CONFIDENCE INTERVALS FOR A BINOMIAL PROPORTION AND EDGEWORTH EXPANSIONS*

by

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

Technical Report #99-18

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

August 1999

⊛

Confidence Intervals for a Binomial Proportion And Edgeworth Expansions*

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

Abstract

We address the classic problem of interval estimation of a binomial proportion. The Wald interval $\hat{p} \pm z_{\alpha/2}\, n^{-1/2}(\hat{p}(1-\hat{p}))^{1/2}$ is currently in near universal use. We first show that the coverage properties of the Wald interval are persistently poor and defy virtually all conventional wisdom. We then proceed to a theoretical comparison of the standard interval and four additional alternative intervals by asymptotic expansions of their coverage probabilities and expected lengths. Fortunately, the asymptotic expansions are remarkably accurate at rather modest sample sizes, such as $n = 40$, or sometimes even $n = 20$.

The expansions show that an interval suggested in Agresti and Coull (1998) dominates the score interval (Wilson (1927)), the Jeffreys prior Bayesian interval, and also the standard interval in coverage probability. However, the asymptotic expansions for expected lengths show that the Agresti-Coull interval is always the longest of these, and the Jeffreys prior interval is always the shortest among these. The standard interval and the Wilson interval, curiously, have identical second order expansions for their average expected length and are in between the Jeffreys and the Agresti-Coull interval in the ranking for length.

These analytical calculations support and complement the findings and the recommendations in Brown, Cai and DasGupta (1999).

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

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

1 Introduction

In this article, we consider a very basic but very important problem of statistical practice, namely, interval estimation of the probability of success in a binomial distribution. There is a interval in virtually universal use. Of course, this is the Wald interval $\hat{p} \pm \kappa\, n^{-1/2}(\hat{p}(1-\hat{p}))^{1/2}$, where $\hat{p} = X/n$ is the sample proportion of successes, and κ is the $100(1-\alpha/2)$th percentile of the standard normal distribution.

Certainly the problem has a lot of literature, and the questionable performance of the standard Wald interval has been sporadically remarked on. Simultaneously, there has also been work on suggesting alternative confidence intervals. For example, alternative intervals have been suggested that use a continuity correction, or intervals that actually guarantee a minimum $1 - \alpha$ coverage probability for all values of the parameter p. In spite of all that literature, there is still a widespread misconception that the problems of the Wald interval are serious only when p is near 0 or 1, or when the sample size n is not that large. The influential texts in statistics provide the perfect testimonial to this misconception. Nearly universally, they recommend the Wald interval when npq is larger than 5 or 10. Inspired by two interesting articles, Santner (1998) and Agresti and Coull(1998), Brown, Cai and DasGupta (1999) recently showed that the performance of this standard interval is far more erratic and miserable than is appreciated. Virtually all of the conventional wisdom and popular prescriptions are false and misplaced. The Wald interval is so poor in this problem that it cannot be trusted.

Brown, Cai and DasGupta (1999) do a fairly comprehensive examination of eight natural alternative confidence intervals for p, and after extensive numerical analysis recommend the score interval of Wilson (Wilson (1927)) or the Jeffreys prior interval for small n, and an interval suggested in Agresti and Coull (1998) for larger n. The principal goal of this article is to present a set of theoretical calculations that reinforce those findings and recommendations with accuracy and consistency. We show that the coverage probability of the standard interval not only exhibits oscillation, but also has a pronounced systematic bias. We also show that the alternative intervals do better in these regards. These theoretical calculations hopefully enable us to get some closure on this obviously important problem. Previously, there had been a nagging suspicion that the Wald interval can be erratically poor, and there had been useful but isolated attempts at finding credible alternatives. Brown, Cai and DasGupta (1999) addressed the issue of credible alternatives in a more comprehensive way and made certain recommendations. By fully reinforcing those practical recommendations through analytical calculations, the present article gives credibility to the findings in that article.

In section 2, we give a few examples to illustrate the extent to which conventional wisdom fails in this problem. Additional examples may be seen in Brown, Cai and DasGupta (1999). In Section 3, first we introduce four alternative confidence intervals; we selected these four for a theoretical study based on our findings in the companion article. The rest of Section 3 deals with Edgeworth expansions for the coverage probabilities of the standard interval and the four alternative intervals. Due to the lattice nature of the Binomial distribution, the Edgeworth expansions here contain certain oscillation terms that typically do not arise for continuous populations. So at the very least, the Edgeworth expansions explain why the Wald interval exhibits such eccentric behavior. We then show that although one term

Edgeworth expansions do not approximate the coverage probabilities with adequate accuracy, the two term expansion provides truly good accuracy at modest sample sizes. The derivations of the two term Edgeworth expansions are somewhat technical, and especially so for the Bayesian intervals. They are derived separately in an appendix.

In Section 4, we use the two term Edgeworth expansions as an analytical tool to compare and rank the various intervals with regard to their coverage probabilities. The two term expansions show that the interval suggested in Agresti and Coull(1998) dominates the standard, the Wilson, and the Jeffreys prior interval in coverage probability. They also show that the Wilson and the Jeffreys prior interval are pretty consistently comparable. But the Jeffreys interval comes the closest to having second order accuracy. The two term expansions even show, on careful closer scrutiny, other subtle and interesting features of this problem. For instance, from these Edgeworth expansions one can see that even the choice of the level α can significantly affect the performance of the standard interval and the alternative intervals in this problem.

As in any interval estimation problem, coverage is only part of the assessment of a confidence interval. Parsimony, naturally measured by expected length, is another important criterion. In Section 5, we derive two term expansions for the expected lengths of the standard and the alternative confidence intervals. Fortunately, the coefficients in the second term are different for different intervals, giving us a basis for comparison of their expected lengths. We then also provide an integrated version of the expansions, the integration being over p on (0, 1). The results are quite interesting; from these expansions one sees that the Agresti-Coull interval is always the longest, and the Wilson and the standard interval have identical two term expansions, and the Jeffreys prior interval is always the shortest. The fact that these rankings are always the same makes the results more valuable as a guide to choosing a credible alternative confidence interval.

As we mentioned before, it is especially gratifying to see that these asymptotic expansions of both the coverage probabilities and the expected lengths reflect the reports in Brown, Cai and DasGupta (1999) with rather remarkable accuracy. Because of these theoretical calculations, we feel assured and comfortable in recommending strongly that the standard interval for this problem should not be used and the suggested alternatives in Brown, Cai and DasGupta (1999) are far better and safer to use.

2 Coverage Properties of the Standard Interval

Although the standard interval is in near universal use the following instructive examples will show that its coverage probabilities are unacceptably erratic and poor. These illustrative examples are given to show that there really is a serious problem here and it deserves to be fully understood by statisticians at large. Specifically, the poor coverage probability is not just for p near the boundaries, and the erratic behavior persists for large and even very large sample sizes. There is therefore a real need for a thorough investigation of alternative confidence intervals in this important problem. Additional examples may be seen in Brown, Cai and DasGupta (1999), Santner (1998), and Agresti and Coull (1998).

Example 1 Consider $p = .5$. Conventional wisdom might suggest that all will be well if n is above 20. Figure 1 plots the coverage probability of the nominal 95% standard interval with

$p = .5$ and $n = 10$ to 100. At $n = 97$, the coverage is still only about .933; in addition, the coverage probability does not at all get steadily closer to the nominal confidence level as n increases. At $n = 17$, the coverage probability is .951, but at the much larger value $n = 40$, the coverage is only .919. The oscillations in this case are caused solely by the discreteness of the distribution. A careful look at the coverage probability shows that it requires $n \geq 194$ to guarantee that the coverage probability stays at .94 or above.

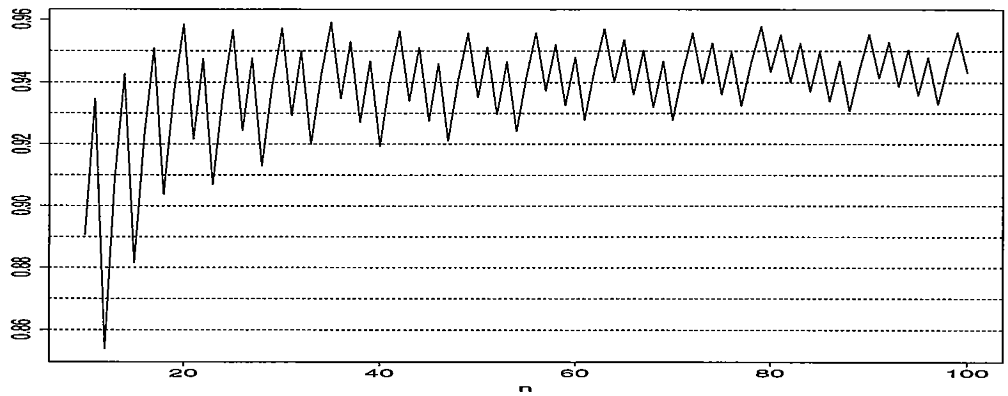

Figure 1: Coverage probability of the standard interval for $p = .5$ and $n = 10$ to 100.

p	.01	.025	.05	.1	.15	.2	.25	.3	.35	.4	.45	.5
n_s	7963	3048	1387	646	399	292	273	245	228	185	195	194
n_J	2906	1543	604	287	182	147	168	112	92	97	52	151
n_W	2605	681	491	231	271	131	112	101	55	77	50	105
n_{AC}	1	5	10	11	58	89	99	78	34	60	50	94

Table 1: Smallest n after which the coverage stays at .94 or above.

Table 1 lists the smallest n after which the coverage stays at .94 or above for selected values of p for the standard interval and three alternative intervals. n_s, n_J, n_W and n_{AC} denote the smallest n required for the standard interval, the equal-tailed Jeffreys prior interval, the Wilson interval, and the Agresti-Coull interval, respectively. See Section 3.2 for the definition of these alternative intervals. When p is small, it takes many thousands of observations for the nominal 95% standard interval to ensure that the coverage probability is at least .94. In certain practical applications, it is common to have a small p. For example, the defective proportions in industrial quality control problems are often very small. Table 1 shows that even if p is not small, the required sample sizes for the approximate validity of the standard interval are much larger than the usual recommendations in popular textbooks. In comparison, the alternative intervals do much better. From Table 1, one may think that the Agresti-Coull interval is the obvious interval of choice. We will see in Section 5 that it tends to be longer than the other intervals and so has a higher coverage probability.

Example 2 This example shows that the standard interval is not just eccentric, but can also be grossly inadequate. There is a systematic bias in the coverage probability of the standard interval. Figure 2 shows the coverage probability of the nominal 99% standard interval with $n = 30$. It is striking that in this case the coverage not only oscillates erratically, but is always smaller than .99. In fact on the average the coverage is only .914. Our evaluations show that for all n up to 45, the coverage of the 99% standard interval is always below the nominal level for all $0 < p < 1$, although certain values of p are of course luckier than others.

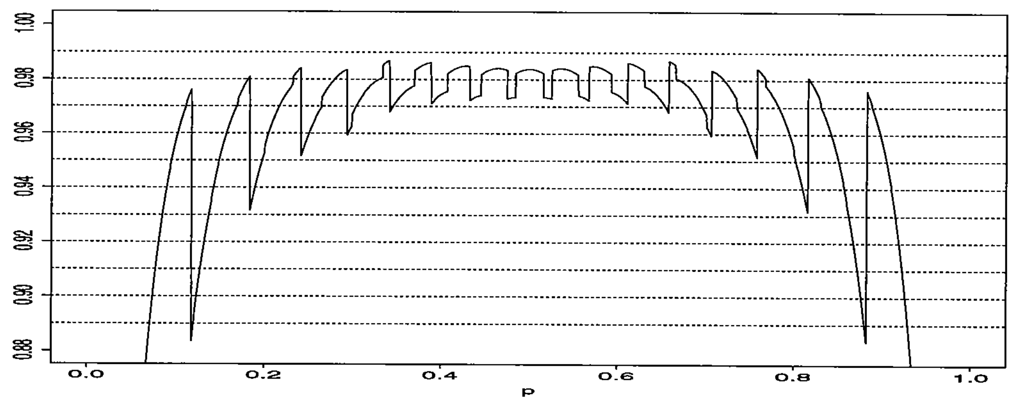

Figure 2: Coverage of the nominal 99% standard interval for $n = 30$ and $0 < p < 1$.

