Statistica Sinica **13**(2003), 19-49

INTERVAL ESTIMATION IN EXPONENTIAL FAMILIES

Lawrence D. Brown[1], T. Tony Cai[1] and Anirban DasGupta[2]

[1] *University of Pennsylvania and* [2] *Purdue University*

Abstract: In this paper we consider interval estimation of the mean in the natural Exponential family with a quadratic variance function; the family comprises the binomial, Poisson, negative binomial, normal, gamma, and a sixth distribution.

For the three discrete cases, the Wald confidence interval and three alternative intervals are examined by means of two term Edgeworth expansions of the coverage probability and a two term expansion of the expected length. The results and additional computation suggest that the equal tailed Jeffreys interval and the likelihood ratio interval are the best overall alternatives to the Wald interval. We also show that the poor performance of the Wald interval is not limited to the discrete cases, and a serious negative bias occurs in the nonnormal continuous cases as well. The results are complemented by various illustrative examples.

Key words and phrases: Bayes, binomial distribution, confidence intervals, coverage probability, edgeworth expansion, expected length, Jeffreys prior, natural exponential family, negative binomial distribution, normal approximation, Poisson distribution, quadratic variance function.

1. Introduction

Interval estimation for a binomial proportion has been widely discussed in the literature for more than fifty years, and it had been generally known that the Wald interval in popular use has poor coverage properties for small n, and for p near 0 or 1. Santner (1998), Agresti and Coull (1998), and more recently Brown, Cai and DasGupta (2001, 2002) showed that the Wald interval is actually far too poor and unreliable, and the problems are not just for p near 0 or 1, or for small n. It has a systematic negative bias in its coverage probability and the coverage is oscillatory in both n and p.

In addition to the binomial case, there are at least two other important discrete distributions that have been extensively used in practice; these are the Poisson and the negative binomial distributions. The applications range from epidemiology to oil exploration and risk assessment studies in nuclear and other industrial plants; see Santner and Duffy (1989), Lui (1995), Clevenson and Zidek (1975), and Kaplan (1983), among many others.

The main purpose of this article is to present a unified treatment of the bias and the oscillation problem in interval estimation covering all of these important discrete cases. Our results are quite a bit more than a technical extension. In fact, possibly the most satisfactory aspects of our results are the constancy of the phenomena and uniformity in the final resolutions of these problems.

Several alternative intervals are proposed and studied theoretically and numerically in the subsequent sections. The intervals are the Rao score interval, the equal tailed Jeffreys interval, and the likelihood ratio interval. The intervals are studied theoretically through a two term Edgeworth expansion of their coverage probability and a two term expansion of their expected length. These theoretical results and additional computation suggest that the Jeffreys interval and the likelihood ratio interval are the best all-round alternatives to the Wald interval in all three cases; moreover they have nearly identical coverage and length properties themselves. The score interval also provides major improvements in coverage, but suffers in parsimony with respect to length; they are a bit too long. Incidentally, the systematic negative bias is not specific to lattice distributions. This is demonstrated briefly in this article as well by consideration of similar results for continuous exponential families having quadratic variance functions.

In Section 2, the Exponential family with quadratic variance functions is introduced, and the relevant facts are summarized. In Section 3, we give some preliminary examples to show that the problems with the Wald interval are real and not limited to the binomial case. We also provide some initial calculations to identify wrong centering as a source of the systematic bias of the coverage of the Wald interval. Section 4 introduces alternative intervals, with a brief motivation and background. In contrast to Brown, Cai and DasGupta (2001, 2002), we now include the likelihood ratio interval in our calculations as well, and the final results show unambiguously that this interval is among the best.

In Section 5, two term Edgeworth expansions for coverage probabilities are provided. The most complex of these are the Edgeworth expansions for the equal tailed Jeffreys and the likelihood ratio interval. In Section 6, the Edgeworth expansions are used to explain what the alternative intervals can do to improve on the Wald interval, and also to compare these alternative intervals among themselves. For instance, from the Edgeworth expansions we see that the systematic bias term is nearly killed in all three cases by the Jeffreys as well as the likelihood ratio interval. That the Wald interval suffers from the same coverage bias problem in continuous cases also is briefly discussed in Section 7.

In Section 8, we present comprehensive length expansions for the Wald as well as each of the alternative intervals. The length expansions also reveal a significant amount of structure. For instance, up to an error of order $O(n^{-2})$, the length expansions show that the likelihood ratio interval is the shortest pointwise for every value of the parameter in the Poisson and the negative binomial cases. This is certainly an exceptionally positive feature of the likelihood interval given that its coverage properties are also generally satisfactory. The expansions also show that in all three cases, the likelihood ratio and the Jeffreys interval are the two shortest among the alternative intervals, in an appropriate sense.

Section 10, a technical appendix, contains the proofs. The results are illustrated by various examples and computations throughout the article.

2. Natural Exponential Family

We consider interval estimation of the mean in the natural exponential family (NEF) with quadratic variance functions (QVF). NEF-QVF families consist of six important distributions, three continuous: normal, gamma, and NEF-GHS distributions; three discrete: binomial, negative binomial, and Poisson (see, e.g., Morris (1982) and Brown (1986)). We consider confidence intervals for both the continuous and the discrete NEF-QVF distributions, although our primary focus is on the discrete case.

First we state some basic facts about the NEF-QVF families for use in the rest of this article. The distributions in a natural exponential family have the form

$$f(x|\xi) = e^{\xi x - \psi(\xi)} h(x);$$

ξ is called the natural parameter. The mean, variance and cumulant generating function are

$$\mu = \psi'(\xi), \quad \sigma^2 = \psi''(\xi), \text{ and } \phi_\xi(t) = \psi(t+\xi) - \psi(\xi)$$

respectively. The cumulants are given as $K_r = \psi^{(r)}(\xi)$. In the subclass with a quadratic variance function (QVF), the variance $\psi''(\xi)$ depends on ξ only through the mean μ, and indeed,

$$\sigma^2 \equiv V(\mu) = a_0 + a_1\mu + a_2\mu^2, \tag{1}$$

for suitable constants a_0, a_1, and a_2. We denote the discriminant by

$$\Delta = a_1^2 - 4a_0a_2. \tag{2}$$

The notation Δ will be later used in the statements of theorems for both the discrete and the continuous cases, although for all the discrete cases Δ happens to be equal to 1.

Discrete NEF-QVF families consist of the binomial, negative binomial, and the Poisson distributions. Let us list the important facts for the three distributions separately.

- Binomial, $B(1,p)$: The pmf is $f(x) = e^{\xi x - \psi(\xi)}h(x)$ with $\xi = \log(p/q)$, $\psi(\xi) = \log(1+e^\xi)$, and $h(x) = 1$. Also $\mu = p$, $V(\mu) = pq = \mu - \mu^2$. Thus here $a_0 = 0$, $a_1 = 1$ and $a_2 = -1$.

- Negative binomial, NB$(1,p)$, the number of successes before the first failure; let p = probability of success. Now $\xi = \log p$, $\psi(\xi) = -\log(1-e^{\xi})$, and $h(x) = 1$. Here $\mu = p/q$, $V(\mu) = p/q^2 = \mu + \mu^2$. Thus in this case, $a_0 = 0$, $a_1 = 1$ and $a_2 = 1$.
- Poisson, Poi(λ): In this case, $\xi = \log \lambda$, $\psi(\xi) = e^{\xi}$, and $h(x) = 1/x!$. Since $\mu = \lambda$, $V(\mu) = \mu$, one has $a_0 = 0$, $a_1 = 1$ and $a_2 = 0$.

We focus on the discrete distributions in most of our discussions. The continuous distributions are considered in Section 7.

The common setup for the discrete cases is that we have i.i.d. observations $X_1, \ldots, X_n \sim f(x|\xi)$ with f as one of the three cases above, and we want to estimate μ. Estimation of monotone functions of μ is certainly a relevant and important related problem, but is not considered here mainly due to space considerations.

3. Performance of the Wald Interval

The Wald interval $\hat{p} \pm z_{\alpha/2}\ n^{-1/2}(\hat{p}(1-\hat{p}))^{1/2}$ for a binomial proportion suffers from a systematic negative bias and oscillation in its coverage probability. These problems are not merely for p near 0 or 1, or for small n. Brown, Cai and DasGupta (2001, 2002) showed that the problems persist for large n and even when p is near or exactly equal to 0.5. The problems are caused by the lattice nature as well as the skewness of the binomial distribution. One would expect that these phenomena of a systematic bias and oscillation are true in lattice problems in general, although the severity might differ. We show by two quick examples that indeed this is the case. These two examples are for the Poisson case.

