Journal of Statistical Modelling: Theory and Applications
Vol. 4, No. 2, 2023, pp. 111-118
Yazd University Press 2023

JSMTA

Research Paper

Comparative evaluation of different confidence intervals for the binomial proportion based on maximum coverage probabilities

Alireza Shirvani*
Department of Mathematics and Statistics, Faculty of Basic Sciences and Engineering, Velayat University, Iranshahr, Iran

Received: April 23, 2024/ Revised: August 18, 2024/ Accepted: August 31, 2024

Abstract: Since the binomial distribution is discrete, finding accurate confidence intervals for its parameters is not easily achievable. Many approximate confidence intervals have been suggested for the binomial distribution's ratio parameter thus far. As we know, these confidence intervals have not been compared based on the maximum coverage probabilities. This article aims to evaluate six widely used confidence intervals for the binomial distribution's ratio parameter, focusing on their accurate maximum coverage probabilities criterion.

Keywords: Binomial distribution; Confidence interval; Coverage probability.
Mathematics Subject Classification (2010): 62Fxx.

1 Introduction

In everyday life, we encounter many experiments that can have only two results. Examples include tossing a coin, medical tests with negative or positive outcomes, acceptance in an exam, penalty shots in football games, and all questions with yes or no answers. One of these two results is considered a success. As we know, human beings are perfectionists who continually seek success and advancement. The ability to take risks and make choices is essential for achieving success. Making a choice can lead to success or failure, and thus, most of the time, we are unconsciously subjected to a Bernoulli trial. Therefore, estimating the probability of success in a Bernoulli trial would greatly aid precise decision-making. The Bernoulli random variable is a special case of the binomial random variable, and the binomial distribution is a discrete distribution for which

*Corresponding author: `a.shirvani@velayat.ac.ir`

the production of accurate confidence intervals is very difficult. Until now, several approximate confidence intervals have been introduced for the binomial proportion. For example, see Clopper and Pearson (1934), Blyth and Still (1983), Kabaila and Lloyd (1997), Brown et al. (2002), Ross (2003), Wang (2007), Guan (2011) and Garthwaite et al. (2024).

Among others, Wang (2007) proposed a methodology to compute the exact confidence coefficient and maximum coverage probabilities of binomial confidence intervals and conducted a comparison of some binomial confidence intervals based on the exact confidence coefficient criterion. Also, based on the algorithm presented by Blyth and Still (1983) to find the optimal confidence interval for the proportion of the binomial distribution in the single-sample case, Peer and Azriel (2023) proposed an algorithm to find optimal confidence intervals for the difference of the proportions of two samples. Furthermore, Sotres-Ramos et al. (2024) assessed the effectiveness of some binomial confidence intervals, aiming to evaluate each interval and make comparisons based on their average coverage probabilities and average expected lengths. In this paper, three confidence intervals considered by Brown et al. (2002), two confidence intervals considered by Blyth and Still (1983), and one considered by Wang (2007) are compared based on the maximum coverage probabilities criterion, proposed by Wang (2007).

The organization of the paper is as follows: Section 2 introduces the Wang (2007) technique for computing accurate maximum coverage probabilities for binomial confidence intervals. Section 3 discusses prominent binomial proportion confidence intervals. In Section 4, we conduct a comparison to evaluate confidence intervals using the accurate maximum coverage probabilities criterion. Lastly, we present our conclusions in Section 5.

2 Methodology

Suppose that X is a binomial random variable with probability mass function $f_p(x)$, where $p \in \Omega = (0,1)$ is an unknown parameter and $x \in S = \{0,1,\ldots,n\}$. If $\big(L(X),U(X)\big)$ is a confidence interval for p, its coverage probability is equal to $P_p\big(p \in \big(L(X),U(X)\big)\big)$, that is, the probability that the random interval $\big(L(X),U(X)\big)$ includes true value of p. In continuous distributions, coverage probability function for all points of parameter space Ω might be the same, but in discrete distributions, coverage probability changes with the changes of unknown parameter p in Ω. Therefore, finding the accurate maximum coverage probabilities for parameters of these distributions is difficult.

In this section, we review the methodology proposed by Wang (2007) to calculate the accurate maximum coverage probabilities of confidence intervals for the ratio parameter of the binomial distribution, under the conditions of Assumption 1.