2.1 The Reason for the Bias

Example 2 indicated that there is a systematic negative bias in the coverage probability of the standard interval. The bias is due mainly to the fact that the standard interval has the "wrong" center. The standard interval is centered at $\hat{p} = X/n$. Although $\hat{p}$ is the MLE and an unbiased estimate of p, as the center of a confidence interval, it causes a systematic negative bias in the coverage. As we will see in Section 3.5, by simply recentering the interval at $\tilde{p} = (X + \kappa^2/2)/(n + \kappa^2)$, one can increase the coverage significantly for p away from 0 or 1 and eliminate the systematic bias.

The standard interval is based on the fact that

$$W_n \equiv \frac{n^{1/2}(\hat{p} - p)}{\sqrt{\hat{p}\hat{q}}} \stackrel{\mathcal{L}}{\Longrightarrow} N(0, 1).$$

However, even for quite large values of n, the actual distribution of W_n is significantly nonnormal. Thus the very premise on which the standard interval is based is seriously flawed for moderate and even quite large values of n. For instance, asymptotically, W_n has bias 0, variance 1, skewness 0, and kurtosis 3. For moderate n, however, the deviations of the bias, variance, skewness and kurtosis of W_n from their respective asymptotic values are often significant and cause a non-negligible negative bias in the coverage probability of the standard confidence interval. Figure 3 plots the very noticeable bias in the distribution of W_n (conditional on $\hat{p} \neq 0$ or 1) for $n = 20$ to 200 and fixed $p = .25$.

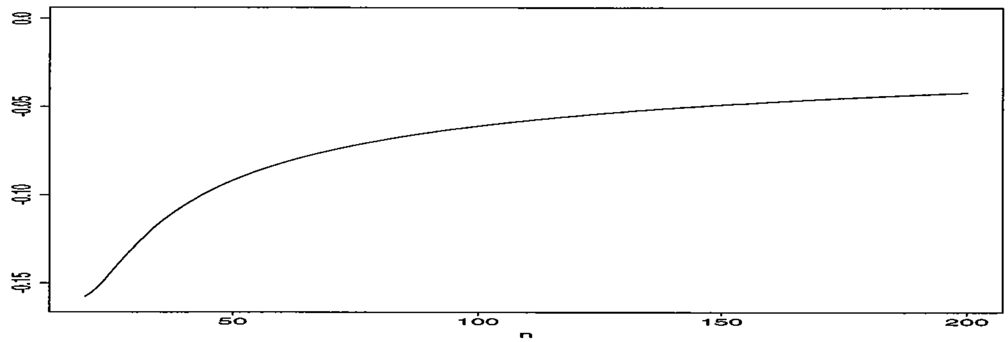

Figure 3: Bias in the distribution of W_n with $p = .25$. Vertical axis is $E(W_n)$.

We can analytically demonstrate the bias in the distribution of W_n by standard expansions. Denote $Z_n = n^{1/2}(\hat{p} - p)/\sqrt{pq}$. Then simple algebra yields

$$W_n \equiv \lambda(Z_n) = \frac{Z_n}{\sqrt{1 + (1-2p)Z_n/\sqrt{npq} - Z_n^2/n}}.$$

A standard Taylor expansion and formulas for central moments of the binomial distribution then yield an approximation to the bias:

$$EW_n = E\lambda(Z_n) = \frac{p - 1/2}{\sqrt{npq}}(1 + \frac{7}{2n} + \frac{9(p-1/2)^2}{2npq}) + o(n^{-3/2}). \qquad (1)$$

It can be seen from (1) that W_n has negative bias for $p < .5$ and positive bias for $p > .5$. Therefore, ignoring the oscillation effect, one can expect to increase the coverage probability by shifting the center of the standard interval towards 1/2. This observation is confirmed in Section 3.5.

Besides the bias, the variance, skewness and kurtosis of W_n often deviate significantly from their respective asymptotic values. See Table 2 below; especially note the high kurtosis values.

n	20	30	40	50	60	70	80	90	100	150	200
Variance	1.36	1.28	1.19	1.14	1.11	1.09	1.08	1.07	1.06	1.04	1.03
Skewness	-0.78	-0.80	-0.61	-0.48	-0.40	-0.35	-0.32	-0.29	-0.27	-0.21	-0.18
Kurtosis	4.41	5.28	4.66	4.03	3.70	3.53	3.43	3.36	3.31	3.19	3.13

Table 2: Variance, skewness and kurtosis of W_n for $p = .25$.

2.2 The Reason for the Oscillation

It is evident from Examples 1 and 2 that the actual coverage probability of the standard interval for p can differ significantly from the nominal confidence level at realistic and even larger than realistic sample sizes. The error, of course, comes from two sources: discreteness and skewness in the underlying distribution. For a two-sided interval, the rounding error due to discreteness is asymptotically dominant. It is of the order $n^{-1/2}$. And the error due to skewness is secondary and is of the order n^{-1}, but still important for even moderately large n. Note that the situation is different for one-sided intervals. There, the error caused by 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 cumulative distribution function contains jumps at integer points and the Edgeworth expansions for the distribution function contain terms that do not appear, typically, in the continuous case (for example, under the Cramer conditions; see Esseen(1945)).

Let us try to understand at a more intuitive level why the coverage probability oscillates so significantly. By a straightforward calculation, one can show that the coverage probability $P_{n,p}(p \in CI_s)$ equals $P_{n,p}(\ell_{n,p} \leq X \leq u_{n,p})$, where $\ell_{n,p}$ is the smallest integer larger than or equal to

$$\frac{n(\kappa^2 + 2np) - \kappa n\sqrt{\kappa^2 + 4npq}}{2(\kappa^2 + n)},$$

and $u_{n,p}$ is the largest integer smaller than or equal to

$$\frac{n(\kappa^2 + 2np) + \kappa n\sqrt{\kappa^2 + 4npq}}{2(\kappa^2 + n)}.$$

What happens is that a small change in n or p can cause $\ell_{n,p}$ and/or $u_{n,p}$ to leap to the next integer value. For example, take the case $p = .5$ and $\alpha = .05$. When $n = 39$, $\ell_{n,p} = 14$ and $u_{n,p} = 25$; but when $n = 40$, $\ell_{n,p}$ leaps to 15 while $u_{n,p}$ remains 25. Thus the set of favorable values of X loses the point $X = 14$ even though n has increased from 39 to 40. This causes $n = 40$ to be an unlucky choice of n. This also happens when n is kept fixed and p changes slightly, and we then begin to see unlucky values of p.

3 Alternative Intervals and Edgeworth Expansions

The preceding discussion demonstrates that the coverage of the standard confidence interval is undesirably unpredictable and poor. Brown, Cai and DasGupta (1999) provides much additional evidence. Due to the obvious methodological importance of the problem, then, we face the undeniable need for alternative intervals. Such alternative intervals would have to be demonstrably better. In addition, it would be desirable to be able to recommend one or two specific alternative intervals for practical use. The theoretical calculations in the rest of this paper address these two important goals.

Three things are of importance here. First, there will have to be an evaluation of the coverage probability of any suggested alternative interval. Second, the intervals have to be assessed for parsimony in terms of their length. And, third, we have to keep in mind the formal simplicity of any recommended alternative interval. Simplicity may well be a dominant factor here because the problem is a basic one and a computationally clumsy procedure is not likely to survive the test of time in such a basic problem.

The theoretical calculations will get rather technical. So it may be helpful to have a preview of what the calculations are, how the calculations are useful, what are the main conclusions from these calculations, and how we process the conclusions to make practical recommendations. For clarity and coherence, we next provide such a preview. A satisfying and salient feature of the theoretical calculations is that all the numerical evidence and conclusions presented in the companion paper Brown, Cai and DasGupta(1999) manifest themselves in these calculations.

3.1 Preview

In Brown, Cai and DasGupta (1999), a number of alternative confidence intervals for a binomial proportion are presented. First, we will present a subset of those intervals with a brief motivation. The coverage properties of these intervals will then be studied by deriving the corresponding Edgeworth expansions of their coverage probabilities. We will see that one term expansions, although simple, are not adequately accurate to address the problem on a serious basis. Therefore we will be compelled to proceed to two term expansions. The two term expansions, rather surprisingly, will be remarkably accurate even for modest sample sizes. Furthermore, comparative examination of the two term Edgeworth expansions will provide a lot of useful information about the alternative intervals. For example, we can see from the two term expansions why the standard interval is so bad and how the alternates compare among themselves. We will also see in the two term expansions some subtle features of the problem itself, e.g., how the choice of α can affect the performance of the confidence intervals.

Next, parsimony of the alternative intervals will be studied by an appropriate expansion of their expected lengths. These are also two term expansions. Moreover, just like the Edgeworth expansions of the coverage probabilities, the expansions for expected length are remarkably accurate at moderate sample sizes, and are directly useful to rank the intervals in terms of parsimony. Together, the Edgeworth expansions for the coverage probabilities and the two term expansions for the expected lengths give us the tools to make an overall comparative assessment of the suggested alternative intervals.

One final point is to be mentioned here. Among the alternative intervals considered are the Bayesian intervals resulting from Beta priors. Ample evidence will be presented that, particularly, the Jeffreys prior interval has excellent coverage and length properties. However, the actual derivation of the Edgeworth expansion is significantly more complex for the Bayesian intervals than for the other alternative intervals. We have presented all of these derivations in an appendix.

We will now present our alternative intervals.

3.2 Alternative Intervals

Specifically, besides the standard interval, we will consider the following intervals.

1. *The recentered interval*: The performance of the standard Wald interval can be much improved by simply moving the center of the interval towards 1/2 to

$$\tilde{p} = (X + \kappa^2/2)/(n + \kappa^2).$$

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 and Coull (1998). The recentered interval has the form

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

2. *The Wilson interval*: This interval is formed by inverting the CLT approximation to the family of equal-tailed 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. The Wilson interval has the form

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

3. *The Agresti-Coull interval*: This interval has the same simple form as the standard interval CI_s, but with a different $\hat{p}$ and a modified value for n. Denote $\tilde{X} = X + \kappa^2/2$ and $\tilde{n} = n + \kappa^2$. Let $\tilde{p} = \tilde{X}/\tilde{n}$ and $\tilde{q} = 1 - \tilde{p}$. The interval is defined as

$$CI_{AC} = \tilde{p} \pm \kappa(\tilde{p}\tilde{q})^{1/2}\tilde{n}^{-1/2}. \quad (3)$$

Again, 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 and Coull (1998). For this reason, we will call it the Agresti-Coull interval.

4. *The equal-tailed Jeffreys interval*: Historically, Bayes procedures under noninformative priors have a track record of good frequentist properties. See, for example, Wasserman (1991). In this problem the Jeffreys prior is $Beta(1/2, 1/2)$; see Berger (1985). The $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval is given by

$$CI_J = [B_{\alpha/2, X+1/2,\, n-X+1/2},\ B_{1-\alpha/2, X+1/2,\, n-X+1/2}], \quad (4)$$

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

3.3 One Term Edgeworth Expansion

Edgeworth expansions are a popular tool for studying complicated probabilistic quantities. See Bhattacharya and Rao(1976), Barndorff-Nielsen and Cox (1989) and Hall (1992) for more details on Edgeworth expansions.

