COMPARISON OF FIVE EXACT CONFIDENCE INTERVALS FOR THE BINOMIAL PROPORTION

[1]Iori Sakakibara, [2]Emiko Haramo,
[2]Akimasa Muto, [1]Ikuya Miyajima and [3]Yohei Kawasaki

[1]Department of Business Solution, Takumi Information Technology, Tokyo, Japan
[2]Department of System Development, Takumi Information Technology, Tokyo, Japan
[3]Department of Mathematics, Tokyo University of Science, Tokyo, Japan

ABSTRACT

The Wald interval is easy to calculate; it is often used as the confidence interval for binomial proportions. However, when using this confidence interval, the actual coverage probability often falls under the nominal coverage probability in small cases. On the other hand, several confidence intervals where the actual cover age probability does not fall under the nominal coverage probability are suggested. In this study, we intro-duce five exact confidence intervals where the actual coverage probability does not fall under the nominal coverage probability and we calculate the expected length of the confidence intervals and compare/verify the accuracy of the coverage probabilities. Further, we examined the characteristics of these five exact confidence intervals at length. Coverage probability of Sterne was significantly closer to 0.95 than the other confidence intervals and stable. Its expected Length are not scattered in the width compared with the other methods. As a result, we found that the quality of the confidence interval based on the Sterne test is its availability for small samples.

Keywords: Binominal Proportion, Confidence Interval, Exact Method

1. INTRODUCTION

Studies on confidence intervals for binomial proportions have been performed since a long time ago and continue to be performed. Because the Wald interval is easy to calculate, it is often used as the confidence interval for binomial proportions. In addition, new confidence intervals have been introduced by (Agresti and Coull, 1998; Newcombe, 1998). However, when using these new confidence intervals, the actual coverage probability often falls under the nominal coverage probability.

On the other hand, several confidence intervals where the actual coverage probability does not fall under the nominal coverage probability are suggested. Clopper and Pearson (1934) suggest a construction method for a confidence interval based on an exact test using binominal proportions. The confidence interval is a method for the actual coverage probability not to fall under the nominal coverage probability at all times, but it has been indicated that this method is extremely conservative (Agresti and Coull, 1998).

In addition, several other exact methods have been suggested. Reiczigel (2003) suggested a method for using the Sterne test to resolve the problems of the method of Clopper and Pearson (1934). This method is easy to understand and program. Fleiss *et al*. (2003) constructed a confidence interval based on an exact test that uses the likelihood ratio test statistic on the binomial distribution test. Hirji (2006) constructed a confidence interval with a method based on an exact test using score test statistics. However, no papers have compared these exact confidence intervals in detail.

Corresponding Author: Iori Sakakibara, Department of Business Solution, Takumi Information Technology, Tokyo, Japan

This study introduces five exact confidence intervals where the actual coverage probability does not fall under the nominal coverage probability; moreover, we calculate the expected length of the confidence interval and compare/verify the accuracy of the coverage probabilities.

This study is organized as follows. In section 2, we present the construction method for the five exact confidence intervals. In section 3, we detect the behavior of the confidence intervals by conducting a simulation. A conclusion is provided in section 4.

2. NOTATION AND METHODS

In this section, we introduce the methodology for the five exact confidence intervals that we discuss in this study.

Let X be independent random variables. Suppose that X follows a binominal distribution with parameters n, π.

2.1. Clopper-Pearson Confidence Interval

The Clopper-Pearson confidence interval is an early and considerably common method for calculating binomial confidence intervals. The Clopper-Pearson confidence interval is commonly called an exact confidence interval because it is based on the cumulative probabilities of the binomial distribution; however, the intervals are not exact in the manner one might assume: The discontinuous nature of the binomial distribution precludes any interval with exact coverage for all population proportions. The Clopper-Pearson confidence interval can be written as Equation 1:

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

where, $F_{a,b}(\alpha)$ is the upper $100(\alpha/2)\%$ quartile from an F-distribution with a and b degrees of freedom.

2.2. Exact Likelihood Ratio Confidence Interval

The exact Likelihood Ratio (LR) confidence interval is based on inverting the acceptance regions for the exact binomial tests of H_0: $\pi = \pi_0$. Following Fleiss *et al*. (2003) and given α and true $\pi = \pi_0$, we define the Generalized Log LR (GLLR) statistic as Equation 2:

$$\begin{aligned} &GLLR\left(\pi_0 \mid x\right) \\ &= x\ln\left\{x/\left(n\pi_0\right)\right\}+(n-x)ln\left\{(n-x)/\left\{n\left(1-\pi_0\right)\right\} \end{aligned} \tag{2}$$

For $\pi_0 = 0$, GLLR is only defined for x = 0; for $\pi_0 = 1$, GLLR is only defined for x = n. We define the attained LR p-value as Equation 3:

$$p-value = \sum_t \binom{n}{t}\pi_0\left(1-\pi_0\right) \tag{3}$$

where, the sum is taken over the set t of xi values for which GLLR $(\pi_0|xi) \geq$ GLLR$(\pi_0|x)$, excluding those values where GLLR is not defined. Then, the exact LR confidence set is the set of all π_0 such that the *p*-value $\geq \alpha$.

