Comparison of interval estimation methods for a binomial proportion

Tuany de Paula Castro[1], Carlos Daniel Paulino[2], and Julio M. Singer[3]

[1]Departamento de Estatística IME, Universidade de São Paulo, São Paulo, SP, Brazil
[2]Departamento de Matemática, Instituto Superior Técnico, Lisboa, Portugal
[3]Departamento de Estatística IME, Universidade de São Paulo, São Paulo, SP, Brazil

ABSTRACT

We compare different confidence and credible intervals for the probability of success in a binomial model with respect to the coverage probability and expected length. The comparison is motivated by the similarity of a confidence interval proposed by Agresti and Coull (The American Statistician, 1998) and a Bayesian credible interval based on a Beta(2,2) prior distribution. Keeping in mind that confidence intervals are random and that credible intervals are numeric, we perform the comparison under the same paradigm, considering the Bayesian intervals (central and HPD) as realizations of random intervals or the latter as numeric intervals. The intervals are compared via simulation studies that show a better performance of the Wilson (score) and HPD intervals with uniform prior distribution and some advantages of Bayesian intervals with respect to the expected and posterior length.

Keywords: Binomial distribution, coverage probability, expected length, Highest posterior density (HPD) intervals.

Introduction

The construction of confidence intervals for a binomial proportion is one of the most basic and important problems in statistical inference. Although this problem has been dealt with for many years, it is still addressed by many recent papers, among which we mention as Agresti & Coull (1998), Agresti & Caffo (2000), Brown et al. (2001) , Thulin (2014) and Jin et al. (2017).

Let X denote a random variable following a binomial distribution with parameter θ and n trials. The most common approach for interval estimation of θ is obtained by inverting the Wald test (Wald, 1943) and is based on the asymptotic normality of the sample proportion of successes and on its standard error. This $100(1-\alpha)\%$ confidence interval is

$$\widehat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{\theta}(1-\widehat{\theta})/n},$$

where $z_{\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution and $\widehat{\theta} = X/n$ is the sample proportion of successes. This approach is available in many statistical software packages. However, many authors show that it behaves poorly in terms of the coverage probability, *i.e.*, the probability that the interval contains a fixed value of θ, even when the sample size is very large [Brown, Cai and DasGupta (2001)].

As an alternative to approximate intervals like the one based on Wald's test, Clopper & Pearson (1934) proposed an "exact" confidence interval for θ obtained by the inversion of equal-tailed binomial tests of $H_0 : \theta = \theta_0$. It has endpoints, θ_L and θ_U, that are the solutions to the equations

$$\sum_{k=x}^{n} \binom{n}{k} \theta_L^k (1-\theta_L)^{n-k} = \alpha/2 \quad \text{and} \quad \sum_{k=0}^{x} \binom{n}{k} \theta_U^k (1-\theta_U)^{n-k} = \alpha/2,$$

with the lower bound set to 0 when $x = 0$ and the upper bound set to 1 when $x = n$. The Clopper-Pearson interval can also be obtained as

$$\theta_L = B(x, n-x+1, \alpha/2) \quad \text{and} \quad \theta_U = B(x+1, n-x, 1-\alpha/2),$$

where $B(a,b,c)$ is the quantile of order $0 < c < 1$ of the Beta(a,b) distribution. The Clopper-Pearson interval guarantees a coverage probability of at least $1-\alpha$ for every possible value of θ. However, it is necessarily conservative, especially for θ near 0 and 1, with the actual coverage probability much larger than the nominal confidence level. This is a consequence of the discreteness of the binomial distribution.

A second approximate but less commonly used confidence interval, proposed in Wilson (1927), is based on the inversion of the score test. The interval corresponds to the values of θ_0 that lead to non-rejection of $H_0 : \theta = \theta_0$ with the score test statistic

$$Q_R(\theta_0) = \frac{(\widehat{\theta} - \theta_0)^2}{\theta_0(1-\theta_0)/n}.$$

Explicitly, the score confidence interval has the form

$$\left(1 + \frac{z^2_{\alpha/2}}{n}\right)^{-1} \left[\widehat{\theta} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2} \sqrt{\frac{1}{n}\left(\widehat{\theta}(1-\widehat{\theta}) + \frac{z^2_{\alpha/2}}{4n}\right)}\right].$$

As discussed in Agresti & Coull (1998), the midpoint of this interval is a weighted average of the sample proportion of successes $\widehat{\theta}$ and $1/2$, with the weight given to $\widehat{\theta}$ approaching 1 asymptotically. The variance used in this interval has the form of a weighted average of the variance of a sample proportion when $\theta = \widehat{\theta}$ and the variance of a sample proportion when $\theta = 1/2$, using $n + z^2_{\alpha/2}$ in lieu of the usual sample size n. As shown in Agresti (2011), the coverage probability of the score confidence interval is closer to the nominal confidence level than the coverage probabilities corresponding to the Wald or to the Clopper-Pearson intervals.

