Statistics Papers

2002

Confidence Intervals for a Binomial Proportion and Asymptotic Expansions

Lawrence D. Brown
University of Pennsylvania

T. Tony Cai
University of Pennsylvania

Anirban DasGupta
Purdue University

Confidence Intervals for a Binomial Proportion and Asymptotic Expansions

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.

The four additional interval methods we study in detail are the score-test interval (Wilson), the likelihood-ratio-test interval, a Jeffreys prior Bayesian interval and an interval suggested by Agresti and Coull. The asymptotic expansions for coverage show that the first three of these alternative methods have coverages that fluctuate about the nominal value, while the Agresti−Coull interval has a somewhat larger and more nearly conservative coverage function. For the five interval methods we also investigate asymptotically their average coverage relative to distributions for p supported within $(0\ 1)$. In terms of expected length, asymptotic expansions show that the Agresti−Coull interval is always the longest of these. The remaining three are rather comparable and are shorter than the Wald interval except for p near 0 or 1.

These analytical calculations support and complement the findings and the recommendations in Brown, Cai and DasGupta (*Statist. Sci.* (2001) **16** 101–133).

CONFIDENCE INTERVALS FOR A BINOMIAL PROPORTION AND ASYMPTOTIC EXPANSIONS[1]

BY LAWRENCE D. BROWN, T. TONY CAI AND ANIRBAN DASGUPTA

University of Pennsylvania, University of Pennsylvania and Purdue University

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.

The four additional interval methods we study in detail are the score-test interval (Wilson), the likelihood-ratio-test interval, a Jeffreys prior Bayesian interval and an interval suggested by Agresti and Coull. The asymptotic expansions for coverage show that the first three of these alternative methods have coverages that fluctuate about the nominal value, while the Agresti–Coull interval has a somewhat larger and more nearly conservative coverage function. For the five interval methods we also investigate asymptotically their average coverage relative to distributions for p supported within $(0, 1)$. In terms of expected length, asymptotic expansions show that the Agresti–Coull interval is always the longest of these. The remaining three are rather comparable and are shorter than the Wald interval except for p near 0 or 1.

These analytical calculations support and complement the findings and the recommendations in Brown, Cai and DasGupta (*Statist. Sci.* (2001) **16** 101–133).

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 an interval in virtually universal use. 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.

The problem has an extensive literature, and the questionable performance of the standard Wald interval has been sporadically remarked on. Simultaneously, there has also been work suggesting alternative confidence intervals. For example, alternative intervals have been suggested that use a continuity correction as well as intervals that actually guarantee a minimum $1 - \alpha$ coverage probability for all

values of the parameter p. In spite of all this 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 rather small. Various widely used texts in statistics provide 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 (2001) (henceforth BCD) recently showed that the performance of this standard interval is far more erratic and inadequate than is appreciated. Virtually all of the conventional wisdom and popular prescriptions are misplaced. The Wald interval is sufficiently poor in this problem that it should not be trusted unless npq is quite large.

We have recently become aware of Schader and Schmid (1990). That paper contains plots very similar to some in BCD, clearly notes the deficiency of the standard interval and makes an alternative proposal which, however, differs from those in BCD.

BCD do a fairly comprehensive examination of several natural alternative confidence intervals for p, and after extensive numerical analysis recommend the score interval of 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. We also investigate the likelihood-ratio-test intervals, which were not treated in detail in BCD. 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.

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 BCD. In Section 3, first we introduce the standard interval and the four alternative confidence intervals. The rest of Section 3 deals with Edgeworth expansions for the coverage probabilities of the standard interval and the 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. 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, especially so for the Bayesian and likelihood 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) has the greatest coverage among the five methods we concentrate on. They

also show that the Wilson, likelihood and the Jeffreys prior interval are pretty consistently comparable. See especially Figure 6.

These Edgeworth expansions are organized to display two types of effects. The principal part of the expansion involves a smooth description of the general value of the coverage. The remainder of the expansion contains oscillating terms related to the effect of discreteness in the binomial distribution. From our perspective, the smooth terms are the more important. Consider any smooth (prior) distribution for p supported within $(0, 1)$. We show that the integrated coverage from the oscillatory terms is of a lower order than that from the smooth terms. Examination of only the smooth terms thus yields a realistic asymptotic comparison of the overall coverage of the interval methods. The notion of looking at such a smoothed evaluation of coverage properties is heuristically related to the concept of a "very weak expansion" as suggested in a different context in Woodroofe (1986). Figure 6 contains a comparison based only on these nonoscillating terms.

A closer scrutiny of the complete expansions also shows other features of interest. For instance, from these expansions one can see how the choice of the level α can also affect the relative performance of the various interval methods. One can also see that the absolute magnitude of the oscillations in expected coverage for the standard method are significantly bigger than those from the other methods. Figure 8 displays this effect and thus shows another respect in which the standard method is inferior to its competitors.

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. 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 with respect to the uniform distribution for p on $(0, 1)$. From these expansions one sees that the Agresti–Coull interval is always the longest, the Wilson and the standard interval have identical two term expansions for integrated length, and the Jeffreys prior interval is always the shortest. The likelihood ratio interval is slightly longer than the Jeffreys interval. Similar results for other one parameter exponential families are presented in Brown, Cai and DasGupta (2000).

As we mentioned before, these asymptotic expansions of both the coverage probabilities and the expected lengths reflect the reports in BCD 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 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 that 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 BCD, Santner (1998), Agresti and Coull (1998) and other references cited there.

EXAMPLE 1. Consider $p = 0.5$. 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 $p = 0.5$ and $n = 10$ to 100. At $n = 97$, the coverage is still only about 0.933; 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 0.951, but at the much larger value $n = 40$, the coverage is only 0.919. The oscillations in this case are related to the discreteness of the binomial distribution. A careful look at the coverage probability shows that it requires $n \geq 194$ to guarantee that the coverage probability stays at 0.94 or above when $p = 0.5$.

Table 1 lists the smallest n after which the coverage stays at 0.93 or above for selected values of p for the standard interval and three alternative intervals. n_s, n_J,
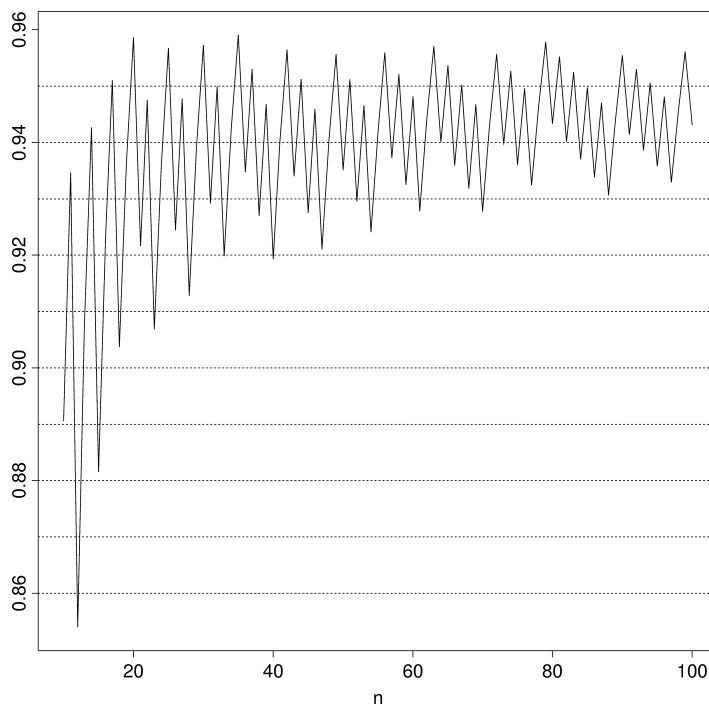

FIG. 1. *Coverage probability of the standard interval for $p = 0.5$ and $n = 10$–100.*

TABLE 1
Smallest n after which the coverage stays at 0.93 or above. The numbers in italic are the corresponding values of npq

p	0.01	0.025	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
n_{S}	2757	1005	526	286	141	118	89	82	56	64	54	71
$n_{\mathrm{S}}pq$	*27.3*	*24.5*	*25.0*	*25.7*	*18.0*	*18.9*	*16.7*	*17.2*	*12.7*	*15.4*	*13.4*	*17.8*
n_{J}	956	384	134	47	62	32	36	25	29	22	22	32
$n_{\mathrm{J}}pq$	*9.5*	*9.4*	*6.4*	*4.2*	*7.9*	*5.1*	*6.8*	*5.2*	*6.6*	*5.3*	*5.4*	*8*
n_{W}	407	42	111	36	24	41	13	31	18	7	13	32
$n_{\mathrm{W}}pq$	*4.0*	*1.0*	*5.3*	*3.2*	*3.1*	*6.6*	*2.4*	*6.5*	*4.1*	*1.7*	*3.2*	*8*
n_{AC}	1	5	10	11	8	4	1	1	10	1	1	32
$n_{\mathrm{AC}}pq$	*0.01*	*0.1*	*0.5*	*1.0*	*1.0*	*0.6*	*0.2*	*0.2*	*2.3*	*0.2*	*0.2*	*8*

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 quite small, it takes thousands of observations for the nominal 95% standard interval to ensure that the coverage probability is at least 0.93.

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 needed to guarantee approximate validity (i.e., 93% coverage) of the standard interval are much larger than the usual recommendations in popular textbooks. Many of those textbooks express requirements in terms of npq. The numbers in italic in Table 1 give the corresponding values of npq needed to guarantee 93% coverage. For the standard interval these numbers can be as large as 27.3 and are never smaller than 12.7. For a minimum coverage of 94% the corresponding minimum and maximum values of npq are for n_{S}: 44.4, 78.8; for n_{J}: 12.9, 37.8; for n_{W}: 12.4, 34.6; for n_{AC}: 0.01, 23.5.

From Table 1, one may think that the Agresti–Coull interval is the obvious interval of choice. However, we will see in Section 5 that it tends to be longer than the other intervals, and so may not be the most desirable.

