Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 131 (2005) 63–88
journal of statistical planning and inference
www.elsevier.com/locate/jspi

One-sided confidence intervals in discrete distributions☆

T. Tony Cai*

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 9 April 2003; accepted 7 January 2004

Abstract

One-sided confidence intervals in the binomial, negative binomial, and Poisson distributions are considered. It is shown that the standard Wald interval suffers from a serious systematic bias in the coverage and so does the one-sided score interval. Alternative confidence intervals with better performance are considered. The coverage and length properties of the confidence intervals are compared through numerical and analytical calculations. Implications to hypothesis testing are also discussed.

Keywords: Bayes; Binomial distribution; Confidence intervals; Coverage probability; Edgeworth expansion; Expected length; Hypothesis testing; Jeffreys prior; Negative Binomial distribution; Normal approximation; Poisson distribution

1. Introduction

The problem of interval estimation of a binomial proportion has a long history and an extensive literature. It had been generally known that the popular two-sided Wald confidence interval was deficient in the coverage probability for p near 0 or 1. See, for example, Cressie (1980), Blyth and Still (1983), Vollset (1993), Santner (1998), Agresti and Coull (1998), and Newcombe (1998).

In two recent articles, Brown et al. (2001, 2002) give a comprehensive treatment of two-sided confidence intervals for a binomial proportion. The Wald interval is shown to suffer from a systematic negative bias in its coverage probability far more persistent than is appreciated. Contrary to common perception, the problems are not just for

☆ Research supported in part by NSF Grant DMS-0296215.
* Tel.: +1-215-898-8222; fax: +1-215-898-1280.
E-mail address: tcai@wharton.upenn.edu (T. Tony Cai).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.01.005

p near 0 or 1, and not just for small n. Alternative intervals with superior coverage properties are recommended. Among them, the score interval, produced by inversion of Rao's score test, always provides major improvements in coverage. Brown et al. (2003) extended the findings on the binomial proportion to the natural exponential family with a quadratic variance function (NEF–QVF). A coherent analytical description of the two-sided interval estimation problem was given. It is shown that the problems and the solutions in the binomial proportion case are common to all the distributions in the NEF–QVF. In particular, the Wald interval has consistent poor coverage problem and the score interval has excellent coverage properties across the NEF–QVF.

In this paper we consider the one-sided interval estimation problem. One-sided confidence intervals are useful in many applications. See Duncan (1986) and Montgomery (2001) for applications in quality control. In the present paper a unified treatment of the one-sided confidence intervals is given for the discrete Exponential family with a quadratic variance function which consists of three most important discrete distributions—the binomial, negative binomial, and Poisson distributions. Although there are some common features, the one-sided interval estimation problem differs significantly from the two-sided problem. In particular, despite the good performance of the score interval in the two-sided problem, the one-sided score interval does not perform well for each of the three distributions both in terms of coverage probability and expected length. Examples given in Section 2.1 show that both the one-sided Wald and score intervals suffer a pronounced systematic bias in the coverage, although the severity and direction differ.

The somewhat surprising fact that the score interval performs well in the two-sided problem but not in the one-sided problem is connected to the issue of probability matching. See Ghosh (1994, 2001) for general discussions on probability matching and confidence sets. In particular, first-order probability matching has no obvious bearing on the coverage for two-sided problem, because all these procedures make compensating one sided errors. It is, however, crucial for one-sided intervals. The Edgeworth expansion of the coverage probabilities given in Section 3 shows that both the one-sided Wald and score intervals are not first-order probability matching.

The deficiency of the Wald and score intervals calls for alternative one-sided intervals with better coverage properties. Two alternative intervals are introduced in Section 2.2 with a brief motivation and background. The one-sided Jeffreys prior credible interval is constructed from a Bayesian perspective. This interval is known to have the first-order probability matching property. See Ghosh (1994). The Jeffreys interval is however not second-order probability matching. A second-order corrected interval is constructed using the Edgeworth expansion. This method of using the Edgeworth expansion for the construction of one-sided intervals has been used for example in Hall (1982). The second-order corrected interval is by construction second-order probability matching.

The properties of the four confidence intervals are compared analytically. The Edgeworth expansions given in Section 3 provide an accurate and useful tool in analyzing the coverage properties. The Edgeworth expansions reveal uniform structure across the three distributions. It is shown that for all three distributions the one-sided Wald and score intervals have the first-order systematic bias in the coverage with different signs. This reinforces the phenomenon observed in the numerical examples given in Section

2.1 that the systematic biases for the two intervals are pronounced and are in the exact opposite directions of each other. In contrast the two alternative intervals have superior coverage properties with nearly vanishing systematic bias for all three distributions.

In addition to the coverage, parsimony in length is also an important issue. The confidence intervals are also compared in terms of the expected distance from the mean. The expansions of the expected distance given in Section 5 also reveal a significant amount of common structure. For instance, up to an error of order $O(n^{-2})$, there is a uniform ranking of the four confidence intervals pointwise for every value of the parameter in the Poisson and negative binomial cases. The ranking is, from the shortest to the longest, the Wald, Jeffreys, second-order corrected, and score intervals. In the binomial case the same ranking holds for $p < \frac{1}{2}$; when $p \geqslant \frac{1}{2}$ the ranking is the score, Jeffreys, second-order corrected, and Wald intervals, from the shortest to the longest.

Section 6 discusses the implications of the confidence interval results to hypothesis testing. The deficiency of the one-sided score interval implies that the actual size of the widely used score test can deviate significantly from the nominal level. Inversion of the Jeffreys and second-order corrected intervals yields better testing procedures than the score test.

2. The one-sided confidence intervals

Throughout the paper we shall assume to have iid observations $X_1, X_2, \ldots, X_n \sim F$ with F as $\mathrm{Bin}(1, p)$ in the binomial case, $\mathrm{Poi}(\lambda)$ in the Poisson case, and $\mathrm{NBin}(1, p)$, the number of successes before the first failure, in the negative binomial case. The objective is to construct one-sided confidence intervals for the mean μ. The focus will be mainly on the upper limit intervals in this paper. The analysis for the lower limit intervals is analogous.

The binomial, negative binomial, and Poisson distributions form the discrete natural exponential family with quadratic variance functions. See Morris (1982) and Brown (1986). A common feature of these distributions is that the variance σ^2 is at most a quadratic function of the mean μ. Indeed

$$\sigma^2 \equiv V(\mu) = \mu + b_* \mu^2, \tag{1}$$

where $\mu = p$ and $b_* = -1$ in the binomial $\mathrm{Bin}(1, p)$ case; $\mu = \lambda$ and $b_* = 0$ in the Poisson $\mathrm{Poi}(\lambda)$ case; and $\mu = p/(1-p)$ and $b_* = 1$ in the negative binomial $\mathrm{NBin}(1, p)$ case.

2.1. Performance of the Wald and score intervals

The Wald interval is the standard confidence interval used in practice. Same as in the two-sided case, the one-sided Wald interval is often the only one-sided confidence procedure given in the introductory statistics texts. In addition to the Wald interval, the score interval is also often used. As mentioned in the introduction the two-sided score interval has satisfactory coverage properties in all three distributions.

Despite the good performance of the score interval in the two-sided problem, we shall show that the score interval does not perform well in the one-sided problem. In this case both the Wald and score intervals suffer a serious systematic bias in the coverage probability. The empirical findings in this section will be reinforced by the theoretical calculations given in Sections 3 and 4. Furthermore, due to the duality between the score interval and the score test, the deficiency in the coverage of the one-sided score interval also has direct implications to the popular one-sided score test. See Section 6 for discussions on hypothesis testing.

Throughout the paper set $X = \sum_{i=1}^{n} X_i$ and $\hat{\mu} = \bar{X} = \sum_{i=1}^{n} X_i/n$. Denote by κ the $100(1-\alpha)$th percentile of the standard normal distribution.

The Wald interval: The Wald interval is constructed based on the normal approximation

$$W_{\rm n} = \frac{\sqrt{n}(\hat{\mu}-\mu)}{V^{1/2}(\hat{\mu})} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1). \tag{2}$$

The $100(1-\alpha)\%$ upper limit Wald interval is defined as

$$\mathrm{CI}_{\rm W}^{\rm u} = [0, \hat{\mu} + \kappa V^{1/2}(\hat{\mu}) n^{-1/2}] = [0, \hat{\mu} + \kappa(\hat{\mu} + b_* \hat{\mu}^2)^{1/2} n^{-1/2}] \tag{3}$$

and the $100(1-\alpha)\%$ lower limit Wald interval is given by

$$\mathrm{CI}_{\rm W}^{\rm l} = [\hat{\mu} - \kappa V^{1/2}(\hat{\mu}) n^{-1/2}, 1]$$

in the binomial case and

$$\mathrm{CI}_{\rm W}^{\rm l} = [\hat{\mu} - \kappa V^{1/2}(\hat{\mu}) n^{-1/2}, \infty)$$

in the Poisson and negative binomial cases. As mentioned earlier, our analysis will be focused on the upper limit intervals. For reason of space we shall combine the three cases and simply write hereafter the upper limit of the lower limit intervals as ∞ for all three distributions with the understanding that it is actually 1 in the binomial case.

The score interval: The score interval is constructed using the normal approximation

$$Z_{\rm n} = \frac{\sqrt{n}(\hat{\mu}-\mu)}{V^{1/2}(\mu)} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1) \tag{4}$$

and the inversion of the score test. In testing the one-sided hypotheses $\mathrm{H}_0\colon \mu \geqslant \mu_0$ against $H_{\rm a}\colon \mu < \mu_0$ at the significance level α, the score test rejects the null hypothesis whenever $n^{1/2}(\bar{X}-\mu_0)/V^{1/2}(\mu_0) < -\kappa$. Solving a simple quadratic equation yield the $100(1-\alpha)\%$ upper limit score interval

$$\mathrm{CI}_{\rm S}^{\rm u} = \left[0, \frac{X+\kappa^2/2}{n-b_*\kappa^2} + \frac{\kappa n^{1/2}}{n-b_*\kappa^2}\left(V(\hat{\mu}) + \frac{\kappa^2}{4n}\right)^{1/2}\right]. \tag{5}$$

The $100(1-\alpha)\%$ lower limit score interval is constructed similarly and has the form

$$\mathrm{CI}_{\rm S}^{\rm l} = \left[\frac{X+\kappa^2/2}{n-b_*\kappa^2} - \frac{\kappa n^{1/2}}{n-b_*\kappa^2}\left(V(\hat{\mu}) + \frac{\kappa^2}{4n}\right)^{1/2}, \infty)\right].$$

Example 1. Consider first the binomial problem. Fig. 1 plots the coverage of the 99% upper limit Wald interval and the upper limit score interval for p with $n = 30$. There

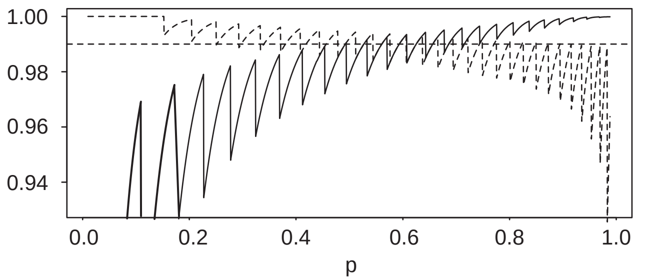

Fig. 1. Coverage probability of the upper limit Wald interval (solid) and the upper limit score interval (dashed) for a binomial proportion p with $n = 30$ and $\alpha = 0.01$.

