INTERVAL ESTIMATION FOR A BINOMIAL PROPORTION*

by

L. Brown, T. Tony Cai, A. DasGupta
University of Pennsylvania, Purdue University and
Purdue University

Technical Report #99-19

Department of Statistics
Purdue University
West Lafayette, IN 47907-1399
USA

August 1999

⊛

Interval Estimation for a Binomial Proportion*

Lawrence D. Brown, T. Tony Cai and Anirban DasGupta
University of Pennsylvania, Purdue University and Purdue University

Abstract

We revisit the problem of interval estimation of a binomial proportion. The erratic behavior of the coverage probability of the standard Wald confidence interval has previously been remarked on in the literature (Blyth & Still (1983), Agresti & Coull (1998), Santner (1998), and others). We begin by showing that the chaotic coverage properties of the Wald interval are far more persistent than is appreciated. Furthermore, common textbook prescriptions regarding its safety are misleading and defective in several respects and cannot be trusted.

This leads us to consideration of alternative intervals. Eight natural alternatives are presented, each with its motivation and context. Each interval is examined as regards its coverage probability and its expected length. Based on this analysis, we recommend the Wilson interval (Wilson (1927)) or the equal tailed Jeffreys prior interval for small n, and the interval suggested in Agresti and Coull for larger n. The theoretical support for these recommendations is available in the companion paper Brown, Cai & DasGupta (1999).

We also explain why the Jeffreys prior interval has appealing coverage properties while maintaining parsimony in its length.

Keywords: Bayes; Binomial distribution; Confidence intervals; Coverage probability; Edgeworth expansion; Expected length; Jeffreys prior; Normal approximation; Posterior.

*Research supported by grants from the National Science Foundation and the National Security Agency.

1 Introduction

This article revisits one of the most basic and methodologically important problems in statistical practice, namely, interval estimation of the probability of success in a binomial distribution. There is a textbook confidence interval for this problem that has acquired nearly universal acceptance in practice. The interval, of course, is $\hat{p} \pm z_{\alpha/2}\, n^{-1/2}(\hat{p}(1-\hat{p}))^{1/2}$, where $\hat{p} = X/n$ is the sample proportion of successes, and $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution. The interval is easy to present and motivate, and easy to compute. With the exceptions of the t test, linear regression, and ANOVA, its popularity in everyday practical statistics is virtually unmatched. The standard interval is known as the Wald interval as it comes from the Wald large sample test for the binomial case.

So at first glance, one may think that the problem is too simple, and has a clear and present solution. In fact, the problem is a difficult one, with unanticipated complexities. It is widely recognized that the actual coverage probability of the standard interval is poor for p near 0 or 1. In fact, even at the level of introductory statistics texts, the standard interval is often presented with the caveat that it should be used only when $n \cdot \min(p,\ 1-p)$ is at least 5 (or 10). Examination of the popular texts reveals that the qualifications with which the standard interval is presented are varied, but they all reflect the concern about poor coverage when p is near the boundaries.

In a series of interesting recent articles, it has also been pointed out that the coverage properties of the standard interval can be erratically poor even if p is not near the boundaries; see, for instance, Vollset (1993), Santner (1998), and Agresti & Coull (1998). Slightly older literature includes Ghosh (1979), Cressie (1980), and Blyth & Still (1983). Agresti & Coull (1998), particularly consider the nominal 95% case, and show the erratic and poor behavior of the standard interval's coverage probability for small n even when p is not near the boundaries. See their Figure 4 for the cases $n = 5$ and 10.

We will show in this article that the eccentric behavior of the standard interval's coverage probability is far deeper than has been explained or is appreciated by statisticians at large. We will show that the popular prescriptions the standard interval comes with are defective in several respects, and are not to be trusted. In addition, we will motivate, present, and analyze several alternatives to the standard interval for a general confidence level. We will ultimately make recommendations about choosing a specific interval for practical use, separately for different intervals of values of n. It will be seen that for small n (40 or less), our recommendation differs from the recommendation Agresti & Coull (1998) made for the nominal 95% case. To facilitate greater appreciation of the seriousness of the problem, we have kept the technical content of this article at a minimal level. The companion article Brown, Cai & DasGupta (1999) presents the associated theoretical calculations on Edgeworth expansions of the various intervals' coverage probabilities and asymptotic expansions for their expected lengths.

In Section 2, we first present a series of examples on the degree of severity of the chaotic behavior of the standard interval's coverage probability. The chaotic behavior does not go away even when n is quite large and p is not near the boundaries. For instance, when n is 100, the actual coverage probability of the nominal 95% standard interval is .952 if p is .106, but only .911 if p is .107. The behavior of the coverage probability can be even more erratic as a function of n. If the true p is .5, the actual coverage probability of the nominal

95% interval is .953 at the rather small sample size $n = 17$, but falls to .919 at the much larger sample size $n = 40$! This eccentric behavior can get downright extreme in certain practically important problems. For instance, consider defective proportions in industrial quality control problems. There it would be quite common to have a true p that is small. For instance, if the true p is .005, then the coverage probability of the nominal 95% interval increases monotonically in n all the way up to $n = 591$ to the level .945, only to drop down to .792 if n is 592. This unlucky spell continues for a while, and then the coverage bounces back to .948 when n is 953, but dramatically falls to .852 when n is 954. Subsequent unlucky spells start off at $n = 1279$, 1583, and on and on. It should be widely known that the coverage of the standard interval can be significantly lower at a much larger sample size, and that all of these happen in an unpredictable and rather random way.

Continuing, also in Section 2, we list a set of common prescriptions that standard texts present while discussing the standard interval. We show what the deficiencies are in some of these prescriptions. For example, Proposition 1 and the subsequent Table 3 illustrate the defects of these common prescriptions.

In Section 3, we begin to address alternative intervals. Eight alternatives are listed. Of these, Agresti & Coull (1998) considered, for the 95% case, the Clopper-Pearson "exact" interval, the "score interval", and an adjusted Wald interval that formally adds two successes and two failures to the observed counts and then uses the standard method. That is, this interval is $\tilde{p} \pm z_{.025}\tilde{n}^{-1/2}(\tilde{p}(1-\tilde{p}))^{1/2}$, where $\tilde{n} = n+4$, and $\tilde{p} = (X+2)/(n+4)$. We refer to this interval as the Agresti-Coull interval. Also, we refer to the score interval as the Wilson interval since Wilson (1927) seems to have introduced it. Two additional intervals we have considered are the arcsine interval, and the Bayesian equal tailed interval resulting from the natural noninformative Jeffreys prior. A very simple interval that simply recenters the standard interval is also presented. The other two intervals are slight modifications of the Wilson and the Jeffrey prior interval in order to correct a disturbing downward spike in their coverages very close to the two boundaries. All the intervals with necessary motivation and additional information are then analyzed in the rest of Section 3.

In section 4, we come to choice of a specific alternative interval. At issue are three things. First, the coverage probability should be close to the target nominal value; second, the interval should be as parsimonious as possible in the sense of expected length. And, of course, simplicity of presentation is also an issue, particularly for class room presentation. On consideration of these factors, we came to the conclusion that for small n (40 or less), we recommend that either the Wilson or the Jeffreys prior interval should be used. They are very comparable, and either may be used depending on taste. The Wilson interval has a closed form formula. The Jeffreys interval does not. One can expect that there would be resistance to using the Jeffreys interval solely due to this reason. We therefore provide a table simply listing the limits of the Jeffrey interval for n up to 30, and in addition also give closed form and very accurate approximations to the limits. These approximations do not need any additional software. For larger n ($n > 40$), the Wilson, Jeffrey, and the Agresti-Coull interval are all very comparable, and so for such n, due to its simplest form, we come to the conclusion that the Agresti-Coull interval should be recommended. Even for smaller sample sizes the Agresti-Coull interval interval is strongly preferable to the standard one, and so might be the choice where simplicity is a paramount objective.

We strongly recommend that introductory texts in statistics present one or more of these alternative intervals, in preference to the standard one. The slight sacrifice in simplicity would be more than worthwhile. The conclusions we make are theoretically supported by the results in Brown, Cai & DasGupta (1999).