EXAMPLE 2. This example emphasizes that the standard interval can be grossly inadequate. It demonstrates that there is a systematic bias in the coverage probability of the standard interval. Figure 2 shows the exact coverage probability of the nominal 99% standard interval with $n = 30$. It is striking that in this case the coverage is always smaller than 0.99. In fact on the average the coverage is only 0.914. 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.
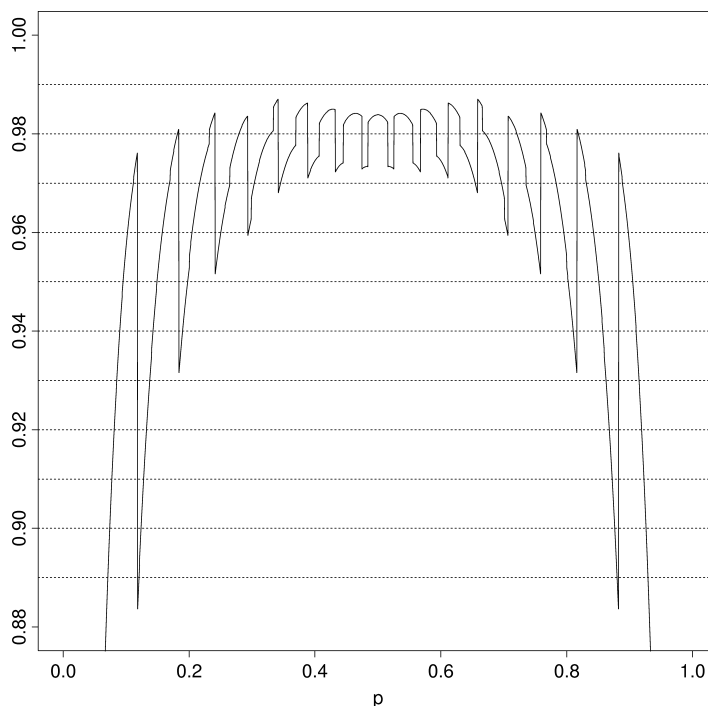
FIG. 2.   *Coverage of the nominal 99% standard interval for $n = 30$ and $0 < p < 1$.*

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

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

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 nonnegligible negative bias in the coverage probability of the standard
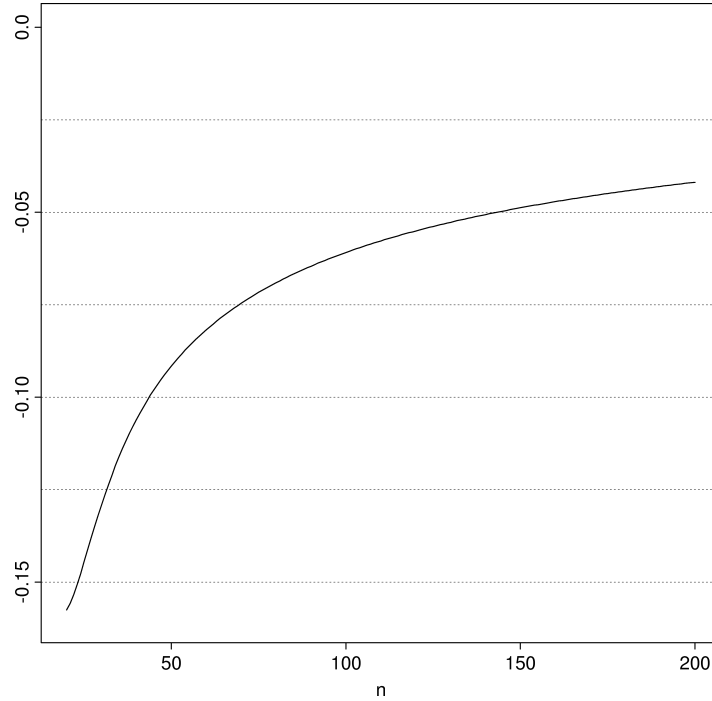
FIG. 3.   *Bias in the distribution of W_n with $p = 0.25$. Vertical axis is $E(W_n)$.*

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 $p = 0.25$.

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:

$$(2.1) \quad EW_n = E\lambda(Z_n) = \frac{p - 1/2}{\sqrt{npq}}\left(1 + \frac{7}{2n} + \frac{9(p - 1/2)^2}{2npq}\right) + o(n^{-3/2}).$$

It can be seen from (2.1) that W_n has negative bias for $p < 0.5$ and positive bias for $p > 0.5$. 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.

TABLE 2
Variance, skewness and kurtosis of W_n for $p = 0.25$

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

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 discussions 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 [e.g., 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 $p = 0.5$ and $\alpha = 0.05$. 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. 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 wish to keep in mind the formal simplicity of any recommended alternative interval. For many uses, simplicity may well be a dominant factor because the problem is a basic one and a computationally clumsy procedure seems not likely to survive the test of time in such a basic problem.

3.1. *Preview.* In BCD 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 alternatives compare among themselves. We will also see in the two term expansions some subtle features of the problem itself, for example, how the choice of α can affect the performance of the confidence intervals. We should mention that other types of asymptotic expansions besides an Edgeworth expansion can also be used; see, for example, Pierce and Peters (1992). But in this problem, Edgeworth expansions seem to be the most appropriate one because they capture the oscillations very effectively, while the other methods do not.

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 expansions for the expected lengths give us the tools to make an overall comparative assessment of the suggested alternative intervals.

3.2. *Alternative intervals.* Besides the standard interval, we will concentrate on the following intervals.

1. *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. 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 Wilson interval has the form

$$(3.1) \qquad CI_W = \tilde{p} \pm \frac{\kappa n^{1/2}}{n + \kappa^2} \left(\hat{p}\hat{q} + \frac{\kappa^2}{4n} \right)^{1/2}.$$

2. *The Agresti–Coull interval.* This interval has the same simple form as the standard interval CI_s, but with a different center, $\tilde{p}$, and a modified value for n. The interval is defined as

$$(3.2) \qquad CI_{AC} = \tilde{p} \pm \kappa(\tilde{p}\tilde{q})^{1/2}\tilde{n}^{-1/2}.$$

   Again, for the case when $\alpha = 0.05$, 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.

3. *The likelihood ratio interval.* This interval is constructed by inversion of the likelihood ratio test which accepts the null hypothesis H_0: $p = p_0$ if $-2\log(\Lambda_n) \le \kappa^2$, where Λ_n is the likelihood ratio

$$\Lambda_n = \frac{L(p_0)}{\sup_p L(p)} = \frac{p_0^X(1 - p_0)^{n-X}}{(X/n)^X(1 - X/n)^{n-X}},$$

   and L denotes the likelihood function. See Rao (1973).

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 thus given by

$$(3.3) \qquad CI_J = [B_{\alpha/2,X+1/2,n-X+1/2}, \, B_{1-\alpha/2,X+1/2,n-X+1/2}],$$

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

REMARK. The so-called exact interval, namely the Clopper–Pearson interval [Clopper and Pearson (1934)], is excessively conservative and inefficient. A much better procedure is to use the interval implied by use of the mid-P value resulting from the exact binomial test. It is interesting that this mid-P interval has a formal connection to the Jeffreys interval introduced above; see BCD.

We should also add that intervals resulting from use of other "normalizing" or "stabilizing" transformations also deserve consideration. In the binomial case, these transformations would be the logit or the arcsine transformation. BCD examined these intervals also, and it was concluded that they do not measure up to the Wilson, Jeffreys or the likelihood ratio interval; they are simply way too long.

3.3. *One term Edgeworth expansion.* Edgeworth expansions are a popular tool for studying complicated probabilistic quantities. See Bhattacharya and Ranga 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

$$(3.4) \qquad h(x) = x - x_-$$

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

$$(3.5) \qquad g(p, z) = g(p, z, n) = h\big(np + z(npq)^{1/2}\big)$$

[we suppress in (3.5) and later the dependence of g on n]. Theorem 23.1 in Bhattacharya and Rao (1976) yields that

$$(3.6) \qquad \begin{aligned} &P\bigg(\frac{n^{1/2}(\hat{p} - p)}{(pq)^{1/2}} \le z\bigg) \\ &= \Phi(z) + \bigg[\bigg(\frac{1}{2} - g(p, z)\bigg) + \frac{1}{6}(1 - 2p)(1 - z^2)\bigg]\phi(z)(npq)^{-1/2} \\ &\quad + O(n^{-1}) \end{aligned}$$

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 (3.7) below.

From (3.6) 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

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

See (A.7) in the Appendix for the exact expressions for ℓ_s and u_s. Correspondingly, the bounds ℓ_{AC}, u_{AC}, etc. are defined similarly.

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

$$(3.7) \quad P_p(p \in CI_s) = (1 - \alpha) + [g(p, \ell_s) - g(p, u_s)]\phi(\kappa)(npq)^{-1/2} + O(n^{-1})$$

The second term in (3.7), due to rounding error, is the principal contributor to the oscillation phenomenon. This 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 (3.7) and gets

$$(3.8) \quad \begin{aligned} P_p(p \in CI_s) &= (1 - \alpha) + [g(p, \ell_s) - g(p, u_s) + 1]\phi(\kappa)(npq)^{-1/2} \\ &\quad + O(n^{-1}). \end{aligned}$$

The same applies to the two-term expansion of the coverage probability of various confidence intervals discussed in Sections 3.5 and 3.6.

Here we would like to point out that there is an error in Ghosh [(1979), Theorem 1, page 895]. The oscillation terms were mistakenly omitted in the expansion. This affects one statement Ghosh [(1979), page 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 0.05 to 0.95. And in Table 3, we compare numerically the coverage probability of the standard interval with the one-term Edgeworth approximation for fixed $p = 0.2$ 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
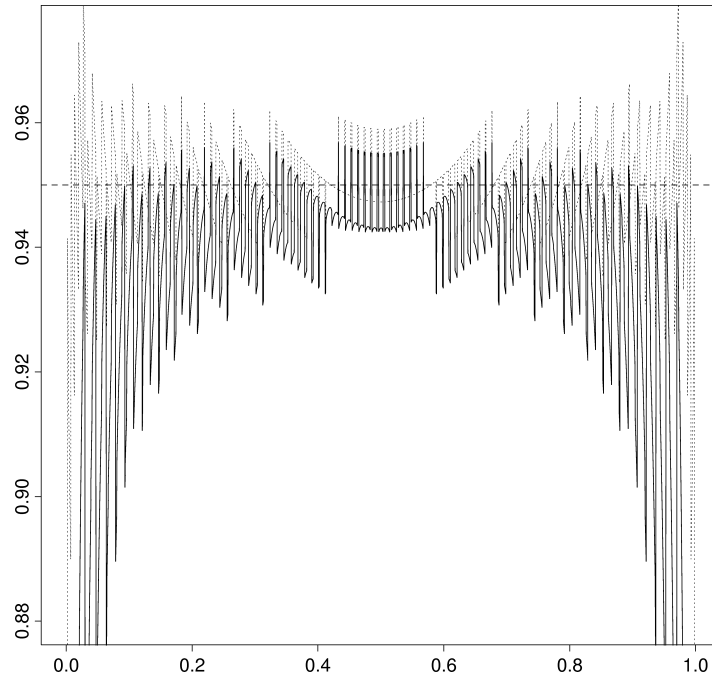
FIG. 4. *Comparison between the actual coverage probability* (*solid*) *and one-term Edgeworth expansion* (*dotted*) *with $n = 100$ and $\alpha = 0.05$.*

the order n^{-1}, is mostly nonoscillating and negative. This can be easily seen from (3.12) in the next section.

Because the $O(n^{-1})$ term is nonnegligible 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.

