Asymptotic infimum coverage probability for interval estimation of proportions

Weizhen Wang & Zhongzhan Zhang

Metrika
International Journal for Theoretical and Applied Statistics

ISSN 0026-1335
Volume 77
Number 5

Metrika (2014) 77:635-646
DOI 10.1007/s00184-013-0457-5


ISSN 0026-1335

Metrika

International Journal
for Theoretical and Applied Statistics

Volume 77 (5)
2014

New robust tests for the parameters of the Weibull distribution for complete and censored data
L. Denecke · C.H. Müller 585

On Kullback–Leibler information of order statistics in terms of the relative risk
S. Park 609

Distributions of stopping times in some sequential estimation procedures
A. Jokiel-Rokita · R. Magiera 617

Asymptotic infimum coverage probability for interval estimation of proportions
W. Wang · Z. Zhang 635

Asymptotic properties of *M*-estimators in linear and nonlinear multivariate regression models
C.S. Withers · S. Nadarajah 647

An empirical likelihood inference for the coefficient difference of a two-sample linear model with missing response data
W. Yu · C. Niu · W. Xu 675

A new bounded log-linear regression model
H. Wang · N. Flournoy · E. Kpamegan 695

Further articles can be found at link.springer.com

Instructions for Authors for *Metrika* are available at http://www.springer.com/184

Springer

Springer

Metrika (2014) 77:635–646
DOI 10.1007/s00184-013-0457-5

Asymptotic infimum coverage probability for interval estimation of proportions

Weizhen Wang · Zhongzhan Zhang

Received: 23 September 2012 / Published online: 1 September 2013

Abstract In this paper, we discuss asymptotic infimum coverage probability (ICP) of eight widely used confidence intervals for proportions, including the Agresti–Coull (A–C) interval (Am Stat 52:119–126, 1998) and the Clopper–Pearson (C–P) interval (Biometrika 26:404–413, 1934). For the A–C interval, a sharp upper bound for its asymptotic ICP is derived. It is less than nominal for the commonly applied nominal values of 0.99, 0.95 and 0.9 and is equal to zero when the nominal level is below 0.4802. The $1-\alpha$ C–P interval is known to be conservative. However, we show through a brief numerical study that the C–P interval with a given average coverage probability $1-\gamma$ typically has a similar or larger ICP and a smaller average expected length than the corresponding A–C interval, and its ICP approaches to $1-\gamma$ when the sample size goes large. All mathematical proofs and R-codes for computation in the paper are given in Supplementary Materials.

Keywords Average expected length · Binomial distribution · Bootstrap percentile interval · Coverage probability

Electronic supplementary material The online version of this article (doi:10.1007/s00184-013-0457-5) contains supplementary material, which is available to authorized users.

W. Wang (✉)
Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, USA
e-mail: weizhen.wang@wright.edu

W. Wang · Z. Zhang
College of Applied Sciences, Beijing University of Technology, Beijing100124,
People's Republic of China
e-mail: zzhang@bjut.edu.cn

1 Introduction

Let $X \sim Bin(n, p)$ be a binomial random variable with n trials and a probability of success p. Interval estimation for p based on X is one of the basic problems in statistical inference, and has many applications in practice. There have been tremendous efforts to derive confidence intervals for p. According to confidence coefficient, intervals fit in two categories: approximate and exact intervals. Several widely used approximate intervals include the Wald interval, the Wilson interval (1927), the Agresti–Coull (A–C) interval (1998), the Jeffreys prior interval and the likelihood-based interval; and some exact intervals are the Clopper–Pearson (C–P) interval (1934) and the Blyth-Still interval (1983). Detailed discussion can be found in New combe (1998), Brown et al. (2001), Pires and Amado (2008) and New combe (2011).

To make our statement precise, several definitions are needed here. Let $C(X) = [L(X), U(X)]$ be a confidence interval for p. The coverage probability function of interval $C(X)$ is given by

$$Cover_C(p) = P_p(p \in C(X)) = \sum_{x=0}^{n} I_{C(X)}(p) bin(x; n, p), \tag{1}$$

where $I_{C(X)}(p)$ is an indicator function and $bin(x; n, p) = n! p^x (1-p)^{(n-x)}/(x!(n-x)!)$ is the probability mass function of X. The infimum coverage probability of $C(X)$ (ICP) is equal to

$$ICP_C = \inf_{p \in [0,1]} Cover_C(p), \tag{2}$$

and is also called the confidence coefficient of $C(X)$. Interval $C(X)$ is said to be of level $1-\alpha$ for $\alpha \in (0, 1)$ if

$$ICP_C \geq 1 - \alpha. \tag{3}$$

i.e., the confidence coefficient of $C(X)$ is at least $1-\alpha$. We also call $C(X)$ an exact interval of level $1-\alpha$.