2 The Standard Interval

When constructing a confidence interval we usually wish the actual coverage probability to be close to the nominal confidence level. Because of the discrete nature of the binomial distribution we cannot always achieve the exact nominal confidence level unless a randomized procedure is used. Thus our objective is to construct non-randomized confidence intervals for p such that the coverage probability $P_p(p \in CI) \approx 1-\alpha$ where α is some prespecified value between 0 and 1. We shall also call $C(p,n) = P_p(p \in CI)$ the confidence coefficient.

A standard confidence interval for p based on normal approximation has gained universal recommendation in the introductory statistics textbooks and in statistical practice. The interval is known to guarantee that for any fixed p, $C(p,n) \to 1-\alpha$ as $n \to \infty$.

Let $\phi(z)$ and $\Phi(z)$ be the standard normal density and distribution functions, respectively. Throughout the paper we denote $\kappa \equiv z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, $\hat{p} = X/n$ and $\hat{q} = 1-\hat{p}$. The standard normal approximation confidence interval CI_s is given by

$$CI_s = \hat{p} \pm \kappa\, n^{-1/2}(\hat{p}\hat{q})^{1/2}. \tag{1}$$

This interval is obtained by inverting the acceptance region of the well known Wald large-sample normal test for a general problem:

$$|(\hat{\theta}-\theta)/\widehat{se}(\hat{\theta})| \le \kappa \tag{2}$$

where θ is a generic parameter, $\hat{\theta}$ is the maximum likelihood estimate of θ and $\widehat{se}(\hat{\theta})$ is the estimated standard error of $\hat{\theta}$. In the binomial case, we have $\theta = p$, $\hat{\theta} = X/n$ and $\widehat{se}(\hat{\theta}) = (\hat{p}\hat{q})^{1/2}n^{-1/2}$.

The standard interval is easy to calculate and is heuristically appealing. In introductory statistics texts and courses, the confidence CI_s is usually presented along with some heuristic justification based on the Central Limit Theorem. The students and users no doubt believe that the larger the number n, the better the normal approximation, and thus the closer the actual coverage would be to the nominal level $1-\alpha$. We will show how badly this is false. Let us take a close look at how the standard interval CI_s really performs.

2.1 Lucky n, Lucky p

An interesting phenomenon for the standard interval is that the actual coverage probability of the confidence interval contains non-negligible oscillation as both p and n vary. There exist some "lucky" pairs $(p,\ n)$ such that the actual coverage probability $C(p,\ n)$ is very close to or larger than the nominal level. On the other hand, there also exist "unlucky" pairs $(p,\ n)$ such that the corresponding $C(p,\ n)$ is much smaller than the nominal level. The phenomenon of oscillation is both in n, for fixed p, and in p, for fixed n. Furthermore, the oscillation is discontinuous. Drastic change in coverage occurs in nearby p for fixed n and in nearby n for fixed p. Let us look at five simple but instructive examples.

The probabilities reported in the following plots and tables, as well as those appearing later in this paper, are the result of direct probability calculations produced in S-PLUS. In all cases their numerical accuracy considerably exceeds the number of significant figures reported and/or the accuracy visually obtainable from the plots. (Plots for variable p are the probabilities for a fine grid of values of p - e.g., 2000 equally spaced values of p for the plots in Figure 6.)

Example 1. Figure 1 plots the coverage probability of the nominal 95% standard interval for $p = .2$. The number of trials n varies from 25 to 100. It is clear from the plot that the oscillation is significant and the coverage probability does not steadily get closer to the nominal confidence level as n increases. For instance, $C(.2, 30) = .946$ and $C(.2, 98) = .928$. So, as hard as it is to believe, the coverage probability is significantly closer to .95 when $n = 30$ than when $n = 98$. We see that the true coverage probability behaves contrary to conventional wisdom in a very significant way.

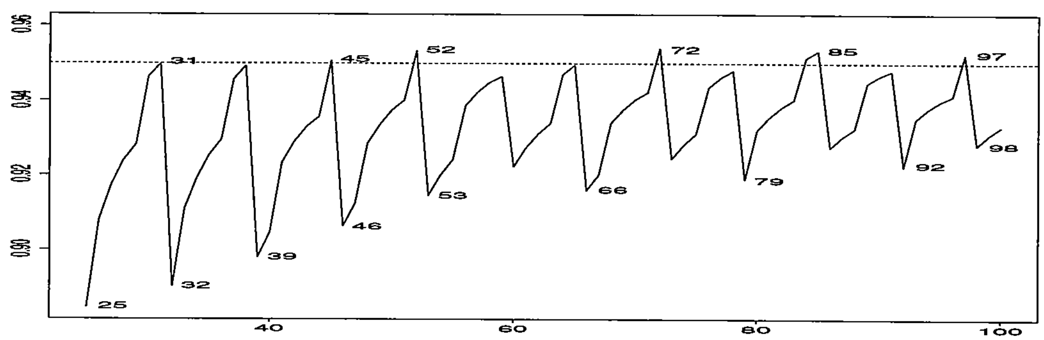

Figure 1: Oscillation phenomenon for fixed $p = .2$ and variable $n = 25$ to 100.

Example 2. Now consider the case of $p = .5$. Since $p = .5$, conventional wisdom might suggest to an unsuspecting user that all will be well if n is about 20. We evaluate the exact coverage probability of the 95% standard interval for $10 \leq n \leq 50$. In Table 1, we list the values of "lucky" n (defined as $C(p,\ n) \geq .95$) and the values of "unlucky" n (defined for specificity as $C(p,\ n) \leq .92$). The conclusions presented in Table 2 are surprising. We note that when $n = 17$ the coverage probability is .951, but the coverage probability equals .904 when $n = 18$. Indeed, the unlucky values of n arise suddenly. Although p is .5, the coverage is still only .919 at $n = 40$. It illustrates the inconsistency, unpredictability and poor performance of the standard interval.

Lucky n	17	20	25	30	35	37	42	44	49
$C(.5, n)$	.951	.959	.957	.957	.959	.953	.956	.951	.956
Unlucky n	10	12	13	15	18	23	28	33	40
$C(.5, n)$	.891	.854	.908	.882	.904	.907	.913	.920	.919

Table 1: Lucky n and Unlucky n for $10 \leq n \leq 50$ and $p = .5$.

Example 3. Now let us move p really close to the boundary, say $p = .005$. We mention in the introduction that such p are relevant in certain practical applications. Since p is so small, now one may fully expect that the coverage probability of the standard interval is poor. Figure 2 and Table 2 show that there are still surprises and indeed we now begin to see a whole new kind of erratic behavior. The oscillation of the coverage probability does not show until rather large n. Indeed, the coverage probability makes a slow ascent all the way until $n = 591$, and then dramatically drops to .792 when $n = 592$. Figure 2 shows that thereafter the oscillation manifests in full force, in contrast to Examples 1 and 2, where the oscillation started early on. Subsequent "unlucky" values of n again arise in the same unpredictable way, as one can see from Table 2.

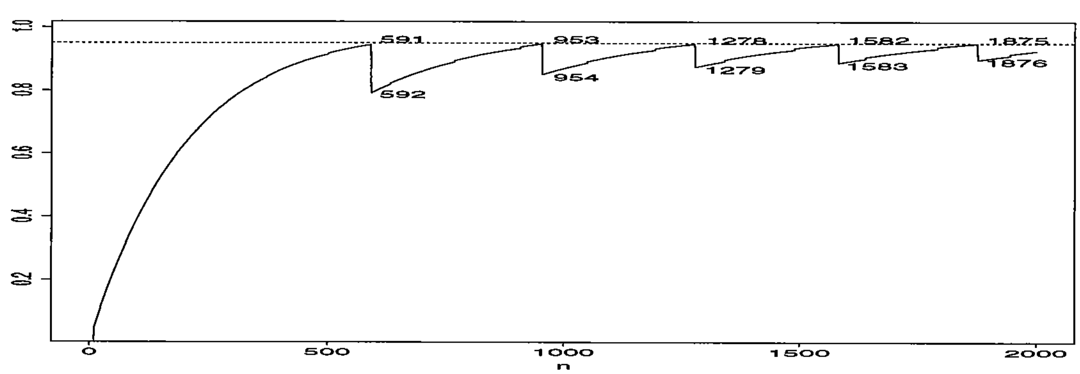

Figure 2: Oscillation in coverage for small p.