Denote by CI a generic confidence interval for p. The coverage probability of CI is defined as

$$C(p,\ n) \equiv P_p(p \in CI) = \sum_{x=0}^{n} I(p, x) \binom{n}{x} p^x (1-p)^{n-x},$$

where $I(p,\ x)$ is the indicator function that equals to 1 if the interval contains p when $X = x$ and equals 0 if it does not contain p.

Define

$$h(x) = x - x_{-} \quad (5)$$

where x_- is the largest integer less than or equal to x. So $h(x)$ is just the fractional part of x. The function h is a periodic function of period 1. Let

$$g(p, z) = g(p, z, n) = h(np + z(npq)^{1/2}) \tag{6}$$

(we suppress in (6) and later the dependence of g on n). Theorem 23.1 in Bhattacharya and Rao(1976) yields that

$$P(\frac{n^{1/2}(\hat{p} - p)}{(pq)^{1/2}} \leq z) = \Phi(z) + [(\frac{1}{2} - g(p,\ z)) + \frac{1}{6}(1 - 2p)(1 - z^2)]\phi(z)(npq)^{-1/2} + O(n^{-1}) \tag{7}$$

where $(1/2 - g(p,\ z))$ takes values in $[-1/2, 1/2]$ and represents the rounding error, and $(1/6)(1 - 2p)(1 - z^2)$ represents the skewness error. For the two-sided confidence intervals under consideration, the rounding error is dominant and the skewness error is reduced to $O(n^{-1})$, as we shall see in (8) below.

From (7) we have a one-term Edgeworth approximation of the coverage probability of the confidence interval CI_s. Let ℓ_s and u_s be defined as functions of p (and n and κ) by

$$\{p \in CI_s\} \equiv \{\ell_s \leq \frac{n^{1/2}(\hat{p} - p)}{(pq)^{1/2}} \leq u_s\}.$$

See (38) in the Appendix for the exact expressions for ℓ_s and u_s. Correspondingly, the bounds ℓ_{AC}, u_{AC}, ℓ_J, and u_J in Section 3.5 are defined similarly.

Suppose $np + \ell_s(npq)^{1/2}$ is not an integer; then the coverage probability of CI_s satisfies

$$P_p(p \in CI_s) = (1 - \alpha) + [g(p,\ \ell_s) - g(p,\ u_s)]\phi(\kappa)(npq)^{-1/2} + O(n^{-1}) \tag{8}$$

The second term in (8), due to rounding error, is the principal contributor to the oscillation phenomenon. The oscillation term is of the order of $n^{-1/2}$. Since $|g(p,\ \ell_s) - g(p,\ u_s)| \leq 1$, this term is bounded by $\phi(\kappa)(npq)^{-1/2}$. Although the $O(n^{-1/2})$ oscillation term can be calculated precisely when p is known, it is clear from the expressions of g, ℓ_s and u_s, the oscillation term is unpredictable when p is unknown. This $O(n^{-1/2})$ term can be significant even for large n, especially when p is close to 0 or 1.

Remark: In the case that $np + \ell_s(npq)^{1/2}$ is an integer, then one needs to add an additional term $P_p(X = np + \ell_s(npq)^{1/2}) = \phi(\kappa)(npq)^{-1/2} + O(n^{-1})$ to (8) and gets

$$P_p(p \in CI_s) = (1 - \alpha) + [g(p,\ \ell_s) - g(p,\ u_s) + 1]\phi(\kappa)(npq)^{-1/2} + O(n^{-1}). \tag{9}$$

The same applies to the two-term expansion of the coverage probability of various confidence intervals discussed in Section 3.5.

Here we would like to point out that there is an error in Ghosh(1979) (Theorem 1, pp. 895). The oscillation terms were mistakenly omitted in the expansion. This affects one statement (Ghosh(1979), pp. 895) made in the paper. Because of this $O(n^{-1/2})$ oscillation term, for any p and α, it is in fact not true that for sufficiently large n, $C(p, n)$ will always exceed $1 - \alpha$ up to the order $n^{-1/2}$. So when p is unknown, there is no guarantee that the coverage probability of the standard interval is larger than the nominal level up to the order $n^{-1/2}$, no matter how large n is.

3.4 One Term Expansion Is Not Accurate Enough

The one-term Edgeworth expansion offers an approximation of the coverage probability and is useful for finding the source of the oscillation. The approximation error of a one-term Edgeworth expansion is $O(n^{-1})$. In Figure 4, we plot the actual coverage probability of the standard interval and the one-term Edgeworth approximation for fixed $n = 100$ and variable p from .05 to .95. And in Table 3, we compare numerically the coverage probability of the standard interval with the one-term Edgeworth approximation for fixed $p = .2$ and some selected values of n from 20 to 200. It is clear that the one-term Edgeworth expansion captures most of the oscillation effect in the true coverage probability. However, it contains a systematic bias. The reason is that the next term in the Edgeworth expansion , which is of the order n^{-1}, is mostly non-oscillating and negative. This can be easily seen from (13) in the next section.

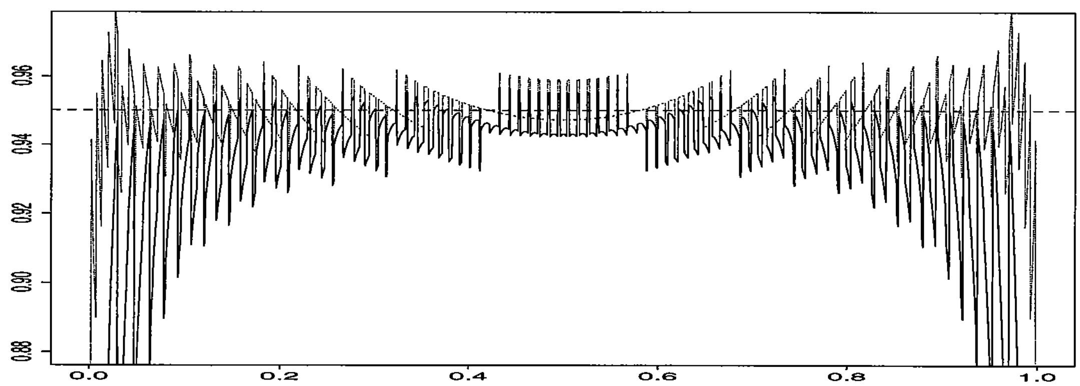

Figure 4: Comparison between the actual coverage probability (solid) and one-term Edgeworth expansion (dotted) with $n = 100$ and $1 - \alpha = .95$.

n	20	30	40	50	60	70	80	90	100	150	200
$C(p,n)$	.921	.946	.905	.938	.922	.940	.932	.947	.933	.944	.941
$e_1(p,n)$	.960	.968	.934	.952	.951	.951	.952	.954	.942	.949	.949
difference	.039	.021	.029	.015	.028	.010	.020	.007	.009	.005	.008

Table 3: A numerical comparison of coverage probability $C(p,n)$ and one-term Edgeworth approximation $e_1(p,n)$ for $p = .2$. The last row is the difference $e_1(p,n) - C(p,n)$.

Because the $O(n^{-1})$ is non-negligible for moderate n, it is usually necessary to look at the two-term Edgeworth expansion. In fact, as we shall see later, several other confidence intervals which have much better performance than the standard interval have almost identical one-term Edgeworth expansions as the standard interval. In these cases, the second order term makes the difference. An expansion of the coverage probability up to the $n^{-1/2}$ term is just not adequately accurate. So let us consider the two term expansion.

3.5 General Two Term Edgeworth Expansion

For a unified derivation of the two term Edgeworth expansion for CI_s and CI_{rs}, it is convenient to define a general confidence interval $CI_*(\beta,\gamma)$ as follows:

$$CI_*(\beta,\ \gamma) = \frac{X+\beta}{n+2\beta} \pm \kappa n^{-1/2}\left(\frac{X+\gamma}{n}\cdot\frac{n-X+\gamma}{n}\right)^{1/2} \tag{10}$$

The standard interval and the recentered interval are just special cases of $CI_*(\beta,\ \gamma)$ with $CI_s = CI_*(0,0)$ and $CI_{rs} = CI_*(\kappa^2/2, 0)$.

The two term Edgeworth expansions are also given separately for the intervals CI_W and CI_{AC}. The following general notation will be repeatedly used for the ensuing Edgeworth expansions.

Notation: Denote, with $g(p,\ \cdot)$ as in (6),

$$\begin{aligned}
t_1 &= (\kappa^2 - 2\beta)(\frac{1}{2} - p)(pq)^{-1/2} \\
t_2 &= (\frac{1}{8pq} - 1)\kappa^3 + (4 - \frac{1}{2pq})\kappa\beta + \frac{\kappa\gamma}{2pq}. \\
w(\kappa) &= (\frac{1}{9} - \frac{1}{36pq})\kappa^5 + (\frac{7}{36pq} - \frac{11}{18})\kappa^3 + (\frac{1}{6} - \frac{1}{6pq})\kappa \\
Q_{21}(\ell,\ u) &= 1 - g(p,\ \ell) - g(p,\ u) \\
Q_{22}(\ell,\ u) &= \frac{1}{2}[-g^2(p,\ \ell) - g^2(p,\ u) + g(p,\ \ell) + g(p,\ u) - \frac{1}{3}]
\end{aligned} \tag{11}$$

Theorem 1 *Let $0 < p < 1$ and $0 < \alpha < 1$. Suppose $np + \ell_*(npq)^{1/2}$ is not an integer. Then the coverage probability of the general confidence interval $CI_*(\beta,\gamma)$ defined in (10) satisfies*

$$\begin{aligned}
P_* &= P_p(p \in CI_*(\beta,\gamma)) = (1-\alpha) + [g(p,\ \ell_*) - g(p,\ u_*)]\phi(\kappa)\cdot(npq)^{-1/2} \\
&+ \{2t_2 - \kappa t_1^2 - (1-2p)(\kappa - \frac{\kappa^3}{3})t_1(pq)^{-1/2} + w(\kappa)\}\phi(\kappa)n^{-1} \\
&+ \{[(1-2p)(\frac{\kappa^2}{6} - \frac{1}{2}) - (pq)^{1/2}t_1]\cdot Q_{21}(\ell_*,\ u_*) + Q_{22}(-\kappa,\ \kappa)\}\kappa\phi(\kappa)\cdot(npq)^{-1} \\
&+ O(n^{-3/2})
\end{aligned} \tag{12}$$

where the quantities ℓ_* and u_* are described immediately above (8) and formally defined in (36) in the appendix.