TABLE 3
A numerical comparison of coverage probability $C(p, n)$ and one-term Edgeworth approximation $e_1(p, n)$ for $p = 0.2$. The last row is the difference $e_1(p, n) - C(p, n)$

n	20	30	40	50	60	70	80	90	100	150	200
$C(p, n)$	0.921	0.946	0.905	0.938	0.922	0.940	0.932	0.947	0.933	0.944	0.941
$e_1(p, n)$	0.960	0.968	0.934	0.952	0.951	0.951	0.952	0.954	0.942	0.949	0.949
difference	0.039	0.021	0.029	0.015	0.028	0.010	0.020	0.007	0.009	0.005	0.008

3.5. *General two term Edgeworth expansion.* For a unified treatment of CI_s and CI_{rs}, to be defined below, it is convenient to define a general confidence interval $CI_*(\beta)$ as follows:

$$(3.9) \qquad CI_*(\beta) = \frac{X + \beta}{n + 2\beta} \pm \kappa n^{-1/2} \left(\frac{X}{n} \frac{n - X}{n} \right)^{1/2}.$$

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

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 (3.5),

$$w(\kappa) = \left(\frac{1}{9} - \frac{1}{36pq} \right) \kappa^5 + \left(\frac{7}{36pq} - \frac{11}{18} \right) \kappa^3 + \left(\frac{1}{6} - \frac{1}{6pq} \right) \kappa,$$

$$(3.10) \quad Q_{21}(\ell, u) = 1 - g(p, \ell) - g(p, u),$$

$$Q_{22}(\ell, u) = \frac{1}{2} \left[-g^2(p, \ell) - g^2(p, u) + g(p, \ell) + g(p, u) - \frac{1}{3} \right].$$

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 $CI_*(\beta)$ defined in* (3.9) *satisfies*

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

where

$$t_1 = (\kappa^2 - 2\beta)\left(\frac{1}{2} - p \right)(pq)^{-1/2}, \qquad t_2 = \left(\frac{1}{8pq} - 1 \right) \kappa^3 + \left(4 - \frac{1}{2pq} \right) \kappa \beta$$

and the quantities ℓ_* and u_* are described immediately above (3.7) *and formally defined in* (A.5) *in the Appendix.*

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

$$
\begin{aligned}
P_{\mathrm{s}} &= P_p(p \in CI_{\mathrm{s}}) \\
&= (1 - \alpha) + [g(p, \ell_{\mathrm{s}}) - g(p, u_{\mathrm{s}})]\phi(\kappa)(npq)^{-1/2} \\
&\quad + \left\{ -\frac{(1 - 2p)^2}{12pq}\kappa^5 - \frac{1}{4pq}\kappa^3 + w(\kappa) \right\}\phi(\kappa)n^{-1} \\
&\quad + \left\{ -(1 - 2p)\left(\frac{\kappa^2}{3} + \frac{1}{2}\right)Q_{21}(\ell_{\mathrm{s}}, u_{\mathrm{s}}) + Q_{22}(\ell_{\mathrm{s}}, u_{\mathrm{s}}) \right\}\kappa\phi(\kappa)(npq)^{-1} \\
&\quad + O(n^{-3/2}).
\end{aligned}
\tag{3.12}
$$

And by setting $\beta = \kappa^2/2$, we have the two term expansion for the recentered interval defined by $CI_{\mathrm{rs}} = \tilde{p} \pm \kappa(\hat{p}\hat{q})^{1/2}n^{-1/2}$ with $\tilde{p} = (X + \kappa^2/2)/(n + \kappa^2)$:

$$
\begin{aligned}
P_{\mathrm{rs}} &= P_p(p \in CI_{\mathrm{rs}}) \\
&= (1 - \alpha) + [g(p, \ell_{\mathrm{rs}}) - g(p, u_{\mathrm{rs}})]\phi(\kappa)(npq)^{-1/2} \\
&\quad + \left\{ \left(2 - \frac{1}{4pq}\right)\kappa^3 + w(\kappa) \right\}\phi(\kappa)n^{-1} \\
&\quad + \left\{ (1 - 2p)\left(\frac{\kappa^2}{6} - \frac{1}{2}\right)Q_{21}(\ell_{\mathrm{rs}}, u_{\mathrm{rs}}) + Q_{22}(\ell_{\mathrm{rs}}, u_{\mathrm{rs}}) \right\}\kappa\phi(\kappa)(npq)^{-1} \\
&\quad + O(n^{-3/2}).
\end{aligned}
\tag{3.13}
$$

REMARK. In (3.12)–(3.13), the first $O(n^{-1})$ term is a key term. It is non-oscillating 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.

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* (3.1) *satisfies*

$$
\begin{aligned}
P_{\mathrm{W}} &= P_p(p \in CI_{\mathrm{W}}) \\
&= (1 - \alpha) + [g(p, -\kappa) - g(p, \kappa)]\phi(\kappa)(npq)^{-1/2} + w(\kappa)\phi(\kappa)n^{-1} \\
&\quad + \left\{ (1 - 2p)\left(\frac{\kappa^2}{6} - \frac{1}{2}\right)Q_{21}(-\kappa, \kappa) + Q_{22}(-\kappa, \kappa) \right\}\kappa\phi(\kappa)(npq)^{-1} \\
&\quad + O(n^{-3/2}).
\end{aligned}
\tag{3.14}
$$

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* (3.2) *satisfies*

$$P_{AC} = P_p(p \in CI_{AC})$$

$$= (1 - \alpha) + [g(p, \ell_{AC}) - g(p, u_{AC})]\phi(\kappa)(npq)^{-1/2}$$

$$(3.15) \quad + \left[\left(\frac{1}{4pq} - 1\right)\kappa^3 + w(\kappa)\right]\phi(\kappa)n^{-1}$$

$$+ \left\{(1 - 2p)\left(\frac{\kappa^2}{6} - \frac{1}{2}\right)Q_{21}(\ell_{AC}, u_{AC}) + Q_{22}(\ell_{AC}, u_{AC})\right\}\kappa\phi(\kappa)(npq)^{-1}$$

$$+ O(n^{-3/2})$$

where the quantities ℓ_{AC} and u_{AC} are explicitly defined in (A.8) *in the Appendix.*

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

3.6. *Two term expansions for the likelihood ratio and beta prior intervals.* Two-term expansions can be derived also for the likelihood ratio and Bayesian intervals. The derivations in these cases, however, are more complex. Unlike the other alternative intervals in Section 3.5, the limits of the likelihood ratio and 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.

First we state the two term expansion for the coverage of the likelihood ratio interval.

THEOREM 4. *Denote by CI_{LR} the likelihood ratio interval. Consider any fixed $0 < p < 1$ and $0 < \alpha < 1$. Suppose $np + \ell_{LR}(npq)^{1/2}$ is not an integer. Then the coverage probability of CI_{LR} satisfies*

$$P_{LR} = P_p(p \in CI_{LR})$$

$$= (1 - \alpha) + [g(p, \ell_{LR}) - g(p, u_{LR})]\phi(\kappa)(npq)^{-1/2}$$

$$(3.16) \quad + \left(\frac{1}{6} - \frac{1}{6pq}\right)\kappa\phi(\kappa)n^{-1}$$

$$+ \left\{\left(p - \frac{1}{2}\right)Q_{21}(\ell_{LR}, u_{LR}) + Q_{22}(\ell_{LR}, u_{LR})\right\}\kappa\phi(\kappa)(npq)^{-1}$$

$$+ O(n^{-3/2})$$

where the quantities ℓ_{LR} and u_{LR} are defined in (A.11) *in the Appendix.*

The next theorem gives 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 5. *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* (3.3) *satisfies*

$$
\begin{aligned}
P_J &= P_p(p \in CI_J) \\
&= (1 - \alpha) + [g(p, \ell_J) - g(p, u_J)]\phi(\kappa)(npq)^{-1/2} - \frac{1}{12pq}\kappa\phi(\kappa)n^{-1} \\
&\quad + \left[\frac{(2p - 1)}{3}Q_{21}(\ell_J, u_J) + Q_{22}(\ell_J, u_J)\right]\kappa\phi(\kappa)(npq)^{-1} \\
&\quad + O(n^{-3/2})
\end{aligned}
$$

(3.17)

where ℓ_J and u_J are defined as in (A.19) *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 expansions for the coverage probabilities to compare the performance of the confidence intervals. We first discuss 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 0.0008 in the range of $0.2 \le p \le 0.8$. The maximum error further is reduced to 0.0002 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 0.0075 for $n = 40$ and $0.2 \le p \le 0.8$. The maximum error decreases to 0.0022 in the same range of p when n increases to 100. The maximum error is 0.0031 between the true coverage of CI_{AC} and its two-term Edgeworth approximation for $n = 40$ and $0.2 \le p \le 0.8$ and the error is reduced to 0.0006 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.
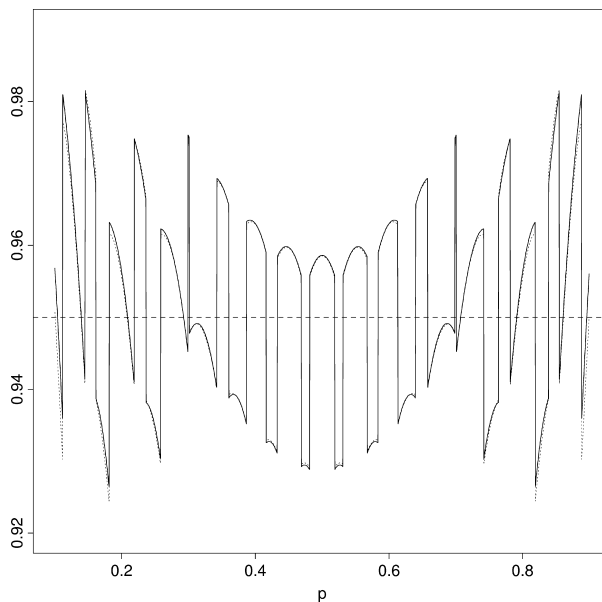
FIG. 5.   *Comparison between the true coverage probability of the Wilson interval* (*solid*) *and two-term Edgeworth expansion* (*dotted*) *with $n = 20$ and $1 - \alpha = 0.95$.*

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}, the likelihood ratio interval CI_{LR}, and the Jeffreys prior interval CI_J. We will show how the nonoscillatory 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, ignoring the $O(n^{-3/2})$ terms, directly from equations (3.12), (3.14)–(3.17) we have:

$$(4.1) \qquad P_{AC} - P_s = \left\{ \frac{(1 - 2p)^2}{12pq} \kappa^5 + \left(\frac{1}{2pq} - 1 \right) \kappa^3 \right\} \phi(\kappa) n^{-1} + \text{osci.},$$

$$(4.2) \qquad P_{AC} - P_W = \left(\frac{1}{4pq} - 1 \right) \kappa^3 \phi(\kappa) n^{-1} + \text{osci.},$$

$$(4.3) \qquad P_{AC} - P_{LR} = \left\{ -\frac{(1 - 2p)^2}{36pq} \kappa^5 + \left(\frac{4}{9pq} - \frac{29}{18} \right) \kappa^3 \right\} \phi(\kappa) n^{-1} + \text{osci.},$$

$$P_{AC} - P_J = \left\{ -\frac{(1 - 2p)^2}{36pq} \kappa^5 + \left(\frac{4}{9pq} - \frac{29}{18} \right) \kappa^3 \right.$$

$$(4.4) \qquad \qquad\qquad\qquad\qquad \left. + \left(\frac{1}{6} - \frac{1}{12pq} \right) \kappa \right\} \phi(\kappa) n^{-1} + \text{osci.},$$
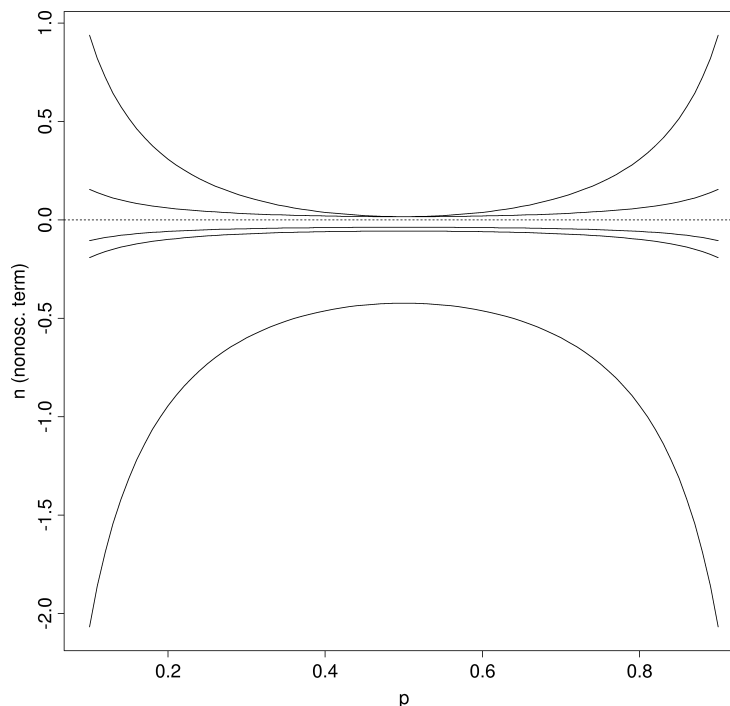
FIG. 6.   *Comparison of the nonoscillating terms. From top to bottom*: *the $O(n^{-1})$ nonoscillating terms of P_{AC}, P_W, P_J, P_{LR} and P_s, with $\alpha = 0.05$.*

where P_s, P_W, P_{AC}, P_{LR} and P_J are the coverage probabilities of CI_s, CI_W, CI_{AC}, CI_{LR} and CI_J, respectively. The most important things to notice in (4.1), (4.2), (4.3) and (4.4) are the following.