Example 1. Suppose we want to estimate a Poisson mean λ on the basis of n i.i.d. observations. Consider the Wald interval $\bar{X} \pm 2.575 n^{-1/2}\sqrt{\bar{X}}$ for the nominal 99% case, with $n = 20$. This is a moderate sample size. The coverage probability is a function of the product $n\lambda$. Figure 1 plots the coverage of the Wald interval for λ from 0.1 to 5. The most striking aspect of the plot is that the coverage never reaches 0.99. A closer inspection of the coverage probability shows that the smallest $n\lambda$ for which the coverage reaches 0.99 is 193.68. We see the clear systematic negative bias. What was previously observed in the binomial proportion problem resurfaces in the Poisson problem. As a matter of fact, this systematic negative bias is arguably the most negative feature of the Wald interval. Subsequent calculations in Section 6 show that the bias problem is substantially less for a number of alternative intervals.

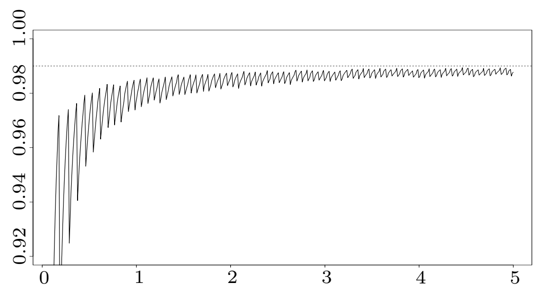

Figure 1. Coverage probability of the Wald interval for $n = 20$ and $0.1 \leq \lambda \leq 5$.

Example 2. Consider again the Poisson mean problem and the coverage of the Wald interval as a function of the sample size n, for a fixed λ, say, $\lambda = 0.5$. Naively, one may expect that the coverage gets systematically closer to the nominal level 95% as n increases. Figure 2 shows that, exactly as in the binomial problem, this is far from the truth. For example, when $n = 9$, the coverage is 0.936; when $n = 16$, the coverage is only 0.892; when $n = 18$, the coverage is 0.940; yet when $n = 72$, the coverage is 0.933, actually smaller than the coverage for $n = 9$! Exactly as seen in Brown, Cai and DasGupta (2001) in the binomial case, the phenomenon of unpredictable arrival of large unlucky values of n reappears in the Poisson problem.

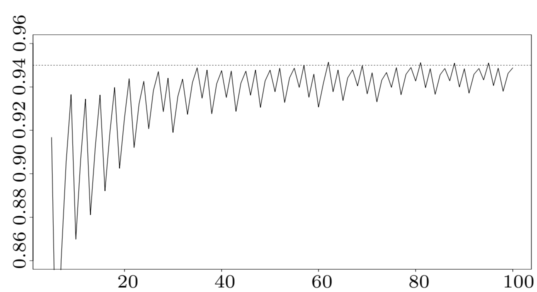

Figure 2. Coverage probability of the Wald interval for fixed $\lambda = 0.5$ and variable n from 5 to 100.

These examples illustrate the poor and erratic behavior of the Wald interval in lattice problems, and make it clear that alternative intervals with better properties are needed. Calculations in Section 6 show that the alternative intervals we propose also have oscillation problems, but to a somewhat lesser degree.

The plots in Figures 1 and 2 have been shown for fixed n and variable λ, and for fixed λ and variable n, respectively, for ease of interpretation. As stated before, in the Poisson case they are actually functions of $n\lambda$.

3.1. Explaining the bias

The standard interval is based on the fact that

$$W_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{a_0 + a_1\hat{\mu} + a_2\hat{\mu}^2}} \stackrel{\mathcal{L}}{\Longrightarrow} N(0,1),$$

where $\hat{\mu} = \bar{X}$; and the interval is constructed by "pretending" that W_n is standard normally distributed. However, as we shall see below, the distribution of W_n could significantly differ from its asymptotic distribution even for moderate to large n. We consider below just the "bias" of W_n. By itself, this bias calculation would be already helpful in understanding part of the reason for the very poor performance of the Wald interval.

Denote $Z_n = \frac{\sqrt{n}(\hat{\mu}-\mu)}{\sqrt{a_0+a_1\mu+a_2\mu^2}}$. Then simple algebra shows

$$W_n = Z_n(1 + (a_1 + 2a_2\mu)\sigma^{-1}n^{-1/2}Z_n + a_2n^{-1}Z_n^2)^{-1/2}.$$

Proceeding as in Brown, Cai, and DasGupta (2002), for the three cases separately one can show the following.

- Binomial (B$(1,p)$):

$$EW_n = \frac{p - 1/2}{\sqrt{npq}}(1 + \frac{7}{2n} + \frac{9(p-1/2)^2}{2npq}) + O(n^{-2}). \quad (3)$$

- Negative binomial (NB$(1,p)$):

$$EW_n = -\frac{1+p}{2\sqrt{np}}(1 + \frac{1}{n} + \frac{9q^2}{8np}) + O(n^{-2}). \quad (4)$$

- Poisson (Poi(λ)):

$$EW_n = -\frac{1}{2\sqrt{n\lambda}}(1 + \frac{9}{8n\lambda}) + O(n^{-2}). \quad (5)$$

3.2. Discussion

These bias expressions give us useful information. From (3), one would suspect that in the binomial case W_n has a negative bias for $p < 1/2$ and a positive bias for $p > 1/2$. This would naturally suggest that the center of the Wald interval for p should be moved towards 1/2. Brown, Cai and DasGupta (2001, 2002) show that indeed recentering does improve the coverage properties in that problem.

Moving on to the Poisson case, we see both similarities and differences of phenomena with the binomial case. First, from (5), we see that W_n appears to have a negative bias for all λ. So the bias problem persists but, unlike the binomial case, the center of the Wald interval for λ should always be moved up. And indeed, our subsequent calculations confirm that by moving the center of the Wald interval up, we can significantly curtail the systematic negative bias in the coverage of the Wald interval (see Figure 6).

Similar disturbing bias is also present in the negative binomial problem, and again examination of (4) would suggest that here too the center of the Wald interval needs to be moved up to address a potential bias problem.

Numerical plots also demonstrate this bias problem. Figure 3 below plots the bias (i.e., $E(W_n)$) as a function of $n\lambda$ for the Poisson case. The clearly significant negative bias even when $n\lambda$ is 40 or so is certainly disconcerting.

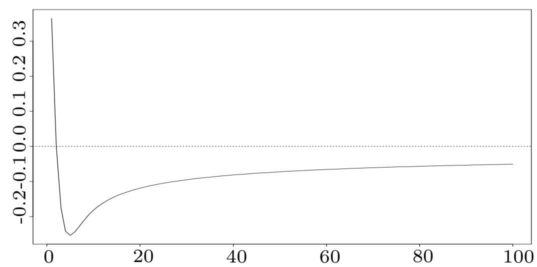

Figure 3. Bias in Poisson case for $n\lambda = 1$ to 100.

4. The Confidence Intervals

Let $X = \sum_{i=1}^{n} X_i$ and let $\hat{\mu} = \bar{X} = X/n$; $\hat{\mu}$ is well known to be the MLE of μ. Then the Central Limit Theorem and Slutsky's Theorem yield

$$W_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma}} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{a_0 + a_1\hat{\mu} + a_2\hat{\mu}^2}} \stackrel{\mathcal{L}}{\Longrightarrow} N(0, 1), \tag{6}$$

$$Z_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{a_0 + a_1\mu + a_2\mu^2}} \stackrel{\mathcal{L}}{\Longrightarrow} N(0, 1). \tag{7}$$

We can construct confidence intervals for μ based on (6) and (7).

Interval # 1: The *Wald interval* is based on (6):

$$CI_s = \hat{\mu} \pm \kappa\hat{\sigma}n^{-1/2} = \hat{\mu} \pm \kappa(a_0 + a_1\hat{\mu} + a_2\hat{\mu}^2)^{1/2}n^{-1/2}. \tag{8}$$

Interval # 2: The *score interval* is based on (7). This interval is formed by inverting Rao's equal tailed score test of $H_0 : \mu = \mu_0$. Hence, one accepts H_0 based on Rao's score test if and only if μ_0 is in this interval. Denote $\tilde{\mu} = (n\hat{\mu} + \kappa^2/2)/(n - \kappa^2 a_2)$. By solving a quadratic equation, one finds the score interval

$$CI_R = \tilde{\mu} \pm \frac{\kappa n^{1/2}}{n - \kappa^2 a_2}(a_0 + a_1\hat{\mu} + a_2\hat{\mu}^2 + \frac{\kappa^2}{4n}\Delta)^{1/2}. \tag{9}$$

4.1. Likelihood ratio intervals

Interval # 3: The Wald and the score interval are obtained by inversion of the acceptance regions of the Wald and the Rao's score test, respectively. The *likelihood ratio interval* is constructed by inversion of the likelihood ratio test which accepts the null hypothesis $H_0: \ \mu = \mu_0$ if $-2\log(\Lambda_n) \le \chi^2_{\alpha,1}$, where Λ_n is the likelihood ratio $L(\mu_0)/\sup_\mu L(\mu)$, with L as the likelihood function based on n i.i.d. observations from the underlying distribution. See Rao (1973) and Serfling (1980).