2.3. Exact Score Confidence Interval

The exact Score Confidence (SC) interval is based on inverting the acceptance regions for the exact score tests of H_0: $\pi = \pi_0$. Following Hirji (2006) and given α and true $\pi = \pi_0$, we define the score statistic as Equation 4:

$$SC\left(\pi_0 \mid x\right) = \left(x-n\pi_0\right)^2 / \left(n\pi_0\left(1-\pi_0\right)\right) \tag{4}$$

For $\pi_0 = 0$, SC is only defined for $x = 0$ and SC = 0. For $\pi_0 = 1$, SC is only defined for $x = n$ and SC = 0. We define the attained score *p*-value as Equation 5:

$$p-value = \sum_t \binom{n}{t}\pi_0\left(1-\pi_0\right) \tag{5}$$

where, the sum is taken over the set t of x_i values for which $SC(\pi_0|x_i) \geq SC(\pi_0|x)$, excluding those values where *SC* is not defined. Then, the exact score confidence set is the set of all π_0 such that the *p*-value $\geq \alpha$.

2.4. Sterne Confidence Interval

The interval proposed by Reiczigel (2003) is defined by inverting the exact binomial test with acceptance regions, including the most probable values of the binomial variable and then taking the most probable, followed by the next most probable, until their total probability reaches the required level, for example, 95%.

Assume that we want to invert a test of H_0: $\pi_1 = \pi_0$ for the binomial parameter π to obtain a 95% confidence interval for π based on $n = 5$ observations. Denote X_1 to be the observed number of successes. The basic idea is that a 95% confidence set should consist

of all such values π_0 of the parameter for which H0: $\pi_1 = \pi_0$ is not rejected by the test at the 95% level. For simplicity, assume that a one-digit precision is sufficient for the interval endpoints, because in such a case, the procedure can be demonstrated using a small table of binomial probabilities (**Table 1**).

For $\pi = 0.4$ and acceptance region X_1 to sum the probability of each exceeds 0.95 the first time up to 0-4. To determine the acceptance region for each of π, we apply the value of π ranging from 0.0 to 1.0. Thus, we observe X_1 by the determined acceptance region. In the case of $X_1 = 3$, the region 0.2 to 0.9 has become the acceptance region (see the underlined portion in **Table 1**); the minimum π is the lower confidence bound and the maximum π is the upper confidence bound.

2.5. Blaker Confidence Interval

Blaker (2000) has proposed a new exact interval that is an excellent alternative to the Sterne interval and that has many commonalities with the Sterne interval. Because the Blaker confidence interval is such an excellent alternative, please refer to Blaker (2000) for the calculation method of this confidence interval.

3. RESULTS AND DISCUSSION

In this study, the coverage probability and the expected length were used as the basis for our evaluation and 95% of each confidence interval was compared.

The coverage probabilities were computed using the proportion with which the confidence interval includes the binominal proportion. A simulation of 100,000 rounds of under defined values of π_1 was conducted. Similarly, the expected lengths were computed using the mean of the difference in the confidence intervals.

3.1. Comparison of Coverage Probability

Figure 1 and 2 show the coverage probabilities of the five exact confidence intervals for $n = 5$ (**Fig. 1**), n

= 10 (**Fig. 2**), a significance level of 0.05 (**Fig. 1 and 2**) and 0.001 $\leq\pi\leq$ 0.999 (**Fig. 1 and 2**). Overall, all methods described a high coverage probability for $\pi = 0$ and 1; in addition, the values were slightly higher near $\pi = 0.5$. **Figure 1 and 2** indicate that Clopper-Pearson is clearly a conservative method compared with the other methods. The results of the exact GLLR and the exact Score showed higher values depending on the value of π. In addition, the coverage probability of Sterne and Blaker were significantly closer to 0.95 than the other confidence intervals.

Figure 3 and 4 show the coverage probability of the five exact confidence intervals for $n = 5$ to 95 (**Fig. 3 and 4**), a significance level of 0.05 (**Fig. 3 and 4**) and $\pi = 0.25$ (**Fig. 3**) and 0.50 (**Fig. 4**). Consequently, **Fig. 3 and 4** describe the coverage probability that appears close to 95% as n increases. The results of these figures indicate that the coverage probability varies for Clopper-Pearson, the exact Score and Blaker by the value of n. Moreover, the Blaker interval was considerably close to 95%. For exact GLLR and Sterne intervals, the value variation is small and it is closer to 95% as n increases.