Unlucky n	592	954	1279	1583	1876
$C(.005, n)$	.792	.852	.875	.889	.898

Table 2: Late arrival of unlucky n for small p.

Example 4. Figure 3 plots the coverage probability of the nominal 95% standard interval with fixed $n = 100$ and variable p. It can be seen from Figure 3 that in spite of the "large" sample size, significant change in coverage probability occurs in nearby p. The magnitude of oscillation increases significantly as p moves toward 0 or 1.

Example 5. Figure 4 shows the coverage probability of the nominal 99% standard interval with $n = 20$ and variable p from 0 to 1. Besides the oscillation phenomenon similar to Figure 3, a striking fact in this case is that the coverage never reaches the nominal level. The coverage probability is ALWAYS smaller than .99, and in fact on the average the coverage is only .883. Our evaluations show that for all $n \leq 45$, the coverage of the 99% standard interval is strictly smaller than the nominal level for all $0 < p < 1$.

It is evident from the preceding presentation that the actual coverage probability of the standard interval can differ significantly from the nominal confidence level for realistic and

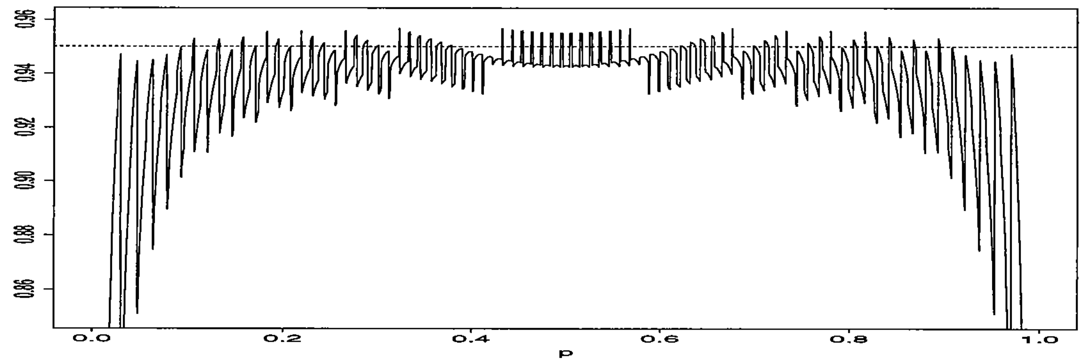

Figure 3: Oscillation phenomenon for fixed $n = 100$ and variable p.

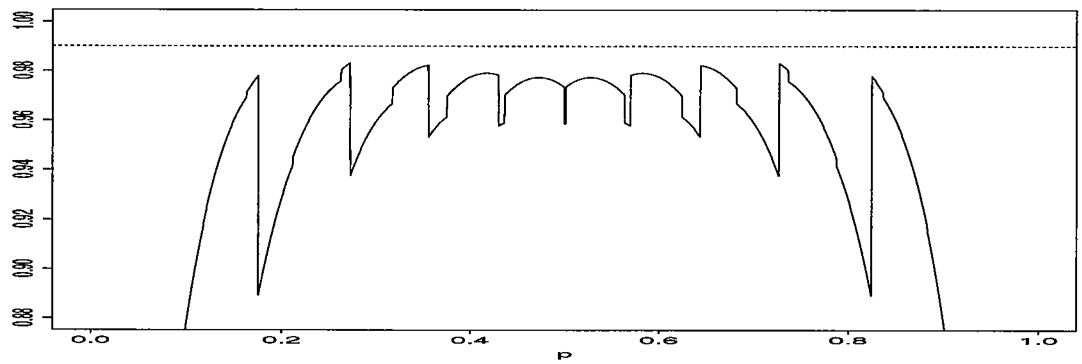

Figure 4: Coverage of the nominal 99% standard interval for fixed $n = 20$ and variable p.

indeed larger than realistic sample sizes. The error comes from two sources: discreteness and skewness in the underlying binomial distribution. For a two-sided interval, the rounding error due to discreteness is dominant, and the error due to skewness is somewhat secondary, but still important for even moderately large n. (See Brown, Cai & DasGupta (1999) for more details.) Note that the situation is different for one-sided intervals. There, the error caused by the skewness can be larger than the rounding error. See Hall (1982) for a detailed discussion on one-sided confidence intervals.

The oscillation in the coverage probability is caused by the discreteness of the binomial distribution, more precisely the lattice structure of the binomial distribution. The oscillations are unavoidable for any nonrandomized procedure.

The erratic and unsatisfactory coverage properties of the standard interval have often been remarked on, but curiously still do not seem to be widely appreciated among statisticians. See, e.g., Ghosh (1979) , Blyth & Still (1983), and Agresti & Coull (1998). Blyth & Still (1983) also shows that the continuity-corrected version still has the same disadvantages.

Here we would like to point an error in Ghosh (1979). It is claimed that, for any p and α, n can be chosen sufficiently large such that the confidence coefficient actually exceeds the nominal level $1 - \alpha$ up to the order $n^{-1/2}$ (Ghosh (1979), pp. 895). This is in fact not true. The oscillation terms were mistakenly omitted in the Edgeworth expansion in Ghosh (1979). See Brown, Cai & DasGupta (1999) for more details on the Edgeworth expansion of the coverage probability.

2.2 Textbook Qualifications

The normal approximation used to justify the standard confidence interval for p can be significantly in error. The error is most evident when the true p is close to 0 or 1. See Lehmann (1999). In fact, it is easy to show that, for any fixed n, the confidence coefficient $C(p,n) \to 0$ as $p \to 0$ or 1. Therefore the most major problems arise as regards coverage probability when p is near the boundaries.

Poor coverage probabilities for p near 0 or 1 are widely remarked on, and generally, in the popular texts, a brief sentence is added qualifying when to use the standard confidence interval for p. It is interesting to see what these qualifications are. A sample of 10 popular texts gives the following qualifications:

The confidence interval may be used if

(a). np, $n(1-p)$ are ≥ 5;	(b). np, $n(1-p)$ are ≥ 10;
(c). $np(1-p) \geq 5$;	(d). $np(1-p) \geq 10$;
(e). $n\hat{p}$, $n(1-\hat{p})$ are ≥ 5;	(f). $n\hat{p}$, $n(1-\hat{p})$ are ≥ 10;
(g). n quite large;	(h). $n \geq 50$ unless p is very small.

It seems clear that the authors are attempting to say that the standard interval may be used if the central limit approximation is accurate. These prescriptions are defective in several respects. In the estimation problem, (a), (b), (c), and (d) are not verifiable. Even when these conditions are satisfied, we see, for instance, from Table 1 in the previous section, that there is no guarantee that the true coverage probability is close to the nominal confidence level. For example, when $n = 40$ and $p = .5$, one has $np = n(1-p) = 20$ and $np(1-p) = 10$, so clearly either of the conditions (a), (b), (c), and (d) is satisfied. But from Table 1, the true coverage probability in this case equals .919 which is certainly unsatisfactory for a confidence interval at nominal level .95.

The qualification (g) is useless and (h) is patently misleading. (e), and (f) are certainly verifiable, but they are not given a meaning. The point is that the standard interval clearly has serious problems and the influential texts caution the readers about that. However, the caution appears to not serve its purpose, for a variety of reasons.

Here is a result that shows that sometimes the qualifications are not correct even in the limit as $n \to \infty$.

Proposition 1 *Let $\gamma > 0$. For the standard confidence interval,*

$$\lim_{n\to\infty} \inf_{p:np,n(1-p)\geq\gamma} C(p,n) = P(a_\gamma < Poisson(\gamma) \leq b_\gamma), \tag{3}$$

where a_γ and b_γ are the integer parts of

$$(\kappa^2 + 2\gamma \pm \kappa\sqrt{\kappa^2 + 4\gamma})/2,$$

where the $-$ sign goes with a_γ and the $+$ sign with b_γ.

Let us use Proposition 1 to investigate the validity of qualifications (a) and (b) in the list above. The nominal confidence level in Table 3 below is .95. It is clear that qualification (a) does not work at all, and (b) is marginal. There are similar problems with qualifications (c) and (d).

γ	5	7	10
$\lim_{n\to\infty} \inf_{p:\,np,n(1-p)\geq\gamma} C(p,n)$	.875	.913	.926

Table 3: Limiting minimum coverage probability when $np, n(1-p) \geq \gamma$.