In particular, by setting $\beta = \gamma = 0$, we have the two term expansion for the standard interval:

$$\begin{aligned}
P_s &= P_p(p \in CI_s) = (1-\alpha) + [g(p,\ \ell_s) - g(p,\ u_s)]\phi(\kappa)\cdot(npq)^{-1/2} \\
&+ \{-\frac{(1-2p)^2}{12pq}\kappa^5 - \frac{1}{4pq}\kappa^3 + w(\kappa)\}\phi(\kappa)n^{-1} \\
&+ \{-(1-2p)(\frac{\kappa^2}{3} + \frac{1}{2})\cdot Q_{21}(\ell_s,\ u_s) + Q_{22}(-\kappa,\ \kappa)\}\kappa\phi(\kappa)\cdot(npq)^{-1} \\
&+ O(n^{-3/2})
\end{aligned} \tag{13}$$

And by setting $\beta = \kappa^2/2$ and $\gamma = 0$, we have the two term expansion for the recentered interval:

$$\begin{aligned} P_{rs} &= P_p(p \in CI_{rs}) = (1-\alpha) + [g(p,\ \ell_{rs}) - g(p,\ u_{rs})]\phi(\kappa)\cdot(npq)^{-1/2} \\ &+ \{(2 - \frac{1}{4pq})\kappa^3 + w(\kappa)\}\phi(\kappa)n^{-1} \\ &+ \{(1-2p)(\frac{\kappa^2}{6} - \frac{1}{2})\cdot Q_{21}(\ell_{rs},\ u_{rs}) + Q_{22}(-\kappa,\ \kappa)\}\kappa\phi(\kappa)\cdot(npq)^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{14}$$

Remark: In (12), the first $O(n^{-1})$ term is nonoscillating and would cause systematic bias if it is omitted. The second $O(n^{-1})$ term represents oscillations from two sources: Q_{22}, taking values between -1/6 and 1/12, contains oscillation caused purely by rounding error; Q_{21} oscillates between -1 and 1 and the term with Q_{21} represents mixed effect of the discreteness and skewness in the underlying distribution. Note that Q_{22} is continuous and the term with Q_{21} vanishes when the binomial distribution is symmetric, i.e., if $p = 1/2$.

The two-term Edgeworth expansion for the coverage probability of the confidence interval CI_W is slightly simpler.

Theorem 2 *Let $0 < p < 1$ and $0 < \alpha < 1$. Suppose $np - \kappa(npq)^{1/2}$ is not an integer. Then the coverage probability of the confidence interval CI_W defined in (2) satisfies*

$$\begin{aligned} P_W &= P_p(p \in CI_W) = (1-\alpha) + [g(p,\ -\kappa) - g(p,\ \kappa)]\phi(\kappa)\cdot(npq)^{-1/2} + w(\kappa)\phi(\kappa)n^{-1} \\ &+ \{(1-2p)(\frac{\kappa^2}{6} - \frac{1}{2})\cdot Q_{21}(-\kappa,\ \kappa) + Q_{22}(-\kappa,\ \kappa)\}\kappa\phi(\kappa)\cdot(npq)^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{15}$$

Similarly, the two-term Edgeworth expansion can be derived for the coverage probability of the confidence interval CI_{AC}.

Theorem 3 *Let $0 < p < 1$ and $0 < \alpha < 1$. Suppose $np - \ell_{AC}(npq)^{1/2}$ is not an integer. Then the coverage probability of the confidence interval CI_{AC} defined in (3) satisfies*

$$\begin{aligned} P_{AC} &= P_p(p \in CI_{AC}) = (1-\alpha) + [g(p,\ \ell_{AC}) - g(p,\ u_{AC})]\phi(\kappa)\cdot(npq)^{-1/2} \\ &+ [(\frac{1}{4pq} - 1)\kappa^3 + w(\kappa)]\phi(\kappa)\cdot n^{-1} \\ &+ \{(1-2p)(\frac{\kappa^2}{6} - \frac{1}{2})\cdot Q_{21}(\ell_{AC},\ u_{AC}) + Q_{22}(-\kappa,\ \kappa)\}\kappa\phi(\kappa)(npq)^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{16}$$

where the quantities ℓ_{AC} and u_{AC} are defined in (39) in the appendix.

The derivation of these expansions is fairly technical and will be given in the appendix.

3.6 Two term Expansion for Beta Prior Intervals

A two-term expansion can be derived also for the Bayesian intervals. The derivation in this case, however, is more complex. Unlike the other alternative intervals in Section 3.5, the limits of Bayesian intervals are not in closed form. So the expansion problem is really two stage: first, an adequate expansion of the limits of the intervals themselves, and then an expansion of the coverage probability.

We state here the two term expansion for the coverage probability of the Jeffreys prior interval. The expansion for general beta prior intervals is given in the appendix.

Theorem 4 *Consider any fixed $0 < p < 1$ and $0 < \alpha < 1$. Suppose $np + \ell_J(npq)^{1/2}$ is not an integer; then the coverage probability of the Jeffreys prior interval CI_J defined in (4) satisfies*

$$\begin{aligned} P_J &= P_p(p \in CI_J) = (1-\alpha) + [g(p,\ \ell_J) - g(p,\ u_J)]\phi(\kappa)\cdot(npq)^{-1/2} + \frac{1}{9}(1-\frac{1}{8pq})\kappa\phi(\kappa)n^{-1} \\ &+[\frac{(2p-1)}{3}\cdot Q_{21}(\ell_J,\ u_J) + Q_{22}(-\kappa,\ \kappa)]\kappa\phi(\kappa)\cdot(npq)^{-1} + O(n^{-3/2}) \end{aligned} \tag{17}$$

where ℓ_J and u_J are defined as in (45) with $a = b = 1/2$.

Again, the proof is given in the appendix.

4 Using the Two Term Expansions

Edgeworth expansions are commonly considered as asymptotic approximations. In our problem, the two term expansion is remarkably accurate even for relatively small n. We will use the two term expansions for the coverage probabilities to compare the performance of the confidence intervals. Let us see some evidence of the accuracy of the two term Edgeworth expansion.

4.1 Accuracy of the Two Term Expansions

The two-term Edgeworth expansions approximate the true coverage probability of a binomial confidence interval with an error of $O(n^{-3/2})$. The approximation is very accurate, even for small to moderate sample sizes.

Figure 5 shows the actual coverage probability of the nominal 95% Wilson interval and the two-term Edgeworth approximation for $n = 20$. The maximum error is only .0008 in the range of $.2 \le p \le .8$. The maximum error further is reduced to .0002 in the same range of p when n increases to 40. The differences are almost indistinguishable on the plot.

Similarly, the two-term Edgeworth approximation is accurate for other intervals. For the standard interval, the maximum error is .0075 for $n = 40$ in the range of $.2 \le p \le .8$. The maximum error decreases to .0022 in the same range of p when n increases to 100. The maximum error is .0031 between the true coverage of CI_{AC} and its two-term Edgeworth approximation for $n = 40$ and $.2 \le p \le .8$ and the error is reduced to .0006 for $n = 100$ in the same range of p. Larger values of n are necessary for very good accuracy if p gets closer to 0 or 1.

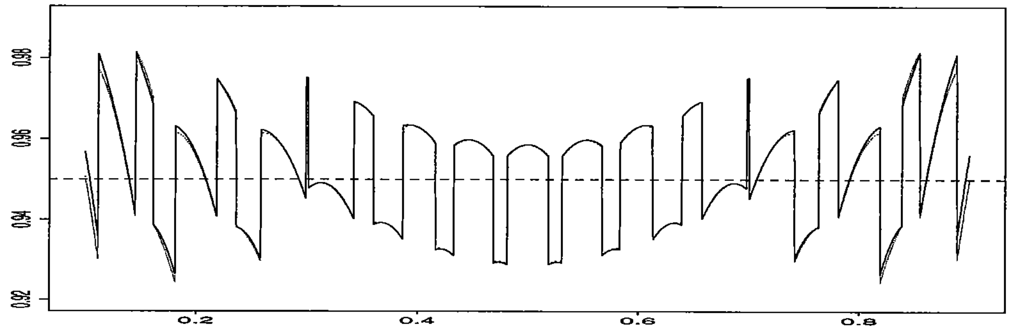

Figure 5: Comparison between the true coverage probability of the Wilson interval (solid) and two-term Edgeworth expansion (dotted) with $n = 20$ and $1 - \alpha = .95$.

4.2 Comparison of Coverage Properties

We will now use the two term Edgeworth expansions presented in Sections 3.5 and 3.6 to compare the coverage properties of the standard interval CI_s, the Wilson interval CI_W, the Agresti-Coull interval CI_{AC} and the Jeffreys prior interval CI_J. We will show how the non-oscillatory part of the second order term can be used to explain the deficiency of the standard procedure and the much better performance of competing ones such as Wilson's procedure. Indeed, directly from equations (13), (15), (16), and (17) we have :

$$P_{AC} - P_s = \{\frac{(1-2p)^2}{12pq}\kappa^5 + (\frac{1}{2pq} - 1)\kappa^3\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (18)$$

$$P_{AC} - P_W = (\frac{1}{4pq} - 1)\kappa^3\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (19)$$

$$P_{AC} - P_J = \{-\frac{(1-2p)^2}{36pq}\kappa^5 + (\frac{4}{9pq} - \frac{29}{18})\kappa^3 + (\frac{1}{18} - \frac{11}{72pq})\kappa\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (20)$$

where P_s, P_W, P_{AC}, and P_J are the coverage probabilities of CI_s, CI_W, CI_{AC}, and CI_J, respectively. The most important things to notice in (18), (19) and (20) are the following.

In (18) and (19), trivially, the coefficient of the n^{-1} term is positive for all p and all κ. In (20), the same coefficient is positive for all p as long as $\kappa \leq 3.95$.

The conclusion is that of the three alternative intervals CI_W, CI_{AC}, and CI_J, CI_{AC} "dominates" the other two intervals as far as coverage is concerned. And, of course, CI_{AC} "dominates" the standard interval CI_s as well. However, coverage is only a part of the story in interval estimation. In Section 5, we will present the corresponding expansions for expected lengths of these intervals and we will then appreciate better the reason for this apparent dominance property of CI_{AC} in coverage. It turns out that CI_{AC} tends to be longer than these competitors, and therefore not very surprisingly has larger coverage probabilities.

Expressions for $P_s - P_W$, $P_J - P_W$, etc. can be obtained from (18), (19) and (20) in an obvious way. Rather than explicitly reporting those expressions, we give a simple plot that might help understand the comparisons a little better. In Figure 6 below, the values of the nonoscillating n^{-1} terms are plotted as a function of p when $n = 40$, a modest value, and $\alpha = .05$. The curves correspond to P_{AC}, P_W, P_J, and P_s. A serious negative bias in the coverage of the standard interval is transparent from this plot. The Wilson interval CI_W does significantly better than the standard interval CI_s, and especially so near the boundaries. However, CI_W and the Jeffrey interval CI_J are pretty comparable and, especially for CI_J, the systematic bias term is very close to zero. So if we consider only the nonoscillating terms, then the Jeffreys interval comes the closest to second order accuracy. On the other hand, the Agresti-Coull interval CI_{AC} has higher coverage probability than CI_W (and likewise the others), and again, the difference is the most noticeable near the boundaries. These conclusions obtained from the two term Edgeworth expansions are very much consistent with numerical reports on the exact coverage probabilities in Brown, Cai and DasGupta (1999). The analytical calculations in this article thus provide a concrete theoretical justification for the practical results in that article.

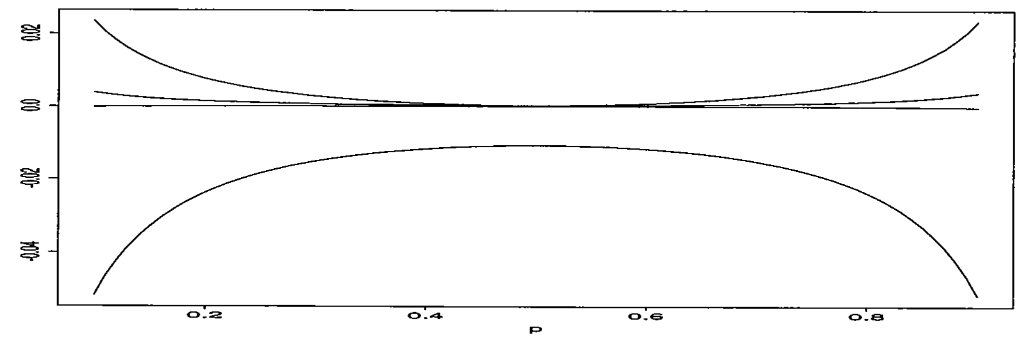

Figure 6: Comparison of the nonoscillating terms. From top to bottom: the nonoscillating $O(n^{-1})$ terms of P_{AC}, P_W, P_J, and P_s, with $n = 40$ and $\alpha = .05$.