3.2. Comparison of Expected Length

Figure 5 shows the expected length of the five exact confidence intervals for $n = 5$ and 10, a significance level of 0.05 and 0.001 $\leq\pi\leq$0.500. Clopper-Pearson is clearly conservative compared with the other methods. For $\pi = 0$, the expected length values are smaller for GLLR; however, for $\pi = 0.5$, the values are larger. For the Score method, the values are smaller compared with the other methods when $\pi = 0.5$; however, for $\pi = 0$, the values are larger and varied. In addition, Sterne and Blaker are not scattered in the width compared with the other methods and their values are similar.

Table 1. Binomial probabilities

| X_1	π										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
5	0.000	0.000	0.000	0.002	0.010	0.031	0.078	0.168	0.328	0.590	1.000
4	0.000	0.000	0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.000
3	0.000	0.008	0.051	0.132	0.230	0.313	0.346	0.309	0.205	0.073	0.000
2	0.000	0.073	0.205	0.309	0.346	0.313	0.230	0.132	0.051	0.008	0.000
1	0.000	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006	0.000	0.000
0	1.000	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
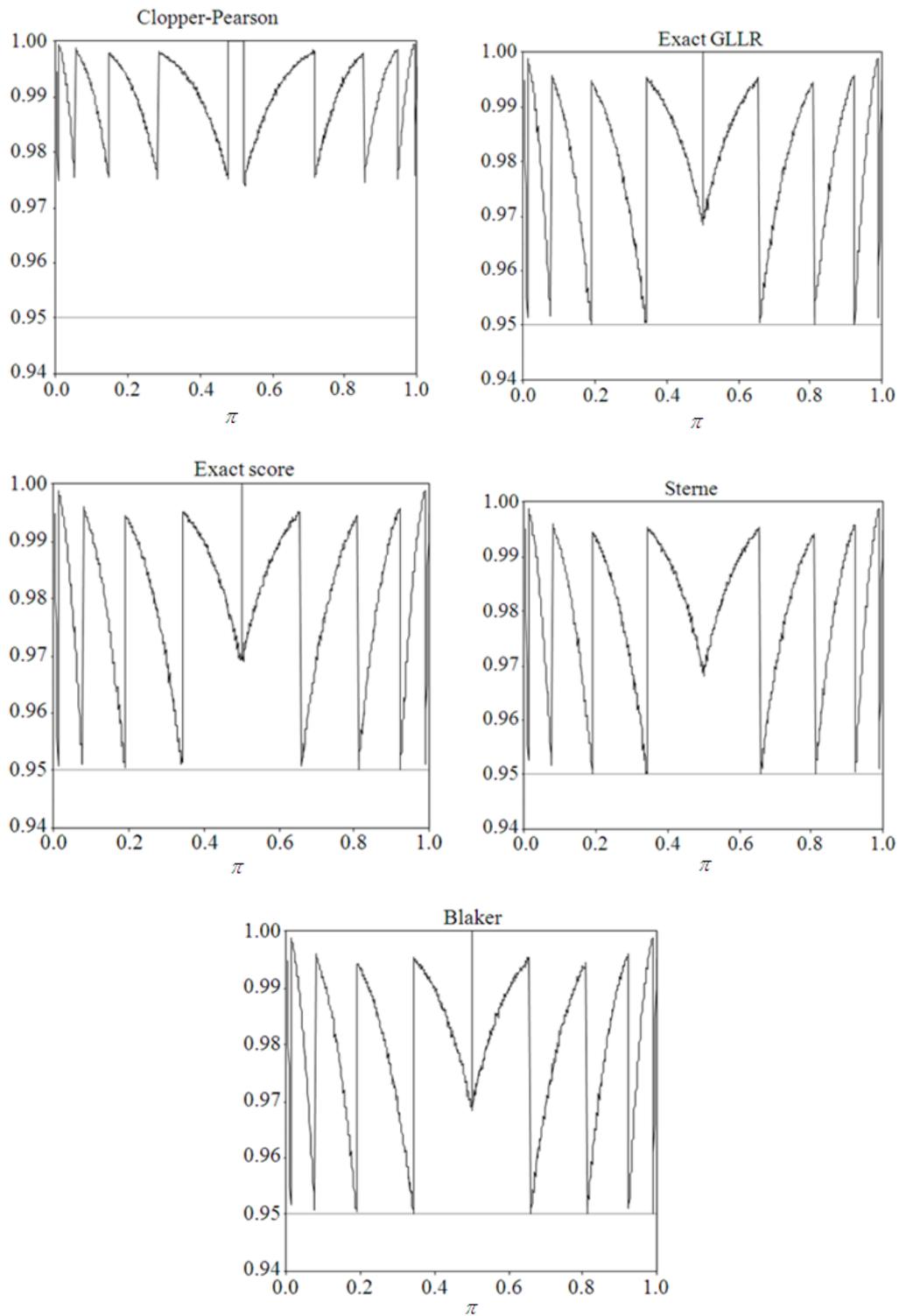
Fig. 1. Coverage probability (*n* = 5)
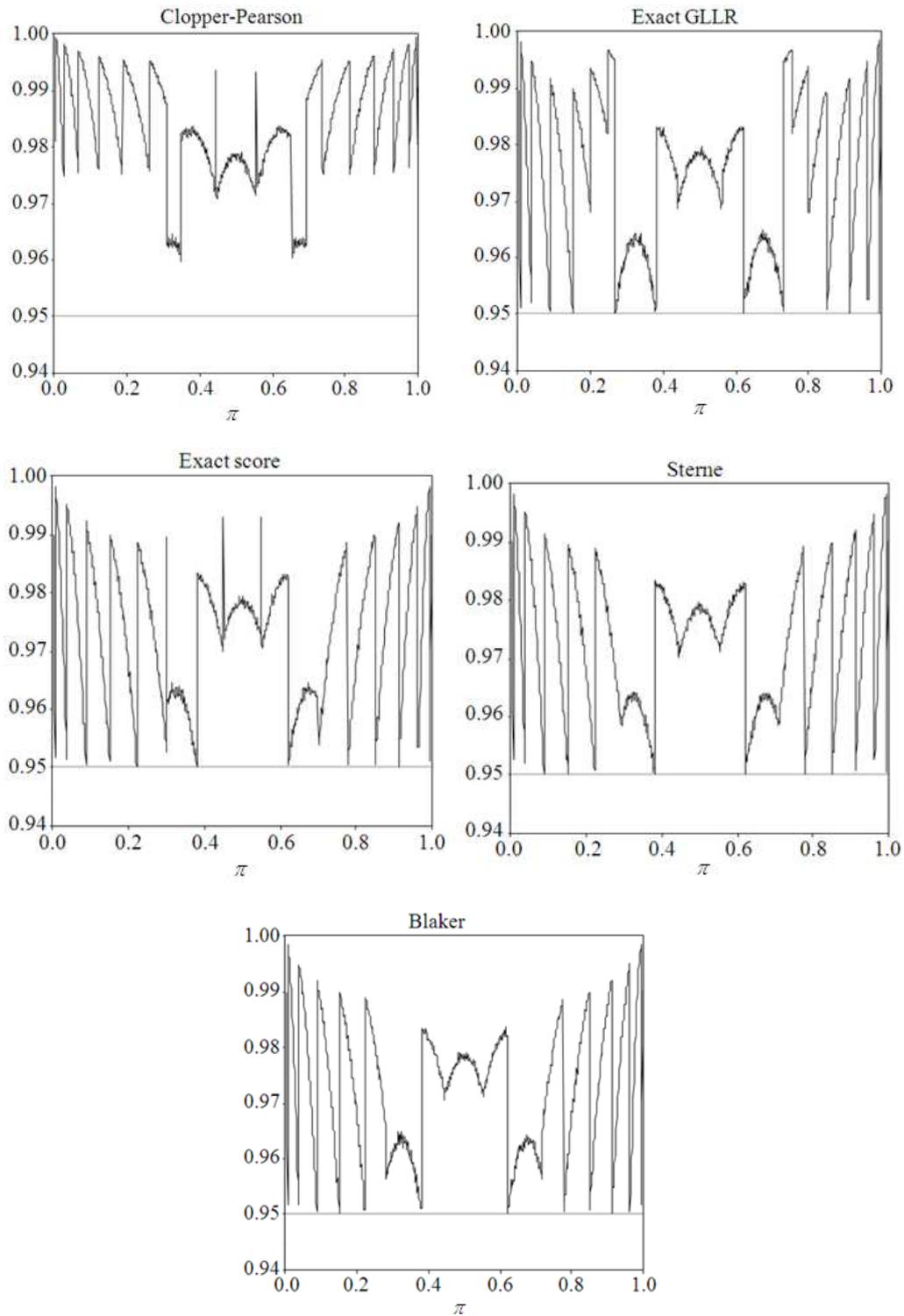
Fig. 2. Coverage probability (*n* = 10)
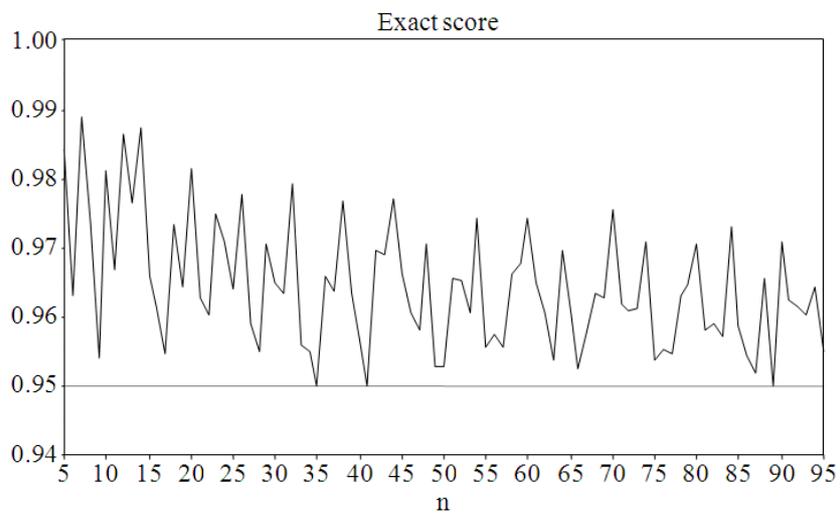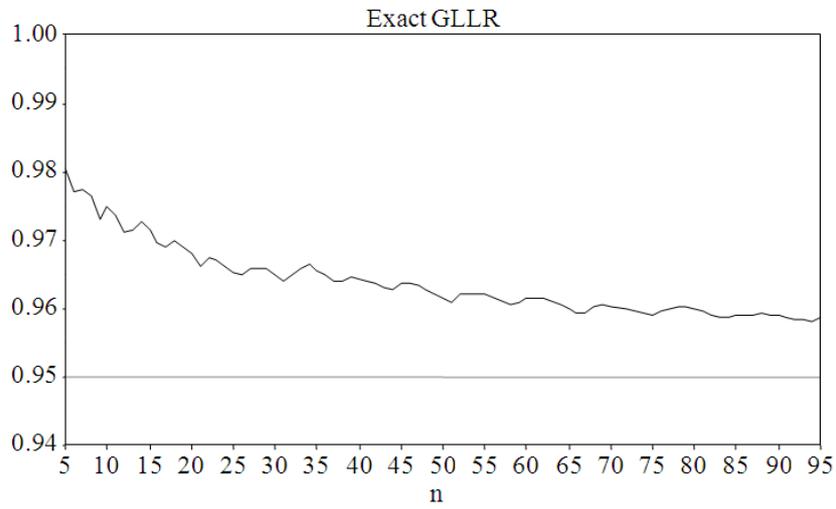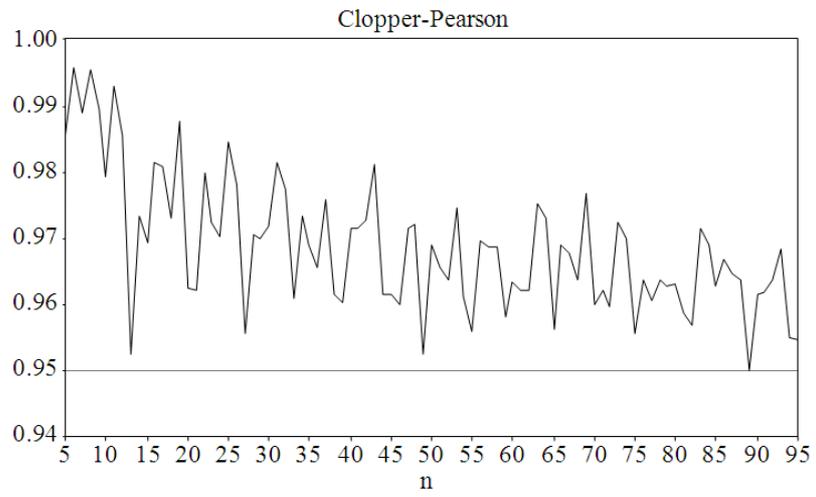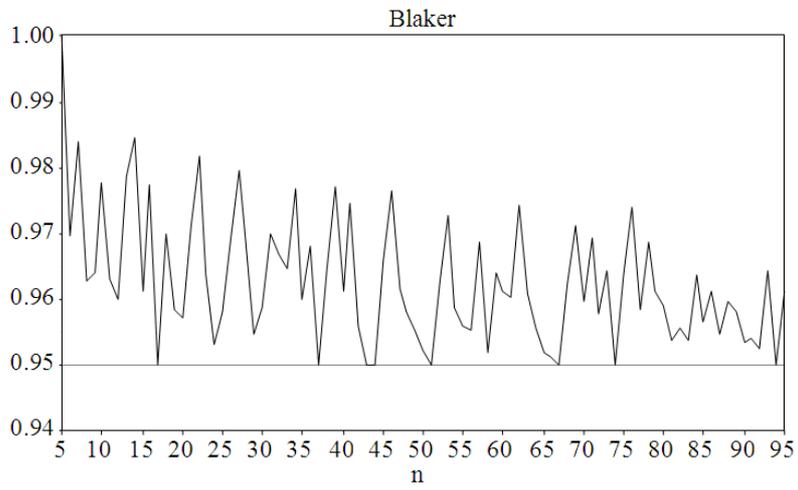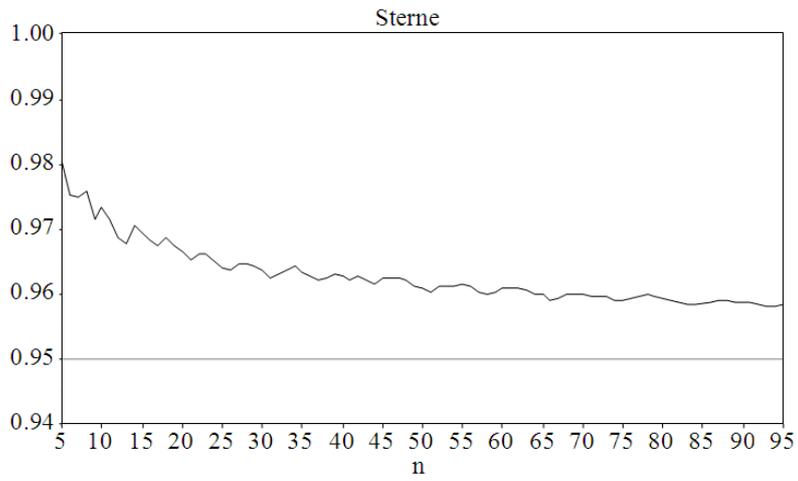
Fig. 3. Coverage probability ($\pi = 0.25$)