The likelihood ratio method of constructing confidence intervals is a well accepted method and so the likelihood ratio intervals merit a theoretical study on their own right. But an example provides additional evidence that the interval deserves very serious consideration.

Figure 4 plots the exact coverage probability of the likelihood ratio interval and three other intervals for $n\lambda$ from 2 to 50 in the Poisson case. We see from these plots that the coverage of the likelihood ratio interval fluctuates acceptably near the nominal level and it clearly outperforms the Wald interval. More interestingly, if we compare the coverage of the LR interval and the score interval as well as the Jeffreys of Section 4.2, we see that the likelihood ratio interval has a substantially smaller oscillation. Much of our subsequent technical calculations will confirm this impressive performance of the likelihood ratio interval.

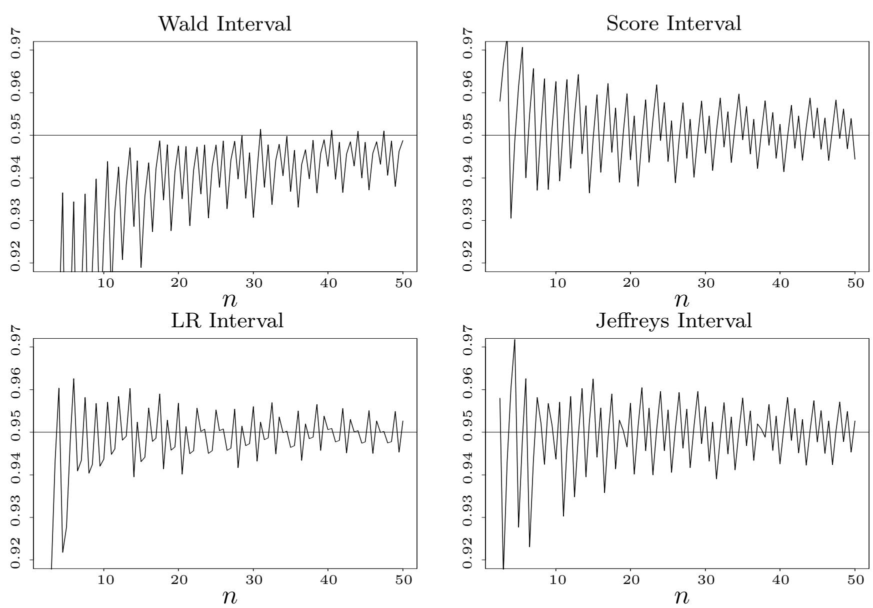

Figure 4. Coverage probability of four intervals for a Poisson mean with $\alpha = 0.05$ and $n\lambda = 2$ to 50.

4.2. The Jeffreys interval

Bayesian procedures constructed from the Jeffreys prior have a long history of providing very good frequentist properties. The Jeffreys prior credible intervals, in particular, are known to have the probability matching property in the absolutely continuous case; see Ghosh (1994). It is thus natural to consider the Jeffreys prior credible interval as well.

Denote $b(\cdot) = (\psi')^{-1}(\cdot)$. Then b is a strictly increasing function and $\xi = b(\mu)$. The Fisher information about μ is

$$\begin{aligned} I(\mu) &= -E_\mu \frac{\partial^2 \log f(X,\mu)}{\partial \mu^2} = -E_\mu[Xb''(\mu) - n\mu b''(\mu) - n\psi''(b(\mu))(b'(\mu))^2] \\ &= n\psi''(b(\mu))(b'(\mu))^2. \end{aligned}$$

Noting that $\psi''(b(\mu)) = \psi''(\xi) = a_0 + a_1\mu + a_2\mu^2$ and $b'(\mu) = 1/\psi''(\xi)$, we have $I(\mu) = n(a_0 + a_1\mu + a_2\mu^2)^{-1}$. The Jeffreys prior is proportional to $I^{1/2}(\mu)$ and thus the posterior satisfies

$$f(\mu|x) \propto \exp\{xb(\mu) - n\psi(b(\mu)) - \frac{1}{2}\log(a_0 + a_1\mu + a_2\mu^2)\}. \tag{10}$$

Interval # 4: The *Jeffreys equal tailed interval* for μ is given by

$$CI_J = [J_{\alpha/2},\ J_{1-\alpha/2}], \tag{11}$$

where $J_{\alpha/2}$ and $J_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior distribution (10) based on n observations, respectively.

Consider the three special distributions for illustration separately.

- Binomial: here $\psi(\xi) = \log(1 + e^\xi)$ and $b(\mu) = \log(\mu/(1-\mu))$. Let $X \sim \mathrm{B}(n,p)$. The Jeffreys prior in this case is $Beta(1/2,\ 1/2)$ and the posterior is $Beta(X + 1/2,\ n - X + 1/2)$. Thus the $100(1-\alpha)\%$ equal tailed Jeffreys interval for p is given by

$$CI_J = [p_l,\ p_u] = [B_{\alpha/2, X+1/2,\, n-X+1/2},\ B_{1-\alpha/2, X+1/2,\, n-X+1/2}]. \tag{12}$$

- Negative binomial: here $\psi(\xi) = -\log(1 - e^\xi)$ and $b(\mu) = \log(\mu/(1+\mu))$. Let $X \sim \mathrm{NB}(n,p)$. The Jeffreys prior for μ is proportional to $\mu^{-1/2}(1 + \mu)^{-1/2}$ and the posterior is a beta-prime distribution (see Johnson, Kotz and Balakrishnan (1995)).

  The Jeffreys interval is transformation-invariant. We can obtain the Jeffreys interval for μ through the Jeffreys interval for p. The Jeffreys prior for p is proportional to $p^{-1/2}(1-p)^{-1}$ and the posterior is $Beta(X+1/2,\ n)$. The $100(1-\alpha)\%$ equal tailed Jeffreys interval for p is given by

$$CI_J^p = [p_l,\ p_u] = [B_{\alpha/2,X+1/2,\ n},\ B_{1-\alpha/2,X+1/2,\ n}]. \tag{13}$$

Since $\mu = p/(1-p)$, the Jeffreys interval for μ is

$$CI_J = [p_l/(1-p_l),\ p_u/(1-p_u)]. \tag{14}$$

- Poisson: here $\psi(\xi) = e^{\xi}$ and $b(\mu) = \log \mu$. Let $X \sim \text{Poi}(n\lambda)$. The Jeffreys prior is proportional to $\lambda^{-1/2}$ which is improper and the posterior is $\text{Gamma}(X+1/2,\ 1/n)$, which is proper. Therefore the $100(1-\alpha)\%$ equal tailed Jeffreys interval for λ is given by

$$CI_J = [\lambda_l,\ \lambda_u] = [G_{\alpha/2,X+1/2,\ 1/n},\ G_{1-\alpha/2,X+1/2,\ 1/n}]. \tag{15}$$

Example 3. We have introduced a number of different confidence intervals as alternatives to the Wald interval. How the limits of these various intervals differ among themselves is of importance to practitioners. If two different intervals have very similar limits, a practitioner is likely to consider them as practically equivalent.

In Figure 5 below, we have plotted the limits of the intervals in the binomial case with $n = 20$ and in the Poisson case with $x \leq 20$. First, in the Poisson plot, we see a clear clustering; the upper and lower limits of the score interval are larger than the corresponding limits of the Jeffreys and the likelihood ratio intervals, which are close to each other. The Wald interval, on the other hand, is all by itself, markedly separated from the other three intervals.

In the binomial plot, we again see that the limits of the Jeffreys and the likelihood ratio interval are virtually indistinguishable and the Wald interval is visibly different. The limits of the score are slightly tilted upward for $\hat{p} < 1/2$ and downward for $\hat{p} > 1/2$.