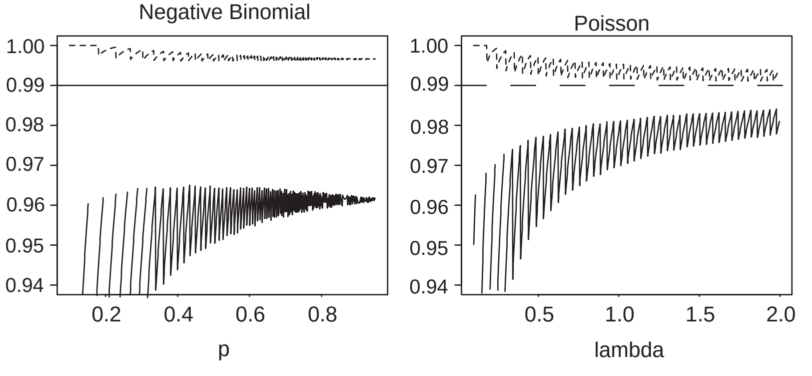

Fig. 2. Coverage probability of the upper limit Wald interval (solid) and the upper limit score interval (dashed) for a negative binomial mean and Poisson mean with $n = 30$.

is a pronounced systematic bias in the coverage for both intervals. It is interesting to note that the systematic biases for the two intervals are in the exact opposite directions of each other. We shall see that this phenomenon is common to all three distributions.

Example 2. Consider the negative binomial and Poisson cases. Fig. 2 plots the coverage probabilities of the 99% Wald and score intervals for a negative binomial mean and Poisson mean with $n=30$. In the Poisson case the coverage is in fact a function of $n\lambda$. The most striking aspect of the plot is that for both distributions the coverage of the Wald interval never reaches 0.99 while the coverage of the score interval always stays above 0.99. In both cases there are serious systematic negative bias in the coverage of the Wald interval and persistent positive bias in the coverage of the score interval. Especially in the negative binomial case the coverage of the Wald interval is far below the nominal level of 0.99 and the coverage of the score interval is close. What was observed in the previous example in the binomial proportion problem resurfaces in a slightly different way in the negative binomial mean and Poisson mean problems.

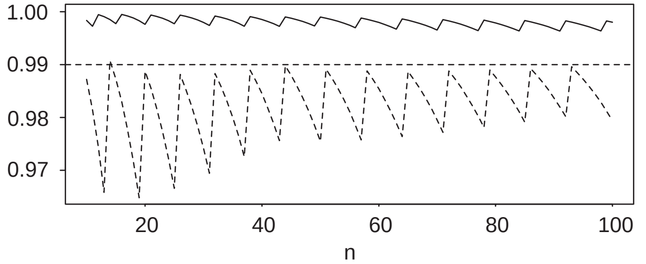

Fig. 3. Coverage probability of the upper limit Wald interval (solid) and the upper limit score interval (dashed) for a binomial proportion p with $p = 0.9$, $n = 10$–100 and $\alpha = 0.01$.

Example 3. Consider again the binomial case and the coverage of the one-sided Wald interval and score interval as a function of the sample size n, for a fixed p, say, $p = 0.9$. Fig. 3 shows that in this case the coverage of the Wald interval is consistently above the nominal level and the coverage of the score interval is nearly always below the nominal level. Once again we see the serious systematic bias in the coverage probabilities of the Wald and score intervals.

2.2. The alternative confidence intervals

The examples in Section 2.1 clearly demonstrate that both the Wald and score intervals perform poorly and erratically and better alternative intervals are needed. In this section we introduce two alternative intervals—the Jeffreys prior interval and the second-order corrected interval. A common feature is that both have good probability matching properties which are particularly important for one-sided intervals.

The Jeffreys interval: The non-informative Jeffreys prior plays a special role in the Bayesian analysis. See e.g. Berger (1985). In particular, the Jeffreys prior is the unique first-order probability matching prior for a real-valued parameter (with no nuisance parameter). See Ghosh (1994). In our setting, simple calculation shows that the Fisher information about μ is $I(\mu) = n(\mu + b_*\mu^2)^{-1}$ and thus the Jeffreys prior is proportional to $I^{1/2}(\mu) = n^{1/2}(\mu + b_*\mu^2)^{-1/2}$. Denote the posterior distribution by J. Then the $100(1-\alpha)\%$ *upper limit and lower limit Jeffreys intervals* for μ are respectively defined as

$$\mathrm{CI}_J^u = [0, J_{1-\alpha}], \quad \text{and} \quad \mathrm{CI}_J^l = [J_\alpha, \infty), \tag{6}$$

where $J_{1-\alpha}$ and J_α are respectively the $1-\alpha$ and α quantiles of the posterior distribution based on n observations. The construction of the one-sided Jeffreys intervals is analogous to that of the two-sided Jeffreys interval given in Brown et al. (2003). Now consider the three distributions separately.

- *Binomial*: The Jeffreys prior is $\mathrm{Beta}(\frac{1}{2}, \frac{1}{2})$ and the posterior is $\mathrm{Beta}(X + \frac{1}{2}, n - X + \frac{1}{2})$. Thus the $100(1-\alpha)\%$ upper limit and lower limit Jeffreys intervals for p are, respectively,

$$\mathrm{CI}_J^u = [0, B_{1-\alpha, X+1/2, n-X+1/2}] \quad \text{and} \quad \mathrm{CI}_J^l = [B_{\alpha, X+1/2, n-X+1/2}, 1]. \tag{7}$$

- *Negative binomial*: The Jeffreys interval is transformation-invariant. The Jeffreys prior for p is proportional to $p^{-1/2}(1-p)^{-1}$ and the posterior is Beta$(X+\frac{1}{2}, n)$. Thus the $100(1-\alpha)\%$ upper limit and lower limit Jeffreys intervals for p are, respectively,

$$\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}(p) = [0, p_{\mathrm{l}}] = [0, B_{1-\alpha, X+1/2, n}] \quad \text{and} \quad \mathrm{CI}_{\mathrm{J}}^{\mathrm{l}}(p) = [p_{\mathrm{u}}, 1] = [B_{\alpha, X+1/2, n}, 1].$$

Therefore, the upper limit and lower limit Jeffreys intervals for $\mu = p/(1-p)$ are, respectively,

$$\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}} = \left[0, \frac{p_{\mathrm{l}}}{1-p_{\mathrm{l}}}\right] \quad \text{and} \quad \mathrm{CI}_{\mathrm{J}}^{\mathrm{l}} = \left[\frac{p_{\mathrm{u}}}{1-p_{\mathrm{u}}}, \infty\right) \tag{8}$$

- *Poisson*: The Jeffreys prior for λ is proportional to $\lambda^{-1/2}$ which is improper and the posterior is Gamma $(X+1/2, 1/n)$. Hence the $100(1-\alpha)\%$ upper limit and lower limit Jeffreys intervals for λ are, respectively,

$$\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}} = [0, G_{1-\alpha, X+1/2, 1/n}] \quad \text{and} \quad \mathrm{CI}_{\mathrm{J}}^{\mathrm{l}} = [G_{\alpha, X+1/2, 1/n}, \infty). \tag{9}$$

The second-order corrected interval: Asymptotic theory has a long history of providing motivation and guidance for the construction of procedures with good finite-sample performance. For the one-sided interval estimation problem the Edgeworth expansion has been used in Hall (1982) for the construction of first-order corrected confidence intervals for a binomial proportion and Poisson mean.

The second-order corrected intervals given below are constructed based on the Edgeworth expansion to explicitly eliminate both the first and second-order systematic bias in the coverage. Although the Edgeworth expansions are mainly regarded as asymptotic approximations, two-term Edgeworth expansions are very accurate for the two-sided problem even for relatively small and moderate n. See Brown et al. (2002, 2003). We will see that this is also true for the one-sided problem and the second-order corrected intervals perform well for small and moderate sample sizes.

Let $\eta = \frac{1}{3}\kappa^2 + \frac{1}{6}$, $\gamma_1 = b_*(\frac{13}{18}\kappa^2 + \frac{17}{18})$ and $\gamma_2 = \frac{1}{18}\kappa^2 + \frac{7}{36}$. Let $\tilde{\mu} = (X+\eta)/(n - 2\eta b_*)$. The $100(1-\alpha)\%$ *upper limit second-order corrected interval* is defined as

$$\mathrm{CI}_2^{\mathrm{u}} = [0, \tilde{\mu} + \kappa(V(\hat{\mu}) + (\gamma_1 V(\hat{\mu}) + \gamma_2)n^{-1})^{1/2} n^{-1/2}] \tag{10}$$

and the $100(1-\alpha)\%$ *lower limit second-order corrected interval* is defined as

$$\mathrm{CI}_2^{\mathrm{l}} = [\tilde{\mu} - \kappa(V(\hat{\mu}) + (\gamma_1 V(\hat{\mu}) + \gamma_2)n^{-1})^{1/2} n^{-1/2}, \infty).$$

Remark. Comparing the second-order corrected interval $\mathrm{CI}_2^{\mathrm{u}}$ with the Wald interval $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, $\tilde{\mu}$ in $\mathrm{CI}_2^{\mathrm{u}}$ can be viewed as a (first-order) correction to $\hat{\mu}$ in the Wald interval $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$ and $(V(\hat{\mu}) + (\gamma_1 V(\hat{\mu}) + \gamma_2)n^{-1})^{1/2}$ in $\mathrm{CI}_2^{\mathrm{u}}$ as a (second-order) correction to $V^{1/2}(\hat{\mu})$ in $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$. Indeed, Hall (1982) showed that the interval $[0, \tilde{\mu} + \kappa V^{1/2}(\hat{\mu})n^{-1/2}]$ eliminate the first-order systematic bias in the binomial and Poisson cases. The reasons for choosing the specific values of $\tilde{\mu}$, γ_1 and γ_2 are given in the proof of Theorem 4.