Agresti & Coull (1998) used the representation of the score interval to obtain an adjusted Wald interval that has a similar performance to that of the score interval, even for small sample sizes. For the 95% case, taking $z^2_{0.025} = 1.96^2 \approx 4$ their approach leads to the adjusted "add two successes and two failures" interval, namely,

$$\left[\widetilde{\theta} \pm z_{0.025} \sqrt{\frac{\widetilde{\theta}(1-\widetilde{\theta})}{n+4}}\right]$$

with midpoint $\widetilde{\theta} = (X+2)/(n+4)$ identical to the midpoint of the 95% score interval. This corresponds to an estimate of the probability θ obtained after adding 4 pseudo observations to the sample, two being successes and two, failures. The corresponding variance is $\widetilde{\theta}(1-\widetilde{\theta})/(n+4)$ instead of the weighted average of the variances considered in the score interval. This generates wider intervals than the score interval. Adjusted Wald intervals can be constructed with confidence levels other than 0.95, as shown in Agresti & Caffo (2000).

The midpoint of the Agresti-Coull interval is identical to the Bayes estimate (mean of the posterior distribution) based on a Beta prior distribution for θ with both parameters equal to 2. For large enough n, the corresponding posterior distribution can be approximated by a normal distribution with mean $\widetilde{\theta}$ and variance $\widetilde{\theta}(1-\widetilde{\theta})/(n+5)$ and an approximate equal-tailed Bayesian credible interval for θ with probability 95% has the form

$$\left[\widetilde{\theta} \pm 1.96\sqrt{\frac{\widetilde{\theta}(1-\widetilde{\theta})}{n+5}}\right]. \tag{1}$$

This interval is similar to the Agresti-Coull interval, except for a small difference in the denominator of the variance. For moderate sample sizes, the two intervals are practically equal.

Although the Agresti-Coull interval is obtained from a frequentist perspective, it also has a Bayesian flavor. This leads to questions about the performance of exact Bayesian credible intervals with respect to that of frequentist ones. Using asymptotic expansions Jin et al. (2017) studied the similarities between frequentist confidence intervals for proportions and equal-tailed Bayesian credible intervals based on low-informative priors. The purpose of this article is to compare the performance of different frequentist intervals to that of exact Bayesian credible intervals (equal-tailed and HPD) under the same paradigm, considering the Bayesian intervals as realizations of random intervals or the frequentist intervals as numeric intervals. In Section 2, we describe the Bayes credible intervals to be investigated. In the third section, the frequentist and the Bayesian intervals will be compared with respect to the average coverage probability, expected length and posterior length under the same paradigm. We conclude with a brief discussion in the last section.

Bayesian credible intervals

Bayesian credible intervals for the binomial proportion are often discussed in literature, as in Agresti & Min (2005) and Brown et al. (2001). It is quite common to use the beta distribution as a prior distribution for θ, since the family of beta densities is the standard conjugate prior for binomial distributions.

An essential part of the Bayesian approach is the possibility of using different prior distributions for θ according to the existing knowledge about the subject matter under investigation. The beta distribution provides a flexible family of prior distributions as an appropriate choice of the prior parameters can generate various shapes with various degrees of dispersion and skewness. Suppose that θ has a $Beta(a,b)$ prior distribution; then the posterior distribution of θ is $Beta(x+a,n-x+b)$. For large enough n, this posterior distribution can be approximated by a normal distribution with mean $\breve{\theta}$ and variance $\breve{\theta}(1-\breve{\theta})/(n+a+b+1)$, where $\breve{\theta} = (x+a)/(n+a+b)$ is the mean of the exact posterior distribution. Based on this approximate posterior distribution, the approximate equal-tailed credible interval of probability $1-\alpha$ has the form

$$\left[\breve{\theta} \pm z_{\alpha/2}\sqrt{\frac{\breve{\theta}(1-\breve{\theta})}{(n+a+b+1)}}\right].$$

Bayesian credible intervals can also be obtained directly from the exact posterior distribution. The corresponding $100(1-\alpha)\%$ equal-tailed exact credible interval is given by the quantiles

$$[B(x+a, n-x+b, \alpha/2), B(x+a, n-x+b, 1-\alpha/2)].$$

The actual endpoints of this interval can be easily computed via standard statistical software.

Alternatively highest posterior density (or HPD) credible interval may be considered. The minimum density of any point within this interval is equal to or larger than the density of any point outside this interval. This interval is the smallest among all the $100(1-\alpha)\%$ Bayesian credible intervals. HPD intervals must be numerically computed but are also available in standard statistical software.