Proof of Proposition 1: By a monotone likelihood ratio argument, it can be seen that for fixed n, the infimum of $C(p,\ n)$ over the set $\{p : np, n(1-p) \geq \gamma\}$ is attained at $p = \gamma/n$ (and $1-\gamma/n$). The sequence of $Bin(n,\ \gamma/n)$ distributions converges weakly to the Poisson(γ) distribution and so the limit of the infimum is the Poisson probability in the Proposition by an easy calculation. ∎

3 Alternative Intervals

From the evidence gathered in Section 2, it seems clear that the standard interval is just too risky. Really, since one can find better alternatives, the standard interval should not be used at all. This brings us to the consideration of alternative intervals. We now present and analyze eight such alternatives, each with its motivation.

3.1 The Re-centered Interval

The standard interval is centered at $\hat{p} = X/n$, the MLE of p. Although $\hat{p}$ is an optimal point estimate of p according to many criteria, it is not the optimal center for a confidence interval.

The performance of the standard interval can be much improved by simply moving the center of the interval towards 1/2 to $\tilde{p} = (X + \kappa^2/2)/(n + \kappa^2)$. The recentered interval has the form

$$CI_{rs} = (X + \kappa^2/2)/(n + \kappa^2) \pm \kappa(\hat{p}\hat{q})^{1/2} n^{-1/2}$$

This interval simply shifts the standard interval CI_s towards the center 1/2 by an amount of $\kappa^2|n/2 - X|/(n(n + \kappa^2))$. Figure 6 plots the coverage for $n = 50$; the dotted line is the coverage of the standard interval. When $\alpha = .05$, if we use the value 2 instead of 1.96 for κ, then $\tilde{p} = (X + 2)/(n + 4)$; this is the Wilson estimator of p. See Wilson (1927) and Agresti & Coull (1998).

3.2 The Wilson Interval

Another possibility is the confidence interval based on inverting the test in equation (2) that uses the null standard error $(pq)^{1/2}n^{-1/2}$, instead of the estimated standard error $(\hat{p}\hat{q})^{1/2}n^{-1/2}$. This confidence interval has the form

$$CI_W = \frac{X + \kappa^2/2}{n + \kappa^2} \pm \frac{\kappa n^{1/2}}{n + \kappa^2}(\hat{p}\hat{q} + \kappa^2/(4n))^{1/2}. \quad (4)$$

This interval was apparently introduced by Wilson (1927) and we will call this interval the Wilson interval.

The Wilson interval has theoretical appeal. The interval is the inversion of the CLT approximation to the family of equal-tail tests of $H_0 : p = p_0$. Hence, one accepts H_0 based on the CLT approximation if and only if p_0 is in this interval.

3.3 Modified Wilson Interval

The lower bound of the Wilson interval is formed by inverting a CLT approximation. The coverage has downward spikes when p is very near 0 or 1. These spikes exist for all n and α. In fact, it can be shown that, when $1 - \alpha = .95$ and $p = .1765/n$, for example,

$$\lim_{n\to\infty} P_p(p \in CI_W) = .838;$$

and when $1 - \alpha = .99$ and $p = .1174/n$, $\lim_{n\to\infty} P_p(p \in CI_W) = .889$. See also Agresti & Coull (1998).

The spikes can be removed by using a one-sided Poisson approximation for x close to 0 or n. Suppose we modify the lower bound for $x = 1, .., x^*$. For a fixed $1 \le x \le x^*$, the lower bound of CI_W should be replaced by a lower bound of λ_x/n where λ_x solves

$$e^{-\lambda}(\lambda^0/0! + \lambda^1/1! + \cdots + \lambda^{x-1}/(x-1)!) = 1 - \alpha. \quad (5)$$

A symmetric prescription needs to be followed to modify the upper bound for x very near n. The value of x^* should be small. Values which work reasonably well for $1 - \alpha = .95$ are

$$x^* = 2 \text{ for } n < 50 \text{ and } x^* = 3 \text{ for } 51 \le n \le 100+.$$

Using the relationship between the Poisson and χ^2 distributions,

$$P(Y \le x) = P(\chi^2_{2(1+x)} \le 2\lambda)$$

where $Y \sim \text{Poisson}(\lambda)$, one can also formally express λ_x in equation (5) in terms of the χ^2 quantiles:

$$\lambda_x = \frac{1}{2}\chi^2_{2x,\alpha}$$

where $\chi^2_{2x,\alpha}$ denotes the 100α-th percentile of the χ^2 distribution with $2x$ degrees of freedom. Table 4 below gives the values of λ_x for selected values of x and α.

For example, consider the case $1 - \alpha = .95$ and $x = 2$. The lower bound of CI_W is $\approx .548/(n + 4)$. The modified Wilson interval replaces this by a lower bound of λ/n where $\lambda = (1/2)\,\chi^2_{4,.05}$. Thus, from a χ^2 table, for $x = 2$ the new lower bound is $.355/n$.

We denote this modified Wilson interval by CI_{M-W}. See Figure 6 for its coverage probability.

$1-\alpha$	$x=1$	$x=2$	$x=3$
.90	0.105	0.532	1.102
.95	0.051	0.355	0.818
.99	0.010	0.149	0.436

Table 4: Values of λ_x for the modified lower bound for the Wilson interval.

3.4 The Agresti-Coull Interval

The standard interval CI_s is simple and easy to remember. For the purposes of classroom presentation and use in texts, it is nice to have an alternative that has the familiar form $\hat{p} \pm z\sqrt{\hat{p}(1-\hat{p})/n}$, with a better and new choice of $\hat{p}$ rather than $\hat{p} = X/n$. Denote $\tilde{X} = X + \kappa^2/2$ and $\tilde{n} = n + \kappa^2$. Let $\tilde{p} = \tilde{X}/\tilde{n}$ and $\tilde{q} = 1 - \tilde{p}$. Define the confidence interval CI_{AC} for p by

$$CI_{AC} = \tilde{p} \pm \kappa(\tilde{p}\tilde{q})^{1/2}\tilde{n}^{-1/2}. \tag{6}$$

For the case when $\alpha = .05$, if we use the value 2 instead of 1.96 for κ, this interval is the "add 2 successes and 2 failures" interval in Agresti & Coull (1998). For this reason, we call it the Agresti-Coull interval. To the best of our knowledge, Samuels & Witmer (1999) is the first introductory statistics textbook that recommends the use of this interval. See Figure 6 for the coverage of this interval. See also Figure 7 for its average coverage probability.

3.5 The Arcsine Interval

Another interval is based on a widely used variance stabilizing transformation for the binomial distribution (see, e.g., Bickel & Doksum (1977)): $T(\hat{p}) = arcsin(\hat{p}^{1/2})$. Anscombe (1948) showed that replacing $\hat{p}$ by $\check{p} = (X + 3/8)/(n + 3/4)$ gives better variance stabilization; furthermore

$$2n^{1/2}[arcsin(\check{p}^{1/2}) - arcsin(p^{1/2})] \to N(0,\ 1), \quad \text{as } n \to \infty.$$

This leads to an approximate $100(1-\alpha)\%$ confidence interval for p:

$$CI_{Arc} = [\sin^2(\arcsin(\check{p}^{1/2}) - \frac{1}{2}\kappa n^{-1/2}),\ \sin^2(\arcsin(\check{p}^{1/2}) + \frac{1}{2}\kappa n^{-1/2})]. \tag{7}$$

See Figure 6 for the coverage probability of this interval for $n = 50$.

3.6 The Clopper-Pearson Interval

The Clopper-Pearson interval is "exact" for all n. If $X = x$ is observed, then the Clopper-Pearson (1934) interval is defined by $CI_{CP} = [L_{CP}(x),\ U_{CP}(x)]$, where $L_{CP}(x)$ and $U_{CP}(x)$ are, respectively, the solutions in p to the equations

$$P_p(X \geq x) = \alpha/2 \quad \text{and} \quad P_p(X \leq x) = \alpha/2.$$

It is easy to show that the lower endpoint is the $\alpha/2$ quantile of a beta distribution $Beta(x,\ n-x+1)$, and the upper endpoint is the $1-\alpha/2$ quantile of a beta distribution $Beta(x+1,\ n-x)$. This interval guarantees that the actual coverage probability is always at least $1-\alpha$. Figure 6 shows the coverage probability for $n = 50$. It can be seen that this interval is quite "conservative", in the sense that the coverage probability can be significantly larger than the nominal value $1-\alpha$.