The individual performance of the intervals themselves also depend somewhat on the value of α. For instance, consider equation (15) representing the two term Edgeworth expansion for the Wilson interval and consider the nonoscillating n^{-1} term there. This is the systematic bias term. The coefficient is $w(\kappa)\phi(\kappa)$, where $w(\kappa)$ defined in (11) actually also involves p. Now simple algebra shows that while for $\kappa = 1.96$, $w(\kappa)\phi(\kappa)$ is always positive, for $\kappa = 2.575$, i.e., when $\alpha = .01$, $w(\kappa)\phi(\kappa)$ is negative for $p < .1$ or $p > .9$. This indicates that the performance of the Wilson interval when p is close to the boundaries is better for the 95% case than for the 99% case. This is also confirmed by exact coverage calculations.

Similar interesting phenomena are seen even for the standard interval. Consider equation (13) giving the two term expansion for the standard interval. The coefficient of the nonoscillating n^{-1} term is significantly negative whenever p is not near .5 for both $\kappa = 1.96$, and 2.575. This corresponds to the previously seen poor coverage of the standard interval.

More interestingly, for $\kappa = 1.96$, the coefficient of that n^{-1} term is uniformly more negative for all p, indicating that overall the nominal 95% interval is generally even more biased than the nominal 99% interval. However, note that the oscillation terms are generally larger for $\kappa = 1.96$ than for $\kappa = 2.575$ because of the presence of the multiplicative factor, $\phi(\kappa)$, which occurs in all those terms. This accounts for the fact that when $n = 30$ there exist values of p for which the 95% interval has coverage over 95% but as shown in Figure 2 there are no values of p for which coverage of the 99% interval exceeds 99%.

5 Expansion for Expected Length

The two term Edgeworth expansions presented in Section 3 show that up to the order $O(n^{-1})$, the Agresti-Coull interval dominates the standard, the Wilson, and the Jeffreys prior interval. However, in mutual comparison of different confidence intervals, parsimony in length in addition to coverage is also always an important issue. Therefore, for the above four intervals, we will now provide an expansion for their expected lengths correct up to the order $O(n^{-3/2})$. As we shall shortly see, the expansion for length differs qualitatively from the two term Edgeworth expansion for coverage probability in that the Edgeworth expansion includes terms involving $n^{-1/2}$ and n^{-1}, whereas the expansion for length includes terms $n^{-1/2}$ and $n^{-3/2}$. The coefficient of the $n^{-1/2}$ term is the same for all the intervals, but the coefficient for the $n^{-3/2}$ term differs. So, naturally, the coefficients of the $n^{-3/2}$ term will be used as a basis for comparison of their length.

Theorem 5 *Let CI be a generic notation for any of the intervals CI_s, CI_W, CI_{AC} and CI_J. Then,*

$$L(n,\, p) \equiv E_{n,p}(\textit{length of } CI) = 2\kappa(pq)^{1/2}n^{-1/2}(1 - \frac{\delta(\kappa, p)}{8npq}) + O(n^{-2}), \tag{21}$$

where

$$\begin{aligned} \delta(\kappa,p) &= 1 \ \textit{for } CI_s; && (22)\\ &= 1+\kappa^2(8pq-1), \ \textit{for } CI_W; && (23)\\ &= 1+\kappa^2(12pq-2), \ \textit{for } CI_{AC}; && (24)\\ &= 1+\frac{2}{9}(13\kappa^2+17)pq - \frac{2}{9}(\kappa^2+2), \ \textit{for } CI_J. && (25)\end{aligned}$$

The expansion given in (21) is very accurate. Figure 7 plots the $O(n^{-2})$ error in (21) for the four intervals with $\alpha = .05$, $n = 40$ and $.1 \le p \le .9$. The maximum error for the standard, the Wilson, the Agresti-Coull, and the Jeffreys prior intervals are only .0013, .0014, .0035, and .0006, respectively.

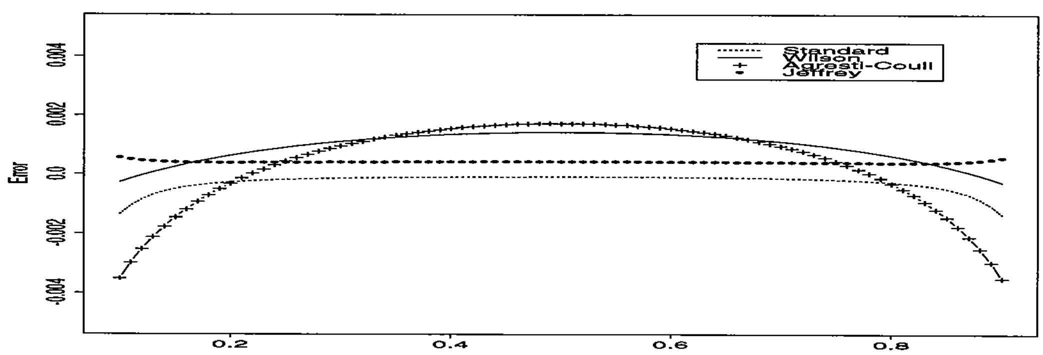

Figure 7: The $O(n^{-2})$ error in the expansion of expected length (21) with $\alpha = .05$ and $n = 40$.

The proof of Theorem 5 is given in the appendix. It is interesting to compare the coefficients $\delta(\kappa, p)$ of the $n^{-3/2}$ term for the four intervals in consideration. First, let us point out that it can be proved directly from their definitions that CI_{AC} always contains CI_W as a subinterval and hence is always longer than CI_W. It is therefore reassuring to see that for all $\kappa > 0$, and all $0 \le p \le 1$, indeed $1 + \kappa^2(8pq - 1) \ge 1 + \kappa^2(12pq - 2)$. For every other pair of intervals, the exact comparison between the corresponding pair of coefficients $\delta(\kappa, p)$ depends on κ and p. For clarity of presentation, we will therefore use the notation $\delta_s(\kappa, p)$, $\delta_W(\kappa, p)$, $\delta_{AC}(\kappa, p)$ and $\delta_J(\kappa, p)$ to denote this coefficient for the four intervals. For convenience, we will say that CI_W is the shortest interval if $\delta_W(\kappa, p)$ is greater than or equal to each of $\delta_s(\kappa, p)$, $\delta_{AC}(\kappa, p)$ and $\delta_J(\kappa, p)$ and similarly for the other intervals. Then (22) – (25) yield the following:

Corollary 1

(i). *CI_W is the shortest interval if $pq \ge \max(\frac{1}{8}, \frac{7\kappa^2-4}{46\kappa^2-34})$;*

(ii). *CI_J is the shortest interval if $\frac{\kappa^2+2}{13\kappa^2+17} \le pq \le \min(\frac{8\kappa^2-2}{41\kappa^2-17}, \frac{7\kappa^2-4}{46\kappa^2-34})$;*

(iii). *CI_s is the shortest interval otherwise.*

Interestingly, for the case $\alpha = .05$, i.e., $\kappa = 1.96$, this works out almost exactly to CI_W being the shortest when $.2 \le p \le .8$, CI_J being the shortest when $.0966 \le p \le .2$ or $.8 \le p \le .9034$, and CI_s being the shortest when $p \le .0966$ or $p \ge .9034$. Of course, it is no surprise that the standard interval is the shortest when p is near the boundaries. CI_s is not really under consideration as a credible choice because of its woefully poor coverage properties. So the conclusion might be that when we look at the coverage-length trade-off, the Jeffreys interval is the most parsimonious for small and large p, and the Wilson interval is the most parsimonious otherwise.

In Brown, Cai and DasGupta (1999), the integrated expected length is discussed as one of the criteria for the performance of the intervals. It is shown, by examples, that the integrated expected length increases in the order of CI_J, CI_W and CI_{AC}. This is also confirmed by integrating (21) over p from 0 to 1.

Corollary 2

(i). $$\int_0^1 E_{n,p}(\textit{length of } CI_s)\, dp = \frac{\kappa\pi}{4} n^{-1/2} - \frac{\kappa\pi}{4} n^{-3/2} + O(n^{-2});$$

(ii). $$\int_0^1 E_{n,p}(\textit{length of } CI_W)\, dp = \frac{\kappa\pi}{4} n^{-1/2} - \frac{\kappa\pi}{4} n^{-3/2} + O(n^{-2});$$

(iii). $$\int_0^1 E_{n,p}(\textit{length of } CI_{AC})\, dp = \frac{\kappa\pi}{4} n^{-1/2} + (\frac{\kappa^2}{2} - 1)\frac{\kappa\pi}{4} n^{-3/2} + O(n^{-2});$$

(iv). $$\int_0^1 E_{n,p}(\textit{length of } CI_J)\, dp = \frac{\kappa\pi}{4} n^{-1/2} - (\frac{37}{36} + \frac{5\kappa^2}{36})\frac{\kappa\pi}{4} n^{-3/2} + O(n^{-2}).$$

Previously we saw that between the standard interval CI_s and the Wilson interval CI_W the standard interval is shorter for p near the boundaries, and the Wilson interval is shorter otherwise. Corollary 2 shows that up to the order n^{-2}, the effects exactly cancel and the integrated expected lengths of the two intervals are always identical. This is not a priori obvious and we find it quite interesting. We also see from Corollary 2 that the integrated expected length is always the smallest for CI_J and always the largest for CI_{AC}. So the ranking is always the same and that is what makes Corollary 2 more valuable.

Among the alternative intervals, CI_W, CI_J and CI_{AC}, the actual choice has to necessarily involve some subjective judgment and we shall return to this issue later. But first we point out another nice feature of the Wilson interval.

5.1 Length Minimization under Coverage Constraint

The interval CI_W, it should be noted, has another natural property. Sometimes one imposes the rigid constraint that a confidence interval must have at least $1-\alpha$ coverage probability for all values of the parameter. If p has a prior density $\pi(p)$ resulting in a marginal pmf $m(x)$ for X, then from Brown, et al. (1995) one has that the confidence set $C_\pi(x)$, that minimizes the expected volume $E_m\{vol(C_\pi(x))\}$ subject to the coverage constraint

$$P_p(p \in C_\pi(x)) \geq 1-\alpha,$$

is a set of the form

$$\left\{p : \binom{n}{x} p^x(1-p)^{n-x} \geq \frac{k(p)}{m(x)}\right\}$$

and $k(p)$ is such that $P_p(p \in C_\pi(x)) \geq 1-\alpha$. Now if $\pi(p)$ is uniform, then $m(x)$ is uniform too. Since the binomial distribution is unimodal with mode at $[(n+1)p]$, the integer part of $(n+1)p$, it follows that $C_\pi(x)$ is formed by inverting inequalities of the form

$$[(n+1)p] - a(n,p) \leq x \leq [(n+1)p] + b(n,p),$$

where $P([(n+1)p] - a(n,p) \leq X \leq [(n+1)p] + b(n,p)) \geq 1-\alpha$. If we do a first order approximation of the distribution of X by the $N(np,\ npq)$ distribution without any continuity correction, and if $a(n,p)$ and $b(n,p)$ are chosen in an equal-tailed way so that

$$P(X > [(n+1)p] + b(n,p)) = P(X < [(n+1)p] - a(n,p)) = \alpha/2,$$

then the set $C_\pi(x)$ that results is

$$\{p : \frac{|x-np|}{\sqrt{np(1-p)}} \leq \kappa\}$$

which is the interval CI_W. Note that the formulation here is a bit different from what was done in Theorem 5. Blyth and Still (1983) have a somewhat related discussion based on Sterne (1954) and Crow (1956). See Casella, et al. (1994) for further discussions on decision theoretic set estimation.