Assumption 1. *For the confidence interval $\big(L(X),U(X)\big)$, if $X_1 < X_2$, then $L(X_1) < L(X_2)$ and $U(X_1) < U(X_2)$, that is $L(x)$ and $U(x)$ should be increasing functions of x.*

For a confidence interval $\big(L(X),U(X)\big)$, there exist $2(n+1)$ endpoints, denoted by

$$L(0), L(1), \ldots, L(n), U(0), U(1), \ldots, U(n),$$

corresponding to $X = 0, \ldots, n$. Sorting these $2(n+1)$ endpoints increasingly, we define v_i as the ith point in sorted mode, for $i = 1, \ldots, 2(n+1)$. We also define

$$w_i = \begin{cases} v_i, & \text{if } 0 < v_i < 1, \\ 0, & \text{if } v_i \leq 0, \\ 1, & \text{if } v_i \geq 1, \end{cases}$$

and for each $p \in (v_i, v_{i+1})$, functions k_0 and k_1 were defined as

$$\begin{aligned} k_0(p) &= \min\{x \mid U(x) > p\}, \\ k_1(p) &= \max\{x \mid L(x) < p\}. \end{aligned}$$

In addition, let

$$a_i = r\left(k_0\left(\frac{w_i + w_{i+1}}{2}\right), k_1\left(\frac{w_i + w_{i+1}}{2}\right)\right),$$

where $r(m_0, m_1) = \frac{1}{1+M}$ and

$$M = \left(\frac{\binom{n}{m_1}(n - m_1)}{\binom{n}{m_0}m_0}\right)^{\frac{1}{m_1 - m_0 + 1}}.$$

Theorem 2.1. *If the confidence interval $\big(L(X), U(X)\big)$ satisfies conditions of Assumption 1, then the maximum coverage probabilities of $\big(L(X), U(X)\big)$ is the maximum of $P_p\big(p \in \big(L(X), U(X)\big)\big)$, where $p \in S_1$ and*

$$S_1 = \{w_1, w_2, \ldots, w_{2n+2}, a_1, \ldots, a_{2n+1}\}.$$

Proof. See Wang (2007). □

Based on Theorem 2.1, the steps for calculating the maximum coverage probabilities are as follows:
1. Investigate whether the confidence interval fulfills Assumption 1.
2. Consider the endpoints of intervals corresponding to $X = 0, \ldots, n$, which are located in the parameter space.
3. Obtain the points $a_1, \ldots, a_{2n+1}$ as stated in Theorem 2.1.
4. Calculate the coverage probabilities corresponding to the points obtained in steps 2 and 3, as well as the upper and lower bounds of the parameter space. The maximum coverage probabilities for the interval is the maximum of these values.

3 Confidence intervals

Let k be the upper $\frac{\alpha}{2}$ quantile of standard normal distribution and $\tilde{X} = X + \frac{k^2}{2}$, $\tilde{n} = n + k^2$, $\tilde{p} = \frac{\tilde{X}}{\tilde{n}}$, $\tilde{q} = 1 - \tilde{p}$, $\hat{p} = \frac{X}{n}$ and $\hat{q} = 1 - \hat{p}$. Six confidence intervals that satisfy conditions of Assumption 1 are introduced in Sections 3.1-3.6.

3.1 Exact interval

Clopper and Pearson (1934) derived a method that computes exact binomial confidence interval boundaries based on cumulative binomial probabilities. Also known as the "exact method", this conservative approach tends to produce wide intervals, thereby reducing the risk of overconfidence but increasing the likelihood of including the true population proportion. Explicit form of this confidence interval was obtained by Blyth and Still (1983) as

$$CI_{CP}(X) = \left(\left(1 + \frac{(n - X + 1)F_1}{X}\right)^{-1}, \left(1 + \frac{n - X}{(X + 1)F_2}\right)^{-1}\right),$$

where $F_1 = F_{(1-\frac{\alpha}{2};2n-2X+2,2X)}$, $F_2 = F_{(\frac{\alpha}{2};2X+2,2n-2X)}$, and $F_{(\alpha;r_1,r_2)}$ is the αth upper quantile of F distribution with r_1 and r_2 degrees of freedom.