It would be reasonable to expect that the Jeffreys and the likelihood ratio interval have comparable coverage and length properties as they seem to have very similar confidence limits. Later in our detailed theoretical calculations in Sections 5 and 8, these visual conjectures will be vindicated.

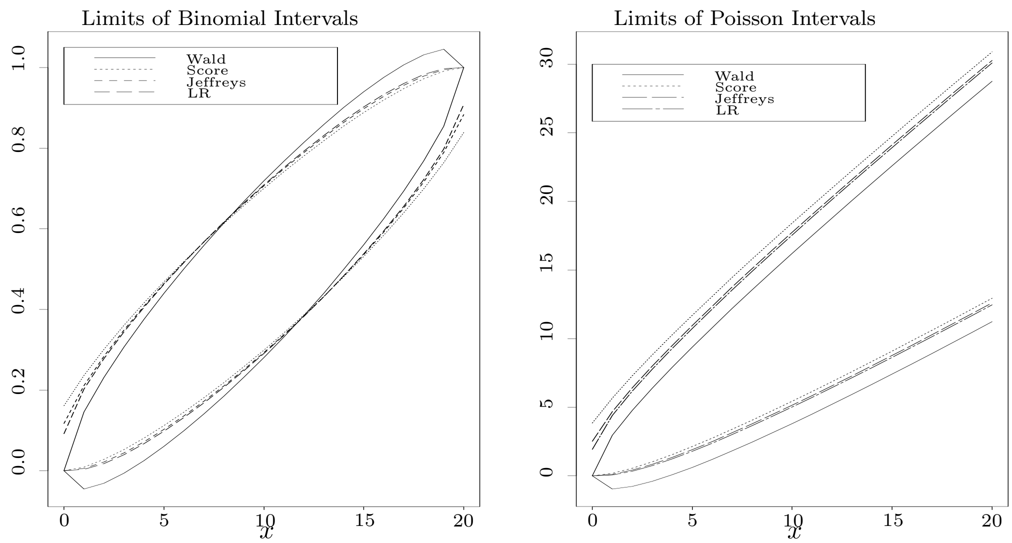

Figure 5. Comparison of the limits of various intervals for $\alpha = 0.05$.

It should be noted that similar to the binomial case, "exact" confidence intervals for the Poisson and the negative binomial cases are available as well. Nearly universally, they are too conservative. See Crow and Gardner (1959), Casella and Robert (1989) and Kabaila and Byrne (2001) for the Poisson case; of these the latter two articles present methods of refining the endpoints for every value of X. The savings in the coverage are marginal; compare Figures 3 and 4 in Casella and Robert (1989). In addition, the refinements are not closed form and tables cannot be produced as the sample space is infinite. We have not studied these exact intervals in this article.

5. The Edgeworth Expansions

Denote by x_- the largest integer less than or equal to x. Define

$$h(x) = x - x_-. \tag{16}$$

So $h(x)$ is the fractional part of x. The function h is a periodic function of period 1. Let

$$g(\mu, z) = g(\mu, z, n) = h(n\mu + n^{1/2}\sigma z). \tag{17}$$

We suppress in (17), and later, the dependence of g on n and denote

$$Q_{21}(\ell,\ u) = 1 - g(\mu,\ \ell) - g(\mu,\ u),$$
$$Q_{22}(\ell,\ u) = \frac{1}{2}[-g^2(\mu,\ \ell) - g^2(\mu,\ u) + g(\mu,\ \ell) + g(\mu,\ u) - \frac{1}{3}].$$

In the theorems below we assume 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.

Theorem 1. *Let $0 < \alpha < 1$. Suppose $n\mu + n^{1/2}\sigma\ell_s$ is not an integer. Then the coverage probability of the confidence interval CI_s defined in (8) satisfies*

$$
\begin{aligned}
P_s = P_\mu(\mu \in CI_s) &= (1-\alpha) + \sigma^{-1}\{g(\mu,\ell_s) - g(\mu,u_s)\} \cdot \phi(\kappa) n^{-\frac{1}{2}} \\
&+ \{-\frac{a_2}{18}(8\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{\Delta}{18\sigma^2}(2\kappa^5 + \kappa^3 + 3\kappa)\} \cdot \phi(\kappa) n^{-1} \\
&+ \{-(a_1 + 2a_2\mu)(\frac{1}{3}\kappa^2 + \frac{1}{2})Q_{21}(\ell_s, u_s) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa) n^{-1} \\
&+ O(n^{-\frac{3}{2}}),
\end{aligned}
\tag{18}
$$

where the quantities ℓ_s and u_s are defined as

$$
(\ell_s,\ u_s) = \left(\frac{(2n\mu + \kappa^2 a_1) \pm \sqrt{(2n\mu + \kappa^2 a_1)^2 - 4(n - \kappa^2 a_2)(n\mu^2 - \kappa^2 a_0)}}{2(n - \kappa^2 a_2)} - \mu \right) \sigma^{-1} n^{\frac{1}{2}}.
$$

with the $-$ sign going with ℓ_s and the $+$ sign with u_s.

Remark. In the case that $n\mu + n^{1/2}\sigma\ell_s$ is an integer, one needs to add an additional term $P_p(X = n\mu + n^{1/2}\sigma\ell_s) = \phi(\kappa) n^{-1/2}\sigma^{-1} + O(n^{-1})$ to (18). The same applies to the two term expansion of the coverage probability of other confidence intervals.

Theorem 2. *Let $0 < \alpha < 1$. Suppose $n\mu - n^{1/2}\sigma\kappa$ is not an integer. Then the coverage probability of the confidence interval CI_R defined in (9) satisfies*

$$
\begin{aligned}
P_R = P_\mu(\mu \in CI_R) &= (1-\alpha) + \sigma^{-1}\{g(\mu,-\kappa) - g(\mu,\kappa)\} \cdot \phi(\kappa) n^{-\frac{1}{2}} \\
&+ \{-\frac{a_2}{18}(2\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{\Delta}{36\sigma^2}(\kappa^5 - 7\kappa^3 + 6\kappa)\} \cdot \phi(\kappa) n^{-1} \\
&+ \{(a_1 + 2a_2\mu)(\frac{1}{6}\kappa^2 - \frac{1}{2})Q_{21}(-\kappa,\kappa) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa) n^{-1} \\
&+ O(n^{-\frac{3}{2}}).
\end{aligned}
\tag{19}
$$

The following theorem gives a unified expression for the two term Edgeworth expansion of the coverage probability of the likelihood ratio interval. We have nevertheless found it necessary for the proofs in the appendix to give separate proofs for the three different distributions.

Theorem 3. *Denote by CI_{LR} a generic LR interval. Let $0 < \alpha < 1$. Suppose $n\mu + n^{1/2}\sigma\ell_{LR}$ is not an integer. Then the coverage probability of CI_{LR} satisfies* *the general representation*

$$P_{LR} = P_\mu(\mu \in CI_{LR}) = (1-\alpha) + \sigma^{-1}\{g(\mu,\ell_{LR}) - g(\mu,u_{LR})\}\cdot\phi(\kappa)n^{-1/2}$$
$$+\{-\frac{a_2}{6}\kappa - \frac{\Delta}{6\sigma^2}\kappa\}\cdot\phi(\kappa)n^{-1}$$
$$+\{-\frac{1}{2}(a_1+2a_2\mu)Q_{21}(\ell_{LR},u_{LR}) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1}$$
$$+O(n^{-3/2}), \tag{20}$$

where the quantities ℓ_{LR} and u_{LR} are defined in (41) *in the appendix.*

The next theorem gives a general expression for the two term Edgeworth expansion of the coverage probability of the Jeffreys interval covering all three cases.

Theorem 4. *Denote by CI_J the equal tailed Jeffreys interval as defined in* (*12*) *in the binomial case,* (*14*) *in the negative binomial case, and* (*15*) *in the Poisson case. Let $0 < \alpha < 1$. Suppose $n\mu + n^{1/2}\sigma\ell_J$ is not an integer. Then the coverage probability of CI_J satisfies*

$$P_J = P_\mu(\mu \in CI_J) = (1-\alpha) + \sigma^{-1}\{g(\mu,\ \ell_J) - g(\mu,\ u_J)\}\phi(\kappa)\cdot n^{-1/2}$$
$$-\frac{\Delta}{12\sigma^2}\kappa\phi(\kappa)n^{-1} + \{-\frac{1}{3}(a_1+2a_2\mu)Q_{21}(\ell_J,u_J) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1}$$
$$+O(n^{-3/2}), \tag{21}$$

where the quantities ℓ_J and u_J are defined as in (46) *for the negative binomial case, in* (57) *for the Poisson case, and in Theorem* 5 *of Brown, Cai and DasGupta* (2002) *for the binomial case.*

The Edgeworth expansions for the three specific distributions, binomial, negative binomial, and Poisson, can be obtained easily from Theorems 1 − 4 by plugging in the corresponding a_2, μ and σ, as given in Section 2.