Approximate interval is often derived following the central limit theorem. Interval $C(X)$ is said to be of level $1-\alpha$ approximately point-wise if

$$\inf_{p \in [0,1]} \lim_{n \to \infty} Cover_C(p) \geq 1 - \alpha, \tag{4}$$

and we simply call $C(X)$ an approximate interval (4). For example, the Wilson interval $C_{WI}(X)$ (1927),

$$C_{WI} = [L_{WI}(X), U_{WI}(X)] = \frac{1}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}} \left(\hat{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}} \right), \tag{5}$$

where $\hat{p} = X/n$ and $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$th percentile of the standard normal distribution, is an approximate interval (4), because, by the central limit theorem,

$$\lim_{n\to+\infty} P\left(\left|\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| < z_{\alpha/2}\right) = 1-\alpha.$$

Similarly, the Wald interval (7) and the A–C interval (12) are approximate intervals (4) following the central limit theorem and Slutsky's theorem.

The concept (4) does not serve the purpose of a confidence interval because the speed of convergence to $1-\alpha$ depends on the value of p which is unknown. For a given sample size, there is no guarantee for such an interval to capture the parameter with a predetermined probability. So we propose a new definition for approximate interval: an interval $C(X)$ is said to be of level $1-\alpha$ approximately if

$$\lim_{n\to\infty} \inf_{p\in[0,1]} Cover_C(p) = \lim_{n\to\infty} ICP_C \geq 1-\alpha, \tag{6}$$

and similarly we call $C(X)$ an approximate interval (6). That is, simply switch the order of lim and inf in (4). A point-wise convergence changes to a uniform convergence. The middle quantity of (6) is called the asymptotic confidence coefficient or the asymptotic ICP. This definition assures that the interval includes p at all possible values with a probability close to or larger than $1-\alpha$ if n is large enough. We distinguish two kinds of approximate interval: an approximate (point-wise) interval (4) and an approximate (uniform) interval (6). The latter implies the former, but the opposite is not true.

Obviously, any exact interval is an approximate interval (6). Several researchers (Brown et al. 2001; Agresti 2002, p. 32) point out the ICP of the Wald interval is zero for any n, so it is not an approximate interval (6). Huwang (1995) provides an upper bound of the asymptotic ICP for the Wilson interval, and (6) does not hold either. One goal of this paper is to show that five more widely used asymptotic intervals and a class of bootstrap intervals (in Example 2) do not satisfy (6).

In practice, researchers often say a 95 % Wald interval, a 90 % A–C interval, or a $1-\alpha$ interval $C(X)$ in general. What this really means is that an interval of level $1-\alpha$ is wanted. However, whether the interval they get is truly of level $1-\alpha$ is not guaranteed at all because the confidence coefficient can be greater than, equal to or smaller than $1-\alpha$. For example, any $1-\alpha$ Wald interval

$$C_W(X) = \hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{7}$$

is indeed of level 0 for any n. Therefore, in this paper we distinguish two terminologies:

(i) a $1-\alpha$ confidence interval $C(X)$,
(ii) a confidence interval $C(X)$ of level $1-\alpha$

The second assures the ICP of $C(X)$ at least $1-\alpha$, while the first does not.

The paper is organized as follows. Asymptotic ICPs of seven widely used intervals for proportions are discussed in Sect. 2. In Sect. 3, we propose a new use of the C–P interval by fixing the average coverage probability. This interval is an approximate interval (6) based on our numerical study. Section 4 contains discussions and all proofs are given in Supplementary Materials.

2 Asymptotic confidence coefficients of several intervals

We discuss asymptotic ICPs of the following intervals for proportions: three Wald type intervals and bootstrap percentile intervals; the Wilson interval (1927) and the A–C interval (1998); and the Jeffreys prior interval and the likelihood-based interval. These intervals all violate (6), the first six are shown by mathematical proofs, and the last two by exact numerical calculation.