In (4.1) and (4.2), trivially, the coefficient of the n^{-1} term is positive for all p and all κ. Also in (4.3), the coefficient is positive for all p and all $\kappa \leq 3.95$. In (4.4) also, the same coefficient is positive for all p as long as $\kappa \leq 3.95$.

The conclusion is that among these intervals CI_{AC} has the largest coverage. 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 (4.1)–(4.4) 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, the values of the nonoscillating n^{-1} terms are plotted as a function of p when $\alpha = 0.05$. The y-axis is $n \cdot$ (nonoscillating term). The curves correspond to P_{AC}, P_W, P_J, P_{LR} and P_s.
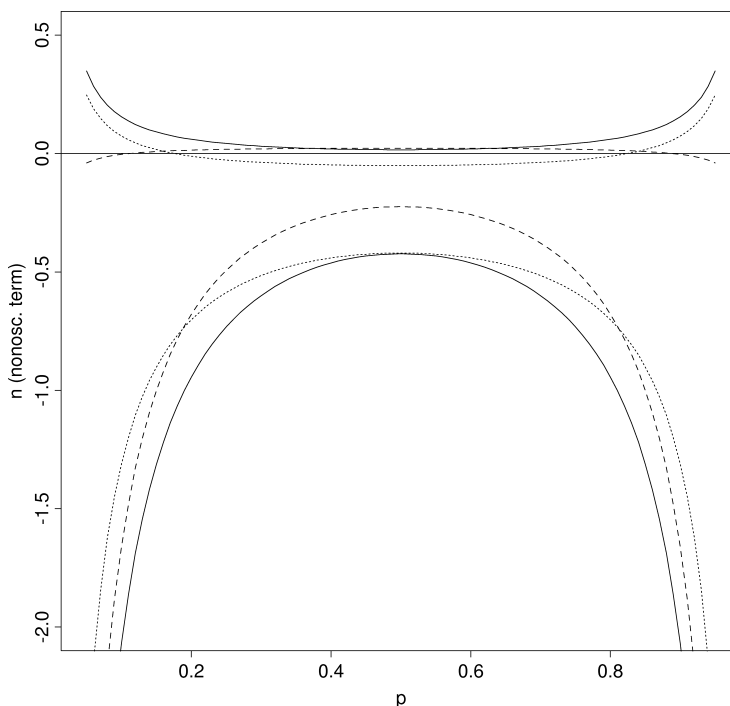
FIG. 7. *Comparison of the nonoscillating terms for different confidence levels. The top three curves are the $O(n^{-1})$ nonoscillating terms of P_W, and the bottom three are those of P_s, with $\alpha = 0.2$ (dotted), $\alpha = 0.05$ (solid) and $\alpha = 0.01$ (dashed).*

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, CI_{LR} and the Jeffreys interval CI_J are pretty comparable. 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 BCD.

The individual performance of the intervals themselves also depends somewhat on the value of α. Figure 7 plots the nonoscillating $O(n^{-1})$ terms of P_W and P_s for $\alpha = 0.2$, 0.05 and 0.01. Consider first the Wilson interval. While for $\alpha = 0.05$ this nonoscillating term is always positive, for $\alpha = 0.2$ this term is negative when $0.18 \leq p \leq 0.82$; and for $\alpha = 0.01$ the term is negative when $p \leq 0.11$ or $p \geq 0.89$. Figure 7 displays that the nonoscillating coverage term for the Wilson interval at $\alpha = 0.01$ is extremely close to the nominal value for the entire range of p, whereas for $\alpha = 0.05$ this coverage term is noticeably conservative for values of p near 0 or 1. In addition, the nonoscillating term for $\alpha = 0.05$ dominates that for $\alpha = 0.2$,

which implies that the Wilson interval is more conservative relative to $(1 - \alpha)$ for $\alpha = 0.05$ than for $\alpha = 0.2$. This is also confirmed by exact coverage calculations.

Consider now the standard interval. The coefficient of the nonoscillating $O(n^{-1})$ term is significantly negative whenever p is not near 0.5 for all three cases. This corresponds to the previously seen poor coverage of the standard interval. More interestingly, the coefficient of this $O(n^{-1})$ term is uniformly more negative for $\alpha = 0.05$ than for $\alpha = 0.01$ and $\alpha = 0.2$, indicating that overall the nominal 95% interval is generally even more biased than the nominal 99% and 80% intervals. 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%.

4.3. *Average coverage properties.*   The two term Edgeworth expansion decomposes the coverage probability into five parts:

$$C(p, n) = (1 - \alpha) + \text{``}O(n^{-1/2}) \text{ oscillation''}$$
$$+ \text{``}O(n^{-1}) \text{ bias''} + \text{``}O(n^{-1}) \text{ oscillation''} + O(n^{-3/2}).$$

We now present a theorem which shows that in an average sense the oscillatory part is of a lower order than the bias part. This adds force to the argument we just made in Section 4.2 that it is sensible to make a comparative evaluation of the intervals through a study of their $O(n^{-1})$ bias terms. In the theorem below, the average is with respect to any smooth compactly supported prior. This is similar to what is called "a very weak expansion" in Woodroofe (1986). We now state the result.

THEOREM 6.   *Let f be any density function supported on a proper subinterval* $[a, b]$ $(0 < a < b < 1)$ *and satisfying the Lipschitz condition*

$$|f(p_1) - f(p_2)| \le M|p_1 - p_2| \quad \text{for all } p_1, p_2 \in [a, b].$$

Then for all the confidence intervals under consideration (*standard, recentered, Wilson, Agresti–Coull, likelihood ratio and Jeffreys intervals*), *the integrated oscillation with respect to the density f is asymptotically negligible. That is,*

$$(4.5) \qquad \int \text{``}O(n^{-1/2}) \text{ oscillation''} \cdot f(p)\, dp = O(n^{-3/2})$$

and

$$(4.6) \qquad \int \text{``}O(n^{-1}) \text{ oscillation''} \cdot f(p)\, dp = O(n^{-3/2}).$$

Hence,

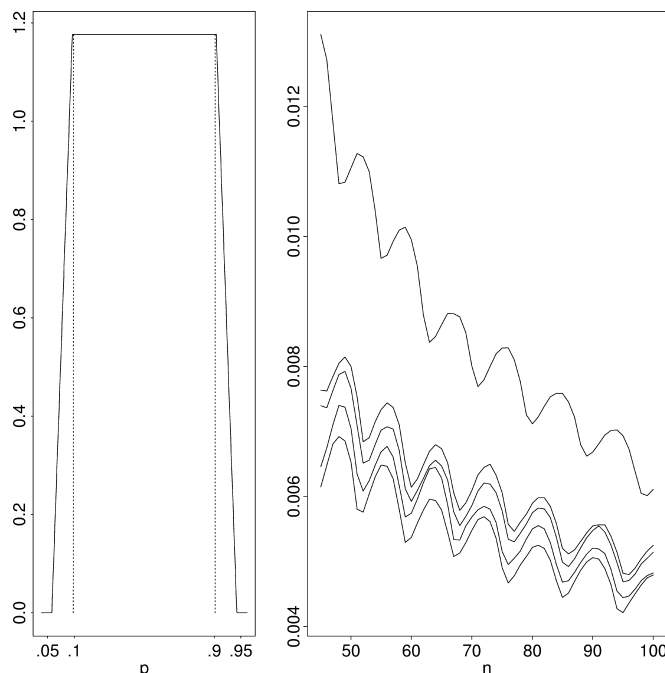$$(4.7) \quad \int \{C(p, n) - (1 - \alpha) - \text{``}O(n^{-1}) \text{ bias''}\} \cdot f(p)\, dp = O(n^{-3/2}).$$

FIG. 8. *Left panel*: *The density function* f. *Right panel*: *Integrated absolute oscillations with respect to* f (*from top to bottom*) *of* P_s, P_{LR}, P_J, P_W *and* P_{AC}, *with* $n = 45$ *to* 100 *and* $\alpha = 0.05$.

4.4. *Magnitude of the oscillations.* It is also of interest to compare the amount of oscillation in the coverage probability of the confidence intervals. We use the integrated absolute oscillation (IAO) as an overall measure of oscillation for an interval. For the intervals under consideration, the IAO with respect to a density function f is defined as

$$\text{IAO}(f, n, \alpha) = \int \left| C(p, n) - (1 - \alpha) - \text{``}O(n^{-1}) \text{ bias''} \right| f(p) \, dp.$$

In Figure 8 we plot the integrated absolute oscillations with respect to a Lipschitz density function f over p from 0.05 to 0.95 of the five intervals for n from 45 to 100. [The $O(n^{-1})$ bias term from the Edgeworth expansion is not accurate for p very close to 0 or 1.] It is interesting to see that there is a fairly clear ranking of the intervals in terms of amount of oscillation in coverage probability. From largest to smallest, the order is the standard interval, the likelihood ratio interval, the Jeffreys interval, the Wilson interval, and the Agresti–Coull interval. The oscillation of the standard interval is much larger and the amounts of oscillation of the other four intervals are comparable.

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 in coverage the standard, the Wilson, the likelihood ratio and the Jeffreys prior intervals. 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 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 of order $n^{-1/2}$ 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 lengths.

THEOREM 7. *Let CI be a generic notation for any of the intervals CI_s, CI_W, CI_{AC}, CI_{LR} and CI_J. Then,*

$$
L(n, p) \equiv E_{n, p}(\text{length of CI})
$$

(5.1)
$$
= 2\kappa (pq)^{1/2} n^{-1/2} \left(1 - \frac{\delta(\kappa, p)}{8npq} \right) + O(n^{-2}),
$$

where

(5.2) $\delta(\kappa, p) = 1$ *for CI_s*

(5.3) $= 1 + \kappa^2 (8pq - 1)$ *for CI_W*

(5.4) $= 1 + \kappa^2 (12pq - 2)$ *for CI_{AC}*

(5.5) $= 1 + \kappa^2 \left(\frac{26}{9} pq - \frac{2}{9} \right)$ *for CI_{LR}*

(5.6) $= 1 + \kappa^2 \left(\frac{26}{9} pq - \frac{2}{9} \right) + \frac{2}{9}(17pq - 2)$ *for CI_J.*

The expansion given in (5.1) is very accurate. For example, with $\alpha = 0.05$, $n = 40$ and $0.1 \le p \le 0.9$, the maximum error for the standard, the Wilson, the Agresti–Coull, the likelihood ratio and the Jeffreys prior intervals is only 0.0013, 0.0014, 0.0035, 0.0003 and 0.0006, respectively.