6. Comparison of Coverage Probability

We now use the two term Edgeworth expansions to compare the coverage properties of the standard interval CI_s and the various alternative intervals. The encouraging part is that we can reach general conclusions for all three distributions. The recommendations therefore carry a unifying character. First we show how the $O(n^{-1})$ nonoscillating term can be used to explain the deficiency of the standard procedure and the much better performance of competing ones such as the likelihood ratio and the Jeffreys procedure. The $O(n^{-1})$ nonoscillating term measures the systematic bias in coverage. Figure 6 displays the nonoscillating $O(n^{-1})$ terms of each interval for binomial, negative binomial, and Poisson cases.

It is transparent that there is a consistent serious negative bias in the coverage of the standard interval for all three distributions. The score interval CI_R does significantly better than the standard interval CI_s, and especially so near the boundaries. The most interesting feature manifested in Figure 6 is the near vanishing bias term in the Edgeworth expansions for the likelihood ratio as well as the Jeffreys interval. The Edgeworth expansions and, in particular, Figure 6 show that both the likelihood ratio and the Jeffreys interval practically annihilate the $O(n^{-1})$ bias term. These two intervals are thus demonstrably superior competitors to the Wald interval in terms of coverage. As regards the score interval, the nonoscillating term is positive in all three cases. Thus, although the coverage is better, it is in fact less parsimonious than the Jeffreys and the likelihood ratio interval. We will revisit this matter in Section 8. See also remarks (a) and (c) in Section 6.1.

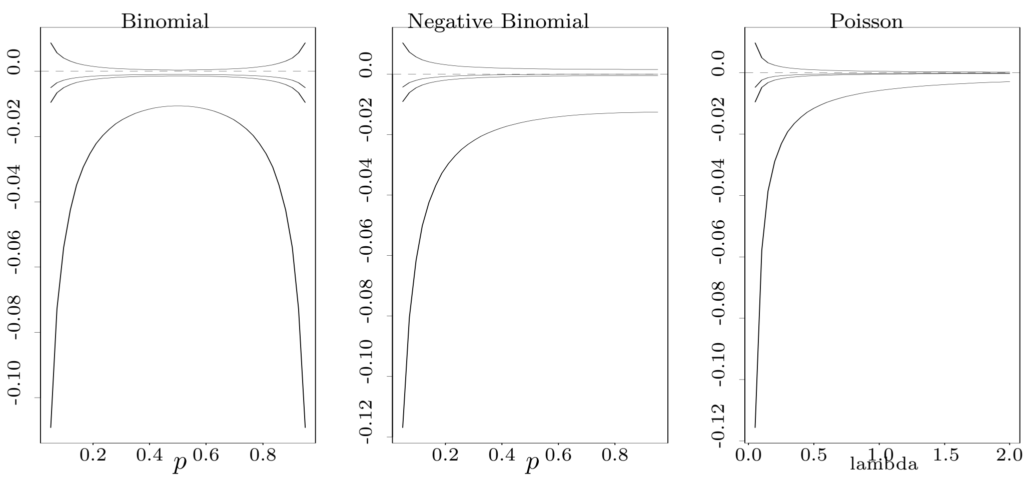

Figure 6. Comparison of the nonoscillating terms with $n = 40$ and $\alpha = 0.05$. From top to bottom: CI_R, CI_J, CI_{LR}, and CI_s.

Directly from equations (18) - (21), ignoring the $O(n^{-3/2})$ terms, we have:

$$P_R - P_s = \{\frac{1}{3}a_2\kappa^5 + \frac{\Delta}{\sigma^2}(\frac{1}{12}\kappa^5 + \frac{1}{4}\kappa^3)\}\phi(\kappa)\cdot n^{-1} + \text{oscillations}, \tag{22}$$

$$P_{LR} - P_s = \{\frac{a_2}{18}(8\kappa^5 - 11\kappa^3) + \frac{\Delta}{18\sigma^2}(2\kappa^5 + \kappa^3)\}\phi(\kappa)\cdot n^{-1} + \text{oscillations}, \tag{23}$$

$$P_J - P_s = \{\frac{a_2}{18}(8\kappa^5 - 11\kappa^3 + 3\kappa) + \frac{\Delta}{36\sigma^2}(4\kappa^5 + 2\kappa^3 + 3\kappa)\}\phi(\kappa)\cdot n^{-1} + \text{oscillations}. \tag{24}$$

These expressions compare the nonoscillating terms in the Edgeworth expansions. The terms can be viewed as smooth approximations to the coverage probability.

Theorem 6 of Brown, Cai, and DasGupta (2002) provides rigorous support for such an interpretation in the binomial case. We believe similar statements are true for the other discrete problems considered here.

6.1. Further discussion

By consideration of the coefficients of the n^{-1} terms in (22)−(24), we can make several more interesting conclusions. These conclusions are borne out in Figure 6. First, recall that a_2 is -1 in the binomial case, $+1$ in the negative binomial case, and 0 in the Poisson case. The coefficient of the n^{-1} term in the expressions (22)−(24) determine if the coverage probability of a specific interval has a smaller bias than another interval it is being compared with. Using the values of a_2 as above, elementary calculations show the following.

(a) For each of the three distributions, the coefficient of the coverage-bias term in $P_R - P_s$ is positive for all κ and all values of the underlying parameter. From this we can expect that the score interval will improve on the standard interval as regards the systematic negative coverage-bias phenomenon for all three distributions. Some of this can in fact be readily seen in Figure 6. The coefficient of the bias term is positive for $P_{LR} - P_s$ and $P_J - P_s$ as well, provided $\kappa > \sqrt{11/8} = 1.17$, which would be true in most practical cases. Thus these two intervals also provide relief to the systematic coverage bias problem of CI_s.

(b) In the Poisson case, the coefficient $a_2 = 0$. Comparison of (23) and (24) immediately reveals the nearly identical coefficients of the $O(n^{-1})$ nonoscillating term for the likelihood ratio and the Jeffreys interval. Even if a_2 is not 0, the coefficients are very similar. We thus have the interesting phenomenon that the two intervals, constructed using totally different methods, have nearly identical coverage properties, in terms of their Edgeworth expansions.

(c) We also see from Figure 6 that for each of the three distributions, the score interval has a slight positive bias in coverage, comparable in magnitude to the negative coverage-bias of the likelihood ratio interval.

7. The Continuous Natural Exponential Family

The significant coverage bias of the Wald interval is not unique to the discrete distributions, although the oscillation phenomenon is. Among NEF-QVF distributions, besides the three discrete distributions we have discussed above, there are also three continuous distributions: the normal, $N(\mu,\ \sigma^2)$ (with σ^2 known), the gamma, Gam$(r,\ \lambda)$ (with r known), and NEF-GHS(r, λ), (with r known). The sixth family NEF-GHS(r, λ) is not a common family of distributions. The NEF-GHS distributions are derived from the generalized hyperbolic secant distributions (see Johnson, Kotz and Balakrishnan (1995)). A NEF-GHS distribution with parameters r and λ has density

$$f_{r,\lambda}(x) = (1+\lambda^2)^{-r/2} exp\{x \tan^{-1}(\lambda)\} f_{r,0}(x), \tag{25}$$

where $f_{r,0}(x)$ is the density of a generalized hyperbolic secant distribution with parameter r which is defined as

$$f_{r,0}(x) = \frac{2^{r-2}}{\Gamma(r)} \prod_{j=0}^{\infty} \{1 + \frac{x^2}{(r+2j)^2}\}^{-1}.$$

The NEF-GHS(r,λ) distribution defined in (25) has mean $r\lambda$ and variance $r(1+\lambda^2)$. So $V(\mu) = r + \mu^2/r$ and the discriminant $\Delta = a_1^2 - 4a_0a_2 = -4$. See Morris (1982) for further details on the properties of the NEF-GHS distributions.