2.1 Several confidence intervals with a zero ICP

As mentioned before, the ICP of the $1-\alpha$ Wald interval in (7) is zero for any $\alpha \in (0, 1)$ and any sample size n (Agresti 2002, p. 32). This is not an isolated phenomenon. Here we give a sufficient condition on when a confidence interval has a zero ICP and then apply this result to intervals for proportions.

Theorem 1 *Let $\underline{X}$ be a random observed vector with a known cumulative distribution function $F(\underline{X}; \underline{\theta})$ with an unknown parameter vector $\underline{\theta} = (\theta, \underline{\eta})$, where $\underline{\theta}$ belongs to a parameter space $\mathcal{H}$. For any confidence interval $C(\underline{X})$ for the parameter of interest θ, if there exists a sample point $\underline{x}_0$ so that*

(i) $C(\underline{x}_0) = \theta_0$ is a single point (i.e., $C(\underline{x}_0) = [\theta_0, \theta_0]$),
(ii) there exists a sequence of $\underline{\theta}_k = (\theta_k, \underline{\eta}_k)$ satisfying $\lim_{k\to+\infty} P_{\underline{\theta}_k}(\underline{X} = \underline{x}_0) = 1$ and $\theta_k \neq \theta_0$, then $C(\underline{X})$ has a zero ICP.

Example 1 It is of great interest to compare two proportions p_1 and p_2 through the difference $\theta = p_1 - p_2$ using: (a) two independent binomials $X \sim Bin(n, p_1)$ and $Y \sim Bin(m, p_2)$; (b) $(N_{11}, N_{12}, N_{21}, N_{22}) \sim Multinomial(n, p_{11}, p_{12}, p_{21}, p_{22})$ from a matched-pair experiment, where $\theta = p_1 - p_2 \overset{def}{=} (p_{11} + p_{12}) - (p_{11} + p_{21})$. For (a), estimate θ using an approximate $1 - \alpha$ interval

$$C_{TI} = \hat{p}_X - \hat{p}_Y \pm z_{\frac{\alpha}{2}} \hat{\sigma}\left(\hat{p}_X - \hat{p}_Y\right), \tag{8}$$

where

$$\hat{p}_X = X/n, \quad \hat{p}_Y = Y/m, \quad \hat{\sigma}(\hat{p}_X - \hat{p}_Y) = \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{m}},$$

see Agresti (2002, p.72). For (b), use an approximate $1 - \alpha$ interval

$$C_{TD} = D \pm z_{\frac{\alpha}{2}} \hat{\sigma}(D), \tag{9}$$

where

$$D = N_{12}/n - N_{21}/n \quad \text{and} \quad \hat{\sigma}(D) = \sqrt{[N_{12}/n + N_{21}/n - (N_{12}/n - N_{21}/n)^2]/n}.$$

See Agresti (2002, p. 410). These two intervals are typically for comparison between a treatment and a control. However, both are of level zero for any sample size(s) following Theorem 1.

For illustration, apply Theorem 1 to interval (9). Let $\underline{\theta} = (\theta, p_{11}, p_{21})$. Pick $\underline{x}_0 = (N_{11}, N_{12}, N_{21}, N_{22}) = (0, n, 0, 0)$, $\theta_0 = p_{12} - p_{21} = 1$, and $\underline{\theta}_k = (1-2/k, 0, 1/k)$. Conditions (i) and (ii) for Theorem 1 hold, and (9) has a zero ICP for any n. □

Example 2 The bootstrap percentile confidence interval (Efron 1979) is widely applied due to its accessibility. However, following Theorem 1, any bootstrap percentile confidence interval, whose limits are equal to the $\frac{\alpha}{2}$th and $(1-\frac{\alpha}{2})$th percentiles of the estimates from bootstrap samples, respectively, for comparing two proportions through difference, relative risk and odds ratio has a zero ICP for any sample size.