The above discussion can be made sharper, and more rigorous, if one introduces the concept of randomized confidence procedures. In general, such a procedure is described by a measurable "inclusion" function $\rho(\cdot|\cdot)$, where $\rho(p|\,\hat{p})$ denotes the probability that the randomized set includes p when $\hat{p}$ is observed. The coverage and expected length of such a procedure are defined, respectively, as

$$P(p) = E_p(\rho(p|\,\hat{p}))$$

and

$$L(p) \equiv E(\text{length at } p \text{ of } CI) = E_p(\int \rho(\theta|\,\hat{p})\, d\theta).$$

For a non-randomized interval $CI_* = CI_*(\hat{p})$ one of course has $\rho_*(p|\,\hat{p}) = I_{CI_*(\hat{p})}(p)$. (For further discussion of such procedures see Brown, et al. (1995) and references therein.)

The following theorem describes a near-optimality property of the Wilson procedure. A near-optimality conclusion this strong is not shared by any of the other procedures in our study. Theorem 6 below says that among all procedures that are as good as the Wilson interval in coverage, all the shortest ones are basically equivalent to the Wilson interval itself, because their inclusion functions coincide with the inclusion function of the Wilson interval. There is a minor qualification needed for this, which is carefully described in (30) and (31) below.

Theorem 6 *Consider the Wilson interval whose nominal coverage is $1-\alpha$. Let $\alpha \geq .015$. For fixed n, let $\mathcal{C}_{W,n}$ denote the collection of randomized confidence procedures whose coverage satisfies*

$$P_{CI}(p) \geq P_W(p), \qquad \textit{for all } p \in (0,\, 1) \tag{26}$$

Let CI_* be any procedure in $\mathcal{C}_{W,n}$ whose average expected length is a minimum, i.e.

$$\int_0^1 L_*(p)\, dp = \min_{CI \in \mathcal{C}_{W,n}} \int_0^1 L_{CI}(p)\, dp. \tag{27}$$

(Such a confidence procedure exists.) Let $\{y\}$ denote the integer for which $-1/2 < \{y\} - y \leq 1/2$. Then, for any $\epsilon > 0$ there is an $n_\epsilon < \infty$ such that for all $n \geq n_\epsilon$ and $\epsilon < p < 1-\epsilon$, except possibly for a Lebesgue-null set of values of p,

(a). if $\{np\} - np = 0$ or 1/2, then

$$\rho_*(p|\,\hat{p}) = \rho_W(p|\,\hat{p}) \quad \textit{for all } \hat{p}; \tag{28}$$

(b). otherwise, i.e. if $\{np\} - np \neq 0$ or 1/2,

$$\rho_*(p|\,\hat{p}) = \rho_W(p|\,\hat{p}) \tag{29}$$

except possibly for the two points, $\hat{p}_a = x_a/n$ and $\hat{p}_r = x_r/n$ where the integers x_a and x_r are defined respectively as:

$$x_a = x_a(p,n) = \arg\max_x\{|\frac{x}{n} - p| : \ \rho_W(p|\,\frac{x}{n}) = 1, \ \ x = 0, ..., n\}, \tag{30}$$

and $$x_r = x_r(p,n) = \arg\min_x\{|\frac{x}{n} - p| : \ \rho_W(p|\,\frac{x}{n}) = 0, \ \ x = 0, ..., n\}. \tag{31}$$

Remark: When $0 < |\{np\} - np| < 1/2$, then x_a and x_r are uniquely defined by (30) and (31) respectively.

6 Conclusions and Summary

Interval estimation of a binomial proportion is certainly one of the most basic problems of statistical practice. We show that the standard method in universal use is riddled with problems; so much so that it cannot be salvaged. This leads us to a search for better alternative intervals. Following the empirical studies in Brown, Cai and DasGupta (1999), in this article we provide the theoretical foundation for choice of an alternative interval. Particularly important is the fact that the theoretical calculations presented here are in remarkable agreement with the extensive numerical reports presented in that companion article. Ordinarily, Edgeworth expansions and indeed asymptotic expansions in general are asymptotic approximations that may not accurately reflect the behavior in moderate samples. However, here, both for coverage and expected length, the two term expansions are remarkably accurate in moderate samples. The theoretical results proved here therefore correctly reinforce the more practical approach and the recommendations in Brown, Cai and DasGupta(1999).

To summarize, the conclusion is that the Agresti-Coull interval dominates the other intervals in coverage, but is also longer on an average and is quite conservative for p near 0 or 1. The Wilson and the Jeffreys prior interval are comparable in both coverage and length, although the Jeffreys interval is a bit shorter on an average. If we also take simplicity of presentation and ease of computation into account, the Agresti-Coull interval, although a bit too long, could be recommended for use in this problem. If simplicity is not a paramount issue, either the Wilson or the Jeffreys interval may be used, depending on taste.

7 Appendix

The binomial distribution belongs to the family of lattice distributions. The asymptotic expansion of the coverage probability contains oscillation terms that do not appear, for example, in the expansion for a continuous distribution. The algebra involved is somewhat tedious. We omit much of the messy algebra in our proofs below.

Lemma 1 *Let $X \sim Bin(n,\ p)$ and $\hat{p} = X/n$. Define $g(p,\ z) = g(p,\ z,\ p,\ n)$ as in (6). Denote $Z_n = n^{1/2}(\hat{p}-p)/(pq)^{1/2}$ and $F_n(z) = P(Z_n \le z)$. Then*

$$\begin{aligned} F_n(z) &= \Phi(z) + \frac{1}{6}(1-2p)(1-z^2)\phi(z)(npq)^{-1/2} + (-g(p,\ z) + \frac{1}{2})\phi(z)(npq)^{-1/2} \\ &\quad + \{(4pq-1)z^5 + (7-22pq)z^3 + (6pq-6)z\}\phi(z)(72npq)^{-1} \\ &\quad + \{\frac{1}{3}(1-2p)(z^2-3)(-g(p,\ z)+\frac{1}{2}) - [g^2(p,\ z) - g(p,\ z) + \frac{1}{6}]\}z(2npq)^{-1} \\ &\quad + O(n^{-3/2}) \end{aligned} \tag{32}$$

If $z = z(n)$ depends on n and can be written as

$$z = \lambda_1 + \lambda_2 n^{-1/2} + \lambda_3 n^{-1} + O(n^{-3/2})$$

where λ_1, λ_2 and λ_3 are constants, then

$$F_n(z) = \Phi(\lambda_1) + [\lambda_1(pq)^{1/2} + \frac{1}{6}(1-2p)(1-\lambda_1^2)]\phi(\lambda_1)(npq)^{-1/2}$$

$$+(-g(p,\ z)+\frac{1}{2})\phi(\lambda_1)(npq)^{-1/2}+\{\lambda_3-\frac{1}{2}\lambda_1\lambda_2^2+\frac{1}{6}(1-2p)(pq)^{-1/2}\lambda_1\lambda_2(\lambda_1^2-3)\}\phi(\lambda_1)n^{-1}$$
$$+\{(4pq-1)\lambda_1^5+(7-22pq)\lambda_1^3+(6pq-6)\lambda_1\}\phi(\lambda_1)(72npq)^{-1}$$
$$+\{[\frac{1}{3}(1-2p)(\lambda_1^2-3)-2(pq)^{1/2}\lambda_2](\frac{1}{2}-g(p,\ z))-[g^2(p,\ \lambda_1)-g(p,\ \lambda_1)+\frac{1}{6}]\}\lambda_1(2npq)^{-1}$$
$$+O(n^{-3/2}) \qquad (33)$$

Proof: The expansion (32) follows, after some algebra, directly from Theorem 23.1 of Bhattacharya and Rao (1976). See also Esseen (1945).

If $z=\lambda_1+\lambda_2 n^{-1/2}+\lambda_3 n^{-1}+O(n^{-3/2})$, we expand $\Phi(z)$ and $\phi(z)$ around λ_1.

$$\Phi(z) = \Phi(\lambda_1)+\lambda_2\phi(\lambda_1)n^{-1/2}+(\lambda_3-\frac{1}{2}\lambda_1\lambda_2^2)\phi(\lambda_1)n^{-1}+O(n^{-3/2}) \qquad (34)$$
$$\phi(z) = \phi(\lambda_1)-\lambda_1\lambda_2\phi(\lambda_1)n^{-1/2}+O(n^{-1}) \qquad (35)$$

Now plugging (34) and (35) into (32) and noting that $g^2(p,\ z)-g(p,\ z)$ admits a one-term Taylor expansion, we obtain (33). ∎

Remark: In (33), the second $O(n^{-1/2})$ and the third $O(n^{-1})$ terms are oscillation terms. Note that in the $O(n^{-1/2})$ oscillation term and first $O(n^{-1})$ oscillation term we still have $g(p,\ z)$ instead of $g(p,\ \lambda_1)$; this is due to the discontinuity in the function g.

Proof of Theorem 1: Denote

$$A = n^3+\kappa^2(n+2\beta)^2$$
$$B = 2n^3[np+\beta(2p-1)]+\kappa^2 n(n+2\beta)^2$$
$$C = n^3[np+\beta(2p-1)]^2-\kappa^2\gamma(n+\gamma)(n+2\beta)^2$$

A few lines of algebra yield

$$P_* = P_p(p\in CI_*) = P(\ell_*\le n^{1/2}(\hat{p}-p)/(pq)^{1/2}\le u_*)$$

where

$$(\ell_*,\ u_*) = (\frac{B\pm\sqrt{B^2-4AC}}{2A}-np)(npq)^{-1/2} \qquad (36)$$

The + sign goes with u_* and the - sign with ℓ_*. Expanding ℓ_* and u_*, one has

$$(\ell_*,\ u_*) = \frac{(\kappa^2-2\beta)(\frac{1}{2}-p)}{\sqrt{npq}}\pm\left\{\kappa+\frac{[(\frac{1}{8}-pq)\kappa^2+(4pq-\frac{1}{2})\beta+\frac{\gamma}{2}]\kappa}{npq}\right\}+O(n^{-3/2}) \qquad (37)$$

Now $P_p(p\in CI_*)=F_n(u_*)-F_n(\ell_*)$, and (12) follows from (33), assuming both β and γ are constants. ∎

Remark: (i). In the case of the standard interval, $\beta=\gamma=0$, and (36) yields

$$(\ell_s,u_s) = \frac{(1/2-p)\kappa^2 n^{1/2}\pm\kappa n(pq+\kappa^2/(4n))^{1/2}}{(pq)^{1/2}(n+\kappa^2)}. \qquad (38)$$

For the recentered interval CI_{rs}, the quantities ℓ_{rs} and u_{rs} are obtained from (36) with $\beta = \kappa^2/2$ and $\gamma = 0$ (ii). When γ is not a constant but a smooth function of $\hat{p}$, (12) still holds with $\gamma = \gamma(p)$.