3.2 Wald interval

Wald interval has been proposed as an interval based on normal approximation by utilizing the sample proportion as an estimator for the true population proportion (p). This method assumes a large sample size and symmetry around the mean. However, it produces inaccurate confidence intervals for small sample sizes or extreme proportion values. The $1 - \alpha$ Wald Confidence interval for p is (see Blyth and Still, 1983)

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

3.3 Wilson interval

The Wilson interval has been introduced as a compromise between exact and normal approximation methods. The Wilson score interval uses a quadratic equation to determine the interval boundaries while correcting for certain biases found in the other methods. It has been widely accepted as a robust method for computing binomial confidence intervals. The $1 - \alpha$ Wilson confidence interval for parameter p is (see Brown et al., 2002)

$$CI_W(X) = \left(\tilde{p} - \frac{k\sqrt{n}}{n + k^2}\sqrt{\hat{p}\hat{q} + \frac{k^2}{4n}}, \tilde{p} + \frac{k\sqrt{n}}{n + k^2}\sqrt{\hat{p}\hat{q} + \frac{k^2}{4n}}\right).$$

3.4 Agresti-Coull interval

The Agresti-Coull interval has been suggested as a modification of the Wald interval by adding "success" and "failure" counts to the sample, which creates an adjusted interval. This method offers more accurate interval estimates in comparison to the Wald interval but may still generate poor results for small or extreme data. The $1 - \alpha$ Agresti-Coull confidence interval for p is (Brown et al., 2002)

$$CI_{AC}(X) = \left(\tilde{p} - k(\tilde{p}\tilde{q})^{\frac{1}{2}}\tilde{n}^{-\frac{1}{2}}, \tilde{p} + k(\tilde{p}\tilde{q})^{\frac{1}{2}}\tilde{n}^{-\frac{1}{2}}\right).$$

3.5 Likelihood ratio interval

The $1-\alpha$ likelihood ratio (LR) confidence interval for parameter p is (Brown et al., 2002)

$$CI_{\Lambda_n}(X) = \left\{ p : \frac{p^X (1-p)^{n-X}}{\left(\frac{X}{n}\right)^X \left(1-\frac{X}{n}\right)^{(n-X)}} > e^{-\frac{k^2}{2}} \right\}.$$

3.6 Jeffreys interval

The Jeffreys interval has been introduced based on a Bayesian method, which works by computing the posterior distribution of the binomial proportion using a non-informative prior distribution. The Jeffreys interval has been widely adopted due to its robustness against different sample sizes and proportions, and its accurate coverage probabilities. The $1-\alpha$ Jeffreys confidence interval for parameter p is

$$\mathrm{CI}_J(X) = \left(\beta_{\left(\frac{\alpha}{2}, X+\frac{1}{2}, n-X+\frac{1}{2}\right)}, \beta_{\left(1-\frac{\alpha}{2}, X+\frac{1}{2}, n-X+\frac{1}{2}\right)} \right),$$

where $\beta_{(\alpha,s_1,s_2)}$ shows the αth upper quantile of $\beta_{(s_1,s_2)}$ distribution (see Wang, 2007).

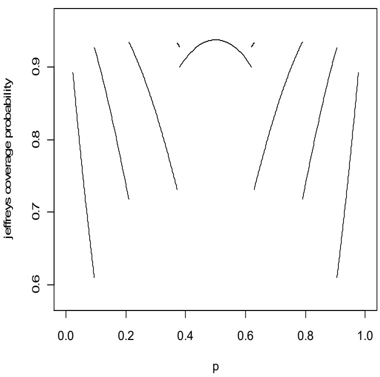
4 Comparison of confidence intervals

As noted in the introduction, Wang (2007) conducted a comparison between binomial confidence intervals based on the exact confidence coefficient criterion. For a given confidence interval, if there exists even one point in the parameter space such that its corresponding coverage probability equals zero, then the confidence coefficient is also zero. For example, the confidence coefficient of the Jeffreys interval is zero (see Wang, 2007). Therefore, this criterion cannot be used to compare all confidence intervals. Additionally, comparing confidence intervals based on the confidence coefficient is very conservative because the coverage probability might be very small at one point but close to the nominal level for other points in the parameter space. Wang (2007) showed that the optimal order of confidence intervals between Wilson, LR, Agresti-Cool, and Jeffreys intervals based on the confidence coefficient is as follows (for $n > 30$): Agresti-Cool, LR, Wilson, and Jeffreys intervals. On the other hand, Sotres-Ramos et al. (2024) conducted a comparison between Wald, exact, Wilson, and Jeffreys intervals in terms of the average coverage probabilities and showed that the optimal order of confidence intervals is as follows: exact, Wilson, Jeffreys, and Wald intervals.