The proof of Theorem 7 is given in the Appendix. It is interesting to compare the coefficients $\delta(\kappa, p)$ of the $n^{-3/2}$ term for the 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 other pairs of intervals, the exact comparison between the corresponding pair of coefficients $\delta(\kappa, p)$ depends on κ and p.

Interestingly, for the case $\alpha = 0.05$, that is, $\kappa = 1.96$, CI_s is the shortest when $0 < p \le 0.084$ or $0.916 \le p < 1$, CI_{LR} is the shortest when $0.084 \le p \le 0.137$ or $0.863 \le p \le 0.916$, CI_J is the shortest when $0.137 \le p \le 0.201$ or $0.799 \le$
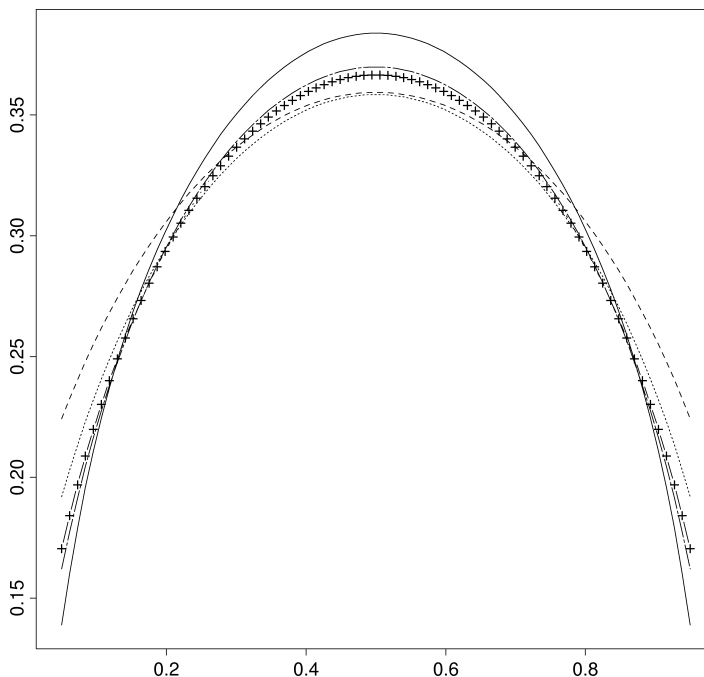
FIG. 9. *Comparison of the expected lengths of the standard (solid), the Wilson (dotted), the Agresti–Coull (dashed), the likelihood ratio $(- \cdot -)$ and the Jeffreys $(+)$ intervals for $n = 25$ and $\alpha = 0.05$.*

$p \leq 0.863$, and CI_W is the shortest when $0.201 \leq p \leq 0.799$. Thus CI_W is the shortest for the longest range of values of p. The comparison does not change qualitatively for other values of α. See Figure 9 for the case of $n = 25$ and $\alpha = 0.05$. 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, among the four procedures with acceptable coverage properties the Jeffreys and the likelihood ratio intervals are the most parsimonious for small and large p, and the Wilson interval is the most parsimonious otherwise.

In BCD 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 (5.1) over p from 0 to 1.

COROLLARY 1.

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

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

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

(iv) $\int_0^1 E_{n,p}(length\ of\ CI_{LR})\,dp = \frac{\kappa\pi}{4}n^{-1/2} - \left(1 + \frac{5\kappa^2}{36}\right)\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2});$

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

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 1 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 1 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 independent of κ and that is what makes Corollary 1 additionally valuable.

Among the alternative intervals, CI_W, CI_{AC}, CI_{LR} and CI_J, 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, Casella and Hwang (1995) one has that the confidence set $C_\pi(x)$, that minimizes the expected volume $E_m\{\mathrm{vol}(C_\pi(x))\}$ subject to the coverage constraint

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

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 $P([(n+1)p] - a(n, p) \leq X \leq [(n+1)p] + b(n, p)) \geq 1 - \alpha$. It turns out that the set $C_\pi(x)$ that results is approximately

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

which is the interval CI_W. Note that the formulation here is a bit different from what was done in Theorem 7. Blyth and Still (1983) have a somewhat related discussion based on Sterne (1954) and Crow (1956). See Casella, Hwang and Robert (1994) for further discussions on decision theoretic set estimation.

The above discussion can be made rigorous if one introduces the concept of randomized confidence procedures. In general, such a procedure is described by a measurable "inclusion" function $\rho(\cdot \mid \cdot)$, where $\rho(p \mid \hat{p})$ 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

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

and

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

For a nonrandomized interval $CI_* = CI_*(\hat{p})$ one of course has $\rho_*(p \mid \hat{p}) = I_{CI_*(\hat{p})}(p)$. [For further discussion of such procedures see Brown, Casella and Hwang (1995) and references therein.]

The following theorem describes a near-optimality property of the Wilson procedure. We believe that such a strong near-optimality conclusion is not shared by any of the other procedures in our study. Theorem 8 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 (5.11) and (5.12) below.

THEOREM 8. *Consider the Wilson interval whose nominal coverage is $1 - \alpha$. Let $\alpha \geq 0.015$. For fixed n, let $\mathcal{C}_{W,n}$ denote the collection of randomized confidence procedures whose coverage satisfies*

(5.7) $$P_{CI}(p) \geq P_W(p) \quad \text{for all } p \in (0, 1).$$

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

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

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

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

$$(5.9) \qquad \rho_*(p \mid \hat{p}) = \rho_W(p \mid \hat{p}) \quad \text{for all } \hat{p};$$

(b) *otherwise, that is, if* $\{np\} - np \neq 0$ *or* $1/2$,

$$(5.10) \qquad \rho_*(p \mid \hat{p}) = \rho_W(p \mid \hat{p}),$$

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*

$$(5.11) \quad x_a = x_a(p, n) = \arg\max_x \left\{ \left| \frac{x}{n} - p \right| : \rho_W\left(p \Big| \frac{x}{n} \right) = 1, \ x = 0, \ldots, n \right\}$$

and

$$(5.12) \quad x_r = x_r(p, n) = \arg\min_x \left\{ \left| \frac{x}{n} - p \right| : \rho_W\left(p \Big| \frac{x}{n} \right) = 0, \ x = 0, \ldots, n \right\}.$$

REMARK. When $0 < |\{np\} - np| < 1/2$, then x_a and x_r are uniquely defined by (5.11) and (5.12) 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 BCD, 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 numerical approach and the recommendations in BCD.

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, the likelihood ratio and the Jeffreys prior interval are comparable in both coverage and length, although the Jeffreys interval is a bit shorter on 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, the likelihood ratio, or the Jeffreys interval may be used, depending on taste.

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 $g(p, z) = g(p, z, n)$ as in (3.5). Denote $Z_n = n^{1/2}(\hat{p} - p)/(pq)^{1/2}$ and $F_n(z) = P(Z_n \leq z)$. Then*

(A.1)
$$
\begin{aligned}
F_n(z) = {}& \Phi(z) + \tfrac{1}{6}(1 - 2p)(1 - z^2)\phi(z)(npq)^{-1/2} \\
& + \left(\tfrac{1}{2} - g(p, z)\right)\phi(z)(npq)^{-1/2} \\
& + \left\{(4pq - 1)z^5 + (7 - 22pq)z^3 + (6pq - 6)z\right\}\phi(z)(72npq)^{-1} \\
& + \left\{\tfrac{1}{6}(1 - 2p)(z^2 - 3)\left(\tfrac{1}{2} - g(p, z)\right)\right. \\
& \quad \left. - \left[\tfrac{1}{2}g^2(p, z) - \tfrac{1}{2}g(p, z) + \tfrac{1}{12}\right]\right\}z\phi(z)(npq)^{-1} \\
& + O(n^{-3/2}).
\end{aligned}
$$

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

(A.2)
$$
\begin{aligned}
F_n(z) = {}& \Phi(\lambda_1) + \left[\lambda_2(pq)^{1/2} + \tfrac{1}{6}(1 - 2p)(1 - \lambda_1^2)\right]\phi(\lambda_1)(npq)^{-1/2} \\
& + \left(\tfrac{1}{2} - g(p, z)\right)\phi(\lambda_1)(npq)^{-1/2} \\
& + \left\{\lambda_3 - \tfrac{1}{2}\lambda_1\lambda_2^2 + \tfrac{1}{6}(1 - 2p)(pq)^{-1/2}\lambda_1\lambda_2(\lambda_1^2 - 3)\right\}\phi(\lambda_1)n^{-1} \\
& + \left\{(4pq - 1)\lambda_1^5 + (7 - 22pq)\lambda_1^3 + (6pq - 6)\lambda_1\right\}\phi(\lambda_1)(72npq)^{-1} \\
& + \left\{\left[\tfrac{1}{6}(1 - 2p)(\lambda_1^2 - 3) - (pq)^{1/2}\lambda_2\right]\left(\tfrac{1}{2} - g(p, z)\right)\right. \\
& \quad \left. - \left[\tfrac{1}{2}g^2(p, z) - \tfrac{1}{2}g(p, z) + \tfrac{1}{12}\right]\right\}\lambda_1\phi(\lambda_1)(npq)^{-1} \\
& + O(n^{-3/2}).
\end{aligned}
$$

PROOF. The expansion (A.1) follows, after some algebra, directly from Theorem 23.1 of Bhattacharya and Ranga 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:

(A.3) $\quad \Phi(z) = \Phi(\lambda_1) + \lambda_2\phi(\lambda_1)n^{-1/2} + \left(\lambda_3 - \tfrac{1}{2}\lambda_1\lambda_2^2\right)\phi(\lambda_1)n^{-1} + O(n^{-3/2}),$

(A.4) $\quad \phi(z) = \phi(\lambda_1) - \lambda_1\lambda_2\phi(\lambda_1)n^{-1/2} + O(n^{-1}).$

Now (A.2) follows by plugging (A.3) and (A.4) into (A.1). $\quad\square$

REMARK.   In (A.2), the second $O(n^{-1/2})$ and the third $O(n^{-1})$ terms are oscillation terms.

PROOF OF THEOREM 1.   Denote

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

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

(A.5) $$(\ell_*, u_*) = \left(\frac{B \pm \sqrt{B^2 - 4AC}}{2A} - np \right)(npq)^{-1/2}.$$

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

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

(A.6)

$$+ O(n^{-3/2}).$$

Now $P_p(p \in CI_*) = F_n(u_*) - F_n(\ell_*)$, and (3.12) follows from (A.2).