Proof of Theorem 2: The Edgeworth expansion for $P_p(p \in CI_W)$ is slightly simpler because

$$P_W = P_p(p \in CI_W) = P(-\kappa \leq n^{1/2}(\hat{p} - p)/(pq)^{1/2} \leq \kappa)$$

And now (15) follows from (33). ∎

Proof of Theorem 3: The proof is similar to the proof of Theorem 1. Denote

$$\begin{aligned} A &= n + 2\kappa^2 \\ B &= 2pn^2 + 4\kappa^2 pn + (2p-1)\kappa^4 \\ C &= p^2 n^3 + \kappa^2 p(3p-1)n^2 + \kappa^4(3p^2 - 2p - \frac{1}{4})n - pq\kappa^6 \end{aligned}$$

It follows from some simple algebra that

$$P_{AC} = P_p(p \in CI_{AC}) = P(\ell_{AC} \leq n^{1/2}(\hat{p} - p)/(pq)^{1/2} \leq u_{AC})$$

where

$$(\ell_{AC},\, u_{AC}) = (\frac{B \pm \sqrt{B^2 - 4AC}}{2A} - np)(npq)^{-1/2} \tag{39}$$

The + sign goes with u_{AC} and the - sign with ℓ_{AC}. Expanding ℓ_{AC} and u_{AC}, one has

$$(\ell_{AC},\, u_{AC}) = \pm\{\kappa + (\frac{1}{8pq} - \frac{1}{2})\kappa^3 n^{-1}\} + O(n^{-3/2}) \tag{40}$$

with the + sign going with u_{AC} and the - sign with ℓ_{AC}. Now $P_{AC} = F_n(u_{AC}) - F_n(\ell_{AC})$, and (16) follows from (33). ∎

Expansion for Beta Prior Intervals

We will prove a more general result than Theorem 4.

Let $X \sim Bin(n,\, p)$. Suppose p has a prior distribution $Beta(a,\, b)$. Then a $100(1-\alpha)\%$ equal-tailed Bayesian interval is given by

$$CI_B = [p_l,\, p_u] = [B_{\alpha/2, X+a,\, n-X+b},\, B_{1-\alpha/2, X+a,\, n-X+b}]. \tag{41}$$

The following gives the two-term Edgeworth expansion of the coverage probability of the interval (41).

Theorem 7 *For any fixed* $0 < p < 1$ *and any* $0 < \alpha < 1$, *the coverage probability of the Beta prior interval (41) satisfies*

$$\begin{aligned} &P_p(p \in CI_B) = (1-\alpha) + [g(p,\, \ell_B) - g(p,\, u_B)]\phi(\kappa) \cdot (npq)^{-1/2} \\ &+ [2T_2 - \kappa T_1^2 - \frac{1}{3}(3\kappa - \kappa^3)(1-2p)(pq)^{-1/2}T_1 + w(\kappa)]\phi(\kappa)n^{-1} \\ &+ [a - \frac{5}{6} + (\frac{5}{3} - a - b)p]Q_{21}(\ell_B,\, u_B) + Q_{22}(-\kappa,\, \kappa)]\kappa\phi(\kappa)(npq)^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{42}$$

where $w(\kappa)$ is defined in (11), ℓ_B and u_B are defined as in (45) and

$$T_1 = [(\frac{1}{6}\kappa^2 + \frac{1}{3} - a) + (a + b - \frac{1}{3}\kappa^2 - \frac{2}{3})p](pq)^{-1/2} \quad (43)$$

$$\begin{aligned} T_2 &= \frac{1}{8}\kappa^3(pq)^{-1} + (a + b - \frac{1}{3}\kappa^2 - \frac{2}{3})\kappa - r_2(p)(pq)^{-1/2}(8pq)^{-1}\kappa^3 \\ &\quad + (a + b - \frac{1}{3}\kappa^2 - \frac{2}{3})\kappa + r_2(p)(pq)^{-1/2} - (\frac{1}{2} - p)(pq)^{-1}r_1(p)\kappa \quad (44) \end{aligned}$$

with $r_1(p)$ and $r_2(p)$ given in (51).

We will use the direct expansion method to derive (42) (see Barndorff-Nielsen and Cox (1989) and Hall (1992)). The expansion can also be derived using asymptotic expansions for posterior distributions (see, e.g., Johnson (1970) and Ghosh (1994)).

Proof of Theorem 7: The posterior distribution of p given $X = x$ is $Beta(x + a, n - x + b)$. Denote by $F(z; m_1, m_2)$ the cdf of the $Beta(m_1, m_2)$ distribution and denote by $B(\alpha; m_1, m_2)$ the inverse of the cdf. Then

$$\begin{aligned} P(p \in CI_B) &= P(B(\alpha/2; X + a, n - X + b) \le p \le B(1 - \alpha/2; X + a, n - X + b)) \\ &= P(\alpha/2 \le F(p; X + a, n - X + b) \le 1 - \alpha/2) \end{aligned}$$

Holding other parameters fixed, the function $F(p; X + a, n - X + b)$ is strictly decreasing in X (see, e.g., Johnson, et al. (1995)). So there exist unique $X_l = \rho_1(1 - \alpha/2, p, a, b)$ and $X_u = \rho_2(\alpha/2, p, a, b)$ satisfying

$$\begin{aligned} F(p; X_l + a, n - X_l + b) \le 1 - \alpha/2 \quad &\text{and} \quad F(p; X_l - 1 + a, n - (X_l - 1) + b) > 1 - \alpha/2, \\ F(p; X_u + a, n - X_u + b) \ge \alpha/2 \quad &\text{and} \quad F(p; X_u + 1 + a, n - (X_u + 1) + b) < \alpha/2 \end{aligned}$$

Therefore

$$P(p \in CI_B) = P(\ell_B \le n^{1/2}(\hat{p} - p)/(pq)^{1/2} \le u_B)$$

with

$$\begin{aligned} \ell_B &= [\rho_1(1 - \alpha/2, p, a, b) - np]/(npq)^{1/2} \\ u_B &= [\rho_2(\alpha/2, p, a, b) - np]/(npq)^{1/2} \quad (45) \end{aligned}$$

The quantities ℓ_B and u_B are defined implicitly in (45) through ρ_1 and ρ_2. The proof of (42) requires an asymptotic expansion for both ℓ_B and u_B. We do this below.

Step 1. Denote

$$\begin{aligned} x_1 &= x + a - 1, \quad n_1 = n + a + b - 2 \\ p_1 &= x_1/n_1, \quad q_1 = 1 - p_1 \\ s &= n_1^{-1}(\frac{1}{x_1} + \frac{1}{n_1 - x_1})^{-1/2} = (p_1q_1)^{1/2}n_1^{-1/2} \\ \gamma &= \frac{\Gamma(n + a + b)}{\Gamma(x + a)\Gamma(n - x + b)} = \frac{\Gamma(n_1 + 2)}{\Gamma(x_1 + 1)\Gamma(n_1 - x_1 + 1)} \end{aligned}$$

Here p_1 is the mode of p under the posterior distribution. Let $Y = (p - p_1)/s$. Then the conditional density of Y given $X = x$ is

$$\psi(y) = \gamma \cdot s(p_1 + sy)^{x_1}(q_1 - sy)^{n_1 - x_1}.$$

Step 2. Let $L(y) = \log \psi(y)$. Then it is easy to see that $L'(0) = 0$, $L''(0) = -1$, $L^{(3)}(0) = 2(1 - 2p_1)(n_1 p_1 q_1)^{-1/2}$, and $L^{(4)}(0) = -6(1 - 3p_1 q_1)(n_1 p_1 q_1)^{-1}$. Applying Stirling's formula to the Gamma functions in $L(0)$, one gets, after some algebra

$$\begin{aligned} L(0) &= \log(\frac{\Gamma(n_1+2)}{\Gamma(x_1+1)\Gamma(n_1-x_1+1)}) + \log(x_1^{1/2}(n_1-x_1)^{1/2}n_1^{-3/2}) \\ &\quad + x_1 \log x_1 + (n_1 - x_1)\log(n_1 - x_1) - n_1 \log n_1 \\ &= -\frac{1}{2}\log(2\pi) + (\frac{13}{12} - \frac{1}{12}(p_1 q_1)^{-1})n_1^{-1} + O(n_1^{-3/2}) \end{aligned}$$

Expanding $L(y)$ at 0, one has

$$L(y) = -\frac{1}{2}\log(2\pi) + c_0 n_1^{-1} - \frac{1}{2}y^2 + c_1 n_1^{-1/2} y^3 + c_2 n_1^{-1} y^4 + O(n_1^{-3/2}) \qquad (46)$$

where $c_0 = \frac{13}{12} - \frac{1}{12}(p_1 q_1)^{-1}$, $c_1 = \frac{1}{3}(1 - 2p_1)(p_1 q_1)^{-1/2}$ and $c_2 = -\frac{1}{4}[(p_1 q_1)^{-1} - 3]$. Then

$$\psi(y) = e^{L(y)} = \phi(y)[1 + c_1 n_1^{-1/2} y^3 + (c_0 + c_2 y^4 + \frac{1}{2}c_1^2 y^6)n_1^{-1}] + O(n_1^{-3/2}) \qquad (47)$$

Step 3. Integrating both sides of (47) from $-\infty$ to z, we have

$$H(z) \equiv \int_{-\infty}^{z} \psi(y)dy = \Phi(z) - v_1(z)\phi(z)n_1^{-1/2} + v_2(z)\phi(z)n_1^{-1} + O(n_1^{-3/2}) \qquad (48)$$

where $v_1(z) = -c_1(z^2 + 2)$ and $v_2(z) = -[\frac{1}{2}c_1^2(z^5 + 5z^3 + 15z) + c_2(z^3 + 3z)]$. (Because the $O(n_1^{-3/2})$ term in (47) is bounded by a polynomial in y times $\phi(y)n_1^{-3/2}$.)

We wish to find an expansion for the quantiles of the distribution H. For fixed $0 < \alpha < 1$, let $\xi_{\alpha,n} = H^{-1}(\alpha)$. It is easy to see that $\xi_{\alpha,n} \to z_\alpha = \Phi^{-1}(\alpha)$ as $n \to \infty$. Let

$$\xi_{\alpha,n} = z_\alpha + \tau_1 n_1^{-1/2} + \tau_* n_1^{-1} + o(n_1^{-1}).$$

Plugging in (48) and solving for τ_1 and τ_*, after some algebra, we get

$$\begin{aligned} \tau_1 &= \frac{1}{3}(1 - 2p_1)(z_\alpha^2 + 2)(p_1 q_1)^{-1/2} \\ \tau_* &= (\frac{1}{36}z_\alpha^3 + \frac{11}{36}z_\alpha)(p_1 q_1)^{-1} - (\frac{13}{36}z_\alpha^3 + \frac{71}{36}z_\alpha) \end{aligned}$$

Step 4. It follows that an approximation to the limits of a $100(1 - \alpha)\%$ interval is

$$\begin{aligned} (p_l,\, p_u) &= p_1 + \frac{1}{3}(1 - 2p_1)(\kappa^2 + 2)n_1^{-1} \pm \Big\{\kappa(p_1 q_1)^{1/2} n_1^{-1/2} \\ &\quad + \kappa(p_1 q_1)^{1/2} n_1^{-3/2}[(\frac{1}{36}\kappa^2 + \frac{11}{36})(p_1 q_1)^{-1} - (\frac{13}{36}\kappa^2 + \frac{71}{36})]\Big\} + O(n_1^{-2}) \qquad (49) \end{aligned}$$

Let

$$r_1(p) = a + \frac{1}{3}(\kappa^2 - 1) - [a + b + \frac{2}{3}(\kappa^2 - 1)]p \qquad (50)$$

$$r_2(p) = \Big\{-(a+b-2)/2 + (1/2 - p)[a - 1 - (a+b-2)p](pq)^{-1} + (\frac{1}{36}\kappa^2 + \frac{11}{36})(pq)^{-1} - (\frac{13}{36}\kappa^2 + \frac{71}{36})\Big\}\,\kappa(pq)^{1/2}. \qquad (51)$$

Rewriting the approximate limits (49) in terms of n, $\hat{p} = x/n$ and $\hat{q} = 1 - \hat{p}$, one has

$$(p_l,\ p_u) = (\hat{p} + r_1(\hat{p})n^{-1}) \pm \{\kappa(\hat{p}\hat{q})^{1/2}n^{-1/2} + r_2(\hat{p})n^{-3/2}\} + O(n^{-2}) \qquad (52)$$

with the + sign going with p_u and the - sign with p_l.

