vol. 9, 1187-1195.

[11] Vollset, S.E.(1993) Confidence intervals for a binomial proportion, *Statistics in Medicine*, Vol. 12, 809-824.

[12] Woodroofe, M. and Jhun, M.(1989) Singh's theorem in the lattice case. *Statistics Probability Letters*, Vol. 7, 201-205.

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