In the case of the standard interval, $\beta = 0$, and (A.5) yields

(A.7) $$(\ell_s, u_s) = \frac{(1/2 - p)\kappa^2 n^{1/2} \pm \kappa n(pq + \kappa^2/(4n))^{1/2}}{(pq)^{1/2}(n + \kappa^2)}.$$

For the recentered interval CI_{rs}, the quantities ℓ_{rs} and u_{rs} are obtained from (A.5) with $\beta = \kappa^2/2$. $\square$

PROOF OF THEOREM 2.   The Edgeworth expansion for $P_p(p \in CI_W)$ is slightly simpler because

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

And now (3.14) follows from (A.2). $\square$

PROOF OF THEOREM 3.   The proof is similar to the proof of Theorem 1. Denote

$$A = n + 2\kappa^2,$$
$$B = 2pn^2 + 4\kappa^2 pn + (2p - 1)\kappa^4,$$
$$C = p^2 n^3 + \kappa^2 p(3p - 1)n^2 + \kappa^4(3p^2 - 2p - \tfrac{1}{4})n - pq\kappa^6.$$

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

(A.8) $$(\ell_{AC}, u_{AC}) = \left(\frac{B \pm \sqrt{B^2 - 4AC}}{2A} - np \right) (npq)^{-1/2}.$$

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

(A.9) $$(\ell_{AC}, u_{AC}) = \pm \left\{ \kappa + \left(\frac{1}{8pq} - \frac{1}{2} \right) \kappa^3 n^{-1} \right\} + O(n^{-3/2})$$

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 (3.15) follows from (A.2). $\square$

Expansion for the likelihood ratio interval.   PROOF OF THEOREM 4.   Simple calculation yields

$$\Lambda_n = \left(\frac{p}{\hat{p}} \right)^{n\hat{p}} \left(\frac{q}{\hat{q}} \right)^{n\hat{q}}.$$

Let $z = \sqrt{n}(\hat{p} - p)/\sqrt{pq}$. Then it follows that $-2 \log \Lambda_n \leq \kappa^2$ is equivalent to

(A.10)
$$v(z) \equiv p\left(1 + (np)^{-1/2}q^{1/2}z\right) \log\left(1 + (np)^{-1/2}q^{1/2}z\right)$$
$$+ q\left(1 - (nq)^{-1/2}p^{1/2}z\right) \log\left(1 - (nq)^{-1/2}p^{1/2}z\right) - \frac{\kappa^2}{2n} \leq 0.$$

It is easy to verify that $v(\cdot)$ is a convex function and so has at most two roots. Denote by ℓ_{LR} and u_{LR} the roots of the equation $v(z) = 0$. So

(A.11) $$v(\ell_{LR}) = v(u_{LR}) = 0.$$

We first approximate ℓ_{LR} and u_{LR}. Note that the function $b(t) = (1 + t) \log(1 + t)$ can be expanded into Taylor series as

(A.12) $$b(t) = t + \frac{1}{2}t^2 - \frac{1}{6}t^3 + \frac{1}{12}t^4 + O(t^5).$$

Now applying (A.12) to (A.10), we have, after some simplification, that $v(z) = 0$ yields

(A.13)
$$z^2 - \frac{1}{3}(1 - 2p)(pq)^{-1/2}n^{-1/2}z^3 + \frac{1}{6}(1 - 3pq)(pq)^{-1}n^{-1}z^4 - \kappa^2$$
$$= O(n^{-3/2}).$$

Let $z = \pm \kappa + b_1 n^{-1/2} + b_2 n^{-1}$. Plugging into (A.13) and solving for b_1 and b_2, we have

$$b_1 = \frac{1}{6}(1 - 2p)(pq)^{-1/2}\kappa^2, \qquad b_2 = \mp \frac{1}{72}\left(\frac{1}{pq} + 2\right)\kappa^3.$$

So the roots of $v(z) = 0$ are

$$(A.14) \quad (\ell_{LR}, u_{LR}) = \pm \left\{ \kappa - \frac{1}{72} \left(\frac{1}{pq} + 2 \right) \kappa^3 n^{-1} \right\} + \frac{1}{6}(1 - 2p)(pq)^{-1/2} \kappa^2 n^{-1/2}$$
$$+ O(n^{-3/2}).$$

Hence,

$$P_p(p \in CI_{LR}) = P \left(\ell_{LR} \le \frac{n^{1/2}(\hat{p} - p)}{\sqrt{pq}} \le u_{LR} \right).$$

Now $P_{LR} = F_n(u_{LR}) - F_n(\ell_{LR})$, and the Edgeworth expansion (3.16) follows from (A.2). $\square$

Expansion for Beta prior intervals. We will prove a more general result than Theorem 5.

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

$$(A.15) \quad CI_B = [p_l, p_u] = [B_{\alpha/2, X+a, n-X+b}, B_{1-\alpha/2, X+a, n-X+b}].$$

The following gives the two-term Edgeworth expansion of the coverage probability of the interval (A.15).

THEOREM 9. *For any fixed* $0 < p < 1$ *and any* $0 < \alpha < 1$, *the coverage probability of the Beta prior interval* (A.15) *satisfies*

$$P_p(p \in CI_B) = (1 - \alpha) + [g(p, \ell_B) - g(p, u_B)]\phi(\kappa)(npq)^{-1/2}$$
$$+ \left[2T_2 - \kappa T_1^2 - \tfrac{1}{3}(3\kappa - \kappa^3)(1 - 2p)(pq)^{-1/2}T_1 + w(\kappa) \right]$$
$$\times \phi(\kappa)n^{-1}$$
$$(A.16)$$
$$+ \left\{ \left[a - \tfrac{5}{6} + (\tfrac{5}{3} - a - b)p \right]Q_{21}(\ell_B, u_B) + Q_{22}(-\kappa, \kappa) \right\}$$
$$\times \kappa\phi(\kappa)(npq)^{-1}$$
$$+ O(n^{-3/2})$$

where $w(\kappa)$ *is defined in* (3.10), ℓ_B *and* u_B *are defined as in* (A.19) *and*

$$(A.17) \quad T_1 = \left[(\tfrac{1}{6}\kappa^2 + \tfrac{1}{3} - a) + (a + b - \tfrac{1}{3}\kappa^2 - \tfrac{2}{3})p \right](pq)^{-1/2},$$

$$(A.18) \quad T_2 = \tfrac{1}{8}\kappa^3(pq)^{-1} + (a + b - \tfrac{1}{3}\kappa^2 - \tfrac{2}{3})\kappa - r_2(p)(pq)^{-1/2}(8pq)^{-1}\kappa^3$$
$$+ (a + b - \tfrac{1}{3}\kappa^2 - \tfrac{2}{3})\kappa + r_2(p)(pq)^{-1/2} - (\tfrac{1}{2} - p)(pq)^{-1}r_1(p)\kappa$$

with $r_1(p)$ *and* $r_2(p)$ *given in* (A.25).

We will use the direct expansion method to derive (A.16) [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 9. 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

$$P(p \in CI_B) = P\big(B(\alpha/2; X + a, n - X + b)$$
$$\leq p \leq B(1 - \alpha/2; X + a, n - X + b)\big)$$
$$= P\big(\alpha/2 \leq F(p; X + a, n - X + b) \leq 1 - \alpha/2\big).$$

Holding other parameters fixed, the function $F(p; X + a, n - X + b)$ is strictly decreasing in X [see, e.g., Johnson, Kotz and Balakrishnan (1995)]. So there exist unique $X_1 = \rho_1(1 - \alpha/2, p, a, b)$ and $X_u = \rho_2(\alpha/2, p, a, b)$ satisfying

$$F(p; X_1 + a, n - X_1 + b) \leq 1 - \alpha/2$$

and

$$F\big(p; X_1 - 1 + a, n - (X_1 - 1) + b\big) > 1 - \alpha/2,$$
$$F(p; X_u + a, n - X_u + b) \geq \alpha/2$$

and

$$F\big(p; X_u + 1 + a, n - (X_u + 1) + b\big) < \alpha/2.$$

Therefore

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

with

(A.19)
$$\ell_B = [\rho_1(1 - \alpha/2, p, a, b) - np]/(npq)^{1/2},$$
$$u_B = [\rho_2(\alpha/2, p, a, b) - np]/(npq)^{1/2}.$$

The quantities ℓ_B and u_B are defined implicitly in (A.19) through ρ_1 and ρ_2. The proof of (A.16) requires an asymptotic expansion for both ℓ_B and u_B. We do this below.

STEP 1. Denote

$$x_1 = x + a - 1, \qquad n_1 = n + a + b - 2,$$
$$p_1 = x_1/n_1, \qquad q_1 = 1 - p_1,$$
$$s = n_1^{-1}\left(\frac{1}{x_1} + \frac{1}{n_1 - x_1}\right)^{-1/2} = (p_1 q_1)^{1/2} n_1^{-1/2},$$
$$\gamma = \frac{\Gamma(n + a + b)}{\Gamma(x + a)\Gamma(n - x + b)} = \frac{\Gamma(n_1 + 2)}{\Gamma(x_1 + 1)\Gamma(n_1 - x_1 + 1)}.$$

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

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

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

$$L(0) = \log\left(\frac{\Gamma(n_1 + 2)}{\Gamma(x_1 + 1)\Gamma(n_1 - x_1 + 1)}\right) + \log(x_1^{1/2}(n_1 - x_1)^{1/2}n_1^{-3/2})$$

$$+ x_1 \log x_1 + (n_1 - x_1)\log(n_1 - x_1) - n_1 \log n_1$$

$$= -\frac{1}{2}\log(2\pi) + \left(\frac{13}{12} - \frac{1}{12p_1 q_1}\right)n_1^{-1} + O(n_1^{-3/2}).$$

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

(A.20) $L(y) = -\frac{1}{2}\log(2\pi) + c_0 n_1^{-1} - \frac{1}{2}y^2 + c_1 n_1^{-1/2}y^3 + c_2 n_1^{-1}y^4 + O(n_1^{-3/2})$

where

$$c_0 = \tfrac{13}{12} - \tfrac{1}{12}(p_1 q_1)^{-1}, \qquad c_1 = \tfrac{1}{3}(1 - 2p_1)(p_1 q_1)^{-1/2}$$

and

$$c_2 = -\tfrac{1}{4}[(p_1 q_1)^{-1} - 3].$$

Then

(A.21) $\begin{aligned}\psi(y) &= e^{L(y)} \\ &= \phi(y)\left[1 + c_1 n_1^{-1/2}y^3 + \left(c_0 + c_2 y^4 + \tfrac{1}{2}c_1^2 y^6\right)n_1^{-1}\right] + O(n_1^{-3/2}).\end{aligned}$

STEP 3.   Integrating both sides of (A.21) from $-\infty$ to z, we have

(A.22) $\begin{aligned}H(z) &\equiv \int_{-\infty}^{z} \psi(y)\,dy \\ &= \Phi(z) - v_1(z)\phi(z)n_1^{-1/2} + v_2(z)\phi(z)n_1^{-1} + O(n_1^{-3/2})\end{aligned}$

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 (A.21) 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

$$\xi_{\alpha,n} = z_\alpha + \tau_1 n_1^{-1/2} + \tau_2 n_1^{-1} + o(n_1^{-1}).$$

Plugging in (A.22) and solving for τ_1 and τ_2, after some algebra, we get

$$\tau_1 = \tfrac{1}{3}(1 - 2p_1)(z_\alpha^2 + 2)(p_1 q_1)^{-1/2},$$

$$\tau_2 = \left(\tfrac{1}{36}z_\alpha^3 + \tfrac{11}{36}z_\alpha\right)(p_1 q_1)^{-1} - \left(\tfrac{13}{36}z_\alpha^3 + \tfrac{71}{36}z_\alpha\right).$$

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

$$
\begin{aligned}
(p_\mathrm{l}, p_\mathrm{u}) = {}& p_1 + \tfrac{1}{3}(1 - 2p_1)(\kappa^2 + 2)n_1^{-1} \\
& \pm \big\{\kappa(p_1 q_1)^{1/2}n_1^{-1/2} \\
& \quad + \kappa(p_1 q_1)^{1/2}n_1^{-3/2}\left[\left(\tfrac{1}{36}\kappa^2 + \tfrac{11}{36}\right)(p_1 q_1)^{-1} - \left(\tfrac{13}{36}\kappa^2 + \tfrac{71}{36}\right)\right]\big\} \\
& + O(n_1^{-2}).
\end{aligned}
$$
(A.23)