Fig. 4 plots the coverage probabilities of the four upper limit confidence intervals for a binomial proportion with $n = 30$ and $\alpha = 0.01$. It shows that the Jeffreys and

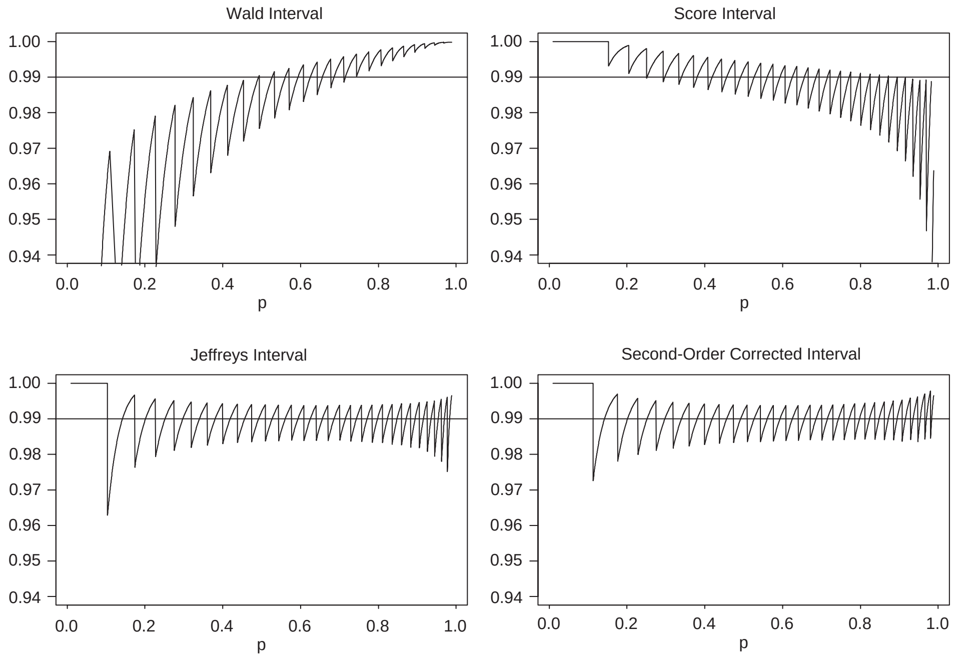

Fig. 4. Coverage of the four intervals for a binomial proportion with $n = 30$ and $\alpha = 0.01$.

second-order corrected intervals have superior performance relative to both the Wald and score intervals. The two alternative intervals nearly eliminate the systematic bias in the coverage probability. In addition, although oscillation in the coverage is unavoidable for non-randomized confidence intervals in these lattice problems, the amount of oscillation in the coverage of the Jeffreys and second-order corrected intervals is smaller than that of the Wald and score intervals.

3. Edgeworth expansions

The Edgeworth expansions provide an accurate and useful tool in analyzing the coverage properties of confidence intervals. See for example Brown et al. (2002). The Edgeworth expansion is particularly useful in understanding analytically why the score interval performs better in the two-sided problem than in the one-sided problem.

Define

$$g(\mu, z) = g(\mu, z, n) = n\mu + n^{1/2}\sigma z - (n\mu + n^{1/2}\sigma z)_-, \tag{11}$$

where $(x)_-$ denotes the largest integer less than or equal to x. So $g(\mu, z)$ is the fractional part of $n\mu + n^{1/2}\sigma z$. We suppress in (11) and later the dependence of g on n and denote

$$Q_1(\mu, z) = g(\mu, z) - \tfrac{1}{2}, \quad \text{and} \quad Q_2(\mu, z) = -\tfrac{1}{2}\, g^2(\mu, z) + \tfrac{1}{2}\, g(\mu, z) - \tfrac{1}{12}. \tag{12}$$

Note that the functions $Q_1(\mu,z)$ and $Q_2(\mu,z)$ are oscillatory functions. They appear in the Edgeworth expansions to precisely capture the oscillation in the coverage probability.

A two-term Edgeworth expansion of the coverage probability has a general form of

$$P(\mu \in \mathrm{CI}) = 1 - \alpha + S_1 \cdot n^{-1/2} + \mathrm{Osc}_1 \cdot n^{-1/2} + S_2 \cdot n^{-1} + \mathrm{Osc}_2 \cdot n^{-1} + \mathrm{O}(n^{-32}), \tag{13}$$

where the first $\mathrm{O}(n^{-1/2})$ term, $S_1 n^{-1/2}$, and the first $\mathrm{O}(n^{-1})$ term, $S_2 n^{-1}$, are respectively the first- and second-order smooth terms, and $\mathrm{Osc}_1 \cdot n^{-1/2}$ and $\mathrm{Osc}_2 \cdot n^{-1}$ are the oscillatory terms. The smooth terms capture the systematic bias in the coverage as seen in the examples in Section 2.1. A one-sided confidence interval is called *first-order probability matching* if the first-order smooth term $S_1 n^{-1/2}$ is vanishing and is called *second-order probability matching* if both the first- and second-order smooth terms are zero. See Ghosh (1994, 2001) for further details on probability matching and confidence sets. See also the discussions in Brown et al. (2001).

We now give the two-term Edgeworth expansions for the four upper limit confidence intervals. Let $0 < \alpha < 1$ and assume that μ is a fixed point in the interior of the parameter spaces. That is, $0 < p < 1$ in the binomial and negative binomial cases and $\lambda > 0$ in the Poisson case. Denote by ϕ and Φ respectively the density function and the cumulative density function of a standard Normal distribution.

Theorem 1. *Let* z_{W} *be defined as in* (38) *in the appendix. Suppose* $n\mu + n^{1/2}\sigma z_{\mathrm{W}}$ *is not an integer. Then the coverage probability of the Wald interval* $\mathrm{CI}^{\mathrm{u}}_{\mathrm{W}}$ *defined in* (3) *satisfies*

$$\begin{aligned} P_{\mathrm{W}} = P(\mu \in \mathrm{CI}^{\mathrm{u}}_{\mathrm{W}}) &= (1-\alpha) - \frac{1}{6}(2\kappa^2+1)(1+2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2} \\ &+ Q_1(\mu, z_{\mathrm{W}})\sigma^{-1}\phi(\kappa)n^{-1/2} \\ &+ \left\{-\frac{b_*}{36}(8\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{1}{36\sigma^2}(2\kappa^5 + \kappa^3 + 3\kappa)\right\}\phi(\kappa)n^{-1} \\ &+ \left\{\frac{1}{6}(2\kappa^2+3)(1+2b_*\mu)Q_1(\mu, z_{\mathrm{W}}) + Q_2(\mu, z_{\mathrm{W}})\right\} \\ &\times \sigma^{-2}\kappa\phi(\kappa)n^{-1} + \mathrm{O}(n^{-3/2}). \end{aligned} \tag{14}$$

Theorem 2. *Suppose* $n\mu - n^{1/2}\sigma\kappa$ *is not an integer. The coverage probability of the score interval* $\mathrm{CI}^{\mathrm{u}}_{\mathrm{S}}$ *defined in* (5) *satisfies*

$$\begin{aligned} P_{\mathrm{S}} = P(\mu \in \mathrm{CI}^{\mathrm{u}}_{\mathrm{S}}) &= (1-\alpha) + \frac{1}{6}(\kappa^2-1)(1+2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2} \\ &+ Q_1(\mu, \kappa)\sigma^{-1}\phi(\kappa)n^{-1/2} \\ &+ \left\{-\frac{b_*}{36}(2\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{1}{72\sigma^2}(\kappa^5 - 7\kappa^3 + 6\kappa)\right\}\phi(\kappa)n^{-1} \end{aligned}$$

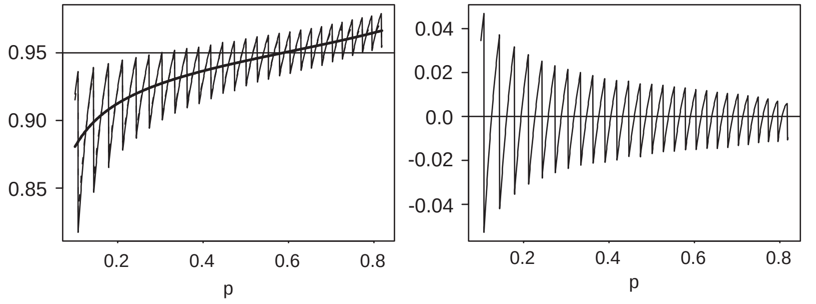

Fig. 5. Two-term Edgeworth expansion for the upper limit Wald interval for a binomial proportion with $n=40$ and $\alpha = 0.05$. Left panel: The solid line is the true coverage, the dotted line is the two-term Edgeworth expansion and the smooth curve is the non-oscillatory terms in the expansion. Right panel: oscillatory terms in the Edgeworth expansion.

$$+\left\{\frac{1}{6}(3-\kappa^2)(1+2b_*\mu)Q_1(\mu,\kappa)+Q_2(\mu,-\kappa)\right\}$$
$$\times\sigma^{-2}\kappa\phi(\kappa)n^{-1}+\mathrm{O}(n^{-3/2}). \tag{15}$$

Remark. It is clear from (14) and (15) that both the Wald and score intervals are not first-order probability matching. The first-order smooth terms $-\frac{1}{6}(2\kappa^2+1)(1+2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2}$ in (14) and $\frac{1}{6}(\kappa^2-1)(1+2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2}$ in (15) are the main contributor of the systematic bias seen in the examples of Section 2.1. See Fig. 5 in Section 4. On the other hand, two-sided intervals make compensating one sided errors and the first-order smooth term is canceled. Thus first-order probability matching has no obvious bearing on the coverage for the two-sided problem. This is why the score interval has much better coverage performance in the two-sided case than in the one-sided case.

The following gives a general expression for the two-term Edgeworth expansion of the coverage probability of the Jeffreys interval in all three cases.

Theorem 3. *Denote by* $\mathrm{CI}_\mathrm{J}^\mathrm{u}$ *the upper limit Jeffreys prior interval as defined in* (7) *in the binomial case*, (8) *in the negative binomial case, and* (9) *in the Poisson case. Let* z_J *be defined as in* (42) *in the appendix. Suppose* $n\mu+n^{1/2}\sigma z_\mathrm{J}$ *is not an integer. Then the coverage probability of* $\mathrm{CI}_\mathrm{J}^\mathrm{u}$ *satisfies*

$$P_\mathrm{J}=P(\mu\in\mathrm{CI}_\mathrm{J}^\mathrm{u})=(1-\alpha)+Q_1(\mu,z_\mathrm{J})\sigma^{-1}\phi(\kappa)n^{-1/2}-\frac{1}{24\sigma^2}\kappa\phi(\kappa)n^{-1}$$
$$+\left[\frac{1}{3}(1+2b_*\mu)Q_1(\mu,z_\mathrm{J})+Q_2(\mu,z_\mathrm{J})\right]\sigma^{-2}\kappa\phi(\kappa)n^{-1}+\mathrm{O}(n^{-3/2}). \tag{16}$$

Theorem 4. *Let* z_2 *be defined as in* (38) *in the appendix. Suppose* $n\mu+n^{1/2}\sigma z_2$ *is not an integer. Then the coverage probability of the second-order corrected interval* CI_2^u

defined in (10) *satisfies*

$$P_2 = P(\mu \in \mathrm{CI}_2^{\mathrm{u}}) = (1-\alpha) + Q_1(\mu, z_2)\sigma^{-1}\phi(\kappa)n^{-1/2} + \left\{ \tfrac{1}{3}(1+2b_*\mu)Q_1(\mu, z_2) + Q_2(\mu, z_2) \right\} \sigma^{-2}\kappa\phi(\kappa)n^{-1} + \mathrm{O}(n^{-3/2}). \quad (17)$$

Recall that for the binomial case $b_* = -1$, $\mu = p$, and $\sigma = (pq)^{1/2}$; for the negative binomial case $b_* = 1$, $\mu = p/q$, and $\sigma = p^{1/2}q$; and for the Poisson case $b_* = 0$, $\mu = \lambda$, and $\sigma = \lambda^{1/2}$. The Edgeworth expansions for the three specific distributions can be obtained easily from Theorems 1–4 by plugging in corresponding b_*, μ, and σ.