3.7 Bayesian Methods

Beta distributions are the standard conjugate priors for binomial distributions and it is quite common to use beta priors for the construction of a Bayesian HPD interval (see Berger (1985) and Robert (1993)).

Suppose $X \sim Bin(n,\ p)$ and suppose p has a prior distribution $Beta(a_1,\ a_2)$; then the posterior distribution of p is $Beta(X + a_1,\ n - X + a_2)$. Thus a $100(1-\alpha)\%$ equal-tailed Bayesian interval is given by

$$[B(\alpha/2;\ X + a_1,\ n - X + a_2),\ B(1 - \alpha/2;\ X + a_1,\ n - X + a_2)],$$

where $B(\alpha; m_1, m_2)$ denotes the α quantile of a $Beta(m_1,\ m_2)$ distribution.

The well-known Jeffreys prior and the uniform prior are each a beta distribution. The non-informative Jeffreys prior is of particular interest to us. Historically, Bayes procedures under noninformative priors have a track record of good frequentist properties. In this problem the Jeffreys prior is $Beta(1/2, 1/2)$ which has the density function

$$f(p) = \pi^{-1} p^{-1/2}(1-p)^{-1/2}.$$

The $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval is defined as

$$CI_J = [L_J(x),\ U_J(x)] \tag{8}$$

where $L_J(0) = 0$, $U_J(n) = 1$ and otherwise

$$L_J(x) = B(\alpha/2;\ X + 1/2,\ n - X + 1/2), \tag{9}$$
$$U_J(x) = B(1 - \alpha/2;\ X + 1/2,\ n - X + 1/2). \tag{10}$$

The interval is formed by taking the central $1-\alpha$ posterior probability interval. This leaves $\alpha/2$ posterior probability in each omitted tail. The exception is for $x = 0$ (n) where the lower (upper) limits are modified to avoid the undesirable result that the coverage probability $C(p, n) \to 0$ as $p \to 0$ or 1.

The actual endpoints of the interval need to be numerically computed. This is very easy to do using softwares such as Minitab, S-Plus or Mathematica. Indeed, in Table 5, we have provided the limits for the case of the Jeffreys prior for $7 \le n \le 30$.

The endpoints of the Jeffreys prior interval are the $\alpha/2$ and $1-\alpha/2$ quantiles of the $Beta(x + 1/2, n - x + 1/2)$ distribution. The psychological resistance to using the interval is the inability to compute them at ease without software.

We provide two avenues to resolving this problem. One is Table 5. The second is a computable approximation to the limits of the Jeffreys prior interval, one that is computable with just a normal table. This approximation is obtained after some algebra from the general approximation to a Beta quantile given in pp. 945 in Abramowitz & Stegun (1970).

The lower limit of the $100(1-\alpha)\%$ Jeffreys prior interval is approximately

$$\frac{x+1/2}{n+1+(n-x+1/2)(e^{2\omega}-1)}, \tag{11}$$

where

$$\omega = \frac{\kappa\sqrt{4\hat{p}\hat{q}}/n + (\kappa^2-3)/(6n^2)}{4\hat{p}\hat{q}} + \frac{(1/2-\hat{p})(\hat{p}\hat{q}(\kappa^2+2)-1/n)}{6n(\hat{p}\hat{q})^2}.$$

The upper limit may be approximated by the same expression with κ replaced by $-\kappa$ in ω. The simple approximation given above is remarkably accurate.

An exact Bayesian solution would involve using the HPD intervals instead of our equal tails proposal. However, HPD intervals are much harder to compute, and do not even do as well in terms of coverage probability. See Figure 5 and compare to the Jeffreys' equal-tailed interval in Figure 6.

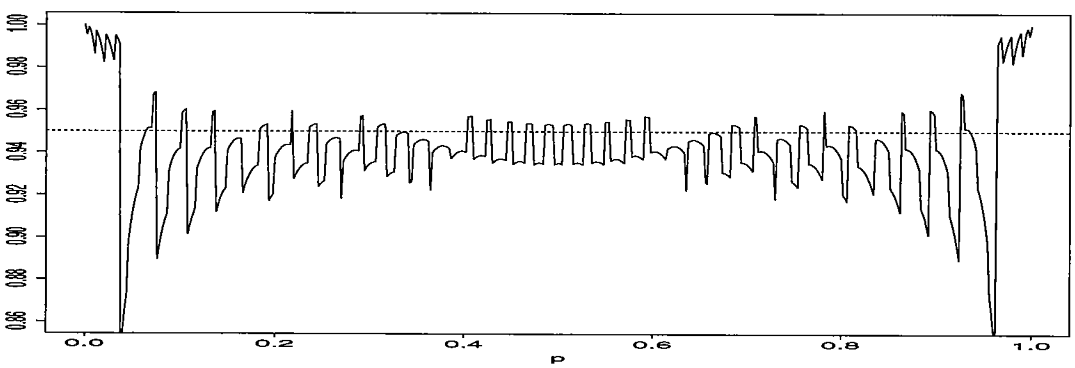

Figure 5: Coverage probability of the Jeffreys HPD interval for $n = 50$.

3.8 Modification of Jeffreys Interval

Evidently, CI_J has an appealing Bayesian interpretation. And, its coverage and length properties are reasonably appealing except for a very narrow downward coverage spike fairly near 0 and 1 (see Figure 6) . The unfortunate downward spikes in the coverage function result because $U_J(0)$ is too small, and symmetrically $L_J(n)$ is too large. To remedy this we revise these two specific limits as

$$U_{M-J}(0) = p_l \quad \text{and} \quad L_{M-J}(n) = 1 - p_l$$

where p_l satisfies $(1-p_l)^n = \alpha/2$ or equivalently $p_l = 1-(\alpha/2)^{1/n}$.

We also made a slight, ad-hoc alteration of $L_J(1)$, and set

$$L_{M-J}(1) = 0 \quad \text{and} \quad U_{M-J}(n-1) = 1.$$

In all other cases, $L_{M-J} = L_J$ and $U_{M-J} = U_J$. We denote the modified Jeffreys interval by CI_{M-J}. This modification seems to remove the two steep downward spikes and the performance of the interval is improved. See Figure 6.