In the next section, Wald, Clopper-Pearson, score and Agresti-Coull intervals will be compared to equal-tailed and HPD intervals. The prior distributions considered for θ are the beta distributions with different parameters. In particular, the $Beta(2,2)$ distribution will be analyzed since it generates credible intervals with similar form to the Agresti-Coull interval. Considering that Jin et al. (2017) showed that the Wilson interval can be viewed as an approximation of the Bayesian interval using the non-informative uniform $[Beta(1,1)]$ prior and that the the well-known non-informative Jeffreys prior $[Beta(1/2,1/2)]$ is a compromise between the likelihood ratio and Wilson intervals, the uniform and Jeffreys priors will also be studied.

Comparison of the interval estimators

The competing intervals will be compared either in terms of frequentist or Bayesian criteria. Since confidence intervals are random and credible intervals are numeric, we must conduct the comparison under the same paradigm, considering Bayesian intervals as realizations of random intervals and confidence intervals as numeric intervals. The frequentist criteria considered are the coverage probability, the average coverage probability and the expected length. The adopted Bayesian criterion is the posterior length.

Coverage probability

The actual coverage probability for a $100(1-\alpha)\%$ confidence interval $[R(X)]$ for estimating θ computed at a fixed value of the parameter is

$$\mathbb{P}[\theta \in R(X)|\theta] = C_n(\theta) = \sum_{k=0}^{n} \mathbb{I}_{R(k)}(\theta)\binom{n}{k}\theta^k(1-\theta)^{n-k},$$

where $\mathbb{I}_{R(k)}(\theta)$ equals 1 if the interval contains θ when $X = k$ and equals 0 otherwise. The average coverage probability is defined as

$$\overline{C}_n = \int_0^1 C_n(\theta) g(\theta) \mathrm{d}\theta,$$

where the function g can be viewed as a weight function over the possible values that θ can take.

Figures 1 and 2 show the coverage probabilities for the 95% frequentist intervals under investigation, for sample sizes $n = 5$ and $n = 10$. The 95% Wald interval behaves poorly, with coverage probability too small converging to zero as θ approaches 0 or 1. On the other hand, the Clopper-Pearson interval is very conservative. For the Wilson interval, the coverage probability is close to the nominal confidence level even for $n = 5$. The Agresti-Coull interval has good minimum coverage probability. However in comparison to the Wilson interval, it is more conservative, especially for θ close to 0 or 1. This is expected since the Agresti-Coull interval contains the Wilson interval.

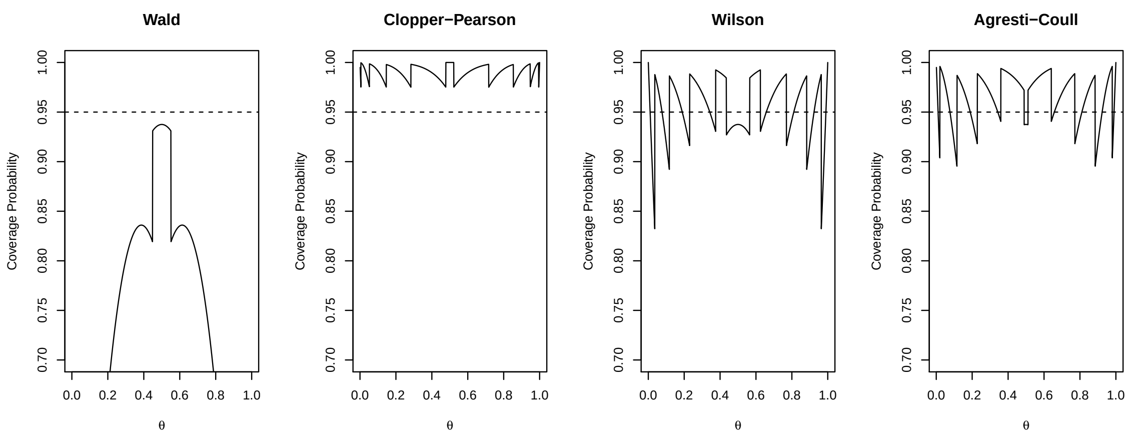

Figure 1. Coverage probabilities for Wald, Clopper-Pearson, Wilson and Agresti-Coull confidence intervals for a binomial proportion when $n = 5$.