Remark. The expansions for the lower limit intervals can be obtained by first replacing α by $1-\alpha$ and κ by $-\kappa$ in the expansion for the coverage of the upper limit intervals, and then subtracting it from 1.

Remark. Inversion of the likelihood ratio test is another common method for the construction of confidence procedures. Brown et al. (2002, 2003) show that the likelihood interval performs very well in the two-sided problem. However, the one-sided likelihood ratio interval is not first-order probability matching. The coverage contains non-negligible first-order systematic bias. In addition, the likelihood ratio interval is relatively difficult to compute. On the other hand, by construction, the confidence interval given in Hall (1982) is first-order probability matching. However, it still contains non-negligible second-order systematic bias in the coverage and does not perform well for parameter values near the boundaries. Since these two confidence intervals do not perform as well as either the Jeffreys interval or the second-order corrected interval, for reason of space they are not discussed in detail in the present paper. See Cai (2003) for more analysis on these two intervals.

4. Comparison of coverage probability

In this section, using the two-term Edgeworth expansions derived in Section 3, we compare the coverage properties of the standard Wald interval $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, the score interval $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$, the Jeffreys interval $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$ and the second-order corrected interval $\mathrm{CI}_2^{\mathrm{u}}$.

The Edgeworth expansions decompose the coverage probability into an oscillatory component and a smooth component which captures the main regularity in the coverage. Two-term Edgeworth expansions of the coverage are very accurate even for relatively small and moderate n. The left panel of Fig. 5 plots the two-term Edgeworth expansion and the exact coverage probability of the upper limit Wald interval for a binomial proportion with $n = 40$ and $\alpha = 0.05$. The two-term approximation is virtually indistinguishable from the exact coverage probability when p is not too close to 0 or 1. The smooth curve in the plot is the non-oscillatory component in the Edgeworth expansion which can be viewed as a smooth approximation to the coverage probability. The right panel plots the oscillation terms in the Edgeworth expansion which vary almost symmetrically around 0.

The smooth terms in the Edgeworth expansion measure the systematic bias in the coverage. They shall be used as the basis for comparison of coverage properties of the four intervals. Denote the sum of the $O(n^{-1/2})$ and $O(n^{-1})$ smooth terms in the two term expansions of the coverage probabilities P_S, P_2, P_J, and P_W by B_S, B_2, B_J, and B_W, respectively. Then directly from (14) to (17), we have

$$B_2 = 0, \tag{18}$$

$$B_S = \frac{1}{6}(\kappa^2 - 1)(1 + 2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2} - \left\{ \frac{b_*}{36}(2\kappa^5 - 11\kappa^3 + 3\kappa) + \frac{1}{72\sigma^2}(\kappa^5 - 7\kappa^3 + 6\kappa) \right\} \phi(\kappa)n^{-1}, \tag{19}$$

$$B_J = -\frac{1}{24\sigma^2}\,\kappa\phi(\kappa)n^{-1}, \tag{20}$$

$$B_W = -\frac{1}{6}(2\kappa^2 + 1)(1 + 2b_*\mu)\sigma^{-1}\phi(\kappa)n^{-1/2} - \left\{ \frac{b_*}{36}(8\kappa^5 - 11\kappa^3 + 3\kappa) + \frac{1}{36\sigma^2}(2\kappa^5 + \kappa^3 + 3\kappa) \right\} \phi(\kappa)n^{-1}. \tag{21}$$

Comparison of the coefficients of the $n^{-1/2}$ and n^{-1} terms in (18)–(21) yields some interesting conclusions. The coefficients of the $n^{-1/2}$ and/or the n^{-1} term in the expressions (18)–(21) determine the direction and the magnitude of the systematic bias in the coverage probability of a specific interval.

First note that the signs of the first-order smooth term in (21) and (19) are different. This shows that the phenomenon observed in Figs. 1 and 2 that the systematic biases for the Wald and score intervals are in the exact opposite directions is true in general so long as $\kappa > 1$. Note also that the coefficient of the $O(n^{-1/2})$ term in B_W always has a larger magnitude than the corresponding term in B_S and hence CI_W^u has more serious systematic bias than CI_S^u.

Now consider the three distributions separately. First the binomial case. In this case the coefficient $b_* = -1$. The $O(n^{-1/2})$ bias term in the coverage of both the Wald and score intervals changes sign at $p = \frac{1}{2}$. For $p < \frac{1}{2}$, the score interval has systematic $O(n^{-1/2})$ positive bias and the Wald interval has serious $O(n^{-1/2})$ negative bias; and for $p > \frac{1}{2}$, the $O(n^{-1/2})$ bias term for the score interval becomes negative and that for the Wald interval turns positive. The $O(n^{-1})$ bias for the Jeffreys interval is not significant. And by construction the interval CI_2^u has both vanishing $O(n^{-1/2})$ and $O(n^{-1})$ bias terms.

Fig. 6 displays the systematic bias in coverage of each interval for the binomial case with $n = 40$ and $\alpha = 0.05$. It is clear that the coverage of the upper limit Wald interval is seriously negatively biased for $p < \frac{1}{2}$ and seriously positively biased for $p > \frac{1}{2}$. The score interval CI_S^u does not perform well either. It behaves in the exact opposite direction as the Wald interval; the coverage has a consistent positive bias for $p < \frac{1}{2}$ and a systematic negative bias for $p > \frac{1}{2}$.

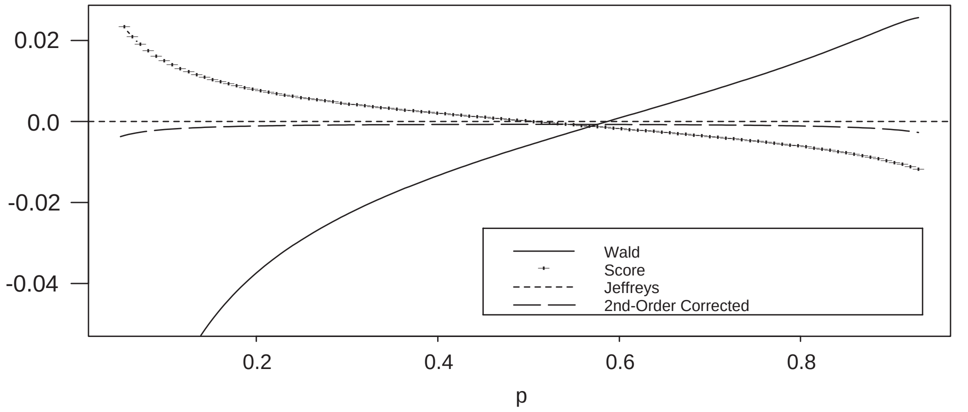

Fig. 6. Comparison of the non-oscillatory terms in the binomial case with $n=40$ and $\alpha=0.05$.

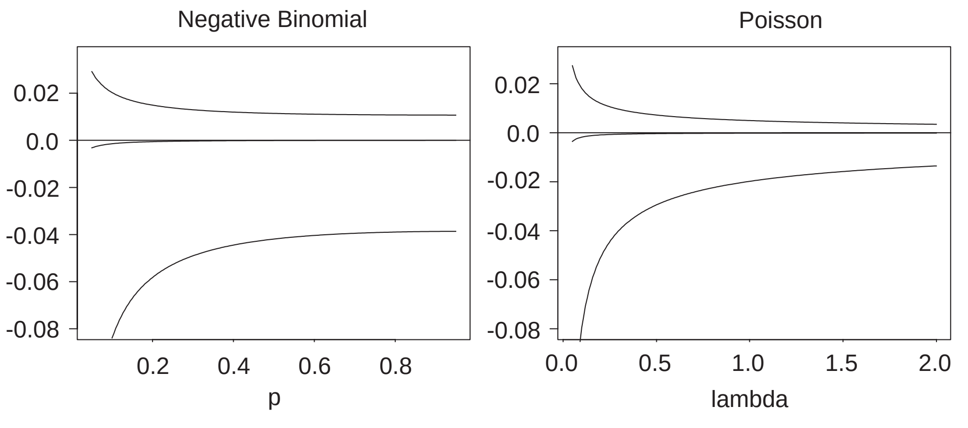

Fig. 7. Comparison of the non-oscillatory terms in the negative binomial and Poisson cases with $n=40$ and $\alpha=0.05$. From top to bottom: B_S, B_2, B_J, and B_W. Note that B_2 is identically 0.

In the negative binomial case the coefficient $b_*=1$ and in the Poisson case $b_*=0$. Comparison of Eqs. (18)–(21) immediately gives the strict ordering of the systematic bias of the coverage, from largest to smallest, of CI_S^u, CI_2^u, CI_J^u, and CI_W^u; the score interval has systematic $O(n^{-1/2})$ positive bias and the Wald interval has serious $O(n^{-1/2})$ negative bias. The alternative intervals CI_2^u and CI_J^u both have vanishing $O(n^{-1/2})$ bias term in the coverage and therefore a much less serious systematic bias problem. In particular, by construction, the interval CI_2^u has vanishing $O(n^{-1})$ bias term as well. So P_2 is centered at the correct nominal level $(1-\alpha)$, up to $O(n^{-1})$ term. The $O(n^{-1})$ bias term of P_J is not zero, but is nearly vanishing. See Fig. 7 below.

Fig. 7 displays the systematic bias for the negative binomial and Poisson cases with $n=40$ and $\alpha=0.05$. It is transparent that there is a consistent serious negative bias in the coverage of the Wald interval in both cases. On the other hand, the coverage of the score interval has a non-negligible positive systematic bias in these two cases.

These comparisons clearly demonstrate that for all three distributions the Jeffreys and second-order corrected intervals have superior coverage performance relative to both the Wald and score intervals.

5. Expansions and comparisons for expected distance from the mean

In addition to the coverage probability, parsimony in length is also important. In this section we provide an expansion for the expected distance from the mean of the four upper limit intervals correct up to the order $O(n^{-3/2})$. Similar to the Edgeworth expansions for the coverage probability, the expansions for the expected distance reveal interesting common structure.