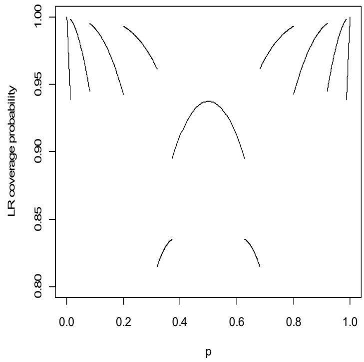
Here we compare the confidence intervals for the proportion parameter of the binomial distribution based on the accurate maximum coverage probabilities criterion. To make these comparisons more tangible, we have first plotted the coverage probability functions of 95% confidence intervals for $n = 5$ in Figure 1.

Based on the figure, the optimal confidence intervals in terms of the minimum coverage probabilities are as follows: exact, Agresti-Coull, Wilson, LR, Jeffreys, and Wald intervals. Additionally, the optimal intervals based on the maximum coverage probabilities are as follows: exact, Agresti-Coull, Wilson, LR, Jeffreys, and Wald intervals.

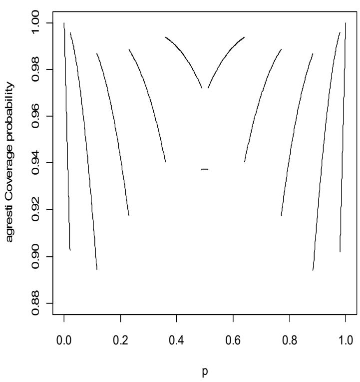
Table 1 displays the maximum coverage probabilities and average lengths of the intervals for various sample sizes. As the table illustrates, until a sample size of 30,



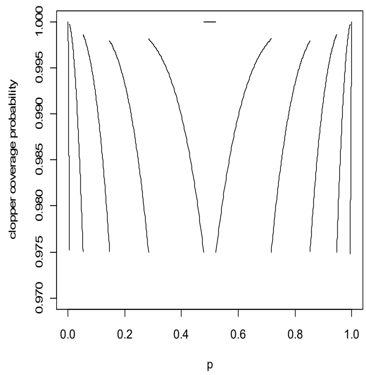
(a) Jeffreys Interval



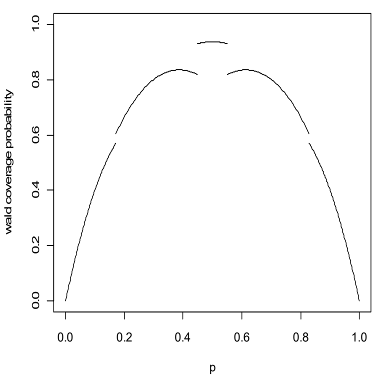
(b) LR Interval



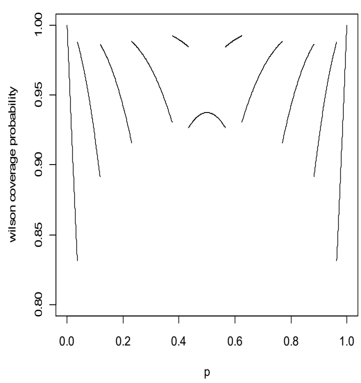
(c) Agresti-Coull Interval



(d) Exact Interval



(e) Wald Interval



(f) Wilson Interval

Figure 1: Coverage probability functions of 95% confidence intervals for $n = 5$.

based on the length of the intervals, confidence intervals in ascending order are: LR, Jeffreys, Wald, Wilson, Agresti-Coull, and exact intervals. However, for larger sample sizes, the intervals do not differ significantly in terms of average lengths. Additionally, the maximum coverage probabilities of Wald intervals for different sample sizes are lower than the corresponding maximum coverage probabilities of Jeffreys intervals. In contrast, the maximum coverage probabilities of Wilson, Agresti-Coull, exact, and LR intervals for various sample sizes remain constant and equal to 1.000. Consequently, the optimality of confidence intervals depends on the selected criterion for comparison, and by changing the comparison criterion, the order of optimality of the intervals also changes.