To compare Bayesian credible intervals to frequentist confidence intervals under the same paradigm, we must regard them as realizations of random intervals. Therefore the ”confidence level” for a Bayesian credible interval is the minimum coverage probability obtained by considering intervals constructed for all possible values of the number of successes. For such purposes, we may use the following algorithm:

- **Step 1:** Fix a value $0 < \theta_0 < 1$;
- **Step 2:** Construct the credible interval (central or HPD), $R(k)$, for θ with probability of 95% for the $n+1$ possible values of X, considering some prior distribution Beta(a,b);
- **Step 3:** Compute the coverage probability of the interval

  $$C_n(\theta_0) = \sum_{k=0}^{n} \mathbb{I}_{R(k)}(\theta_0) \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k}$$

  where $\mathbb{I}_{R(k)}(\theta)$ is equal to 1 if the interval $R(k)$ contains the value of θ_0;

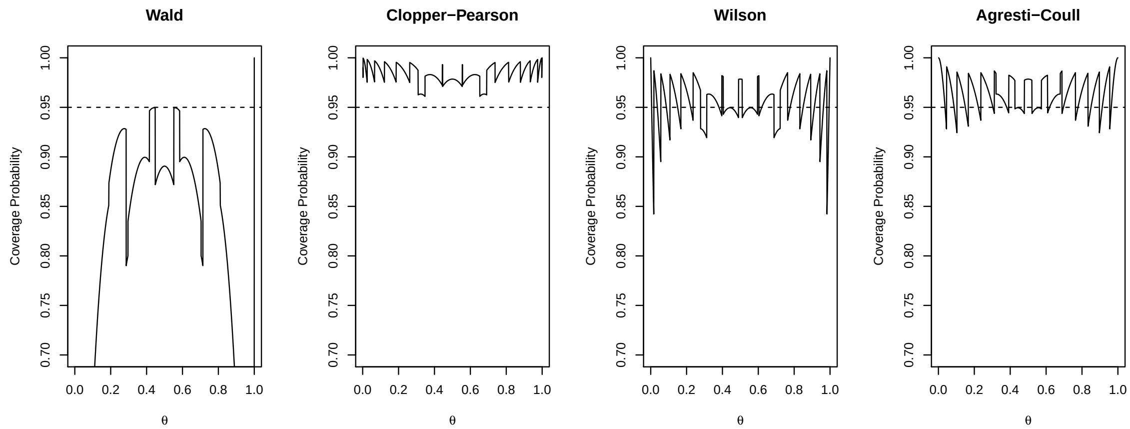

Figure 2. Coverage probabilities for Wald, Clopper-Pearson, Wilson and Agresti-Coull confidence intervals for a binomial proportion when $n = 10$.

- **Step 4:** Repeat the previous two steps for a large number N of fixed values of θ;
- **Step 5:** Finally, compute the coverage probability as

$$\gamma = \min_{1 \le i \le N} C_n(\theta_i).$$

Table 1 shows the minimum coverage probabilities (that we consider as nominal confidence levels for the credible intervals) and the average coverage probabilities for both the frequentist and Bayesian intervals. The HPD intervals have uniformly better minimum coverage probabilities than the Bayesian equal-tailed intervals (central intervals). The average coverage probabilities are similar for the HPD and equal-tailed intervals. The HPD interval with uniform prior distribution is comparable to the Wilson interval in terms of average coverage probability. In all the cases, the nominal confidence levels of the Bayesian intervals under an uniform prior distributions are larger than the minimum coverage probability of the Wilson interval, which theoretically should not fall below its nominal 95% confidence level. All the Bayesian intervals have a better performance than the Wald interval.

In figures 3 and 4, we display the coverage probability of the Bayesian intervals for $n = 5$ and $n = 10$. Again the HPD interval with uniform prior distribution shows a better performance, exhibiting a nominal confidence level closer to 95% (the desired confidence level of the frequentist intervals).

Expected length

The expected length of a $100(1-\alpha)\%$ confidence interval $[R(X) = (\theta_L(X), \theta_U(X)]$ for a binomial proportion is defined as

$$T_n(\theta) = \sum_{k=0}^{n} [\theta_U(k) - \theta_L(k)] \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

For Bayesian intervals, it is again necessary to regard them as random intervals and compute their expected lengths by obtaining intervals with probability 95% for all $(n+1)$ possible values of X.

Method	$n = 5$		$n = 10$		$n = 30$		$n = 50$		$n = 100$	
	Min	Average	Min	Average	Min	Average	Min	Average	Min	Average
Wald	0.005	0.643	0.010	0.771	0.030	0.877	0.049	0.903	0.095	0.925
Clopper-Pearson	0.975	0.991	0.961	0.984	0.951	0.974	0.953	0.970	0.951	0.965
Wilson	0.832	0.956	0.842	0.955	0.860	0.953	0.860	0.953	0.905	0.952
Agresti-Coull	0.895	0.967	0.924	0.965	0.934	0.961	0.935	0.959	0.939	0.956
HPD Beta(2,2)	0.718	0.936	0.722	0.940	0.739	0.945	0.740	0.947	0.740	0.948
Central Beta(2,2)	0.641	0.927	0.598	0.926	0.562	0.933	0.575	0.937	0.547	0.941
HPD Jeffreys	0.806	0.930	0.830	0.932	0.844	0.939	0.845	0.942	0.841	0.945
Central Jeffreys	0.806	0.939	0.830	0.942	0.818	0.947	0.815	0.948	0.811	0.948
HPD uniform	0.834	0.950	0.889	0.950	0.914	0.950	0.914	0.950	0.921	0.950
Central uniform	0.803	0.950	0.800	0.950	0.810	0.950	0.818	0.950	0.818	0.950

Table 1. Minimum and average coverage probabilities for nominal 95% confidence and credible intervals for the binomial proportion.