The expansion for the expected distance includes terms of the order $n^{-1/2}$, n^{-1} and $n^{-3/2}$. The coefficient of the $n^{-1/2}$ term is the same for all the intervals, but the coefficients for the n^{-1} and $n^{-3/2}$ terms differ. So, naturally, the coefficients of the n^{-1} and $n^{-3/2}$ terms will be used as a basis for comparison of their expected length.

Theorem 5. *Let U be a generic notation for the upper limit of any of the four intervals, $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$, $\mathrm{CI}_{2}^{\mathrm{u}}$, and $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$ for estimating the mean μ. Then the expected distance from the mean*

$$E(U-\mu)=\kappa(\mu+b_*\mu^2)^{1/2}n^{-1/2}+\delta_1(\kappa,\mu)n^{-1}+\delta_2(\kappa,\mu)n^{-3/2}+\mathrm{O}(n^{-2}), \quad (22)$$

where

$$\delta_1(\kappa,\mu)=0 \;\; for \;\; \mathrm{CI}_{\mathrm{W}}^{\mathrm{u}} \quad (23)$$

$$=\left(\tfrac{1}{3}\kappa^2+\tfrac{1}{6}\right)\cdot(1+2b_*\mu) \;\; for \;\; \mathrm{CI}_{\mathrm{J}}^{\mathrm{u}} \;\; and \;\; \mathrm{CI}_{2}^{\mathrm{u}} \quad (24)$$

$$=\tfrac{1}{2}\kappa^2\cdot(1+2b_*\mu) \;\; for \;\; \mathrm{CI}_{\mathrm{S}}^{\mathrm{u}} \quad (25)$$

and

$$\delta_2(\kappa,\mu)=-\tfrac{1}{8}\kappa(\mu+b_*\mu^2)^{-1/2} \;\; for \;\; \mathrm{CI}_{\mathrm{W}}^{\mathrm{u}} \quad (26)$$

$$=\tfrac{1}{72}[(2\kappa^3-5\kappa)(\mu+b_*\mu^2)^{-1/2} + (26\kappa^3+34\kappa)b_*(\mu+b_*\mu^2)^{1/2}] \;\; for \;\; \mathrm{CI}_{\mathrm{J}}^{\mathrm{u}} \quad (27)$$

$$=\tfrac{1}{72}[(2\kappa^3-2\kappa)(\mu+b_*\mu^2)^{-1/2} + (26\kappa^3+34\kappa)b_*(\mu+b_*\mu^2)^{1/2}] \;\; for \;\; \mathrm{CI}_{2}^{\mathrm{u}} \quad (28)$$

$$=\tfrac{1}{72}[(9\kappa^3-9\kappa)(\mu+b_*\mu^2)^{-1/2}+72\kappa^3b_*(\mu+b_*\mu^2)^{1/2}] \;\; for \;\; \mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}. \quad (29)$$

It is interesting to note that between the two alternative intervals, up to an error of order $O(n^{-2})$, CI_J^u is always slightly shorter than CI_2^u across all three distributions.

First consider the binomial case. In this case there are two different rankings of the four confidence intervals in terms of the expected length for $p < \frac{1}{2}$ and $p \geqslant \frac{1}{2}$. Denote the expected distance from μ of the upper limit of CI_W^u, CI_J^u, CI_2^u, and CI_S^u by L_W, L_J, L_2 and L_S, respectively.

Corollary 1. *Consider the special binomial case. Then the expected distance of the upper limit of* CI_W^u, CI_J^u, CI_2^u, *and* CI_S^u *from the mean* μ *admit the expansions*:

$$E(L_W) = \kappa(pq)^{1/2}n^{-1/2} - \tfrac{1}{8}\kappa(pq)^{-1/2}n^{-3/2} + O(n^{-2}),$$

$$E(L_J) = \kappa(pq)^{1/2}n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)(1-2p)n^{-1} + \tfrac{1}{72}[(2\kappa^3 - 5\kappa)(pq)^{-1/2} - (26\kappa^3 + 34\kappa)(pq)^{1/2}]n^{-3/2} + O(n^{-2}),$$

$$E(L_2) = \kappa(pq)^{1/2}n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)(1-2p)n^{-1} + \tfrac{1}{72}[(2\kappa^3 - 2\kappa)(pq)^{-1/2} - (26\kappa^3 + 34\kappa)(pq)^{1/2}]n^{-3/2} + O(n^{-2}),$$

$$E(L_S) = \kappa(pq)^{1/2}n^{-1/2} + \tfrac{1}{2}\kappa^2(1-2p)n^{-1} + \tfrac{1}{8}[(\kappa^3 - \kappa)(pq)^{-1/2} - 8\kappa^3(pq)^{1/2}]n^{-3/2} + O(n^{-2}).$$

The ranking of the expected distances depends on the value of p. Assume that $\kappa \geqslant 1$. For every $p < \frac{1}{2}$ comparing the coefficients in the $O(n^{-1})$ term in Corollary 1 immediately yields that the ranking is CI_W^u, CI_J^u, CI_2^u, and CI_S^u, from the shortest to the longest. For $p \geqslant \frac{1}{2}$ the ranking is CI_S^u, CI_J^u, CI_2^u, and CI_W^u, from the shortest to the longest. For all $0 < p < 1$, CI_2^u is always slightly longer than the Jeffreys interval and the expected distances of these two intervals are always between those of the Wald and score intervals.

Now consider the cases of Poisson and negative binomial distributions. In these two cases there is an even stronger uniform ranking of the confidence intervals in terms of the expected distance from the mean pointwise for every value of the parameter.

Corollary 2. *Consider the special Poisson case. Then the expected distance of the upper limit of* CI_W^u, CI_J^u, CI_2^u, *and* CI_S^u *from the mean* μ *admit the expansions*:

$$E(L_W) = \kappa\lambda^{1/2}n^{-1/2} - \tfrac{1}{8}\kappa\lambda^{-\frac{1}{2}}n^{-\frac{3}{2}} + O(n^{-2}),$$

$$E(L_J) = \kappa\lambda^{1/2}n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)n^{-1} + \tfrac{1}{72}(2\kappa^3 - 5\kappa)\lambda^{-1/2}n^{-3/2} + O(n^{-2}),$$

$$E(L_2) = \kappa\lambda^{1/2}n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)n^{-1} + \tfrac{1}{36}(\kappa^3 - \kappa)\lambda^{-1/2}n^{-3/2} + O(n^{-2}),$$

$$E(L_S) = \kappa\lambda^{1/2}n^{-1/2} + \tfrac{1}{2}\kappa^2 n^{-1} + \tfrac{1}{8}(\kappa^3 - \kappa)\lambda^{-1/2}n^{-3/2} + O(n^{-2}).$$

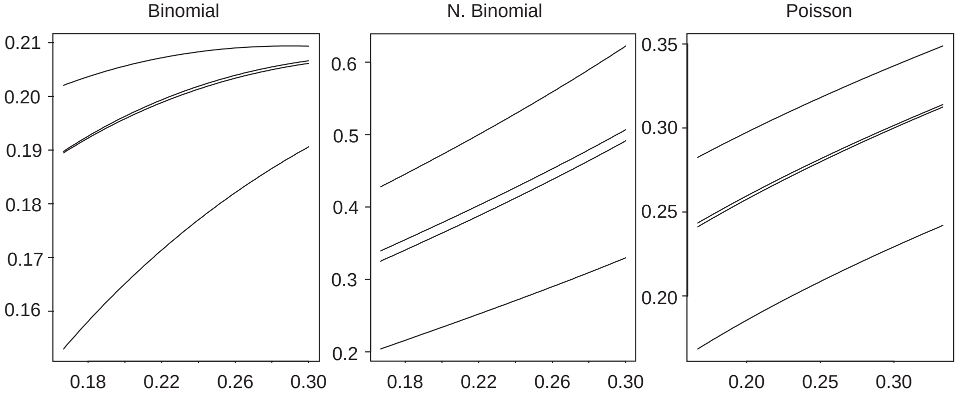

Fig. 8. Expected distance of the upper limit of the four confidence intervals from the mean μ for $n=30$ and $\alpha=0.01$. For all three distributions, from bottom to top, the expected distance of the upper limit of CI_{W}, $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$, $\mathrm{CI}_{2}^{\mathrm{u}}$, and $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$. The expected distances of the upper limits of $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$ and $\mathrm{CI}_{2}^{\mathrm{u}}$ (dark lines in the middle) are almost indistinguishable in the binomial and Poisson cases.

Hence, up to the error n^{-2}, very interestingly, for every $\lambda > 0$, the ranking of the intervals is $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$, $\mathrm{CI}_{2}^{\mathrm{u}}$, and $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$, from the shortest to the longest, as long as $\kappa \geqslant 1$. So we have a uniform ranking of the intervals for all $\lambda > 0$.

The exact same ranking holds in the negative binomial case.

Corollary 3. *Consider the upper limit confidence intervals for $\mu = p/q$ in the negative binomial case. Then the expected distance of the upper limit of* $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$, $\mathrm{CI}_{2}^{\mathrm{u}}$, *and* $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$ *from the mean μ admit the expansions*:

$$E(L_{\mathrm{W}}) = \kappa p^{1/2} q^{-1} n^{-1/2} - \tfrac{1}{8}\kappa p^{-1/2} q n^{-3/2} + \mathrm{O}(n^{-2}),$$

$$E(L_{\mathrm{J}}) = \kappa p^{1/2} q^{-1} n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)(1+p) q^{-1} n^{-1} + \tfrac{1}{72}[(2\kappa^3 - 5\kappa) p^{-1/2} q - (26\kappa^3 + 34\kappa) p^{1/2} q^{-1}] n^{-3/2} + \mathrm{O}(n^{-2}),$$

$$E(L_{2}) = \kappa p^{1/2} q^{-1} n^{-1/2} + \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)(1+p) q^{-1} n^{-1} + \tfrac{1}{72}[(2\kappa^3 - 2\kappa) p^{-1/2} q - (26\kappa^3 + 34\kappa) p^{1/2} q^{-1}] n^{-3/2} + \mathrm{O}(n^{-2}),$$

$$E(L_{\mathrm{S}}) = \kappa p^{1/2} q^{-1} n^{-1/2} + \tfrac{1}{2}\kappa^2 (1+p) q^{-1} n^{-1} + \left[\tfrac{1}{8}(\kappa^3 - \kappa) p^{-1/2} q + \kappa^3 p^{1/2} q^{-1}\right] n^{-3/2} + \mathrm{O}(n^{-2}).$$

Fig. 8 plots the expected distance of the upper limit of the four confidence intervals from the mean μ for $n=30$ and $\alpha=0.01$. The values of p (for the binomial and negative binomial cases) and λ (for the Poisson case) vary from $\frac{1}{6}$ to $\frac{1}{3}$. For these values of p and λ the ranking of the intervals from the shortest to the longest is $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$, $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$, $\mathrm{CI}_{2}^{\mathrm{u}}$, and $\mathrm{CI}_{\mathrm{S}}^{\mathrm{u}}$ for all three distributions. In the cases of binomial and Poisson distributions the expected distances of $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$ and $\mathrm{CI}_{2}^{\mathrm{u}}$ are almost indistinguishable.