For illustration purpose, also consider case (b) in Example 1 with a parameter of interest: the relative risk

$$\theta_{rd} = \frac{p_1}{p_2} = \frac{p_{11} + p_{12}}{p_{11} + p_{21}}.$$

In the ith trial of a matched-pair experiment for $i = 1, \ldots, n$, we observe four possible outcomes (s_1, s_2), (s_1, f_2), (f_1, s_2), (f_1, f_2), where s and f stand for success and failure, respectively, and 1 is the treatment and 2 is the control. Let $(I_{11i}, I_{12i}, I_{21i}, I_{22i})$ be the indicator vector of the four outcomes. Then

$$N_{jk} = \sum_{i=1}^{n} I_{jki}, \quad j = 1, 2, \quad k = 1, 2,$$

and the observed sample is

$$OS = \{(I_{11i}, I_{12i}, I_{21i}, I_{22i})\}_{i=1}^{n}.$$

Let $\hat{\theta}_{rd}(N_{11}, N_{12}, N_{21}, N_{22})$ be any statistic that estimates θ_{rd} and let $C_B(N_{11}, N_{12}, N_{21}, N_{22}, \hat{\theta}_{rd}, m)$ be the $1-\alpha$ bootstrap percentile interval for θ_{rd} based on observation $(N_{11}, N_{12}, N_{21}, N_{22})$, statistic $\hat{\theta}_{rd}$ and m bootstrap samples.

To apply Theorem 1 to C_B, pick $\underline{x}_0$=$(N_{11}, N_{12}, N_{21}, N_{22}) = (n, 0, 0, 0)$. With this observation, any bootstrap sample of size n is a sequence of $(1, 0, 0, 0)$'s and has no variation. So the bootstrap estimates $\hat{\theta}_{rd}$ computed from m bootstrap samples are a sequence of m constants $\theta_0 = \hat{\theta}_{rd}(n, 0, 0, 0)$. Any percentile of these bootstrap estimates is equal to θ_0. Therefore, $C_B(n, 0, 0, 0, \hat{\theta}_{rd}, m) = [\theta_0, \theta_0]$ satisfies (i) in Theorem 1. To meet (ii), let $\underline{\theta} = (\theta_{rd}, p_{11}, p_{21})$. Pick $\underline{\theta}_k = (\theta_{rdk}, p_{11k}, p_{21k})$ so that θ_{rdk} is not equal to θ_0 and (p_{11k}, p_{21k}) approaches to $(1, 0)$ as k goes to infinity.

Therefore,

$$\lim_{k\to+\infty} P_{\underline{\theta}_k}((N_{11}, N_{12}, N_{21}, N_{22}) = \underline{x}_0) = 1.$$

Hence C_B has a zero ICP for any n. Similar results hold for bootstrap intervals for the difference and odds ratio of proportions in cases (a) and (b). □

2.2 The Wilson interval and the Agresti–Coull interval

Back to the case of a single proportion p, the $1-\alpha$ Wilson interval (5) is believed to have a good confidence coefficient (close to $1-\alpha$). In Table 1, we report the ICP for the Wilson interval at different nominal levels with a sample size up to 5,000, and the corresponding ICP is much smaller than the desired nominal level. The ICP of an interval $C(X) = [L(X), U(X)]$ satisfying

$$U(X) = 1 - L(n - X) \tag{10}$$

is computed, following Wang (2007), at a finite set $\{p : p = L(x)^- \text{ for any } L(x) > 0\}$, where a^- denote the limit of y when y approaches to a from the side of $y < a$. So no need to find the infimum on $p \in [0, 1]$. Huwang (1995) proved

$$\lim_{n\to\infty} \inf_{\{p\in[0,1]\}} Cover_{C_{WI}}(p) \le \begin{cases} e^{-\left(1+\frac{z^2_{\frac{\alpha}{2}}}{2} - z_{\alpha/2}\sqrt{1+\frac{z^2_{\frac{\alpha}{2}}}{4}}\right)}, & \text{if } z_{\frac{\alpha}{2}} \ge \frac{1}{\sqrt{2}}; \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The right hand side of (11) is obtained by taking the limit of $Cover_{C_{WI}}(p_n)$ at a sequence of p_n, where each p_n is just below $L_{WI}(1)$, the lower limit of interval C_{WI} at $X = 1$ when the sample size is equal to n. In particular, when $1-\alpha = 0.99$, 0.95, and 0.9, the right hand side of (11) is equal to 0.8892, 0.8382 and 0.8000, respectively. So C_{WI} is not an approximate interval (6). Are these upper bounds sharp? Yes, because these bounds are very close to the ICPs presented in Table 1 for $n = 5{,}000$.