In Figures 5 and 6 we display the expected lengths of all the frequentist and Bayesian intervals under investigation for $n = 10$, $n = 30$ and $\alpha = 0.05$.

The Wilson interval is the shortest for θ between 0.21 and 0.79 and the Wald interval is the shortest for θ outside that region. However, the Wald interval is the shortest near the boundaries because it is degenerate at 0 or at 1 when or $x = 0$ or $x = n$. The expected lengths of the Wald and Agresti-Coull intervals have similar expected lengths as n increases and the difference becomes irrelevant when n is larger than 30. We may therefore conclude that the Wilson interval is the best among the frequentist intervals in terms of expected length.

Under a Bayesian approach, HPD intervals with $Beta(2,2)$ prior distributions are the shortest when θ lies between 0.26 and 0.74. For θ outside this region, HPD intervals with Jeffreys prior are the shortest. Again, the differences become irrelevant as n increases.

We now compare the best frequentist intervals (Wilson and Agresti-Coull) with the Bayesian alternatives. Figure 7 shows their expected lengths for $n = 10$ and $\alpha = 0.05$. The Agresti-Coull interval has the worst performance, since it has the biggest expected length over almost all the parametric space. The HPD interval with a $Beta(2,2)$ prior distribution is shorter than the Wilson interval for $0.16 \leq \theta \leq 0.84$. On the other hand, the HPD interval with Jeffreys prior is shorter than Wilson's interval for $\theta \leq 0.19$ and $\theta \geq 0.81$. Finally the HPD interval with an uniform prior distribution outperforms the Wilson interval for $\theta \leq 0.39$ and $\theta \geq 0.61$; they are comparable outside this region.

This analysis indicates that if information about θ is available, the HPD interval with $Beta(2,2)$ or Jeffreys prior distribution can be better choices, since they will be more accurate intervals. But if there is no information about θ, the Wilson interval and the HPD interval with an uniform prior distribution are both good choices, although the Bayesian interval has a

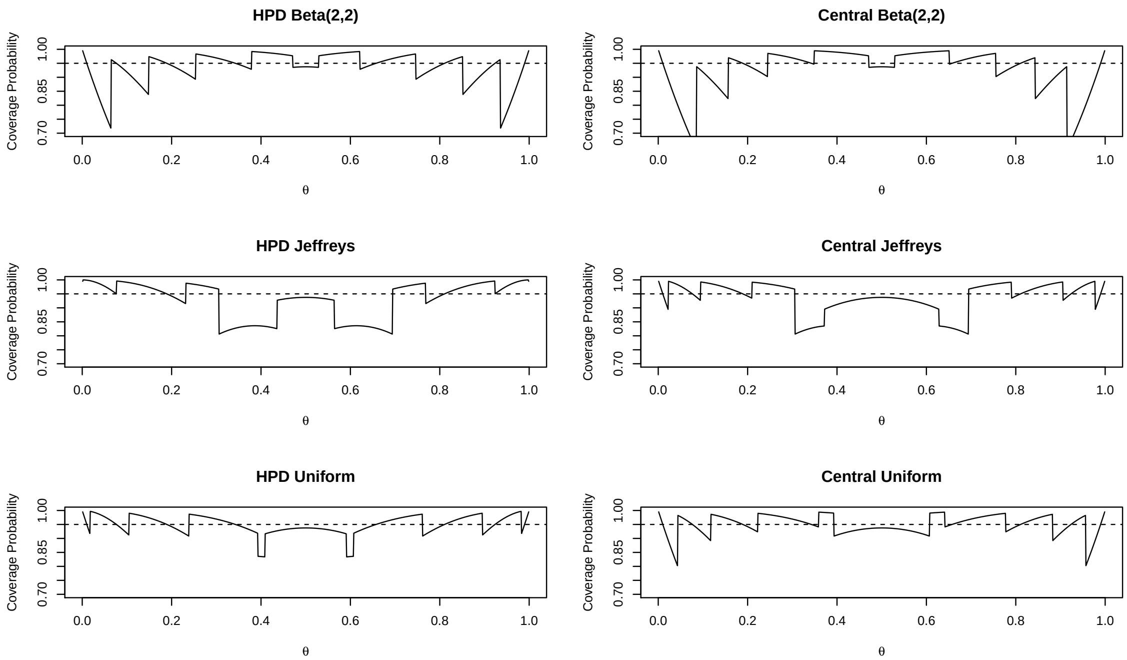

Figure 3. Coverage probabilities for HPD and central intervals with $Beta(2,2)$, Jeffreys and uniform prior distributions ($n = 5$).

small advantage.