Considering together with the coverage properties discussed in the earlier sections, we can conclude that for all three distributions the expected distances of the upper limit of the Wald and score intervals are either too short or too long, which is not desirable in either case. The Jeffreys and second-order corrected intervals are better alternatives.

6. One-sided hypothesis testing

The results on one-sided confidence intervals discussed in the earlier sections have direction implications on testing one-sided hypotheses. In the cases of the binomial, negative binomial and Poisson distributions the score test occupies a particularly important position in hypothesis testing. It is often the only test given in many introductory textbooks. Due to the duality between the one-sided score interval and the one-sided score test, the fact that the one-sided score interval contains significant systematic bias in the coverage probability implies that the actual size of the one-sided score test may be far from the nominal level.

Recall that in testing the one-sided hypotheses:

$$\mathrm{H}_0\colon \mu \geqslant \mu_0 \quad \text{versus} \quad \mathrm{H}_\mathrm{a}\colon \mu < \mu_0 \tag{30}$$

at the significance level α, the score test rejects the null hypothesis whenever

$$\frac{n^{1/2}(\bar{X} - \mu_0)}{V^{1/2}(\mu_0)} < -\kappa. \tag{31}$$

The true size of the score test equals type I error probability under $\mu = \mu_0$, which is the same as the non-coverage of the upper limit score interval $1 - P_{\mu_0}(\mu_0 \in \mathrm{CI}_{\mathrm{S}}^{\mathrm{u}})$.

Fig. 9 plots the size of the upper limit score test at the nominal $\alpha = 0.01$ level for a binomial proportion, negative binomial mean, and Poisson mean with $n = 30$. It is clear that in all three cases the actual size not only oscillates as a function of μ, but contains a serious systematic bias as well. In the binomial case, oscillations aside, the

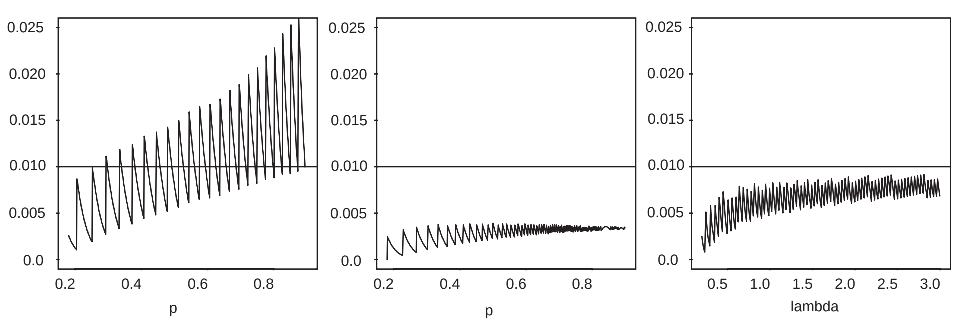

Fig. 9. The size of the upper limit score test at nominal $\alpha = 0.01$ level with $n = 30$. From left to right, binomial, negative binomial, and Poisson.

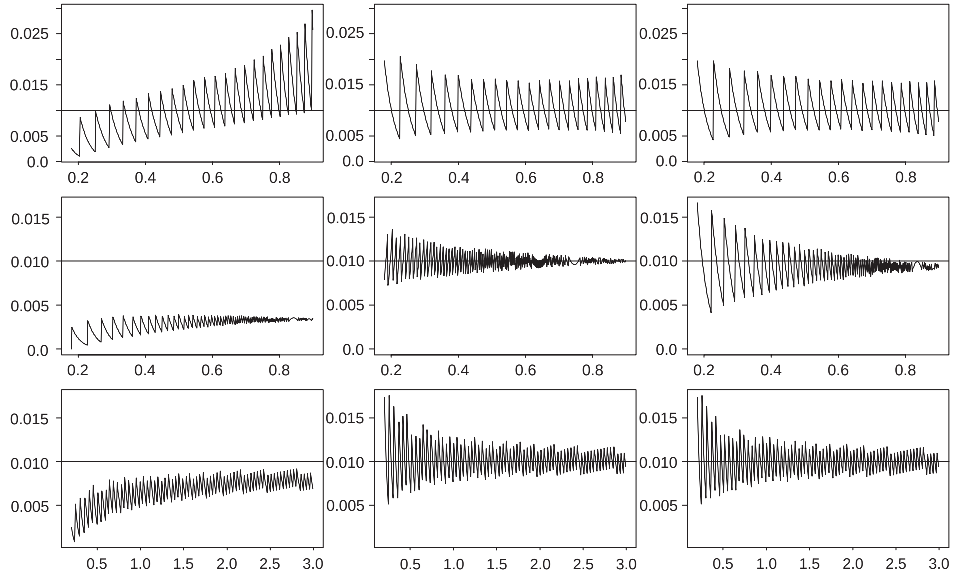

Fig. 10. Actual sizes of the upper limit score, Jeffreys, and second-order corrected tests (from left to right) with nominal level $\alpha = 0.01$ and $n = 30$. From top to bottom: binomial, negative binomial, and Poisson.

size is biased below the nominal level for $p < \frac{1}{2}$ and biased above the nominal level for $p > \frac{1}{2}$. In the cases of negative binomial and Poisson distributions, the score test is very conservative. The size is consistently below the nominal level and the difference is especially significant in the negative binomial case. For small to moderate sample size n, the actual size of the one-sided score interval can be far from the nominal significance level.

Just as inversion of hypothesis tests yields confidence intervals, inversion of the confidence intervals also yields tests. In testing the one-sided hypotheses (30) a nominal level α test can be obtained by rejecting H_0 whenever the null value μ_0 is not in a level $100(1-\alpha)\%$ upper limit confidence interval. We shall called the tests inverted from the upper limit Jeffreys interval and the second-order corrected interval *the Jeffreys test* and *the second-order corrected test* respectively.

Since all three distributions have monotone likelihood ratio, it follows from the Karlin–Rubin Theorem (see, e.g., Casella and Berger, 1990, p. 370) that for testing $H_0\colon \mu \geqslant \mu_0$ versus $H_a\colon \mu < \mu_0$ a test that rejects H_0 if and only if $\bar{X} < c$ is a UMP level β test, where $\beta = P_{\mu_0}(\bar{X} < c)$.

The score, Jeffreys, and second-order corrected tests can all be expressed in the form of $I(\bar{X} < c)$, thus all three are UMP tests at the level of their actual size. Therefore the accuracy and performance of these tests can simply be measured by the size of the test. One would prefer the test which has actual level closest to the nominal level.

Fig. 10 plots the sizes, as a function of the null value μ_0, of the upper limit score, Jeffreys, and second-order corrected tests at the nominal $\alpha = 0.01$ level for a binomial proportion, negative binomial mean, and Poisson mean with $n = 30$. In comparison to the score test, it is clear that overall the Jeffreys and second-order corrected tests have actual sizes much closer to the nominal significance level.

7. Discussion and conclusions

7.1. The Clopper–Pearson interval

In the previous sections we give a unified treatment of the one-sided confidence intervals for the discrete Exponential family with a quadratic variance function which consists of the binomial, negative binomial, and Poisson distributions. In the case of binomial proportion, there is a well known "exact" interval, the Clopper–Pearson interval, which has received special attention.

The one-sided Clopper–Pearson interval is the inversion of the one-sided binomial test rather than its normal approximation. If $X \sim \text{Bin}(n, p)$ and $X = x$ is observed, then the upper limit Clopper–Pearson interval (Clopper and Pearson, 1934) is defined by $\text{CI}^{\text{u}}_{\text{CP}} = [0, U_{\text{CP}}(x)]$, where $U_{\text{CP}}(x)$, which is the solution in p to the equation $P_p(X \leqslant x) = \alpha$, equals the $1 - \alpha$ quantile of a beta distribution $\text{Beta}(x + 1, n - x)$. Similarly, the lower limit Clopper–Pearson interval is defined by $\text{CI}^{\text{l}}_{\text{CP}} = [L_{\text{CP}}(x), 1]$, where $L_{\text{CP}}(x)$, which is the solution in p to the equation $P_p(X \geqslant x) = \alpha$, equals the α quantile of a beta distribution $\text{Beta}(x, n - x + 1)$.

By construction, the Clopper–Pearson interval has guaranteed coverage probability of at least $1 - \alpha$. However, the actual coverage probability can be far above the nominal level and the expected distance of the upper/lower limit from p is much larger than those of Jeffreys or the second-order corrected intervals unless n is very large. Therefore the Clopper–Pearson interval is too conservative and is not a good choice for practical use, unless strict adherence to the prescription that the coverage is at least $1 - \alpha$ is required. Fig. 11 compares the coverage probability and the expected distance from p of the upper limit Clopper–Pearson and Jeffreys intervals. The expected distance from p of the upper limit of the Clopper–Pearson interval is about 12–18% larger than that of the Jeffreys interval.

7.2. Conclusions

We show through numerical and analytical calculations that the standard one-sided Wald interval and to a slightly lesser degree the one-sided score interval are uniformly poor in the binomial, negative binomial, and Poisson distributions. The results show that the Jeffreys and second-order corrected intervals provide significant improvements over both the Wald and score intervals. These two alternative intervals nearly completely eliminate the systematic bias in the coverage probability. The one-sided Jeffreys and second-order corrected intervals can be resolutely recommended. In testing one-sided hypotheses, the actual size of the score test deviate systematically from the nominal significance level. The inversion of the Jeffreys and second-order corrected intervals

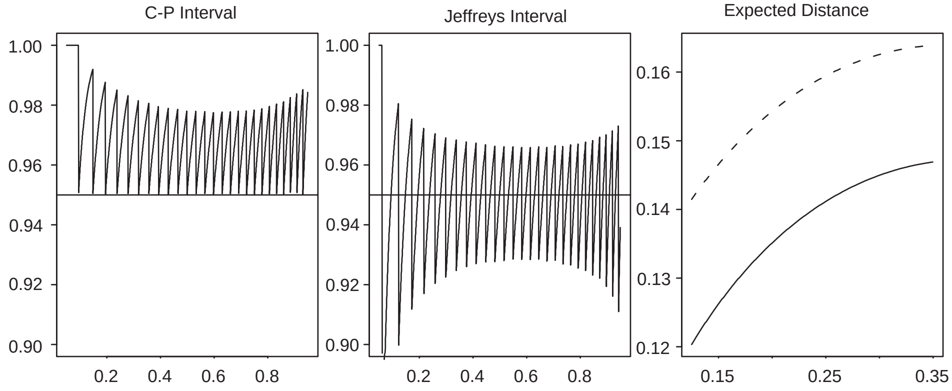

Fig. 11. Coverage probability of the upper limit Clopper–Pearson interval (left panel) and the upper limit Jeffreys interval (middle panel); expected distance from p of the upper limit Clopper–Pearson interval (right panel, dotted line) and the upper limit Jeffreys interval (right, solid line) for a binomial proportion p with $n = 30$ and $\alpha = 0.05$.

yields UMP tests which have actual sizes much closer to the nominal level. The Jeffreys and second-order corrected tests are preferable to the one-sided score test.