Another widely used $1-\alpha$ approximate interval for p is given by Agresti and Coull (1998) with a closed form

$$C_{AC}(X) = [L_{AC}(X), U_{AC}(X)] = \tilde{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{12}$$

where $\tilde{n} = n + z^2_{\frac{\alpha}{2}}$ and $\tilde{p} = (X + z^2_{\frac{\alpha}{2}}/2)/\tilde{n}$. This interval has a simple form, does not reduce to a point estimate at any value of X, and has a coverage probability closer to $1-\alpha$ than the Wald and Wilson intervals. In Table 2, we find the ICP approaching to the nominal level in the cases of 99 and 95 % when n goes large. However, as the nominal level decreases, the A–C interval is falling apart evidenced in Table 2. In particular, when the nominal level is chosen <48.02 %, Lemma 1 concludes that the

Table 1 Infimum coverage probability (ICP) for the Wilson interval C_{WI} and its infimum point p at different sample sizes and confidence levels

n	99 % interval		95 % interval		90 % interval	
	ICP	p	ICP	p	ICP	p
50	0.8889	0.0023524	0.8376	0.0035392	0.7991	0.0044745
100	0.8891	0.0011751	0.8379	0.0017674	0.7996	0.0022340
200	0.8892	0.0005872	0.8380	0.0008831	0.7998	0.0011162
300	0.8892	0.0003914	0.8381	0.0005886	0.7999	0.0007439
500	0.8892	0.0002348	0.8381	0.0003531	0.8000	0.0004463
1,000	0.8892	0.0001174	0.8382	0.0001765	0.8000	0.0002231
5,000	0.8892	2.348×10^{-5}	0.8382	3.530×10^{-5}	0.8000	4.4619×10^{-5}

Table 2 Infimum coverage probability (ICP) for Agresti–Coull interval C_{AC} and its infimum point p at different sample sizes and confidence levels

n	99 % interval		95 % interval		90 % interval	
	ICP	p	ICP	p	ICP	p
50	0.9872	0.3827965	0.9345	0.3115578	0.8685	0.3305323
100	0.9880	0.3016692	0.9380	0.2454214	0.8753	0.0754247
200	0.9880	0.4201003	0.9408	0.1501686	0.8790	0.0451723
300	0.9884	0.4134312	0.9423	0.0484458	0.8833	0.0435604
500	0.9887	0.3240625	0.9431	0.0420841	0.8808	0.0260270
1,000	0.9889	0.0665205	0.9432	0.0160102	0.8827	0.0105319
5,000	0.9891	0.0125020	0.9440	0.0045054	0.8816	0.0021014
	85 % interval		80% interval		70 % interval	
	ICP	p	ICP	p	ICP	p
50	0.8113	0.0822589	0.7270	0.0414742	0.5871	0.0324345
100	0.8053	0.0403563	0.7243	0.0204128	0.5846	0.0160502
200	0.8102	0.0126322	0.7229	0.0101223	0.5834	0.0079825
300	0.8097	0.0083917	0.7224	0.0067292	0.5830	0.0053121
500	0.8096	0.0050206	0.7220	0.0040283	0.5826	0.0031826
1,000	0.8094	0.0025048	0.7217	0.0020107	0.5824	0.0015895
5,000	0.8092	0.0005001	0.7215	0.0005015	0.5822	0.0003176

ICP is zero when n is large. This is because the upper limit of $C_{AC}(0)$ is less than the lower limit of $C_{AC}(1)$. Fortunately, the A–C interval works fine at 95 and 99 %, two commonly used levels in practice. Next, we provide a calculation of its asymptotic ICP for a general $1-\alpha$.

Lemma 1 *The asymptotic ICP of the* $1-\alpha$ *A–C interval is zero if* $z_{\frac{\alpha}{2}} < \sqrt{\sqrt{2}-1}$ (*equivalent to* $1-\alpha < 0.4802$).

Theorem 2 *For any positive integer* $x > a \overset{def}{=} \frac{z_{\frac{\alpha}{2}}^2}{2}$, *let*

$$y_x = \left(\sqrt{x+a} - z_{\frac{\alpha}{2}}\right)^2 - a,$$

and let

$$y_0 = max\left\{0, [y_x] + 1\right\},$$

where $[u]$ *denotes the largest integer no larger than u. Also define*

$$c(x) = e^{-\left(x+a-z_{\frac{\alpha}{2}}\sqrt{x+a}\right)} \sum_{i=y_0}^{x-1} \frac{\left(x+a-z_{\frac{\alpha}{2}}\sqrt{x+a}\right)^i}{i!}. \tag{13}$$

Then the asymptotic ICP of the $1-\alpha$ *A–C interval satisfies*

$$\lim_{n\to\infty} \inf_{p\in[0,1]} Cover_{C_{AC}}(p) \le \inf_{x>a}\{c(x)\}. \tag{14}$$

In consequence we obtain the following.