Posterior length

A Bayesian measure of evaluation of credible interval is the posterior length, defined as

$$L_n(x) = \theta_U(x) - \theta_L(x),$$

where $\theta_L(x)$ and $\theta_U(x)$ are, respectively the lower and upper limits of a $100(1-\alpha)\%$ interval obtained from the posterior distribution of θ given $X = x$.

To compare confidence intervals to credible intervals with given posterior probability under a Bayesian perspective, it is necessary to regard the frequentist intervals as numeric intervals and to compute their ”posterior probabilities” and ”posterior lengths”. The ”posterior probability” of a confidence interval under a $Beta(a,b)$ prior distribution can be computed as

- **Step 1:** Fix values for n, x and for the parameters of the prior distribution $Beta(a,b)$;
- **Step 2:** Construct the $100(1-\alpha)\%$ confidence interval for θ, namely, $[\theta_L(x), \theta_U(x)]$;
- **Step 3:** Obtain the corresponding ”posterior probability” as $\int_{\theta_L(x)}^{\theta_U(x)} \pi(\theta|x)d\theta$, where $\pi(\theta|x)$ is the posterior distribution of θ obtained with the fixed prior distribution $Beta(a,b)$.

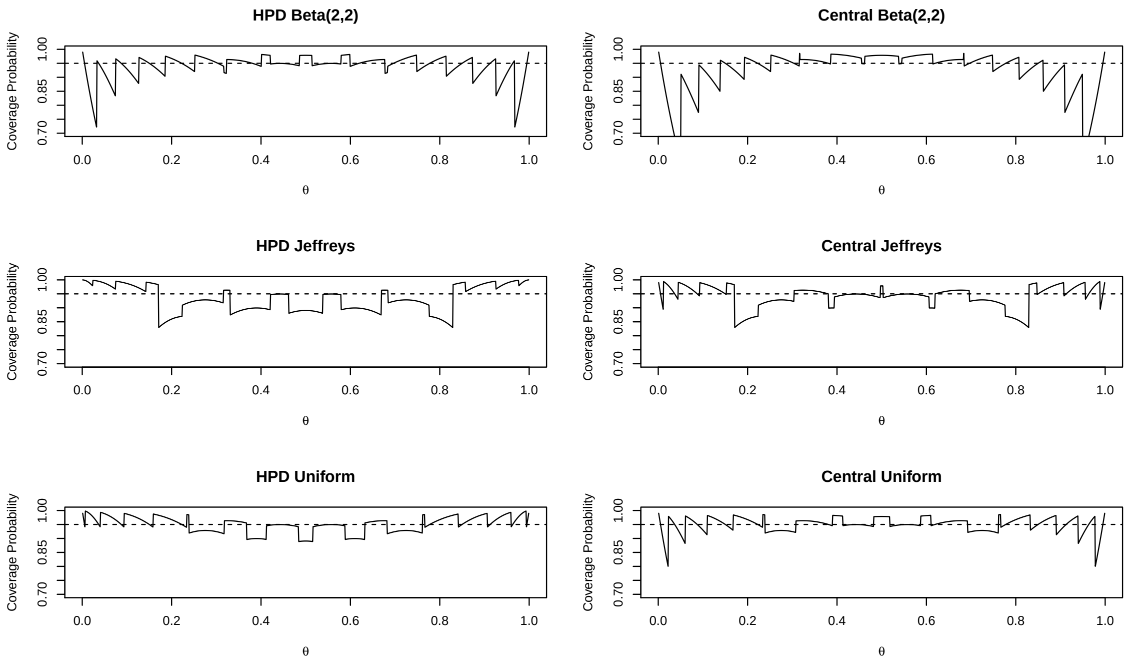

Figure 4. Coverage probabilities for HPD and central intervals with $Beta(2,2)$, Jeffreys and uniform prior distributions ($n = 10$).

Given the discrete nature of the binomial distribution, for fixed values of n and x, different types of confidence intervals will have different posterior probabilities according to the prior distribution under investigation. So that they must be compared to credible interval with the same posterior probability.

In tables 2 to 7 we show the posterior lengths for the Agresti-Coul, Wilson, HPD and Central intervals considering the $Beta(2,2)$ distribution, Jeffreys and uniform prior distributions when $n = 5$. In general, the central intervals have the worst performance for all the studied priors. Wilson and HPD intervals are comparable in terms of posterior lengths. On the other hand, the posterior length of the Agresti-Coull interval is larger than of HPD interval by 0.06 to 0.02 when the observed value, x, is near the boundaries 0 and n. This difference can be of definite practical relevance. The differences start to wear off when x takes central values or when n is greater than 30.

x	Posterior Probability	Agresti-Coull	HPD	Central
0	0.96	0.54	0.49	0.51
1	0.97	0.62	0.61	0.62
2	0.97	0.65	0.65	0.65
3	0.97	0.65	0.66	0.65
4	0.97	0.62	0.60	0.61
5	0.96	0.54	0.48	0.52

Table 2. Posterior lengths of Agresti-Coull interval, HPD and Central intervals with $Beta(2,2)$ prior distribution for $n = 5$.