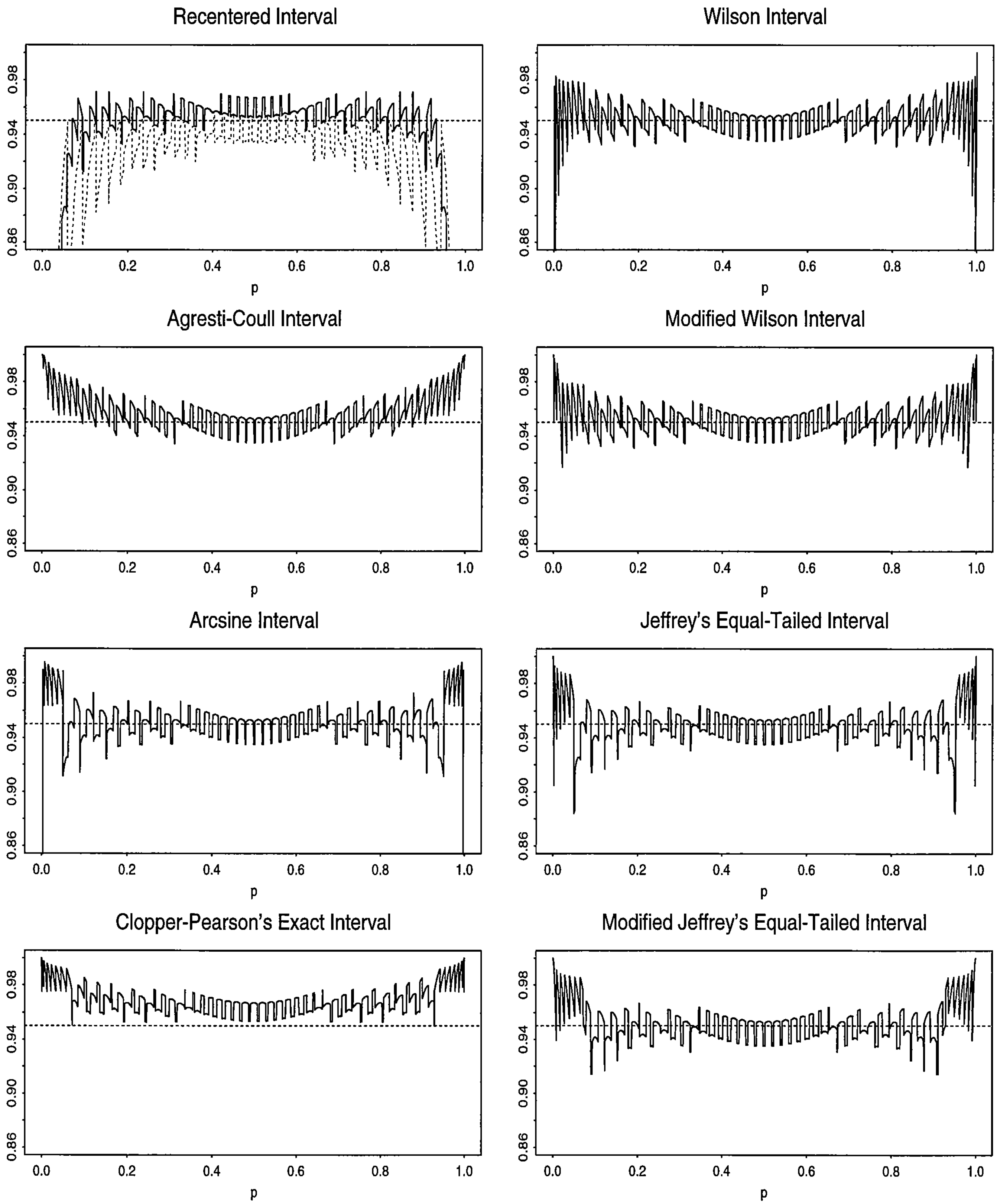

Figure 6: Coverage probability for $n = 50$. The dotted line on the upper-left plot is the coverage of the standard interval.

3.9 Connections Between Jeffreys Intervals and Exact Intervals

The equal-tailed Jeffreys prior interval has some interesting connections to the Clopper-Pearson "exact" interval. As we mentioned earlier, the Clopper-Pearson interval CI_{CP} can be written as

$$CI_{CP} = [B(\alpha/2;\ X,\ n-X+1),\ B(1-\alpha/2;\ X+1,\ n-X)].$$

It therefore follows immediately that CI_J is always contained in CI_{CP}. Thus CI_J corrects the conservativeness of CI_{CP}.

It turns out that the Jeffreys prior interval, although Bayesianly constructed, has a clear and convincing frequentist motivation. It is thus no surprise that it does well from a frequentist perspective. As we explain, the Jeffreys prior intervals CI_J can be regarded as a continuity corrected version of the Clopper-Pearson intervals CI_{CP}. The interval CI_{CP} inverts the inequality $P_p(X \le L(p)) \le \alpha/2$ to obtain the lower limit and similarly for the upper limit. Thus, for fixed x, the upper limit of the interval for p, $U_{CP}(x)$, satisfies

$$P_{U_{CP}(x)}(X \le x) \le \alpha/2, \tag{12}$$

and symmetrically for the lower limit.

This interval is very conservative; undesirably so for most practical purposes. A familiar proposal to eliminate this over-conservativeness is to instead invert

$$P_p(X \le L(p)-1) + (1/2)P_p(X = L(p)) = \alpha/2, \tag{13}$$

This amounts to solving

$$(1/2)\{P_{U_{CP}(x)}(X \le x-1) + P_{U_{CP}(x)}(X \le x)\} = \alpha/2, \tag{14}$$

which is the same as

$$U_{midP}(X) = (1/2)B(1-\alpha/2; x, n-x+1) + (1/2)B(1-\alpha/2; x+1, n-x), \tag{15}$$

and symmetrically for the lower endpoint. These are the "Mid-P Clopper-Pearson" intervals. They are known to have good coverage and length performance. U_{midP} given in (15) is a weighted average of two incomplete Beta functions. The incomplete Beta function, $B(1-\alpha/2; x, n-x+1)$, is continuous and monotone in x if we formally treat x as a continuous argument. Hence the average of the two functions defining U_{midP} is approximately the same as the value at the halfway point, $x+1/2$. Thus

$$U_{midP}(X) \approx B(1-\alpha/2; x+1/2, n-x+1/2) = U_J(x).$$

This is exactly the upper limit for the equal-tailed Jeffrey interval. Similarly, the corresponding approximate lower endpoint is the Jeffreys' lower limit.

Another way to frequentistly interpret the Jeffreys prior interval is to say that $U_J(x)$ is the upper limit for the Clopper-Pearson rule with $x-1/2$ successes and $L_J(x)$ is the lower limit for the Clopper-Pearson rule with $x+1/2$ successes.

3.10 Performance of the Intervals

The performance of the recentered interval is already significantly better than that of the standard interval for p away from the boundaries. However, the coverage probability is still unsatisfactory when p is close to 0 or 1, because it does not adjust the estimate of the standard error.

Average coverage of the Wilson interval CI_W is very close to the nominal level $1-\alpha$. See Figure 7. Coverage fluctuates acceptably near $1-\alpha$, except for p very near 0 or 1. It can be shown that, when $1-\alpha = .95$ and $1/n \leq p \leq 1-1/n$,

$$.92 \leq \lim_{n\to\infty} P_p(p \in CI_W) \leq .98.$$

The modification CI_{M-W} removes the first few deep downward spikes of the coverage function for CI_W. The resulting coverage function is overall somewhat conservative for p very near 0 or 1. Both CI_W and CI_{M-W} have the same coverage functions away from 0 or 1. Heuristically, CI_W nearly minimizes average length of intervals (among procedures with coverage "approximately" $\geq 1-\alpha$). This minimization property can be precisely formulated and proved using a first order Edgeworth expansion and the main result in Brown, Casella & Hwang (1994). See Brown, Cai & DasGupta (1999) for further details.

The Agresti-Coull interval has a very simple form and is easy to remember - "Add $\kappa^2/2$ successes and $\kappa^2/2$ failures; then use the standard method". The interval also has good minimum coverage probability. The coverage probability of the interval is quite conservative for p very close to 0 or 1. In comparison to the Wilson interval it is more conservative, especially for small n. This, by the way, is not surprising because it is easy to show that CI_{AC} always contains CI_W as a proper subinterval.

The arcsine interval performs well for p not too close to the boundaries. The coverage also has downward spikes near the two edges (see Figure 6). Modifications can also be made to correct the problem. For example, by setting the lower limit to 0 when $x = 0$ and the upper limit to 1 when $x = n$, the sharp downward spikes in the two boundaries can be eliminated. This simple modification avoids the awkward result that the coverage probability $C(p,\ n) \to 0$ as $p \to 0$ or 1. We also note that our evaluations show that the performance of the arcsine interval with the standard $\hat{p}$ in place of $\check{p}$ in (7) is much worse than that of CI_{Arc}.

The Clopper-Pearson interval guarantees that the actual coverage probability is above the nominal confidence level. However, for any fixed p, the actual coverage probability can be much larger than $1-\alpha$ unless n is quite large, and thus the confidence interval is rather inaccurate in this sense. See Figure 6. See also Figure 7 for its average coverage probability. The Clopper-Pearson interval is wastefully conservative and is not a good choice for practical use, unless strict adherence to the prescription $C(p,\ n) \geq 1-\alpha$ is demanded.

The coverage of the Jeffreys interval is qualitatively similar to that of CI_W over most of the parameter space $[0,\ 1]$. See Figure 6. Correspondingly, the average coverage is amazingly close to $1-\alpha$. See Figure 7. In addition, as we just saw in Section 3.9, CI_J has an appealing connection to the mid-P corrected version of the Clopper-Pearson "exact" intervals. These are very similar to CI_J, over most of the range, and have similar appealing properties. CI_J is a serious and credible candidate for practical use. The coverage has an unfortunate fairly deep spike near $p = 0$ and, symmetrically, another near $p = 1$. However, our simple modification of CI_J removes these two deep downward spikes. The modified Jeffreys interval CI_{M-J} performs well.