Let

$$r_1(p) = a + \tfrac{1}{3}(\kappa^2 - 1) - \left[a + b + \tfrac{2}{3}(\kappa^2 - 1)\right]p,$$
(A.24)

$$
\begin{aligned}
r_2(p) = {}& \big\{-(a + b - 2)/2 + (1/2 - p)[a - 1 - (a + b - 2)p](pq)^{-1} \\
& + \left(\tfrac{1}{36}\kappa^2 + \tfrac{11}{36}\right)(pq)^{-1} - \left(\tfrac{13}{36}\kappa^2 + \tfrac{71}{36}\right)\big\}\kappa(pq)^{1/2}.
\end{aligned}
$$
(A.25)

Rewriting the approximate limits (A.23) in terms of n, $\hat{p} = x/n$ and $\hat{q} = 1 - \hat{p}$, one has

$$(p_\mathrm{l}, p_\mathrm{u}) = \left(\hat{p} + r_1(\hat{p})n^{-1}\right) \pm \left\{\kappa(\hat{p}\hat{q})^{1/2}n^{-1/2} + r_2(\hat{p})n^{-3/2}\right\} + O(n^{-2})$$
(A.26)

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

STEP 5. Now we expand the coverage probability by using (A.2). In order to use (A.2) we invert the inequalities $p_\mathrm{l} \le p \le p_\mathrm{u}$ into the form of

$$\ell_\mathrm{B} \le n^{1/2}(\hat{p} - p)/(pq)^{1/2} \le u_\mathrm{B}.$$

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*

$$x \pm \kappa[x(1 - x)]^{1/2}n^{-1/2} + w_1(x)n^{-1} + w_2(x)n^{-3/2} - p = 0$$
(A.27)

can be expressed as

$$
\begin{aligned}
x_* = {}& p \mp \kappa(pq)^{1/2}n^{-1/2} + \left[\left(\tfrac{1}{2} - p\right)\kappa^2 - w_1\left(p \mp \kappa(pq)^{1/2}n^{-1/2}\right)\right]n^{-1} \\
& - w_2(p)n^{-3/2} \\
& \mp \left\{\left[\tfrac{1}{8}(pq)^{-1/2} - (pq)^{1/2}\right]\kappa^3 - \left(\tfrac{1}{2} - p\right)(pq)^{-1/2}w_1(p)\kappa\right\}n^{-3/2} \\
& + O(n^{-2}).
\end{aligned}
$$
(A.28)

All the $-(+)$ signs in $\mp$ in (A.28) go with the $+(-)$ sign in $\pm$ in (A.27).

Applying Lemma 2 to (A.26), we obtain

$$P(p \in CI_{\mathrm{B}}) = P\big(\ell_{\mathrm{B}} \le n^{1/2}(\hat{p} - p)/(pq)^{1/2} \le u_{\mathrm{B}}\big)$$

with

$$(\ell_{\mathrm{B}}, u_{\mathrm{B}}) = \pm\kappa + \big[\big(\tfrac{1}{6}\kappa^2 + \tfrac{1}{3} - a\big) + (a + b - \tfrac{1}{3}\kappa^2 - \tfrac{2}{3})p\big](npq)^{-1/2}$$

$$\pm \big\{(8pq)^{-1}\kappa^3 + (a + b - \tfrac{1}{3}\kappa^2 - \tfrac{2}{3})\kappa + r_2(p)(pq)^{-1/2}$$

(A.29)

$$- (\tfrac{1}{2} - p)(pq)^{-1}r_1(p)\kappa\big\}n^{-1}$$

$$+ O(n^{-3/2}).$$

The expansion (A.16) now follows from (A.2).   □

PROOF OF THEOREM 5.   In the special case of Jeffreys prior, $a = b = 1/2$. Simple calculations show that

(A.30) $$r_1(p) = \big(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\big)(1 - 2p),$$

(A.31) $$r_2(p) = \big[\big(\tfrac{1}{36}\kappa^2 + \tfrac{1}{18}\big)(pq)^{-1} - \big(\tfrac{13}{36}\kappa^2 + \tfrac{17}{36}\big)\big]\kappa(pq)^{1/2}.$$

Plugging (A.30) and (A.31) into (A.18), after some algebra, the expansion for Jeffreys prior interval (3.17) follows from (A.16).   □

PROOF OF THEOREM 6.   We prove (4.5). The proof of (4.6) is similar.

STEP 1.   Suppose $0 < a < b < 1$ and $z = z_n(p) = \kappa + w_n(p)(npq)^{-1/2}$ with $w_n(p)$ satisfying

(A.32) $$|w_n(p)|, |w'_n(p)| \text{ and } |w''_n(p)| \le C \quad \text{for } p \in [a, b] \text{ and all } n.$$

Let $r_n(p, w) = np + \kappa(npq)^{1/2} + w_n(p)$. Then for sufficiently large n, r_n is a strictly increasing function of p on $[a, b]$. Let

$$\Gamma_n = \big\{\gamma : a \le \gamma \le b \text{ and } r_n(\gamma, w_n(\gamma)) \text{ is an integer}\big\}.$$

It is easy to see that $Card(\Gamma_n) = (b - a)n(1 + o(1))$. Let $\gamma_i \in \Gamma_n$ be defined by $r_n(\gamma_i, w) = i$, where i is an integer. Denote $\gamma_* = \min\{\gamma : \gamma \in \Gamma_n\}$ and $\gamma^* = \max\{\gamma : \gamma \in \Gamma_n\}$. Consider $p \in [\gamma_i, \gamma_{i+1}]$. Let $\delta = p - \gamma_i$. Expanding $r_n(p, w)$ at $p = \gamma_i$, one has

$$r_n(p, w) = i + \big\{n + (1/2 - \gamma_i)[\gamma_i(1 - \gamma_i)]^{-1/2}n^{1/2} + w'_n(\gamma_i)\big\}\delta$$

$$- \tfrac{1}{8}[\gamma_i(1 - \gamma_i)]^{-3/2}n^{1/2}\delta^2 + O(\delta^2)$$

$$\equiv i + (n + A_1 n^{1/2} + A_2)\delta + A_3 n^{1/2}\delta^2 + O(\delta^2).$$

Let $\delta_i^* = \gamma_{i+1} - \gamma_i$; then

$$(n + A_1 n^{1/2} + A_2)\delta_i^* + A_3 n^{1/2}(\delta_i^*)^2 + O((\delta_i^*)^2) = 1.$$

Solving for δ_i^*, one has, after some algebra,

$$\delta_i^* = n^{-1} - A_1 n^{-3/2} + (A_1^2 - A_2)n^{-2} + O(n^{-5/2}).$$

Let $g(p, z, n)$ be defined as in (3.5). Then

$$\int_{\gamma_i}^{\gamma_{i+1}} \left(g(p, z, n) - \tfrac{1}{2}\right) dp = \int_{\gamma_i}^{\gamma_{i+1}} \left(r_n(p, w) - i - \tfrac{1}{2}\right) dp$$

$$= \int_0^{\delta_i^*} \{(n + A_1 n^{1/2} + A_2)\delta + A_3 n^{1/2}\delta^2 + O(\delta^2)\}d\delta$$

(A.33)
$$- \int_0^{\delta_i^*} \tfrac{1}{2}d\delta$$

$$= \tfrac{1}{2}(n + A_1 n^{1/2} + A_2)(\delta_i^*)^2 - \tfrac{1}{2}\delta_i^* + O\left(n^{1/2}(\delta_i^*)^3\right)$$

$$= O(n^{-5/2}).$$

Similarly,

$$\int_a^{\gamma_*} \left(g(p, z, n) - \tfrac{1}{2}\right) dp = O(n^{-5/2})$$

and

$$\int_{\gamma^*}^{b} \left(g(p, z, n) - \tfrac{1}{2}\right) dp = O(n^{-5/2}).$$

STEP 2.   We now show that for any Lipschitz function f supported on $[a, b]$,

(A.34) $$n^{-1/2} \int_a^b \left(g(p, z, n) - \tfrac{1}{2}\right)(p(1 - p))^{-1/2} f(p) \, dp = O(n^{-3/2}).$$

Let $v(p) = (p(1 - p))^{-1/2} f(p)$. Then $v(p)$ is bounded and also Lipschitz on $[a, b]$. Then

$$n^{-1/2} \int_{\gamma_*}^{\gamma^*} \left(g(p, z, n) - \tfrac{1}{2}\right)(p(1 - p))^{-1/2} f(p) \, dp$$

$$= n^{-1/2} \sum_i \int_{\gamma_i}^{\gamma_{i+1}} \left(g(p, z, n) - \tfrac{1}{2}\right)v(p) \, dp$$

$$= n^{-1/2} \sum_i \int_{\gamma_i}^{\gamma_{i+1}} \left(g(p, z, n) - \tfrac{1}{2}\right)v(\gamma_i) \, dp$$

$$+ n^{-1/2} \sum_i \int_{\gamma_i}^{\gamma_{i+1}} \left(g(p, z, n) - \tfrac{1}{2}\right)(v(p) - v(\gamma_i)) \, dp$$

$$\equiv T_1 + T_2.$$

It follows from the boundedness of $v(p)$ and (A.33) that $T_1 = O(n^{-2})$. For T_2, since $v(p)$ is Lipschitz on $[a, b]$ and $|g(p, z, n) - 1/2| \leq 1/2$, we have, for some constant $C > 0$,

$$|T_2| \leq n^{-1/2} \sum_i \int_{\gamma_i}^{\gamma_{i+1}} |g(p, z, n) - \tfrac{1}{2}||v(p) - v(\gamma_i)| \, dp$$

$$\leq C n^{-1/2} \sum_i \int_{\gamma_i}^{\gamma_{i+1}} (p - \gamma_i) \, dp$$

$$\leq \frac{C}{2} n^{-1/2} \sum_i (\delta_i^*)^2 = O(n^{-3/2}).$$

It is easy to see that

$$n^{-1/2} \int_a^{\gamma*} \left(g(p, z, n) - \tfrac{1}{2}\right)(p(1 - p))^{-1/2} f(p) \, dp = O(n^{-5/2})$$

and

$$n^{-1/2} \int_{\gamma*}^b \left(g(p, z, n) - \tfrac{1}{2}\right)(p(1 - p))^{-1/2} f(p) \, dp = O(n^{-5/2}).$$

Hence,

$$n^{-1/2} \int_a^b \left(g(p, z, n) - \tfrac{1}{2}\right)(p(1 - p))^{-1/2} f(p) \, dp = n^{-3/2}.$$

STEP 3. For any of the five intervals under consideration,

$$O(n^{-1/2}) \text{ oscillation} = \left(g(p, \ell) - g(p, u)\right)\phi(\kappa)(npq)^{-1/2}$$

$$= n^{-1/2}\phi(\kappa)\left(g(p, \ell) - 1/2\right)(p(1 - p))^{-1/2}$$

$$- n^{-1/2}\phi(\kappa)\left(g(p, u) - 1/2\right)(p(1 - p))^{-1/2},$$

with $\ell = \ell_n(p)$ and $u = u_n(p)$ satisfying (A.32). Now it follows directly from (A.34) that

$$\int \{O(n^{-1/2}) \text{ oscillation}\} f(p) \, dp = O(n^{-3/2}). \qquad \square$$

PROOF OF THEOREM 7. 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

$$L_s = 2\kappa n^{-1/2}\left[\frac{X}{n}\left(1 - \frac{X}{n}\right)\right]^{1/2}$$

(A.35) $$= 2\kappa n^{-1/2}(pq)^{1/2}(1 + \omega)^{1/2}(1 - \tau\omega)^{1/2}$$

$$= 2\kappa n^{-1/2}(pq)^{1/2}\left\{1 + \frac{1 - \tau}{2}\omega - \frac{(1 + \tau)^2}{8}\omega^2 + R_s(\omega)\right\},$$

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 (A.36) that

$$(A.36) \qquad E(L_s) = 2\kappa n^{-1/2}(pq)^{1/2}\left(1 - \frac{1}{8npq}\right) + O(n^{-2}),$$

which establishes (5.2).