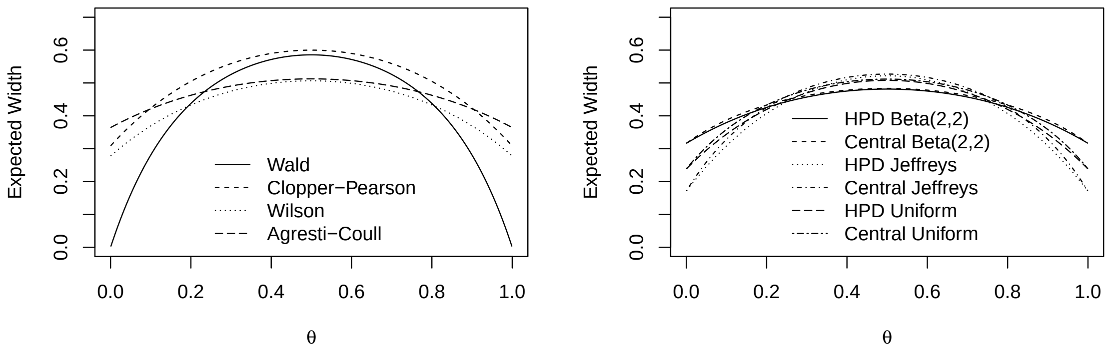

Figure 5. Expected lengths of frequentist and Bayesian intervals ($n = 10$).

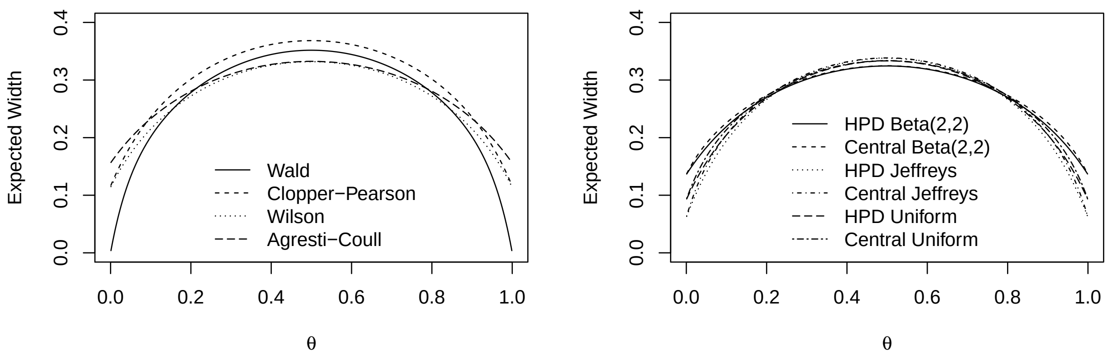

Figure 6. Expected lengths of the frequentist and Bayesian intervals ($n = 10$).

Discussion

We show that the Clopper-Pearson interval, as expected, is very conservative and not so accurate. Although it has good minimum coverage probability, the Agresti-Coull interval is more conservative than the Wilson interval and has larger expected length when $n \leq 30$. The Wilson interval has a good performance in terms of coverage probability, expected length and posterior length even for small sample sizes.

The HPD and central intervals show similar values of mean coverage probability, but the HPD intervals are better in terms of minimum coverage probability. The HPD interval with uniform prior distribution has a performance of coverage probability comparable to the Wilson interval. In addition, it has the advantage of better minimum coverage probability and smaller expected length. The similarity between those intervals was studied in Jin et al. (2017), where they show that the Wilson interval can be viewed as an approximation of the equal-tailed Bayesian interval with uniform prior.

Based on this analysis, the Wilson or the HPD uniform prior interval could be used when $n \leq 30$. The HPD or central