Exact GLLR

Exact score

Sterne
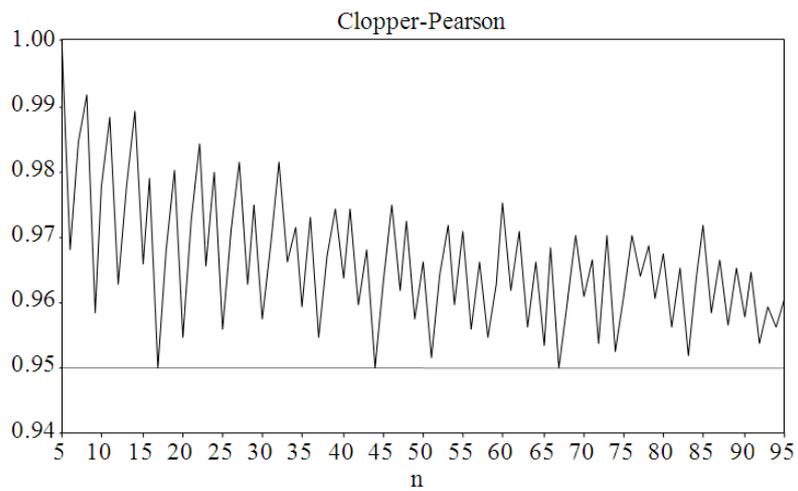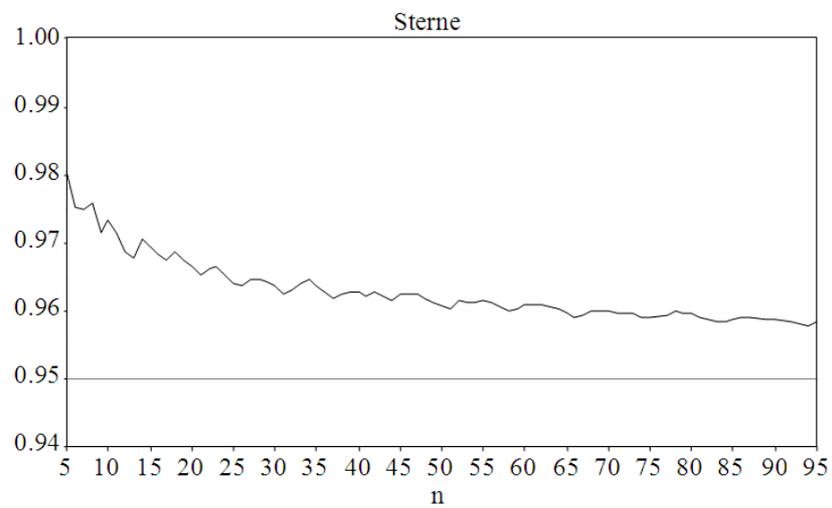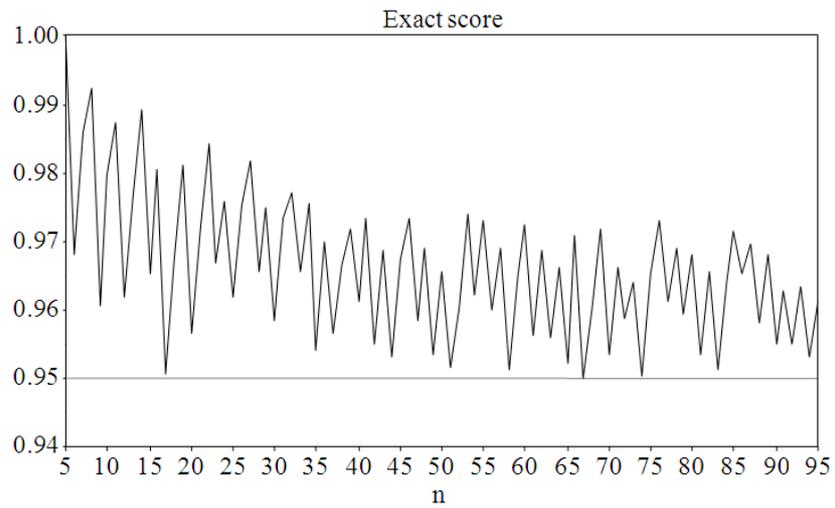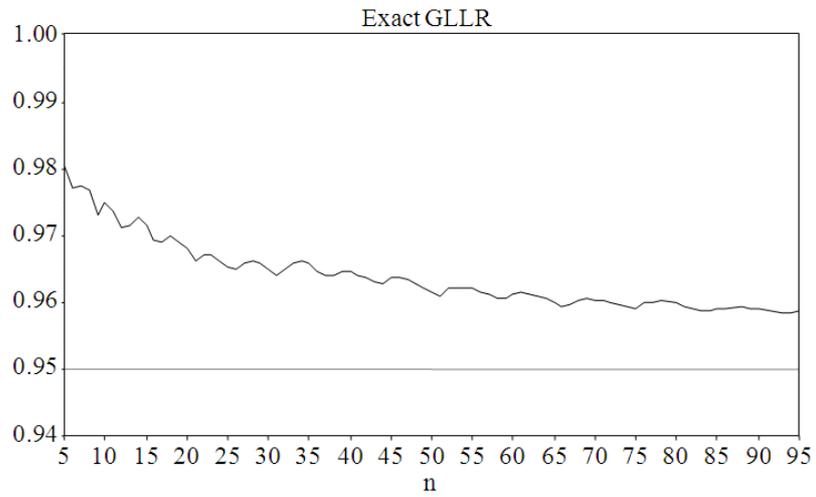