The specific choices of the values of n and p in the examples and figures are artifacts. The theoretical results in Brown, Cai & DasGupta (1999) show that the same comparative phenomena as regards coverage and length hold for general n and p. (Those results are asymptotic as $n \to \infty$, but we argue they are also sufficiently accurate for realistically moderate n.)

4 Choosing An Interval

We compare the performance of the various intervals in terms of the average coverage probability, the mean absolute error of the coverage, and the expected length of the interval.

Figure 7 demonstrates the striking difference in average coverage probability among five intervals: the Clopper-Pearson interval, the Agresti-Coull interval, the Wilson interval, the Jeffreys prior interval, and the standard interval. The Clopper-Pearson interval is very conservative and the convergence to the nominal confidence level is extremely slow. The standard interval performs poorly. The interval CI_{AC} is slightly conservative in terms of average coverage probability, but the convergence is much faster than the Clopper-Pearson interval. Both the Wilson interval and the Jeffreys prior interval have excellent performance in terms of the average coverage probability; that of the Jeffreys prior interval is, if anything, slightly superior. The average coverage of the Jeffreys interval is really very close to the nominal level even for quite small n.

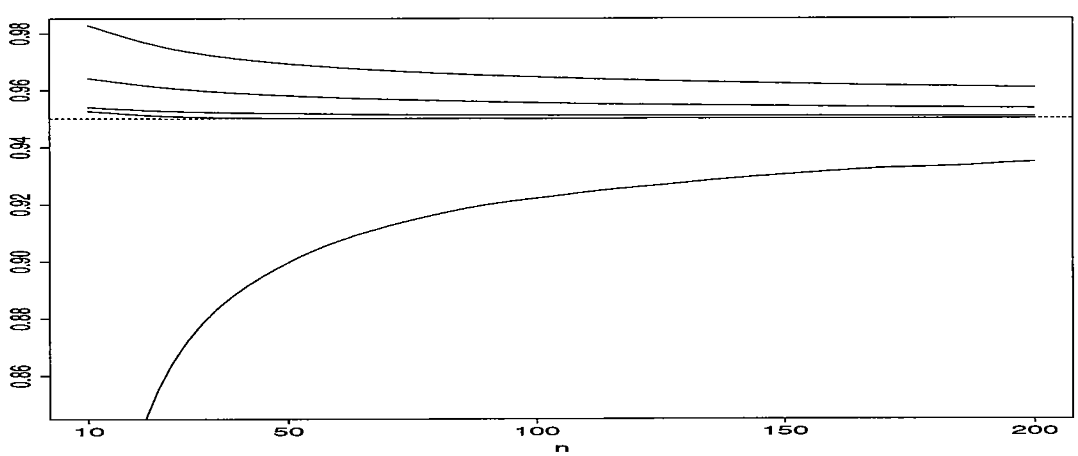

Figure 7: Comparison of the average coverage probabilities. From top to bottom: the Clopper-Pearson interval CI_{CP}, the Agresti-Coull interval CI_{AC}, the Wilson interval CI_W, the Jeffreys prior interval CI_J, and the standard interval CI_s. The nominal confidence level is .95.

Ideally, it would be nice to compare the intervals for individual n up to some limit. But actually it would only add to the confusion. A summary is often more instructive than excessive detail. We group the values of n into four intervals: $10 \le n \le 25$, $26 \le n \le 40$, $41 \le n \le 60$, and $61 \le n \le 100$. Admittedly, the grouping is somewhat subjective. But it has to be and these intervals of n seem to be reasonable.

Two additional criteria we use for comparison of the alternative intervals are:

$$\text{Mean Absolute Error} = \int_0^1 |C(p,\, n) - (1-\alpha)|\; dp,$$

and

$$\begin{aligned}\text{Average Expected Length} &= \int_0^1 E_{n,p}(\text{length}(CI))\; dp \\ &= \int_0^1 \sum_{x=0}^{n} (U(x,\, n) - L(x,\, n)) \binom{n}{x} p^x (1-p)^{n-x}\; dp,\end{aligned}$$

where U and L are the upper and lower limits of the confidence interval CI, respectively.

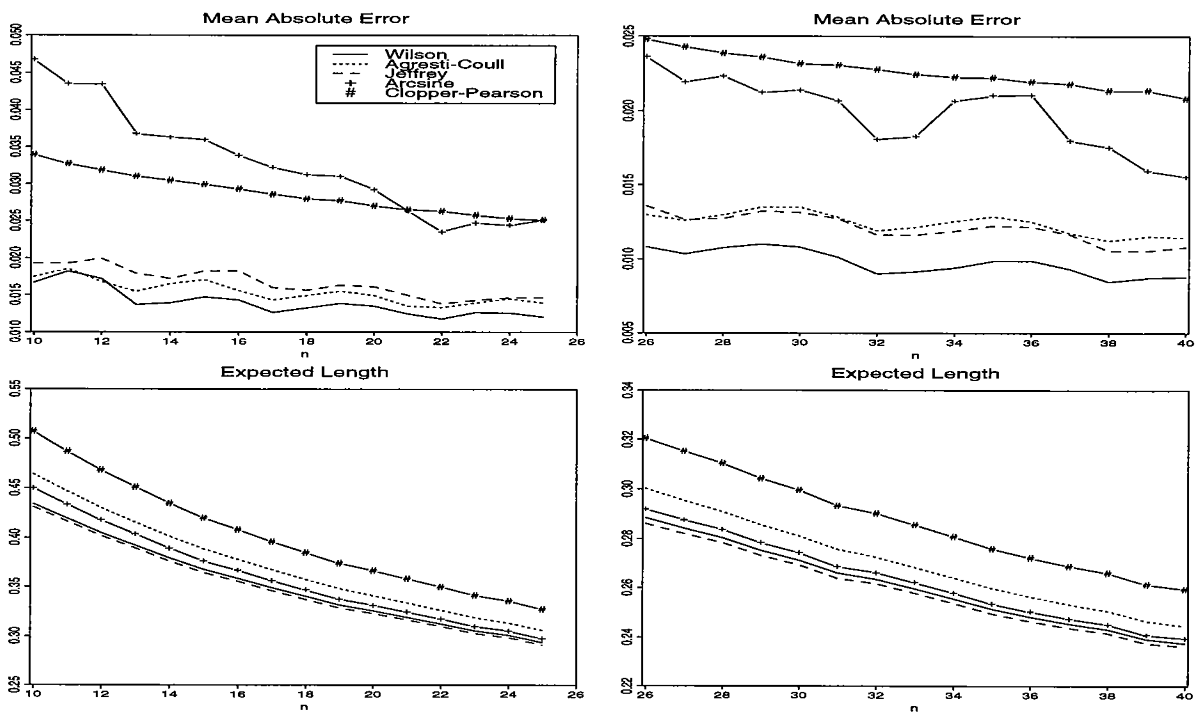

Figure 8: The mean absolute errors of the coverage and the expected lengths for n = 10 to 25, and n = 26 to 40.

The top panel of Figure 8 displays the mean absolute errors for n = 10 to 25, and n = 26 to 40. It is clear from the plots that among the five intervals, CI_W, CI_{AC} and CI_J are comparable, but the mean absolute errors of CI_{CP} and CI_{Arc} are significantly larger. The bottom panel of Figure 8 shows the average expected lengths of the five intervals for n = 10 to 25, and n = 26 to 40. Interestingly, the comparison is clear and consistent as n changes. Always, the Jeffreys interval CI_J and the Wilson interval CI_W are comparable, and CI_J is just slightly more parsimonious. But the difference is not of practical relevance. However, especially when n is small, the average expected length of CI_{AC} is noticeably larger than that of CI_J and CI_W. In fact, for n till about 20, the average expected length of CI_{AC} is larger than that of CI_J by .04 to .02, and this difference can be of definite practical relevance. The difference starts to wear off when n is larger than 30 or so. The Clopper-Pearson interval

CI_{CP} is significantly longer than the other intervals. Based on these plots, we recommend the Wilson interval or the Jeffreys interval for small n ($n \leq 40$). These two intervals are comparable in both absolute error and length for $n \leq 40$, and we believe that either could be used, depending on taste.

For larger n, the Wilson, Jeffreys and the Agresti-Coull intervals are all comparable, and the Agresti-Coull interval is the simplest to present. It is generally true in statistical practice that only those methods that are easy to describe, remember and compute are widely used. Keeping this in mind, we recommend the Agresti-Coull interval for practical use when $n \geq 40$.