For the continuous distributions, the Edgeworth expansions for the coverage probability of various intervals are the same as those in the cases of the discrete distributions as given in (18)$-$(21), except that now there are no oscillation terms. For example, the two term expansion for the coverage of the Wald interval in the three continuous cases is

$$P_s = (1-\alpha) + \{-\frac{a_2}{18}(8\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{\Delta}{18\sigma^2}(2\kappa^5 + \kappa^3 + 3\kappa)\} \cdot \phi(\kappa) n^{-1} + O(n^{-3/2}), \tag{26}$$

where the discriminant $\Delta = 0$ for the normal and gamma distributions and $\Delta = -4$ for NEF-GHS(r,λ).

For our problem, the case of normal $N(\mu,\ \sigma^2)$ with σ^2 known is not interesting; the Wald interval and the score interval coincide and attain the exact nominal coverage. For the other two distributions the systematic bias in coverage of the Wald interval is very similar to what we have found in the discrete cases in the previous sections. Figure 7 below plots the two term Edgeworth expansion of the coverage probability of the Wald and the score intervals in the Gamma$(r,\ \lambda)$ and NEF-GHS$(r,\ \lambda)$ distributions. Clearly, the coverage bias of the Wald interval is significant and much larger than the corresponding score interval. Interestingly, in the case of the sixth family, NEF-GHS$(r,\ \lambda)$, the coverage bias of the Wald interval is not always negative. For some values of $r,\ \lambda$, and α, the coverage bias can be positive. The primary reason is that the discriminant $\Delta = -4 < 0$ in this case, whereas in all other cases Δ is nonnegative. But even in the case that the Wald interval has positive coverage bias, the corresponding score interval still has smaller bias in coverage and is preferable. For reason of space, we do not get into more details on this phenomenon.

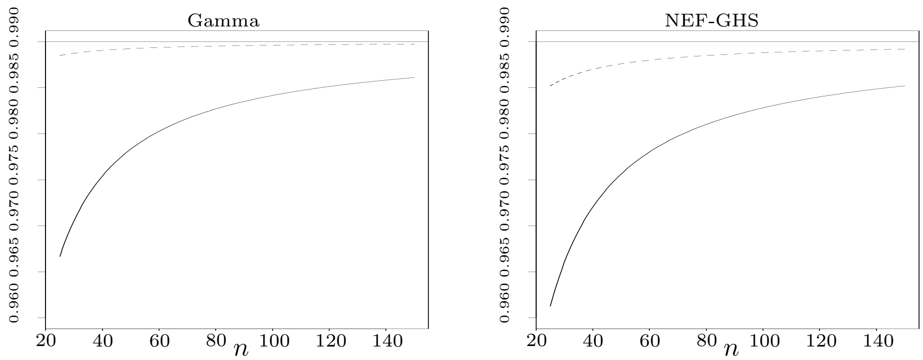

Figure 7. Two term Edgeworth expansion of the coverage of the Wald (solid) and the score (dashed) intervals, with $n = 25$ to 150 and $\alpha = 0.01$. For the Gamma distribution (left) the parameters are $r = \lambda = 1$ and for the NEF-GHS distribution (right) $r = 1/4$ and $\lambda = 1$.

8. Expansions for Expected Length

The two term Edgeworth expansions presented in Section 5 compare the coverage property of the various intervals. However, in addition to coverage, parsimony in length is also an important issue. Therefore for the intervals we discussed in Section 4 we now provide an expansion for their expected lengths correct up to the order $O(n^{-3/2})$. The theoretical calculations are somewhat technical. However, the main conclusions from these calculations are clean and very structured. For ease of comparison, it might be helpful to have a glimpse into what these conclusions are prior to the actual calculations.

8.1. Preview

In the Poisson and the negative binomial case, up to an error of order $O(n^{-2})$, there is a uniform ranking of the four intervals in expected length pointwise for every value of the parameter. The intervals are ranked as CI_s, CI_{LR}, CI_J, and CI_R from the shortest to the longest. Thus, among the three alternative intervals, the likelihood ratio interval is pointwise the shortest. In the binomial case, there is no such uniform ranking pointwise for every value of p. But if we take the integrated version of the length expansion, then the ranking is CI_J, CI_{LR}, and $CI_R = CI_s$ from the shortest to the longest. Furthermore, CI_J and CI_{LR} have virtually identical integrated length expansions. Note that it was already observed in Brown, Cai and DasGupta (2002) that CI_s and CI_R have exactly identical integrated length expansions.

To put it all together, the combined lesson is that among the alternative intervals, the likelihood ratio and the Jeffreys intervals are always the shortest.

Simplicity of computation aside, the likelihood ratio and the Jeffreys interval may be the most credible alternatives to the Wald interval in all three cases.

Let us look at some particular examples. Figure 8 displays the expected lengths of various intervals for the mean of negative binomial and Poisson distributions. The left panel is the comparison for the negative binomial case with $n = 20$ and $0.1 \leq p \leq 0.9$, and the right panel compares the expected lengths in the Poisson case for $n\lambda$ from 5 to 30. There is a clear ranking of the intervals in terms of expected length from the shortest to the longest, CI_s, CI_{LR}, CI_J, and CI_R. In both cases, the expected lengths of CI_J and CI_{LR} are almost identical. The asymptotic results are consistent with this ranking.

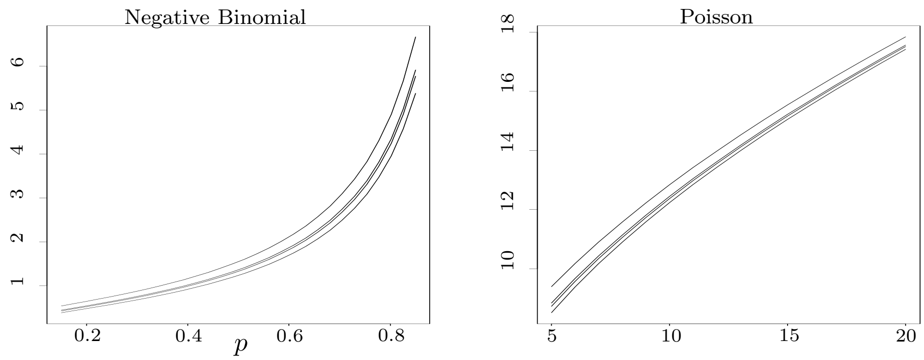

Figure 8. Expected length of various intervals for the mean. From bottom to top, the expected length of CI_s, CI_{LR}, CI_J, and CI_R.

8.2. Expansions and comparisons

The expansion for length differs qualitatively from the two term Edgeworth expansion for coverage in that the Edgeworth expansion includes terms involving $n^{-1/2}$ and n^{-1}, whereas the expansion for length includes terms involving $n^{-1/2}$ 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.

Theorem 5. *Let CI be a generic notation for any of the four intervals, CI_s, CI_R, CI_{LR} and CI_J, for estimating the mean μ, as defined in equations (8)–(11). Then,*

$$L(n,\mu) = E(\text{length of } CI) = 2\kappa(\mu + a_2\mu^2)^{1/2} n^{-1/2} \left(1 - \frac{\delta(\kappa,\mu)}{72n(\mu + a_2\mu^2)}\right) + O(n^{-2}), \qquad (27)$$

$$\delta(\kappa,\ \mu) = 9\Delta \quad \text{for } CI_s; \qquad (28)$$

$$= 9(1-\kappa^2)\Delta - 72\kappa^2 a_2(\mu + a_2\mu^2) \quad \textit{for } CI_R; \qquad (29)$$
$$= (9-2\kappa^2)\Delta - 26\kappa^2 a_2(\mu + a_2\mu^2) \quad \textit{for } CI_{LR}; \qquad (30)$$
$$= (5-2\kappa^2)\Delta - 2(13\kappa^2+17) a_2(\mu + a_2\mu^2) \quad \textit{for } CI_J; \qquad (31)$$

Corollary 1. *Consider the Poisson case. Then the expected lengths of CI_s, CI_{LR}, CI_J, and CI_R admit the expansions*

$$E(L_s) = 2\kappa\lambda^{1/2}n^{-1/2}[1 - \frac{9}{72n\lambda}] + O(n^{-2}),$$
$$E(L_{LR}) = 2\kappa\lambda^{1/2}n^{-1/2}[1 + \frac{9(\kappa^2-1) - 7\kappa^2}{72n\lambda}] + O(n^{-2}),$$
$$E(L_J) = 2\kappa\lambda^{1/2}n^{-1/2}[1 + \frac{9(\kappa^2-1) + 4 - 7\kappa^2}{72n\lambda}] + O(n^{-2}),$$
$$E(L_R) = 2\kappa\lambda^{1/2}n^{-1/2}[1 + \frac{9(\kappa^2-1)}{72n\lambda}] + O(n^{-2}).$$

Remark. Hence, up to the error n^{-2}, for every $\lambda > 0$ the ranking of the intervals is CI_s, CI_{LR}, CI_J, and CI_R from the shortest to the longest, as long as $\kappa > 2/\sqrt{7} = 0.76$. In practice, κ will certainly be larger than 0.76 and so, we have the quite remarkable fact that pointwise in λ, a uniform ranking of the intervals is possible. Furthermore, we see from the above Corollary that among the alternative intervals, the likelihood ratio interval is the shortest. It is particularly worth noting that it is shorter than the Jeffreys interval CI_J at every λ.