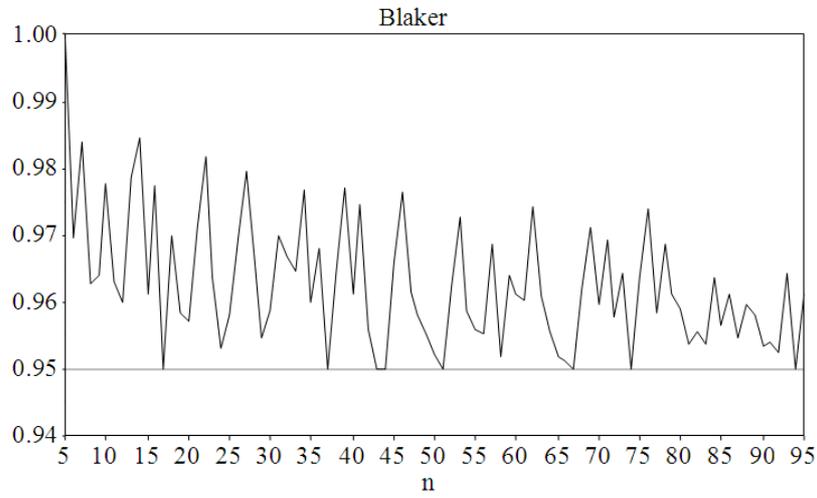
**Fig. 4.** Coverage probability ($\pi = 0.50$)
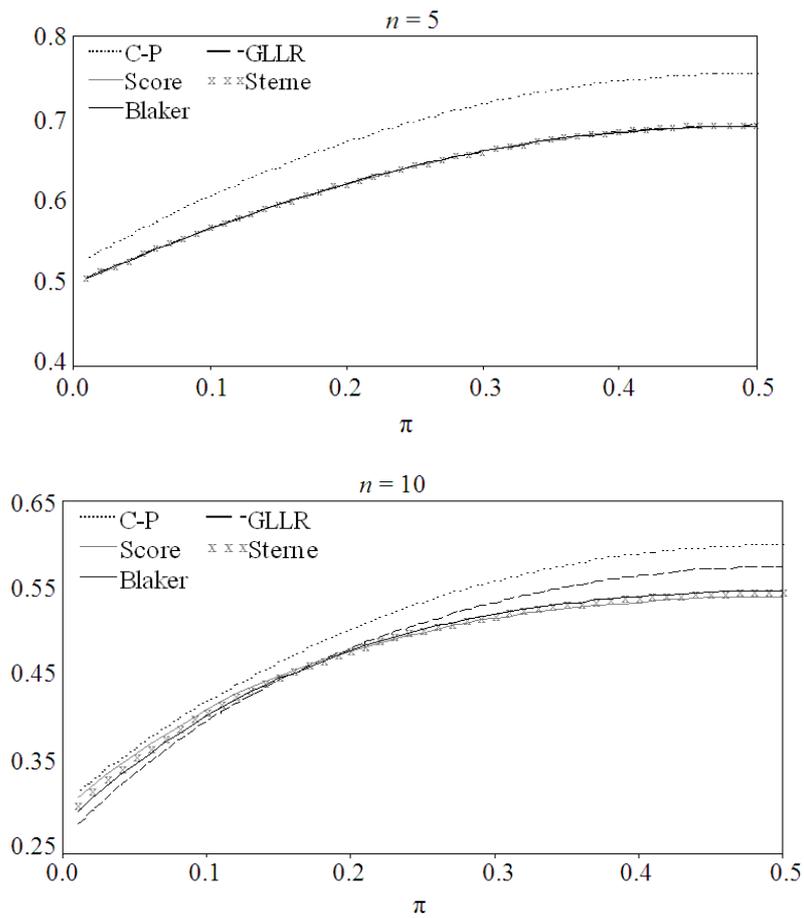
Fig. 5. Expected length ($n = 5, 10$)