Corollary 1 *The asymptotic ICP of the* $1-\alpha$ *A–C interval is less than or equal to*

$$\begin{cases} 0.9891, & if\ 1-\alpha = 0.99; \\ 0.9436, & if\ 1-\alpha = 0.95; \\ 0.8813, & if\ 1-\alpha = 0.90; \\ 0.8189, & if\ 1-\alpha = 0.85; \\ 0.7215, & if\ 1-\alpha = 0.80; \\ 0.5821, & if\ 1-\alpha = 0.70; \\ 0, & if\ 1-\alpha < 0.4802. \end{cases} \tag{15}$$

The bounds given in (15) are quite sharp since these are very close to the true ICPs for $n = 5,000$ shown in Table 2. The only suspicious case is $1-\alpha = 0.85$, which shows a difference of about 1 % (0.8189 vs. 0.8092). However, we calculate cases of larger n, like $n = 6,000$ and 7,000, and obtain the ICPs, 0.8191 and 0.8190, respectively. This confirms the sharp bounds in (15).

Corollary 1 delivers two aspects of information: the A–C interval is quite safe to apply at two commonly used levels: 0.99 and 0.95 due to small differences between the two levels and their corresponding asymptotic confidence coefficients; on the other hand, the $1-\alpha$ A–C interval, just like the Wald and Wilson intervals, is not an approximate interval (6) either. Practitioners may question the value of a large sample since a large sample still does not guarantee a correct confidence coefficient. Also the asymptotic confidence coefficient is getting worse (smaller than the nominal level) as the nominal level decreases. From the proof of Corollary 1, the asymptotic ICP of the A–C interval is not achieved right below $C_{AC}(1)$, which occurs for the Wilson interval, and the locations of the asymptotic ICP depend on the level $1-\alpha$.

2.3 The Jeffreys prior interval and likelihood-based interval

The $1-\alpha$ Jeffreys prior interval $C_J(X) = [L_J(X), U_J(X)]$ is the Bayesian interval with equal tails, and is recommended for practice by some researchers. Let $Beta(r, s)$ denote a $Beta$-distribution with parameters r and s. Brown et al. (2001) recommended using $Beta(0, 5, 0.5)$ as a prior distribution on p, while Pires and Amado (2008) used $Beta(1, 1)$ (the uniform prior). In general, for a prior $Beta(r, s)$ on p, $L_J(0) = 0$, $U_J(n) = 1$, and

$$L_J(x) = q_{beta}\left(\frac{\alpha}{2}, x+r, n-x+s\right),$$
$$U_J(x) = q_{beta}\left(1-\frac{\alpha}{2}, x+r, n-x+s\right), \tag{16}$$

where $q_{beta}(c; r, s)$ is the cth percentile of $Beta(r, s)$. Although a mathematical proof is not available, our numerical study shows that C_J, similar to C_{WI}, also violates (6) as shown in Table 3. For example, if $Beta(0.5, 0.5)$ and n=5,000, the 95 % Jeffreys interval (16) has a ICP 0.8977, which is achieved at $L_J(1)^-$.

Table 3 Infimum coverage probability (ICP) for the Jeffreys prior interval C_J and its infimum point p at different sample sizes and confidence levels using two prior distributions

n	99 % interval		95 % interval		90% interval	
	ICP	p	ICP	p	ICP	p
$Beta(1, 1)$ *prior*						
50	0.9026	0.0020474	0.7868	0.0047849	0.7034	0.0070125
100	0.9021	0.0010292	0.7858	0.0024071	0.7022	0.0035297
300	0.9018	0.0003443	0.7852	0.0008056	0.7013	0.0011818
500	0.9018	0.0002067	0.7851	0.0004838	0.7012	0.0007097
1,000	0.9017	0.0001034	0.7850	0.0002420	0.7010	0.0003551
5,000	0.9017	2.069×10^{-5}	0.7849	4.843×10^{-5}	0.7009	7.106×10^{-5}
	99 % interval		95 % interval		90% interval	
	ICP	p	ICP	p	ICP	p
$Beta(0.5, 0.5)$ *prior*						
50	0.9646	0.0007205	0.8838	0.0487585	0.8132	0.0375012
100	0.9647	0.0003594	0.8801	0.0247452	0.8105	0.0189769
300	0.9647	0.0001196	0.8774	0.0083313	0.8087	0.0063767
500	0.9648	7.175×10^{-5}	0.8769	0.0050088	0.8083	0.0038322
1,000	0.9648	3.586×10^{-5}	0.8977	0.0001079	0.8080	0.0019185
5,000	0.9648	7.171×10^{-6}	0.8977	2.158×10^{-5}	0.8078	0.0003841