Step 5. Now we expand the coverage probability by using (33). In order to use (33) we invert the inequalities $p_l \le p \le p_u$ into the form of

$$\ell_B \le n^{1/2}(\hat{p} - p)/(pq)^{1/2} \le u_B.$$

We need the following lemma. The proof, which we omit here, is straightforward.

Lemma 2 *Let w_1 and w_2 be two functions with continuous first derivative. Then the roots x_* of the equations*

$$x \pm \kappa[x(1-x)]^{1/2}n^{-1/2} + w_1(x)n^{-1} + w_2(x)n^{-3/2} - p = 0 \qquad (53)$$

can be expressed as

$$x_* = p \mp \kappa(pq)^{1/2}n^{-1/2} + [(\frac{1}{2} - p)\kappa^2 - w_1(p \mp \kappa(pq)^{1/2}n^{-1/2})]n^{-1} - w_2(p)n^{-3/2} \mp \{[\frac{1}{8}(pq)^{-1/2} - (pq)^{1/2}]\kappa^3 - (\frac{1}{2} - p)(pq)^{-1/2}w_1(p)\kappa\}n^{-3/2} + O(n^{-2}) \qquad (54)$$

All the $-$ $(+)$ signs in $\mp$ in (54) go with the $+$ $(-)$ sign in $\pm$ in (53).

Applying Lemma 2 to (52), we obtain

$$P(p \in CI_B) = P(\ell_B \le n^{1/2}(\hat{p} - p)/(pq)^{1/2} \le u_B)$$

with

$$(\ell_B, u_B) = \pm\kappa + [(\frac{1}{6}\kappa^2 + \frac{1}{3} - a) + (a + b - \frac{1}{3}\kappa^2 - \frac{2}{3})p](npq)^{-1/2} \pm \Big\{(8pq)^{-1}\kappa^3 + (a + b - \frac{1}{3}\kappa^2 - \frac{2}{3})\kappa + r_2(p)(pq)^{-1/2} - (\frac{1}{2} - p)(pq)^{-1}r_1(p)\kappa\Big\}\,n^{-1} + O(n^{-3/2}) \qquad (55)$$

The expansion (42) now follows from (33). ∎

Proof of Theorem 4: In the special case of Jeffreys prior, $a = b = 1/2$. Simple calculations show that

$$r_1(p) = (\frac{1}{3}\kappa^2 + \frac{1}{6})(1-2p) \tag{56}$$

$$r_2(p) = [(\frac{1}{36}\kappa^2 + \frac{1}{18})(pq)^{-1} - (\frac{13}{36}\kappa^2 + \frac{17}{36})]\kappa(pq)^{1/2} \tag{57}$$

Plugging (56) and (57) into (44), after some algebra, the expansion for Jeffreys prior interval (17) follows from (42). ∎

Proof of Theorem 5: Throughout this proof, we will use the notation $\omega = (X/n - p)/p$ and $\tau = p/q$.

The interval CI_s. The length of the standard interval, denoted by L_s, is

$$\begin{aligned} L_s &= 2\kappa n^{-1/2}[\frac{X}{n}(1-\frac{X}{n})]^{1/2} = 2\kappa n^{-1/2}(pq)^{1/2}(1+\omega)^{1/2}(1-\tau\omega)^{1/2} \\ &= 2\kappa n^{-1/2}(pq)^{1/2}\{1 + \frac{1-\tau}{2}\cdot\omega - \frac{(1+\tau)^2}{8}\omega^2 + R_s(\omega)\}, \end{aligned} \tag{58}$$

where $R_s(\omega) \le C_1|\omega|^3 + C_2|\omega|^4 + C_3|\omega|^6$ for universal constants C_1, C_2 and C_3, depending on p, but not n. Since $E|X-np|^3 = O(n^{3/2})$, $E|X-np|^4 = O(n^2)$, and $E|X-np|^6 = O(n^3)$, it follows from (58) that

$$E(L_s) = 2\kappa n^{-1/2}(pq)^{1/2}(1 - \frac{1}{8npq}) + O(n^{-2}), \tag{59}$$

which establishes (22).

The interval CI_W. The length of the Wilson interval, L_W, is

$$\begin{aligned} L_W &= 2\kappa n^{-1/2}\frac{n}{n+\kappa^2}[\frac{X}{n}(1-\frac{X}{n}) + \frac{\kappa^2}{4n}]^{1/2} \\ &= 2\kappa n^{-1/2}[1 - \kappa^2 n^{-1} + O(n^{-2})](pq)^{1/2}\cdot \\ &\quad \{1 + \frac{1-\tau}{2}\cdot\omega - \frac{(1+\tau)^2}{8}\omega^2 + \frac{\kappa^2}{8npq} + R_W(\omega)\}, \end{aligned} \tag{60}$$

where $R_W(\omega) \le C_1|\omega|^3 + C_2|\omega|^4 + C_3|\omega|^6 + C_4 n^{-2}$ for universal constants C_1, C_2, C_3 and C_4. As in (59), it now follows from (60) that

$$\begin{aligned} E(L_W) &= 2\kappa n^{-1/2}(pq)^{1/2}[1 - \kappa^2 n^{-1} + O(n^{-2})](1 + \frac{\kappa^2-1}{8npq} + O(n^{-3/2})) \\ &= 2\kappa n^{-1/2}(pq)^{1/2}[1 - \frac{8\kappa^2 pq + 1 - \kappa^2}{8npq}] + O(n^{-2}), \end{aligned} \tag{61}$$

which establishes (23). The proof for the Agresti-Coull interval is very similar to the proof of (23) and so we will omit it.

The interval CI_J. Using Equation (52), the length of the Jeffreys interval, L_J, is

$$L_J = 2\kappa n^{-1/2}[\frac{X}{n}(1-\frac{X}{n})]^{1/2} + 2r_2(\hat{p})n^{-3/2} + O(n^{-2}), \tag{62}$$

with $\hat{p} = X/n$ and the function $r_2(\cdot)$ as defined in equation (57). Note that for any $0 < p < 1$, $r_2(p)$ is differentiable. The first term in (62) exactly equals the length of the standard interval CI_s. Therefore, from (22) and the mean value theorem,

$$\begin{aligned} E(L_J) &= 2\kappa n^{-1/2}(pq)^{1/2}(1 - \frac{1}{8npq}) + 2r_2(p)n^{-3/2} + O(n^{-2}) \\ &= 2\kappa n^{-1/2}(pq)^{1/2}\{1 - \frac{(2/9)(13\kappa^2 + 17)pq - (2/9)(\kappa^2 + 2)}{8npq}\} + O(n^{-2}), \end{aligned}$$

by algebra from the definition of $r_2(p)$ in equation (57). This establishes (25) and completes the proof of Theorem 5. ∎

Proof of Theorem 6: Brown, et al. (1995) shows that CI_* essentially uniquely satisfies the following rule

$$\rho_*(p|\,\hat{p}) = \begin{cases} 1 & \text{if } B(n,p;n\hat{p}) > k(p) \\ 0 & \text{if } B(n,p;n\hat{p}) < k(p) \end{cases} \tag{63}$$

where $k(p)$ is chosen so that CI_* satisfies (26). Here $B(n,p;x)$ denotes the binomial probability mass function. Randomization is allowed when $B(n,p;n\hat{p}) = k(p)$, so long as (26) is satisfied. Note that for any $K < \infty$

$$B(n,p;\,x) = \phi(z) + \frac{1}{6}(z^3 - 3z)(1-2p)\phi(z)(npq)^{-1/2} + O(n^{-1}) \tag{64}$$

where $z = z(x) = (x - np)(npq)^{-1/2}$, uniformly as $n \to \infty$ for $\epsilon < p < 1-\epsilon$, $|z| < K$. Note next that for $x \geq np$, we have $z(x) \geq 0$, and

$$\phi(z(x+1)) - \phi(z(x)) = z\phi(z)(npq)^{-1/2} + O(n^{-1}) \tag{65}$$

uniformly for $\epsilon < p < 1-\epsilon$, $|z| < K$. A similar expression holds for $x < np$.

Now, suppose $\{np\} - np = 0$ or $1/2$. Then (for $\epsilon < p < 1-\epsilon$ and $n \geq n_\epsilon$) there are two distinct points satisfying (30) and (31) respectively and two satisfying (63). Denote the corresponding x values as $x_{a,1}$, $x_{a,2}$ and $x_{r,1}$, $x_{r,2}$. With appropriate labeling

$$x_{r,1} + 1 = x_{a,1} < np < x_{a,2} = x_{r,2} - 1.$$

Also

$$-z(x_{a,1}) = z(x_{a,2}) = \kappa + O(n^{-1/2}).$$

From (64)

$$\min_{i=1,2} B(n,p;x_{a,i}) \;\geq\; \phi(z(x_{a,2})) - \frac{1}{6}|\kappa^3 - 3\kappa||1-2p|\phi(\kappa)(npq)^{-1/2} + O(n^{-1}), \tag{66}$$

$$\text{and } \max_{i=1,2} B(n,p;x_{r,i}) \;\leq\; \phi(z(x_{a,2})+1) - \frac{1}{6}|\kappa^3 - 3\kappa||1-2p|\phi(\kappa)(npq)^{-1/2} + O(n^{-1}). \tag{67}$$

Hence, from (66) and (67)

$$\begin{aligned} & \min\{B(n,p;x):\ \rho_W(p|\,\frac{x}{n}) = 1\} - \max\{B(n,p;x):\ \rho_W(p|\,\frac{x}{n}) = 0\} \\ = \;& \min_{i=1,2} B(n,p;x_{a,i}) - \max_{i=1,2} B(n,p;x_{r,i}) \\ \geq \;& (\kappa - \frac{1}{3}|\kappa^3 - 3\kappa|)\phi(\kappa)(npq)^{-1/2} + O(n^{-1}) > 0 \end{aligned} \tag{68}$$

for $\kappa < \sqrt{6}$ and $n \geq n_\epsilon$. Since $\alpha > .015$ implies $\kappa < 2.44 < \sqrt{6}$, it follows from (68) and (63) that $\rho_*(p|\,\hat{p}) = \rho_W(p|\,\hat{p})$ a.e., as claimed.

When $\{np\} - np \neq 0$ or $1/2$ similar reasoning shows that for any integer $y_1 \neq x_a$ having $\rho_W(p|\,\frac{y_1}{n}) = 1$ and any integer $y_2 \neq x_r$ having $\rho_W(p|\,\frac{y_2}{n}) = 0$

$$B(n,p;\, y_1) - B(n,p;\, y_2) > 0 \tag{69}$$

for $\epsilon < p < 1 - \epsilon$, $n \geq n_\epsilon$. Conclusion (30) then follows from this, (63) and the unimodality of ϕ. ∎

References

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

[2] Barndorff-Nielsen, O.E. & Cox, D.R. (1989). *Asymptotic techniques for use in statistics.* Chapman and Hall, London.

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

[4] Bhattacharya, R.N. & Rao, R.R. (1976). *Normal approximation and asymptotic expansions.* Wiley, New York.

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

[6] Brown, L.D., Cai, T. & DasGupta, A. (1999). Interval estimation for a binomial proportion. Technical Report, Department of Statistics, Purdue University.

[7] 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.

[8] 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.

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

[10] Esseen, C.G. (1945). Fourier analysis of distribution functions: a mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1 - 125.

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

[12] Ghosh, J.K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series, Institute of Mathematical Statistics, Hayward.

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

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

[15] Johnson, N.L., Kotz. S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Second Edition, Vol. 2. Wiley, New York.

[16] Johnson, R.A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851-64.

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

[18] Sterne, T.E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41**, 275-278.

[19] Wasserman, L. (1991). An inferential interpretation of default priors. Technical Report, Carnegie Mellon University.