5 Concluding Remarks

Interval estimation of a binomial proportion is a very basic problem in practical statistics. The standard Wald interval is in nearly universal use. We first show that the performance of this standard interval is persistently chaotic and unacceptably poor. Indeed its coverage properties defy all conventional wisdom, much more than is presently widely understood. The performance is so erratic and the qualifications given in the influential texts are so defective, that the standard interval should not be used. We provide a fairly comprehensive evaluation of many natural alternative intervals. Based on this analysis, we recommend the Wilson or the equal-tailed Jeffrey prior interval for small n ($n \leq 40$), and the Agresti-Coull interval for $n \geq 40$. Even for small sample sizes the easy to present Agresti-Coull interval is much preferable to the standard one.

We would be satisfied if this article contributes to a greater appreciation of the severe flaws of the popular standard interval and an agreement that it deserves not to be used at all. We also hope that the recommendations as regards alternative intervals will provide constructive suggestions as to what may be used in preference to the standard method.

Acknowledgments

We wish to thank Xuefeng Li for performing some helpful computations and Jim Berger and Ron Thisted for urging us to consider the Jeffreys and the arcsine intervals.

References

[1] Abramowitz, M. & Stegun, I.A. (1970). *Handbook of Mathematical Functions.* Dover, New York.

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

[3] Anscombe, F.J. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika* **35**, 246-254.

[4] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition. Springer-Verlag, New York.

[5] Bickel, P. & Doksum, K. (1977). *Mathematical Statistics.* Prentice-Hall, Englewood Cliffs, NJ.

[6] Blyth, C.R. & Still, H.A. (1983). Binomial confidence intervals. *J. Amer. Stat. Assoc.* **78**, 108-116.

[7] Blyth, C.R. (1986). Approximate binomial confidence limits. *J. Amer. Stat. Assoc.* **81**, 843-55.

[8] Brown, L.D., Cai, T. & DasGupta, A. (1999). Confidence intervals for a binomial proportion and Edgeworth expansions. Technical Report.

[9] Brown, L.D., Casella, G. & Hwang, J.T.G. (1995). Optimal confidence sets, bioequivalence, and the limacon of Pascal. *J. Amer. Stat. Assoc.* **90**, 880-889.

[10] Casella, G., Hwang, J.T.G. & Robert, C.P. (1994). Loss functions for set estimation. In *Statistical Decision Theory and Related Topics V,* S.S. Gupta and J. Berger (Eds.). Academic Press, New York, 237-251.

[11] Clopper, C.J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-13.

[12] Clunies-Ross, C.W. (1958). Interval estimation for the parameter of a binomial distribution. *Biometrika* **45**, 275-9.

[13] Cox, D. R. (1970). The continuity correction. *Biometrika* **57**, 217-9.

[14] Cressie, N. (1980). A finely tuned continuity correction. *Ann. Inst. Statist. Math.* **30**, 435-42.

[15] Crow, E.L. (1956). Confidence intervals for a proportion. *Biometrika* **43**, 423-35.

[16] Ghosh, B.K. (1979). A comparison of some approximate confidence intervals for the binomial parameter *J. Amer. Stat. Assoc.* **74**, 894-900.

[17] Hall, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika* **69**, 647-52.

[18] Hall, P. (1992). *The Bootstrap and Edgeworth expansion.* Springer-Verlag, New York.

[19] Lehmann, E.L. (1999). *Elements of Large-Sample Theory.* Springer, New York.

[20] Robert, C.P. (1993). *The Bayesian Choice.* Springer-Verlag, New York.

[21] Samuels, M.L. & Witmer, J.W. (1999). *Statistics for the Life Sciences*, 2nd Edition. Prentice Hall.

[22] Santner, T.J. (1998). A note on teaching binomial confidence intervals. *Teaching Statistics* **20**, 20-23.

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

[24] Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *J. Amer. Stat. Assoc.* **22**, 209-212.

x	$n=7$		$n=8$		$n=9$		$n=10$		$n=11$		$n=12$	
0	0	.292	0	.262	0	.238	0	.217	0	.200	0	.185
1	.016	.501	.014	.454	.012	.414	.011	.381	.010	.353	.009	.328
2	.065	.648	.056	.592	.049	.544	.044	.503	.040	.467	.036	.436
3	.139	.766	.119	.705	.104	.652	.093	.606	.084	.565	.076	.529
4	.234	.861	.199	.801	.173	.746	.153	.696	.137	.652	.124	.612
5					.254	.827	.224	.776	.200	.730	.180	.688
6									.270	.800	.243	.757

x	$n=13$		$n=14$		$n=15$		$n=16$		$n=17$		$n=18$	
0	0	.173	0	.162	0	.152	0	.143	0	.136	0	.129
1	.008	.307	.008	.288	.007	.272	.007	.257	.006	.244	.006	.232
2	.033	.409	.031	.385	.029	.363	.027	.344	.025	.327	.024	.311
3	.070	.497	.064	.469	.060	.444	.056	.421	.052	.400	.049	.381
4	.114	.577	.105	.545	.097	.517	.091	.491	.085	.467	.080	.446
5	.165	.650	.152	.616	.140	.584	.131	.556	.122	.530	.115	.506
6	.221	.717	.203	.681	.188	.647	.174	.617	.163	.589	.153	.563
7	.283	.779	.259	.741	.239	.706	.222	.674	.207	.644	.194	.617
8					.294	.761	.272	.728	.254	.697	.237	.668
9									.303	.746	.284	.716

x	$n=19$		$n=20$		$n=21$		$n=22$		$n=23$		$n=24$	
0	0	.122	0	.117	0	.112	0	.107	0	.102	0	.098
1	.006	.221	.005	.211	.005	.202	.005	.193	.005	.186	.004	.179
2	.022	.297	.021	.284	.020	.272	.019	.261	.018	.251	.018	.241
3	.047	.364	.044	.349	.042	.334	.040	.321	.038	.309	.036	.297
4	.076	.426	.072	.408	.068	.392	.065	.376	.062	.362	.059	.349
5	.108	.484	.102	.464	.097	.446	.092	.429	.088	.413	.084	.398
6	.144	.539	.136	.517	.129	.497	.123	.478	.117	.461	.112	.444
7	.182	.591	.172	.568	.163	.546	.155	.526	.148	.507	.141	.489
8	.223	.641	.211	.616	.199	.593	.189	.571	.180	.551	.172	.532
9	.266	.688	.251	.662	.237	.638	.225	.615	.214	.594	.204	.574
10	.312	.734	.293	.707	.277	.681	.263	.657	.250	.635	.238	.614
11					.319	.723	.302	.698	.287	.675	.273	.653
12									.325	.713	.310	.690

x	$n=25$		$n=26$		$n=27$		$n=28$		$n=29$		$n=30$	
0	0	.095	0	.091	0	.088	0	.085	0	.082	0	.080
1	.004	.172	.004	.166	.004	.160	.004	.155	.004	.150	.004	.145
2	.017	.233	.016	.225	.016	.217	.015	.210	.015	.203	.014	.197
3	.035	.287	.034	.277	.032	.268	.031	.259	.030	.251	.029	.243
4	.056	.337	.054	.325	.052	.315	.050	.305	.048	.295	.047	.286
5	.081	.384	.077	.371	.074	.359	.072	.348	.069	.337	.067	.327
6	.107	.429	.102	.415	.098	.402	.095	.389	.091	.378	.088	.367
7	.135	.473	.129	.457	.124	.443	.119	.429	.115	.416	.111	.404
8	.164	.515	.158	.498	.151	.482	.145	.468	.140	.454	.135	.441
9	.195	.555	.187	.537	.180	.521	.172	.505	.166	.490	.160	.476
10	.228	.594	.218	.576	.209	.558	.201	.542	.193	.526	.186	.511
11	.261	.632	.250	.613	.239	.594	.230	.577	.221	.560	.213	.545
12	.295	.669	.282	.649	.271	.630	.260	.611	.250	.594	.240	.578
13	.331	.705	.316	.684	.303	.664	.291	.645	.279	.627	.269	.610
14					.336	.697	.322	.678	.310	.659	.298	.641
15									.341	.690	.328	.672

Table 5: 95% Limits of the Jeffreys prior interval.