The next corollary deals with the negative binomial case.

Corollary 2. *Consider the problem of estimating $\mu = p/q$ in the negative binomial case. Then the expected lengths of CI_s, CI_{LR}, CI_J, and CI_R admit the expansions*

$$E(L_s) = 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2}{72np}] + O(n^{-2}),$$
$$E(L_{LR}) = 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 2\kappa^2(1+11p+p^2)}{72np}] + O(n^{-2}),$$
$$E(L_J) = 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 2\kappa^2(1+11p+p^2) - 2(2+13p+2p^2)}{72np}]$$
$$+O(n^{-2}),$$
$$E(L_R) = 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 9\kappa^2(1+6p+p^2)}{72np}] + O(n^{-2}).$$

Remark. From the expressions in Corollary 2, one can verify that, up to an error of $O(n^{-2})$ and pointwise at every $p > 0$, the ranking of the intervals is CI_s, CI_{LR}, CI_J, and CI_R from the shortest to the longest, provided $\kappa > \sqrt{17/23} = 0.86$. Note that this ranking exactly coincides with the ranking in the Poisson case. Again we see the quite impressive performance of the likelihood ratio interval.

Unlike the Poisson and the negative binomial cases, a uniform ranking in length pointwise for all p is not possible in the binomial case. However, if the expansions are integrated over p, then a clear ranking still emerges.

Consideration of the integrated expected length is natural as well as reasonable since we cannot have a uniform ranking of the different intervals pointwise. An alternative would be to consider the supremum over p, but averaging seems more natural and technically more feasible.

Corollary 3. *Consider the binomial case. The integrated expected lengths of CI_J, CI_{LR}, CI_R, and CI_s admit the expansions*

$$\int_0^1 E(L_J)dp = \frac{\kappa\pi}{4}n^{-1/2} - (\frac{37}{36} + \frac{5\kappa^2}{36})\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}),$$
$$\int_0^1 E(L_{LR})dp = \frac{\kappa\pi}{4}n^{-1/2} - (1 + \frac{5\kappa^2}{36})\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}),$$
$$\int_0^1 E(L_R)dp = \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}),$$
$$\int_0^1 E(L_s)dp = \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}).$$

Remark. Hence, up to the error n^{-2}, the ranking of the intervals is CI_J, CI_{LR}, CI_R, and CI_s from the shortest to the longest in integrated expected length. Note specifically the almost identical expansions for CI_J and CI_{LR}. Thus again we see that the likelihood ratio interval delivers solid performance in the binomial case as well.

9. Summary and Conclusions

The examples and theoretical results we have presented in this article demonstrate that the popular Wald interval is uniformly poor in a number of important lattice distributions, and better alternatives are needed. In fact, the systematic coverage bias is not confined to the discrete case but exists in the continuous case as well. Our comprehensive comparisons show that, fortunately, a number of alternative intervals provide significant improvements with respect to the disturbing negative bias in the coverage of the Wald interval. However, in coverage as well as length, two intervals always stand out. The likelihood ratio interval and the equal tailed Jeffreys interval are the best overall alternatives in all these cases. It is certainly true that the Rao score interval is easier to present and compute in an informal environment. But in the absence of an overriding need for very easy computation, the likelihood ratio and the Jeffreys interval can be resolutely recommended because of better length properties. Ultimately, the choice will no doubt be influenced by a user's personal preferences, and either one of the score, Jeffreys and the likelihood ratio intervals is a decisive improvement over the Wald interval. That is the principal message of this article.

10. Proofs

We present here the detailed proofs of the results for the likelihood ratio and the Jeffreys intervals. The proofs for the other intervals are slightly easier and are omitted for the reason of space. For interested readers, please see Brown, Cai, and DasGupta (2000) for the detailed proofs.

All of the distributions in the discrete natural exponential families 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).

Proposition 1. *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with mean μ, standard deviation $\sigma > 0$ and $E(|X_1|^3) < \infty$. Suppose that X_1 takes only integer values with a maximal span of one. Let $Z_n = n^{1/2}(\bar{X} - \mu)/\sigma$ where $\bar{X} = \sum_1^n X_i/n$ and let $F_n(z) = P(Z_n \le z)$. The two term Edgeworth expansion for $F_n(z)$ is given by*

$$\begin{aligned} F_n(z) = {} & \Phi(z) + p_1(z)\phi(z)n^{-1/2} + \sigma^{-1}(\frac{1}{2} - g(\mu, z))\phi(z)n^{-1/2} + p_2(z)\phi(z)n^{-1} \\ & + \{(\frac{1}{2} - g(\mu, z))\sigma p_3(z) - (\frac{1}{2}g^2(\mu, z) - \frac{1}{2}g(\mu, z) + \frac{1}{12})\}\sigma^{-2} z\phi(z)n^{-1} \\ & + O(n^{-3/2}), \end{aligned} \tag{32}$$

with $p_1(z) = \frac{1}{6}\beta_3(1 - z^2)$, $p_2(z) = -\frac{1}{24}\beta_4(z^3 - 3z) - \frac{1}{72}\beta_3^2(z^5 - 10z^3 + 15z)$, $p_3(z) = \frac{1}{6}\beta_3(z^2 - 3)$, where $\beta_3 = K_3/\sigma^3$ and $\beta_4 = K_4/\sigma^4$ are the skewness and the kurtosis of X_1, respectively.

If $z = z(n)$ depends on n and can be written as $z = z_0 + c_1 n^{-1/2} + c_2 n^{-1} + O(n^{-3/2})$ where z_0, c_1 and c_2 are constants, then

$$\begin{aligned} F_n(z) = {} & \Phi(z_0) + \tilde{p}_1(z)\phi(z_0)n^{-1/2} + \sigma^{-1}(\frac{1}{2} - g(\mu, z))\phi(z)n^{-1/2} + \tilde{p}_2(z)\phi(z_0)n^{-1} \\ & + \{\sigma(\frac{1}{2} - g(\mu, z))\tilde{p}_3(z_0) - (\frac{1}{2}g^2(\mu, z) - \frac{1}{2}g(\mu, z) + \frac{1}{12})\}\sigma^{-2} z_0\phi(z_0)n^{-1} \\ & + O(n^{-3/2}), \end{aligned} \tag{33}$$

$$\tilde{p}_1(z) = c_1 + \frac{1}{6}\beta_3(1 - z_0^2), \tag{34}$$

$$\tilde{p}_2(z) = c_2 - \frac{1}{2}z_0c_1^2 + \frac{1}{6}(z_0^3 - 3z_0)\beta_3c_1 - \frac{1}{24}\beta_4(z_0^3 - 3z_0) - \frac{1}{72}\beta_3^2(z_0^5 - 10z_0^3 + 15z_0), \quad (35)$$

$$\tilde{p}_3(z) = -c_1 + \frac{1}{6}\beta_3(z_0^2 - 3). \quad (36)$$

Proof. The expansion (32) follows from Theorem 23.1 of Bhattacharya and Rao (1976). See also Esseen (1945).

If $z = z_0 + c_1n^{-1/2} + c_2n^{-1} + O(n^{-3/2})$, we expand $\Phi(z)$, $\phi(z)$ and z^2 around z_0:

$$\Phi(z) = \Phi(z_0) + c_1\phi(z_0)n^{-1/2} + (c_2 - \frac{1}{2}z_0c_1^2)\phi(z_0)n^{-1} + O(n^{-3/2}), \quad (37)$$

$$\phi(z) = \phi(z_0) - z_0c_1\phi(z_0)n^{-1/2} + O(n^{-1}), \quad (38)$$

$$z^2 = z_0^2 + 2z_0c_1n^{-1/2} + O(n^{-1}). \quad (39)$$

We obtain (33) by plugging (37) - (39) into (32).

Remark. In (33), the second $O(n^{-1/2})$ and the second $O(n^{-1})$ terms are oscillation terms.

Expansion for the Likelihood Ratio Interval

We now prove Theorem 3. The proofs of the three cases are similar. We give the proof of the negative binomial case in detail; the proof for the binomial and Poisson cases are slightly simpler and are omitted.