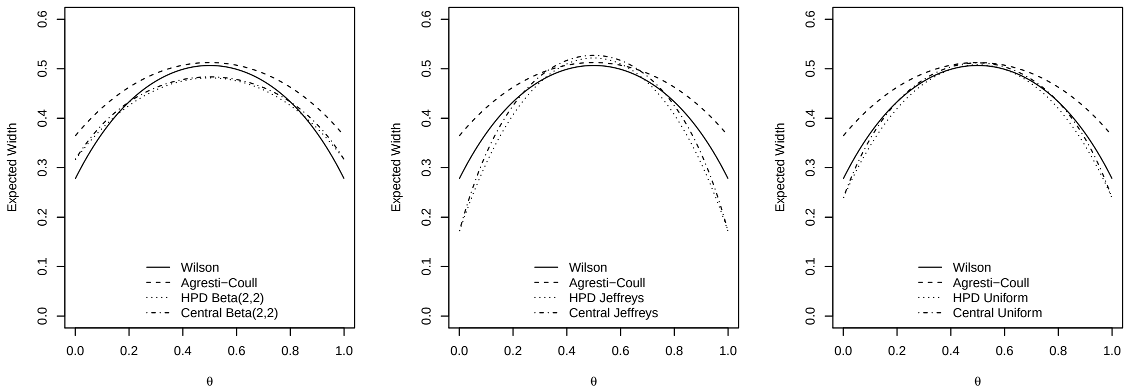

Figure 7. Expected lengths of the Wilson and Bayesian intervals ($n = 10$).

x	Posterior Probability	Wilson	HPD	Central
0	0.93	0.43	0.42	0.46
1	0.96	0.59	0.59	0.59
2	0.97	0.65	0.65	0.64
3	0.97	0.65	0.64	0.65
4	0.96	0.59	0.58	0.60
5	0.93	0.43	0.42	0.46

Table 3. Posterior lengths of Wilson interval, HPD and Central intervals with $Beta(2,2)$ prior distribution for $n = 5$.

Jeffreys prior intervals are also good choices, specially when there is information that $\theta \leq 0.26$ or $\theta \geq 0.74$, since it is the most accurate interval in this case. For larger sample sizes, the Agresti-Coull interval is comparable to those intervals and it could also be used.

References

1. Agresti, A. & Coull, B.A. Approximate is better than ”exact” for interval estimation of binomial proportions. *The Am. Stat.* 52, 119-126 (1998).
2. Agresti, A. & Caffo, B. Simple and effective confidence intervals for proportions and difference of proportions result from adding two successes and two failures. *The Am. Stat.* 54, 280-288 (2000).
3. Agresti, A. & Min, Y. Frequentist performance of Bayesian confidence intervals for comparing proportions in 2x2 contingency table. *Biom.* 61, 515-523 (2005).
4. Agresti, A. Score and pseudo-score confidence intervals for categorical data analysis. *Statistics in Biopharmaceutical Research* 3, 163-172 (2011).

x	Posterior Probability	Agresti-Coull	HPD	Central
0	0.99	0.54	0.50	0.56
1	0.96	0.62	0.58	0.62
2	0.93	0.65	0.64	0.66
3	0.93	0.65	0.65	0.66
4	0.96	0.62	0.58	0.62
5	0.99	0.54	0.49	0.57

Table 4. Posterior lengths of Agresti-Coull interval, HPD and Central intervals with Jeffreys prior distribution for $n = 5$.

x	Posterior Probability	Wilson	HPD	Central
0	0.99	0.43	0.42	0.51
1	0.92	0.59	0.51	0.56
2	0.92	0.65	0.64	0.65
3	0.92	0.65	0.64	0.66
4	0.92	0.59	0.51	0.55
5	0.99	0.43	0.45	0.52

Table 5. Posterior lengths of Wilson interval, HPD and Central intervals with Jeffreys prior distribution for $n = 5$.

5. Brown, L.D., Cai, T.T. & DasGupta, A. Interval estimation for a binomial proportion. *Statistical Science* 16, 101-117 (2001)
6. Clopper, C. & Pearson, E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biom.* 26, 404-413 (1934).
7. Jin, S., Thulin, M. & Larsson, R. Approximate bayesianity of frequentist confidence intervals for a binomial proportion. *The Am. Stat.* 71, 106-111 (2017).
8. Thulin, M. Converage-adjusted confidence intervals for a binomial proportion. *Scand. J. Stat.* 41, 291-300 (2014).
9. Thulin, M. *et al.* The cost of using exact confidence intervals for a binomial proportion. *Electron. J. Stat.* 8, 817-840 (2014).
10. Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*. 54, 426-482 (1943)
11. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *Journal of the American Association*. 22. 209-212. (1927)

x	Posterior Probability	Agresti-Coull	HPD	Central
0	0.98	0.54	0.48	0.54
1	0.97	0.62	0.62	0.65
2	0.95	0.65	0.65	0.65
3	0.95	0.65	0.65	0.66
4	0.97	0.62	0.61	0.64
5	0.98	0.54	0.49	0.54

Table 6. Posterior lengths of Agresti-Coull interval, HPD and Central intervals with uniform prior distribution for $n = 5$.

x	Posterior Probability	Wilson	HPD	Central
0	0.97	0.43	0.44	0.50
1	0.95	0.59	0.56	0.61
2	0.95	0.65	0.65	0.66
3	0.95	0.65	0.65	0.65
4	0.95	0.59	0.58	0.60
5	0.97	0.43	0.43	0.49

Table 7. Posterior lengths of Wilson interval, HPD and Central intervals with uniform prior distribution for $n = 5$.