Appendix A. Proofs

Except for the second-order corrected interval, most of the essential algebraic derivation is similar to that given in Brown et al. (2002, 2003) for the two-sided confidence intervals.

All three distributions under consideration are lattice distributions with the maximal span of one. Formulas of Edgeworth expansion for lattice distributions can be found, for example, in Esseen (1945) and Bhattacharya and Rao (1976). The following result is from Brown et al. (2003).

Proposition 1. *Let* $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} F$ *with* F *as one of the* $\text{Bin}(1,p)$, $\text{NBin}(1,p)$, *and* $\text{Poi}(\lambda)$ *distributions. Denote* $Z_n = n^{1/2}(\hat{\mu} - \mu)/\sigma$ *and* $F_n(z) = P(Z_n \leqslant z)$. *Then the two-term Edgeworth expansion of* $F_n(z)$ *is given as*

$$F_n(z) = \Phi(z) + p_1(z)\phi(z)n^{-1/2} - Q_1(\mu,z)\sigma^{-1}\phi(z)n^{-1/2} + p_2(z)\phi(z)n^{-1} + \{Q_1(\mu,z)\sigma p_3(z) + Q_2(\mu,z)\}\sigma^{-2}z\phi(z)n^{-1} + O(n^{-3/2}), \quad \text{(A.1)}$$

where Q_1 *and* Q_2 *are given as in* (12) *and*

$$p_1(z) = \tfrac{1}{6}(1 - z^2)(1 + 2b_*\mu)\sigma^{-1},$$

$$p_2(z) = -\tfrac{1}{36}(2z^5 - 11z^3 + 3z)b_* - \tfrac{1}{72}(z^5 - 7z^3 + 6z)\sigma^{-2},$$

$$p_3(z) = -\tfrac{1}{6}(z^2 - 3)(1 + 2b_*\mu)\sigma^{-1}.$$

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

$$z=z_0+c_1n^{-1/2}+c_2n^{-1}+\mathrm{O}(n^{-3/2}),$$

where z_0, c_1 and c_2 are constants, then

$$F_n(z)=\Phi(z_0)+\tilde{p}_1(z)\phi(z_0)n^{-1/2}-Q_1(\mu,z)\sigma^{-1}\phi(z)n^{-1/2}+\tilde{p}_2(z)\phi(z_0)n^{-1}$$
$$+\{Q_1(\mu,z)\sigma\tilde{p}_3(z_0)+Q_2(\mu,z)\}\sigma^{-2}z_0\phi(z_0)n^{-1}+\mathrm{O}(n^{-3/2}), \tag{A.2}$$

where

$$\tilde{p}_1(z)=c_1+\tfrac{1}{6}(1-z_0^2)(1+2b_*\mu)\sigma^{-1}, \tag{A.3}$$

$$\tilde{p}_2(z)=c_2-\tfrac{1}{2}z_0c_1^2+\tfrac{1}{6}c_1(z_0^3-3z_0)(1+2b_*\mu)\sigma^{-1}+p_2(z_0), \tag{A.4}$$

$$\tilde{p}_3(z)=c_1-\tfrac{1}{6}(z_0^2-3)(1+2b_*\mu)\sigma^{-1}. \tag{A.5}$$

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

Proof of Theorems 1 and 4. We consider the coverage of a general upper limit interval of the form:

$$\mathrm{CI}_*=\left[0,\frac{X+s_1}{n-b_*s_2}+\kappa\{V(\hat{\mu})+(r_1V(\hat{\mu})+r_2)n^{-1}\}^{1/2}n^{-1/2}\right], \tag{A.6}$$

where s_1, s_2, r_1 and r_2 are constants. The confidence intervals $\mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}$ and $\mathrm{CI}_2^{\mathrm{u}}$ are special cases of CI_*. Denote

$$A=n-b_*\kappa^2(1+r_1n^{-1})(1-b_*s_2n^{-1})^2,$$

$$B=2n\mu-2(s_1+b_*s_2\mu)+\kappa^2(1+r_1n^{-1})(1-b_*s_2n^{-1})^2,$$

$$C=n(\mu-(s_1+b_*s_2\mu)n^{-1})^2-r_2\kappa^2n^{-1}(1-b_*s_2n^{-1})^2.$$

By solving a quadratic equation, after some algebra, we have

$$P(\mu\in\mathrm{CI}_*)=P\left(\frac{n^{1/2}(\hat{\mu}-\mu)}{\sigma}\geqslant z_*\right),$$

where

$$z_*=\left(\frac{B-\sqrt{B^2-4AC}}{2A}-\mu\right)\sigma^{-1}n^{1/2}. \tag{A.7}$$

Expanding z_*, one has

$$z_* = -\kappa - (s_1 + b_* s_2 \mu - \tfrac{1}{2}\kappa^2(1+2b_*\mu))\sigma^{-1}n^{-1/2} - \left\{ \left(\tfrac{1}{2}r_1 + b_*\kappa^2 - b_* s_2\right)\kappa - \tfrac{1}{2}\kappa(s_1 + b_* s_2\mu)(1+2b_*\mu)\sigma^{-2} + \left(\tfrac{1}{2}r_2\kappa + \tfrac{1}{8}\kappa^3\right)\sigma^{-2} \right\} n^{-1} + \mathrm{O}(n^{-3/2}). \tag{A.8}$$

Denote

$$c_1^* = -(s_1 + b_* s_2\mu - \tfrac{1}{2}\kappa^2(1+2b_*\mu))\sigma^{-1},$$

$$c_2^* = -\left\{ \left(\tfrac{1}{2}r_1 + b_*\kappa^2 - b_* s_2\right)\kappa - \tfrac{1}{2}\kappa(s_1 + b_* s_2\mu)(1+2b_*\mu)\sigma^{-2} + \left(\tfrac{1}{2}r_2\kappa + \tfrac{1}{8}\kappa^3\right)\sigma^{-2} \right\}.$$

Then $z_* = -\kappa + c_1^* n^{-1/2} + c_2^* n^{-1} + \mathrm{O}(n^{-3/2})$. It follows from (A.2) that the coefficient of the $\mathrm{O}(n^{-1/2})$ non-oscillatory term in the Edgeworth expansion of the coverage $P(\mu \in \mathrm{CI}_*) = 1 - F_n(z_*)$ is

$$p_1^*(z_*) = -\tilde{p}_1(z_*) = -c_1^* + \tfrac{1}{6}(\kappa^2 - 1)(1+2b_*\mu)\sigma^{-1}$$
$$= \left\{ \left(s_1 - \left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)\right) + \left(s_2 - 2\left(\tfrac{1}{3}\kappa^2 + \tfrac{1}{6}\right)\right) b_*\mu \right\} \sigma^{-1}.$$

Thus, to make $p_1^*(z_*)$ vanishing for all μ, one needs

$$s_1 = \tfrac{1}{6}(2\kappa^2 + 1) \text{ and } s_2 = \tfrac{1}{3}(2\kappa^2 + 1) \tag{A.9}$$

With s_1 and s_2 given as in (A.9), one has

$$c_2^* = -\tfrac{1}{2}\kappa r_1 + \tfrac{1}{3}(\kappa^3 + 2\kappa)b_* - \tfrac{1}{2}\kappa\sigma^{-2}r_2 + \tfrac{1}{24}(\kappa^3 + 2)\sigma^{-2}.$$

It follows from (A.4) that the coefficient of the $\mathrm{O}(n^{-1})$ non-oscillatory term in the Edgeworth expansion of $P(\mu \in \mathrm{CI}_*)$ is

$$p_2^*(z_*) = -\tilde{p}_2(z_*) = -c_2^* - \tfrac{1}{36}(\kappa^3 - 7\kappa)b_* + \tfrac{1}{72}(\kappa^3 - \kappa)\sigma^{-2}$$
$$= \tfrac{1}{2}\kappa\left\{r_1 - \tfrac{1}{18}(13\kappa^2 + 17)b_*\right\} + \tfrac{1}{2}\kappa\sigma^{-2}\left\{r_2 - \tfrac{1}{36}(2\kappa^2 + 7)\right\}.$$

Therefore we choose

$$r_1 = \tfrac{1}{18}(13\kappa^2 + 17)b_* \text{ and } r_2 = \tfrac{1}{36}(2\kappa^2 + 7) \tag{A.10}$$

to make the $\mathrm{O}(n^{-1})$ non-oscillatory term vanishing.

- For the standard interval, we have z_{W} defined as in (A.7) with $s_1 = s_2 = r_1 = r_2 = 0$. From (A.8) we have

  $$z_{\mathrm{W}} = -\kappa + \tfrac{1}{2}\kappa^2(1+2b_*\mu)\sigma^{-1}n^{-1/2} - (b_* + \tfrac{1}{8}\sigma^{-2})\kappa^3 n^{-1} + \mathrm{O}(n^{-3/2}).$$

  Now, $P(\mu \in \mathrm{CI}_{\mathrm{W}}^{\mathrm{u}}) = 1 - F_n(z_{\mathrm{W}})$ and (14) follows from (A.2).
- For the second-order corrected interval, we have z_2 defined as in (A.7) with $s_1 = \frac{1}{6}(2\kappa^2+1)$, $s_2 = 2s_1$, $r_1 = \frac{1}{18}(13\kappa^2 + 17)b_*$, and $r_2 = \frac{1}{36}(2\kappa^2 + 7)$. It follows

from (A.8) that

$$z_2 = -\kappa + \tfrac{1}{6}(\kappa^2 - 1)(1 + 2b_*\mu)\sigma^{-1}n^{-1/2}$$
$$- \left\{ \tfrac{1}{18}(7\kappa^3 + 5\kappa)b_* - \tfrac{1}{72}(\kappa^3 - \kappa)\sigma^{-2}) \right\} n^{-1} + O(n^{-3/2}).$$

Now (17) follows from (A.2). □

Proof of Theorem 2. The Edgeworth expansion for $P(\mu \in \mathrm{CI}_S^u)$ is simple because

$$P_S = P(\mu \in \mathrm{CI}_S^u) = P\left(\frac{n^{1/2}(\hat{\mu} - \mu)}{\sigma} \geqslant -\kappa \right).$$