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

$$L_W = 2\kappa n^{-1/2}\frac{n}{n + \kappa^2}\left[\frac{X}{n}\left(1 - \frac{X}{n}\right) + \frac{\kappa^2}{4n}\right]^{1/2}$$

$$(A.37) \qquad = 2\kappa n^{-1/2}[1 - \kappa^2 n^{-1} + O(n^{-2})](pq)^{1/2}$$

$$\times \left\{1 + \frac{1 - \tau}{2}\omega - \frac{(1 + \tau)^2}{8}\omega^2 + \frac{\kappa^2}{8npq} + R_W(\omega)\right\},$$

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 (A.36), it now follows from (A.38) that

$$E(L_W) = 2\kappa n^{-1/2}(pq)^{1/2}[1 - \kappa^2 n^{-1} + O(n^{-2})]$$

$$(A.38) \qquad \times \left(1 + \frac{\kappa^2 - 1}{8npq} + O(n^{-3/2})\right)$$

$$= 2\kappa n^{-1/2}(pq)^{1/2}\left[1 - \frac{8\kappa^2 pq + 1 - \kappa^2}{8npq}\right] + O(n^{-2}),$$

which establishes (5.3). The proof for the Agresti–Coull interval is very similar to the proof of (5.3) and so we will omit it.

The interval CI_{LR}. With $X \sim Bin(n, p)$, the likelihood ratio is given by

$$(A.39) \qquad \Lambda_n = \frac{p^X(1 - p)^{n-X}}{(X/n)^X(1 - X/n)^{n-X}} = \left(\frac{p}{\hat{p}}\right)^{n\hat{p}}\left(\frac{q}{\hat{q}}\right)^{n\hat{q}}.$$

We may assume that $\hat{p} = X/n > 0$, and for $\hat{p} > 0$, the equation $-\log\Lambda_n = \kappa^2/2$ has two roots in p, which are the limits of the interval CI_{LR}. Writing $t = p/\hat{p} - 1$, the roots of $-\log\Lambda_n = \kappa^2/2$ satisfy

$$\log(1 + t) - \frac{\hat{q}}{\hat{p}}\log\left(1 - \frac{\hat{p}}{\hat{q}}t\right) = \kappa^2/(2n\hat{p}).$$

The roots $\underline{t}$ and $\bar{t}$ satisfy

$$(A.40) \qquad \underline{t} = -\kappa(\hat{q}/\hat{p})^{1/2}n^{-1/2} + \tfrac{1}{3}\kappa^2(1 - 2\hat{p})(n\hat{p})^{-1}$$

$$- \tfrac{1}{36}\kappa^3(1 - 13\hat{p}\hat{q})\hat{q}^{-1/2}(n\hat{p})^{-3/2} + R_{1,n}$$

and

$$\bar{t} = \kappa(\hat{q}/\hat{p})^{1/2}n^{-1/2} + \tfrac{1}{3}\kappa^2(1 - 2\hat{p})(n\hat{p})^{-1}$$

(A.41)

$$+ \tfrac{1}{36}\kappa^3(1 - 13\hat{p}\hat{q})\hat{q}^{-1/2}(n\hat{p})^{-3/2} + R_{2,n}$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1, 2$. From (A.40) and (A.41),

$$E(L_{\mathrm{LR}}) = 2\kappa E[(\hat{p}\hat{q})^{1/2}]n^{-1/2} + \tfrac{1}{18}\kappa^3(1 - 13pq)(pq)^{-1/2}n^{-3/2}$$

(A.42)

$$+ O(n^{-2}).$$

Writing $Z_n = n^{1/2}(\hat{p} - p)/(pq)^{1/2}$, by a straightforward expansion,

$$E[(\hat{p}\hat{q})^{1/2}] = (pq)^{1/2}[1 - (8npq)^{-1}] + O(n^{-3/2}),$$

and so from (A.42) one obtains

$$E(L_{\mathrm{LR}}) = 2\kappa(pq)^{1/2}[1 - (8npq)^{-1}]n^{-1/2} + \tfrac{1}{18}\kappa^3(1 - 13pq)(pq)^{-1/2}n^{-3/2}$$
$$+ O(n^{-2}),$$

which simplifies to

$$E(L_{\mathrm{LR}}) = 2\kappa(pq)^{1/2}n^{-1/2}\left[1 - \frac{9 + \kappa^2(26pq - 2)}{72npq}\right] + O(n^{-2}).$$

The interval CI_{J}.   Using equation (A.26), the length of the Jeffreys interval, L_{J}, is

(A.43) $$L_{\mathrm{J}} = 2\kappa n^{-1/2}\left[\frac{X}{n}\left(1 - \frac{X}{n}\right)\right]^{1/2} + 2r_2(\hat{p})n^{-3/2} + O(n^{-2}),$$

with $\hat{p} = X/n$ and the function $r_2(\cdot)$ as defined in equation (A.31). Note that for any $0 < p < 1$, $r_2(p)$ is differentiable. The first term in (A.43) exactly equals the length of the standard interval CI_{s}. Therefore, from (5.2) and the mean value theorem,

$$E(L_{\mathrm{J}}) = 2\kappa n^{-1/2}(pq)^{1/2}\left(1 - \frac{1}{8npq}\right) + 2r_2(p)n^{-3/2} + O(n^{-2})$$

$$= 2\kappa n^{-1/2}(pq)^{1/2}\left\{1 - \frac{(2/9)(13\kappa^2 + 17)pq - (2/9)(\kappa^2 + 2)}{8npq}\right\}$$

$$+ O(n^{-2}),$$

by algebra from the definition of $r_2(p)$ in equation (A.31). This establishes (5.6) and completes the proof of Theorem 7.   □

PROOF OF THEOREM 8.   Brown, Casella and Hwang (1995) shows that CI_* essentially uniquely satisfies the following rule:

(A.44) $$\rho_*(p \mid \hat{p}) = \begin{cases} 1, & \text{if } B(n, p; n\hat{p}) > k(p), \\ 0, & \text{if } B(n, p; n\hat{p}) < k(p), \end{cases}$$

where $k(p)$ is chosen so that CI_* satisfies (5.7). Here $B(n, p; x)$ denotes the binomial probability mass function. Note that for any $K < \infty$, by the local limit theorem [Breiman (1992)],

$$(A.45) \qquad B(n, p; x) = \phi(z) + \tfrac{1}{6}(z^3 - 3z)(1 - 2p)\phi(z)(npq)^{-1/2} + O(n^{-1})$$

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

$$(A.46) \qquad \phi(z(x + 1)) - \phi(z(x)) = z\phi(z)(npq)^{-1/2} + O(n^{-1})$$

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

Now, suppose $\{np\} - np = 0$ or $1/2$. Then (for $\varepsilon < p < 1 - \varepsilon$ and $n \geq n_\varepsilon$) there are two distinct points satisfying (5.11) and two satisfying (5.12), respectively. 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 (A.45),

$$(A.47) \qquad \min_{i=1,2} B(n, p; x_{a,i}) \geq \phi(z(x_{a,2})) - \tfrac{1}{6}|\kappa^3 - 3\kappa||1 - 2p|\phi(\kappa)(npq)^{-1/2}$$
$$+ O(n^{-1})$$

and

$$(A.48) \qquad \max_{i=1,2} B(n, p; x_{r,i}) \leq \phi(z(x_{a,2} + 1)) + \tfrac{1}{6}|\kappa^3 - 3\kappa||1 - 2p|\phi(\kappa)(npq)^{-1/2}$$
$$+ O(n^{-1}).$$

Hence, from (A.47) and (A.48),

$$\min\left\{ B(n, p; x) : \rho_W\left(p \Big| \frac{x}{n} \right) = 1 \right\} - \max\left\{ B(n, p; x) : \rho_W\left(p \Big| \frac{x}{n} \right) = 0 \right\}$$

$$(A.49) \qquad = \min_{i=1,2} B(n, p; x_{a,i}) - \max_{i=1,2} B(n, p; x_{r,i})$$

$$\geq \left(\kappa - \frac{1}{3}|\kappa^3 - 3\kappa| \right)\phi(\kappa)(npq)^{-1/2} + O(n^{-1}) > 0$$

for $\kappa < \sqrt{6}$ and $n \geq n_\varepsilon$. Since $\alpha > 0.015$ implies $\kappa < 2.44 < \sqrt{6}$, it follows from (A.49) and (A.44) that $\rho_*(p \mid \hat{p}) = \rho_W(p \mid \hat{p})$ 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 $\rho_W(p \mid y_1/n) = 1$ and any integer $y_2 \neq x_r$ having $\rho_W(p \mid y_2/n) = 0$,

$$(A.50) \qquad\qquad\qquad B(n, p; y_1) - B(n, p; y_2) > 0$$

for $\varepsilon < p < 1 - \varepsilon$, $n \geq n_\varepsilon$. Conclusion (5.11) then follows from this, (A.44) and the unimodality of ϕ. $\square$

REFERENCES

AGRESTI, A. and COULL, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statist.* **52** 119–126.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

BLYTH, C. R. and STILL, H. A. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc.* **78** 108–116.

BREIMAN, L. (1992). *Probability.* SIAM, Philadelphia.

BROWN, L. D., CAI, T. and DASGUPTA, A. (2000). Interval estimation in exponential families. Technical report, Dept. Statistics, Univ. Pennsylvania. Available at www-stat.wharton. upenn.edu/~tcai/.

BROWN, L. D., CAI, T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statist. Sci.* **16** 101–133.

BROWN, L. D., CASELLA, G. and HWANG, J. T. G. (1995). Optimal confidence sets, bioequivalence, and the limacon of Pascal. *J. Amer. Statist. Assoc.* **90** 880–889.

CASELLA, G., HWANG, J. T. G. and ROBERT, C. P. (1994). Loss functions for set estimation. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. Berger, eds.) 237–251. Academic Press, New York.

CLOPPER C. J. and PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** 404–13.

CROW, E. L. (1956). Confidence intervals for a proportion. *Biometrika* **43** 423–35.

ESSEEN, C. G. (1945). Fourier analysis of distribution functions: a mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125.

GHOSH, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter *J. Amer. Stat. Assoc.* **74** 894–900.

GHOSH, J. K. (1994). *Higher Order Asymptotics.* IMS, Hayward, CA.

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

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions* **2**, 2nd ed. Wiley, New York.

JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–64.

PIERCE, D. A. and PETERS, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *J. Roy. Statist. Soc. Ser. B* **54** 701–725.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

SANTNER, T. J. (1998). Teaching large-sample binomial confidence intervals. *Teaching Statistics* **20** 20–23.

SCHADER, M. and SCHMID, F. (1990). Charting small sample characteristics of asymptotic confidence intervals for the binomial parameter *p*. *Statist. Papers* **31** 251–264.

STERNE, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41** 275–278.

WASSERMAN, L. (1991). An inferential interpretation of default priors. Technical report, Dept. Statistics, Carnegie Mellon Univ.

WILSON, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc.* **22** 209–212.

WOODROOFE, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14** 1049–1067.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: lbrown@stat.wharton.upenn.edu
      tcai@stat.wharton.upenn.edu

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907
E-MAIL: dasgupta@stat.purdue.edu