We also study a likelihood-based interval $C_L(X) = [L_L(X), U_L(X)]$ that "has been suggested as theoretically most appealing" (New combe 1998). This interval is obtained by inverting a family of level-α likelihood ratio tests for $H_0 : p = p_0$ for any $p_0 \in [0, 1]$. Thus, it meets (4). The lower limit $L_L(x)$ is the smallest solution of

$$f(p){=}xln(p) + (n - x)ln(1 - p) - xln(x/n) - (n - x)ln(1 - x/n) - z^2_{\frac{\alpha}{2}}/2{=}0;$$

for $x > 0$ and $L_L(0) \stackrel{def}{=} 0$, and the upper limit satisfies (10).

The asymptotic ICPs are reported in Table 4. Again, they are much less than the corresponding nominal levels. So the interval does not satisfy (6) based on the numerical study.

3 An alternative use of the Clopper–Pearson interval

The $1-\alpha$ C–P interval (1934), denoted by $C_{CP}(X)$, is the first interval of level $1-\alpha$ for the proportion p. The ICP of the $1 - \alpha$ C–P interval may be higher than $1 - \alpha$ especially when n and/or $1 - \alpha$ is small see Wang (2006). When n is large, although the ICP is almost equal to (just a little larger than) the nominal level, the criticism is that "For any fixed parameter value (p), the actual coverage probability can be much larger than the nominal confidence level" (Agresti and Coull 1998). So practitioners choose approximate intervals (including those in Sect. 2) which, however, may not serve the purpose of confidence interval well since they violate (6). To overcome these drawbacks, we propose:

(A) A desired confidence interval $C(X)$ for p has a predetermined average coverage probability (ACP) $1 - \gamma$. i.e.,

$$ACP = \int_0^1 Cover_C(p)dp = 1 - \gamma, \tag{17}$$

Table 4 Infimum coverage probability (ICP) for the likelihood-based interval C_L and its infimum point p at different sample sizes and confidence levels

n	99 % interval		95 % interval		90% interval	
	ICP	p	ICP	p	ICP	p
50	0.9596	0.9358042	0.8425	0.9623139	0.6972	0.9733073
100	0.9582	0.9673697	0.8411	0.9809760	0.6951	0.9865634
300	0.9572	0.9890028	0.8401	0.9936181	0.6937	0.9955009
500	0.9570	0.9933870	0.8399	0.9961659	0.6934	0.9972981
1,000	0.9568	0.9966881	0.8398	0.9980811	0.6932	0.9986481
5,000	0.9615	0.9993368	0.8396	0.9996160	0.6930	0.9997295

(B) Between two intervals $C(X)$ and $C^*(X)$ satisfying A), $C(X)$ is preferred if it has a larger ICP than $C^*(X)$.
(C) The ICP of $C(X)$ converges to $1-\gamma$ when n goes to infinity.

Condition (B) rules out approximate intervals, including the Wald interval, with a small ICP, and (C) assures a correct overall coverage for a large sample.

ACP, as a function of $1-\alpha$, is continuous and increasing. We use the Newton method in R to solve the unknown level $1-\alpha$ for a given ACP $1-\gamma$ for the C–P and A–C intervals. ACP can be computed more efficiently for any confidence interval $C(X) = [L(X), U(X)]$ using the identity

$$ACP = \sum_{x=0}^{n} \frac{H(U(x); x+1, n-x+1) - H(L(x); x+1, n-x+1)}{n+1}, \tag{18}$$

where $H(p; r, s)$ is the cumulative distribution function for a Beta-distribution with parameters r and s. Also Wang (2007) shows that the ICP is achieved on a finite many values of p, which simplifies the computation of the ICP.