4. CONCLUSION

We examined five exact confidence intervals that do not fall under the nominal coverage probability in order to determine the most useful method for small sample sizes. In this study, we calculated the expected length of the confidence intervals and compared/verified the accuracy of the coverage probabilities.

The results indicated that for all five exact confidence intervals, when $\pi = 0$ and close to 1, the values of the coverage probability are higher and the expected lengths are larger. For the Clopper-Pearson method, we found that the expected length and coverage probability of the Clopper-Pearson method is even more conservative than the other methods. For the exact GLLR method, the evaluated values were near the edge of the expected length; however, for $\pi = 0.5$, the values were conservative, similar to the Clopper-Pearson method. However, the values were stable and varied less with respect to n. For the exact Score method, the values in the coverage probability for $n = 5$ and 10 tended to appear high depending on π and the variation was significantly related to variations of n. Moreover, the calculated expected length appeared scattered; for example, the values near the end were larger and the values were smaller for $\pi = 0.5$. For the Sterne and Blaker methods, the values were comparatively close to 95% for the coverage probabilities of $n = 5$ and 10 and for the expected lengths; however, the Blaker method showed scattering values of the coverage probabilities related to the variance of n, whereas the Sterne method was stable. In summary, we considered the Sterne confidence interval method to be more useful than the other methods in small sample sizes.

5. ACKNOWLEDGEMENT

6. REFERENCES

Agresti, A and B.A. Coull, 1998. Approximate is better than "exact" for interval estimation of binomial pro-portions. Am. Stat., 52: 119-126. DOI: 10.2307/2685469

Blaker, H., 2000. Confidence curves and improved exact confidence intervals for discrete distributions. Canadian J. Stat., 28: 783-798. DOI: 10.2307/3315916

Clopper, C.J. and E.S. Pearson, 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26: 404-413. DOI: 10.1093/biomet/26.4.404

Fleiss, J.L., B. Levin and M.C. Paik, 2003. Statistical Methods for Rates and Proportions. 3rd Edn. John Wiley and Sons, New York, ISBN-10: 0-471-52629- 0, pp: 41.

Hirji, K., 2006. Exact Analysis of Discrete Data. 1st Edn., CRC Press, ISBN-10: 142003619X, pp: 552.

Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Stat. Med., 17: 857-872. DOI: 10.1002/sim.2164

Reiczigel, J., 2003. Confidence intervals for the binomial parameter: Some new considerations. Stat. Med., 22: 611-621. DOI: 10.1002/sim.1320