Table 1: Maximum coverage probabilities (M.C.) and average lengths (A.L.) of confidence intervals (C.I.) for different sample sizes.

	Jeffreys		Exact		LR	
n	A.L.	M.C.	A.L.	M.C.	A.L.	M.C.
5	0.560	0.937	0.678	1.000	0.552	1.000
20	0.323	0.964	0.366	1.000	0.322	1.000
30	0.269	0.965	0.299	1.000	0.268	1.000
50	0.212	0.962	0.231	1.000	0.212	1.000
70	0.180	0.962	0.194	1.000	0.180	1.000
90	0.160	0.962	0.170	1.000	0.160	1.000
100	0.152	0.961	0.161	1.000	0.152	1.000
300	0.088	0.959	0.092	1.000	0.088	1.000
600	0.063	0.959	0.064	1.000	0.063	1.000
900	0.051	0.958	0.052	1.000	0.051	1.000
	Wilson		A-Coull		Wald	
n	A.L.	M.C.	A.L.	M.C.	A.L.	M.C.
5	0.558	1.000	0.606	1.000	0.520	0.937
20	0.325	1.000	0.341	1.000	0.324	0.959
30	0.271	1.000	0.281	1.000	0.270	0.957
50	0.213	1.000	0.218	1.000	0.213	0.955
70	0.181	1.000	0.185	1.000	0.181	0.953
90	0.160	1.000	0.163	1.000	0.160	0.956
100	0.152	1.000	0.154	1.000	0.152	0.956
300	0.088	1.000	0.089	1.000	0.088	0.955
600	0.063	1.000	0.063	1.000	0.063	0.954
900	0.051	1.000	0.051	1.000	0.051	0.954

5 Conclusion

According to Section 4, the optimality of confidence intervals depends on the selected criterion for comparison, and by changing the comparison criterion, the order of optimality of the intervals also changes. In terms of the maximum coverage probabilities criterion, the Jeffreys confidence interval is better than the Wald confidence interval, and both have lower optimality compared to other intervals. However, comparing the optimality of the Wilson, Agresti-Coull, exact, and LR intervals in terms of maximum coverage probabilities is not possible; they should be compared using another criterion, such as the confidence coefficient or average coverage probabilities. For example, based on the confidence coefficient criterion, the Agresti-Cool confidence interval is better than the LR confidence interval (see Wang, 2007), or based on the average coverage probabilities criterion, the exact confidence interval is better than the Wilson confidence interval (see Sotres-Ramos et al., 2024).

References

Blyth, C.R. and Still, H.A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, **78**(381):108–116.

Brown, L.D., Cai, T.T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics*, **30**(1):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**(4):404–413.

Garthwaite, P.H., Moustafa, M.W. and Elfadaly, F.G. (2024). Locally correct confidence intervals for a binomial proportion: A new criteria for an interval estimator. *Scandinavian Journal of Statistics*, **51**(1):220–244.

Guan, Y. (2011). Moved score confidence intervals for means of discrete distributions. *Open Journal of Statistics*, **1**(2):81–86.

Kabaila, P. and Lloyd, C.J. (1997). Tight upper confidence limits from discrete data. *Australian Journal of Statistics*, **39**(2):193–204.

Peer, A. and Azriel, D. (2023). Optimal confidence interval for the difference of proportions. *arXiv preprint arXiv:2308.16650.*

Ross, T.D. (2003). Accurate confidence intervals for binomial proportion and Poisson rate estimation. *Computers in Biology and Medicine*, **33**(6):509–531.

Sotres-Ramos, D.A., Ramírez-Guzmán, M.E., Mora-Aguilera, G. and Rodríguez-Bravo, O.T. (2024). Comparison of confidence intervals for the Bernoulli parameter. *Far East Journal of Mathematical Education*, **26**(1):1–13.

Wang, H. (2007). Exact confidence coefficients of confidence intervals for a binomial proportion. *Statistica Sinica*, **17**(1):361–368.