And now (15) follows from (A.1). □

A.1. Expansion for Jeffreys prior intervals

We now prove Theorem 3. Denote the cdf of the posterior distribution of θ ($\theta = p$ in the binomial and negative binomial case and $\theta = \lambda$ in the Poisson case) given $X = x$ by $F(\cdot; x, n)$ and denote by $B(\alpha; x, n)$ the inverse of the cdf. Then

$$P(\theta \in \mathrm{CI}_J^u) = P(\theta \leqslant B(1 - \alpha; X, n)) = P(F(\theta; X, n) \leqslant 1 - \alpha).$$

Holding other parameters fixed, the function $F(\theta; x, n)$ is strictly decreasing in x in all three cases. So there exist a unique $Z_l = \rho(1 - \alpha, \theta)$ satisfying

$$F(\theta; Z_l, n) \leqslant 1 - \alpha \text{ and } F(\theta; Z_l - 1, n) > 1 - \alpha.$$

Therefore

$$P(\theta \in \mathrm{CI}_J^u) = P\left(\frac{n^{1/2}(\bar{X} - \mu)}{\sigma} \geqslant z_J \right)$$

with

$$z_J = (\rho(1 - \alpha, \theta) - n\mu)\sigma^{-1}n^{-1/2} \tag{A.11}$$

Here z_J is defined implicitly in (A.11) through ρ. The proof of (16) requires an asymptotic expansion for z_J. Using (A.29) in Brown et al. (2002) for the binomial case, and (56) and (65) in Brown et al. (2003) respectively for the negative binomial and the Poisson case, we have a unified expression for the approximation of z_J:

$$z_J = -\kappa + \tfrac{1}{6}(\kappa^2 - 1)(1 + 2b_*\mu)\sigma^{-1}n^{-1/2}$$
$$- \tfrac{1}{72}\{(2\kappa^3 - 14\kappa)b_* - (\kappa^3 + 2\kappa)\sigma^{-2}\}n^{-1} + O(n^{-3/2}). \tag{A.12}$$

Now we can obtain expansion (16) by plugging in (A.12) into (A.2). □

A.2. Expansions for expected distance from the mean

We now prove Theorem 5. Denote below $Z_n = (\bar{X} - \mu)(\mu + b_*\mu^2)^{-1/2}n^{1/2}$. Then $E(Z_n) = 0$ and $E(Z_n^2) = 1$.

The interval CI_W^u: The expected distance of the upper limit of the Wald interval CI_W^u from the mean μ is

$$L_W = \bar{X} + \kappa(\bar{X} + b_*\bar{X}^2)^{1/2}n^{-1/2} - \mu = Z_n(\mu + b_*\mu^2)^{1/2}n^{1/2}$$
$$+\kappa(\mu + b_*\mu^2)^{1/2}n^{-1/2}(1 + Z_n(1 + 2b_*\mu)(\mu + b_*\mu^2)^{-1/2}n^{-1/2} + b_*Z_n^2n^{-1})^{1/2},$$

on some algebra by using the definition of Z_n. Hence,

$$L_W = Z_n(\mu+b_*\mu^2)^{1/2}n^{1/2}+\kappa(\mu+b_*\mu^2)^{1/2}n^{-1/2}+\tfrac{1}{2}\kappa Z_n(1 + 2b_*\mu)n^{-1}$$
$$- \tfrac{1}{8}Z_n^2(\mu+b_*\mu^2)^{-1/2}\kappa n^{-3/2}+R_W(Z_n), \qquad \text{(A.13)}$$

where $E(|R_W(Z_n)|)=\mathrm{O}(n^{-2})$. see Brown et al. (2002) for more details. Now expansion (22) follows from (A.13) on some algebra.

The interval CI_S^u: For the Rao score interval CI_S^u, the distance is

$$L_S = (\bar{X}+\tfrac{1}{2}\kappa^2)(1-b_*\kappa^2n^{-1})^{-1}+\kappa(1-b_*\kappa^2n^{-1})^{-1}(\bar{X}+b_*\bar{X}^2+\tfrac{1}{4}\kappa^2n^{-1})^{1/2}n^{-1/2}-\mu$$
$$= \tfrac{1}{2}(1+2b_*\mu)\kappa^2n^{-1}+Z_n(\mu+b_*\mu^2)^{1/2}n^{-1/2}+\kappa(\mu + b_*\mu^2)^{1/2}(1 + b_*\kappa^2n^{-1})n^{-1/2}$$
$$+ \tfrac{1}{2}\kappa Z_n(1 + 2b_*\mu)n^{-1} + \tfrac{1}{8}(\kappa^3 - \kappa)(\mu + b_*\mu^2)^{-1/2}Z_n^2n^{-3/2} + R_S(Z_n),$$

where exactly as in (A.13) above, $E(|R_S(Z_n)|) = \mathrm{O}(n^{-2})$. Thus

$$E(L_S) = \tfrac{1}{2}(1 + 2b_*\mu)\kappa^2n^{-1} + \kappa(\mu + b_*\mu^2)^{1/2}(1 + b_*\kappa^2n^{-1})n^{-1/2}$$
$$+\tfrac{1}{8}(\kappa^3 - \kappa)(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + \mathrm{O}(n^{-2})$$

which yields expressions (25) and (29).

The interval CI_2^u: Using the definition of Z_n,

$$\tilde{\mu} = (\mu + Z_n(\mu + b_*\mu^2)^{1/2}n^{-1/2} + \eta n^{-1})(1 - 2\eta b_*n^{-1})^{-1}$$

and hence, after some algebra,

$$L_2 = (1 + 2b_*\mu)\eta n^{-1} + Z_n(1 + 2\eta b_*n^{-1})(\mu + b_*\mu^2)^{1/2}n^{-1/2}$$
$$+\kappa(1 + \tfrac{1}{2}\gamma_1n^{-1})(\mu + b_*\mu^2)^{1/2}n^{-1/2} + \tfrac{1}{2}\kappa Z_n(1 + 2b_*\mu)n^{-1}$$
$$- \tfrac{1}{8}\kappa Z_n^2(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + \tfrac{1}{2}\kappa\gamma_2(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + R_2(Z_n), \qquad \text{(A.14)}$$

where $E(|R_2(Z_n)|) = \mathrm{O}(n^{-2})$. Thus, finally, from (A.14),

$$E(L_2) = (1 + 2b_*\mu)\eta n^{-1} + \kappa(1 + \tfrac{1}{2}\gamma_1n^{-1})(\mu + b_*\mu^2)^{1/2}n^{-1/2}$$
$$- \tfrac{1}{8}\kappa(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + \tfrac{1}{2}\kappa\gamma_2(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + \mathrm{O}(n^{-2})$$

which simplifies to Eq. (28) on a few steps of algebra.

The interval $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$: The upper limit of $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$ admit the general representation

$$\bar{X} + w_1(\bar{X})n^{-1} + \{\kappa(\bar{X} + b_*\bar{X}^2)^{1/2}n^{-1/2} + w_2(\bar{X})n^{-3/2}\} + R_{\mathrm{J}}(n),$$

where the remainder $R_{\mathrm{J}}(n)$ satisfies $E(|R_{\mathrm{J}}(n)|) = \mathrm{O}(n^{-2})$, and

$$w_1(\mu) = \tfrac{1}{6}(2\kappa^2 + 1)(1 + 2b_*\mu),$$

$$w_2(\mu) = \tfrac{1}{36}(\mu + b_*\mu^2)^{-1/2}\{(\kappa^3 + 3\kappa) + b_*(\mu + b_*\mu^2)(13\kappa^3 + 17\kappa)\}.$$

Thus, directly, the distance L_{J} of $\mathrm{CI}_{\mathrm{J}}^{\mathrm{u}}$ satisfies

$$\begin{aligned} E(L_{\mathrm{J}}) &= \tfrac{1}{6}(2\kappa^2 + 1)(1 + 2b_*\mu)n^{-1} \\ &\quad + E[\kappa(\bar{X} + b_*\bar{X}^2)^{1/2}n^{-1/2} + w_2(\bar{X})n^{-3/2}] + \mathrm{O}(n^{-2}) \\ &= \kappa(\mu + b_*\mu^2)^{1/2}n^{-1/2} + \tfrac{1}{6}(1 + 2b_*\mu)(2\kappa^2 + 1)n^{-1} \\ &\quad - \tfrac{1}{8}\kappa(\mu + b_*\mu^2)^{-1/2}n^{-3/2} + w_2(\mu)n^{-3/2} + \mathrm{O}(n^{-2}) \end{aligned}$$

which yields expressions (24) and (27) after some algebra.

References

Agresti, A., Coull, B.A., 1998. Approximate is better than "exact" for interval estimation of binomial proportions. Amer. Statist. 52, 119–126.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd Edition. Springer, New York.

Bhattacharya, R.N., Rao, R.R., 1976. Normal Approximation And Asymptotic Expansions. Wiley, New York.

Blyth, C.R., Still, H.A., 1983. Binomial confidence intervals. J. Amer. Statist. Assoc. 78, 108–116.

Brown, L.D., 1986. Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward.

Brown, L.D., Cai, T., DasGupta, A., 2001. Interval estimation for a binomial proportion (with discussion). Statist. Sci. 16, 101–133.

Brown, L.D., Cai, T., DasGupta, A., 2002. Confidence intervals for a binomial proportion and Edgeworth expansions. Ann. Statist. 30, 160–201.

Brown, L.D., Cai, T., DasGupta, A., 2003. Interval estimation in exponential families. Statist. Sinica 13, 19–49.

Cai, T., 2003. One-sided confidence intervals in discrete distributions. Technical Report, Department of Statistics, University of Pennsylvania.

Casella, G., Berger, R.L., 1990. Statistical Inference. Wadsworth, Brooks Cole, CA.

Clopper, C.J., Pearson, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.

Cressie, N., 1980. A finely tuned continuity correction. Ann. Inst. Statist. Math. 30, 435–442.

Duncan, A.J., 1986. Quality Control and Industrial Statistics, 5th Edition. Irwin, Homewood, IL.

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

Ghosh, J.K., 1994. Higher Order Asymptotics. NSF-CBMS Regional Conference Series, Institute of Mathematical Statistics, Hayward.

Ghosh, M., 2001. Comment on "Interval estimation for a binomial proportion" by L. Brown, T. Cai and A. DasGupta. Statist. Sci. 16, 124–125.

Hall, P., 1982. Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. Biometrika 69, 647–652.

Montgomery, D.C., 2001. Introduction to Statistical Quality Control, 4th Edition. John Wiley, New York.

Morris, C.N., 1982. Natural exponential families with quadratic variance functions. Ann. Statist. 10, 65–80.
Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion; comparison of several methods. Statist. Med. 17, 857–872.
Santner, T.J., 1998. A note on teaching binomial confidence intervals. Teaching Statistics 20, 20–23.
Vollset, S.E., 1993. Confidence intervals for a binomial proportion. Statist. Med. 12, 809–824.