Table 5 contains a small study on ICPs for the C–P and A–C intervals at different $1-\gamma$'s. This is done by exact numerical computation and a R-code is available from the authors. It is evident that for large sample sizes ($n \geq 100$) the C–P interval has a larger ICP than the A–C interval especially when the nominal level gets lower, More importantly, when n goes large, the difference between the ICP and the ACP is vanishing for the C–P interval at all levels of $1-\gamma$ including those in Table 5, and the same does not occur for the A–C interval. This suggests that the C–P interval with an ACP $1-\gamma$ is an approximate interval (6) of level $1-\gamma$. Also the C–P interval always

Table 5 Infimum coverage probability (ICP) and average expected length (AEL) for the C–P and A–C intervals at different sample sizes n and average coverage probabilities (ACP = $1-\gamma$)

$1-\gamma$	99 %		90 %		80 %		70 %	
	C–P	A–C	C–P	A–C	C–P	A–C	C–P	A–C
n	ICP	ICP	ICP	ICP	ICP	ICP	ICP	ICP
	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$
	AEL	AEL	AEL	AEL	AEL	AEL	AEL	AEL
100	0.9853	0.9851	0.8668	0.8649	0.7455	0.7227	0.6278	0.5837
	0.9850	0.9882	0.8665	0.8921	0.7440	0.7919	0.6264	0.6938
	0.1974	0.1992	0.1262	0.1263	0.0981	0.0986	0.0790	0.0800
500	0.9879	0.9879	0.8856	0.8803	0.7755	0.7216	0.6675	0.5824
	0.9879	0.9893	0.8855	0.8975	0.7755	0.7976	0.6675	0.6983
	0.0898	0.0900	0.0573	0.0574	0.0446	0.0448	0.0360	0.0362
5,000	0.9894	0.9890	0.8955	0.8815	0.7923	0.7214	0.6898	0.5821
	0.9894	0.9899	0.8955	0.8996	0.7923	0.7996	0.6898	0.6996
	0.0286	0.0286	0.0182	0.0182	0.0142	0.0142	0.0115	0.0115

has a shorter average expected length than the A–C interval. A more detailed study is available in Wang and Zhang (2013).

4 Discussion

In this paper, we compare two definitions for approximate intervals: (4) and (6). The former is widely used in the current statistical practice, but does not serve the purpose of confidence interval: capture the parameter no matter where it is in the parameter space with a certain probability. The Wald interval is a typical example for such failure. The latter is newly proposed, overcomes this problem and should be adopted in practice, however, it is not easy to establish due to a uniform convergence. We study seven commonly used intervals and bootstrap intervals for proportions and show that they do not meet (6). Therefore, practitioners should realize that a large sample does not guarantee a correct asymptotic confidence coefficient even with an asymptotic normality established. On the other hand, a new use of the C–P interval is proposed and the resultant satisfies (6) based on a numerical study. It is of great interest for a future research to provide a mathematical proof and seek for other intervals that satisfy (6).

Acknowledgments The first author was partially supported by NSF (US) Grant No. DMS-0906858. The second author was partially supported by NSFC Grant No. 10971007. The authors are grateful to an Associate Editor for the helpful comments.

References

Agresti A (2002) Categorical data analysis, 2nd edn. Wiley, Hoboken

Agresti A, Coull BA (1998) Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 52:119–126

Blyth CR, Still HA (1983) Binomial confidence intervals. J Am Stat Assoc 78:108–116

Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. Stat Sci 16:101–133

Clopper CJ, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404–413

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26

Huwang L (1995) A note on the accuracy of an approximate interval for the binomial parameter. Stat Probab Lett 24:177–180

Newcombe RG (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 17:857–872

Newcombe RG (2011) Measures of location for confidence intervals for proportions. Commun Stat Theory Methods 40:1743–1767

Pires AM, Amado C (2008) Interval estimators for a binomial proportion: comparison of twenty methods. REVSTAT Stat J 6:165–197

Wang W (2006) Smallest confidence intervals for one binomial proportion. J Stat Plan Inference 136: 4293–4306

Wang H (2007) Exact confidence coefficients of confidence intervals for a binomial proportion. Stat Sin 17:361–368

Wang W, Zhang Z (2013) Asymptotic infimum coverage probability for interval estimation of proportions (manuscript). Wright State Unviersity Dayton

Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22:209–212