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathrm{NB}(1, p)$. Then the MLE for p is $\hat{p} = \bar{X}/(1 + \bar{X})$ and $\Lambda_n = (p/\hat{p})^{n\bar{x}}(q/\hat{q})^n$. Let $z = \sqrt{n}(\bar{x} - p/q)/\sqrt{p/q^2}$. Then it follows, after some algebra, that $-2\log\Lambda_n \le \kappa^2$ is equivalent to

$$p(1+(pn)^{-1/2}z)\log(1+(pn)^{-1/2}z) - (1+p^{1/2}n^{-1/2}z)\log(1+p^{1/2}n^{-1/2}z) - \frac{q\kappa^2}{2n} \le 0. \quad (40)$$

Denote the LHS of (40) by $d(z)$. It is easy to verify that $d(\cdot)$ is a convex function and so has at most two roots. Denote by ℓ_{LR} and u_{LR} the roots of the equation $d(z) = 0$. So

$$d(\ell_{LR}) = d(u_{LR}) = 0. \quad (41)$$

We need to find an approximation to ℓ_{LR} and u_{LR}. Let $b(t) = (1 + t)\log(1 + t)$. Then $b(t)$ can be expanded into Taylor series as

$$b(t) = t + \frac{1}{2}t^2 - \frac{1}{6}t^3 + \frac{1}{12}t^4 + O(t^5). \quad (42)$$

Now applying (42) to (40), we can see, after some simplification, that $d(z) = 0$ is equivalent to

$$z^2 - \frac{1}{3}(1 + p)p^{-1/2}n^{-1/2}z^3 + \frac{1}{6}(1 + p + p^2)p^{-1}n^{-1}z^4 - \kappa^2 = O(n^{-3/2}). \quad (43)$$

Let now $z = \pm\kappa + b_1 n^{-1/2} + b_2 n^{-1}$. Plugging into (43) and solving for b_1 and b_2, we have $b_1 = \frac{1}{6}(1+p)p^{-1/2}\kappa^2 = \frac{1}{6}(1+2\mu)\sigma^{-1}\kappa^2$, $b_2 = \mp\frac{1}{72}(1-4p+p^2)p^{-1}\kappa^3 = \mp\frac{1}{72}(\sigma^{-2}-2)\kappa^3$. So the roots of $d(z)=0$ are

$$(\ell_{LR},\ u_{LR}) = \frac{1}{6}(1+p)p^{-1/2}\kappa^2 n^{-1/2} \pm \{\kappa - \frac{1}{72}(1-4p+p^2)p^{-1}\kappa^3 n^{-1}\} + O(n^{-3/2}). \tag{44}$$

The $+$ sign goes with u_{LR} and the $-$ sign with ℓ_{LR}. Hence,

$$P_p(p \in CI_{LR}) = P(\ell_{LR} \le \frac{n^{1/2}(\bar{x} - p/q)}{(p/q^2)^{1/2}} \le u_{LR}).$$

The binomial and the Poisson cases can be worked out similarly. The three cases together admit a unified expression $P_\mu(\mu \in CI_{LR}) = P(\ell_{LR} \le n^{1/2}(\hat{\mu}-\mu)/\sigma \le u_{LR})$ with

$$(\ell_{LR},\ u_{LR}) = \frac{1}{6}(1+2a_2\mu)\sigma^{-1}\kappa^2 n^{-1/2} \pm \{\kappa - \frac{1}{72}(\sigma^{-2}-2a_2)\kappa^3 n^{-1}\} + O(n^{-3/2}). \tag{45}$$

Now $P_{LR} = F_n(u_{LR}) - F_n(\ell_{LR})$, and the Edgeworth expansion (20) follows from (33).

Expansion for Jeffreys Intervals

We now prove Theorem 4. We use the direct expansion method to derive (21) (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)).

Contrary to the all at one stroke derivations for the other intervals in the entire discrete natural Exponential family with a quadratic variance function, for the Jeffreys interval a general Edgeworth expansion of the coverage probability seems to be basically impossible. So we are forced to consider the negative binomial and Poisson cases separately (The binomial case was derived in Brown, Cai, and DasGupta (2002)). These specific cases are already very complex, as is seen in the proof below.

Negative binomial case: The posterior distribution of p given $X=x$ is $Beta(x+1/2, n)$. Denote by $F(z; m_1, m_2)$ the cdf of the $Beta(m_1, m_2)$ distribution and denote by $B(\alpha; m_1, m_2)$ the inverse of the cdf. Then

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

Holding other parameters fixed, the function $F(p; X+1/2, n)$ is strictly decreasing in X (see, e.g., Johnson, Kotz and Balakrishnan (1995)). So there exist unique

$X_l = \rho_1(1-\alpha/2, p)$ and $X_u = \rho_2(\alpha/2, p)$ satisfying

$$F(p; X_l + 1/2, n) \le 1 - \alpha/2 \ \text{ and } \ F(p; X_l - 1/2, n) > 1 - \alpha/2,$$
$$F(p; X_u + 1/2, n) \ge \alpha/2 \ \text{ and } \ F(p; X_u + 3/2, n) < \alpha/2.$$

Therefore $P(p \in CI_J) = P(\ell_J \le \frac{n^{1/2}(\bar{X} - p/q)}{(p/q^2)^{1/2}} \le u_J)$ with $\bar{X} = X/n$ and

$$\ell_J = [\rho_1(1-\alpha/2, p) - np/q]/(np/q^2)^{1/2},$$
$$u_J = [\rho_2(\alpha/2, p) - np/q]/(np/q^2)^{1/2}. \tag{46}$$

The quantities ℓ_J and u_J are defined implicitly in (46) through ρ_1 and ρ_2. The proof of (21) for the negative binomial case requires an expansion for both ℓ_J and u_J. We do this below.

Step 1. Denote $x_1 = x - 1/2$, $n_1 = n + x - 3/2$, $p_1 = x_1/n_1$, $q_1 = 1 - p_1$, $s = (p_1 q_1)^{1/2} n_1^{-1/2}$, and $\gamma = \frac{\Gamma(n_1+2)}{\Gamma(x_1+1)\Gamma(n_1-x_1+1)}$. 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 $\psi(y) = \gamma \cdot s(p_1 + sy)^{x_1}(q_1 - sy)^{n_1 - x_1}$.

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

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

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

$$L(y) = -\frac{1}{2}\log(2\pi) + c_0 n_1^{-1} - \frac{1}{2}y^2 + c_1 n_1^{-1/2} y^3 + c_2 n_1^{-1} y^4 + O(n_1^{-3/2}), \tag{47}$$

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

$$\psi(y) = e^{L(y)} = \phi(y)[1 + c_1 n_1^{-1/2} y^3 + (c_0 + c_2 y^4 + \frac{1}{2} c_1^2 y^6) n_1^{-1}] + O(n_1^{-3/2}). \tag{48}$$

Step 3. Integrating both sides of (48) from $-\infty$ to z, we have

$$H(z) \equiv \int_{-\infty}^{z} \psi(y)dy = \Phi(z) - v_1(z)\phi(z) n_1^{-1/2} + v_2(z)\phi(z) n_1^{-1} + O(n_1^{-3/2}), \tag{49}$$

where $v_1(z) = -c_1(z^2+2)$ and $v_2(z) = -[\frac{1}{2}c_1^2(z^5+5z^3+15z)+c_2(z^3+3z)]$. (Note that the $O(n_1^{-3/2})$ term in (48) 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 (49) and solving for τ_1 and τ_2, after some algebra, we get $\tau_1 = \frac{1}{3}(1-2p_1)(z_\alpha^2+2)(p_1q_1)^{-1/2}$, $\tau_2 = (\frac{1}{36}z_\alpha^3 + \frac{11}{36}z_\alpha)(p_1q_1)^{-1} - (\frac{13}{36}z_\alpha^3 + \frac{71}{36}z_\alpha)$.

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

$$(p_l,\ p_u) = p_1 + \frac{1}{3}(1-2p_1)(\kappa^2+2)n_1^{-1}$$
$$\pm\left\{\kappa(p_1q_1)^{1/2}n_1^{-1/2} + \kappa(p_1q_1)^{1/2}n_1^{-3/2}[(\frac{1}{36}\kappa^2+\frac{11}{36})(p_1q_1)^{-1} - (\frac{13}{36}\kappa^2+\frac{71}{36})]\right\}$$
$$+O(n_1^{-2}). \tag{50}$$

Let

$$w_1(\mu) = (\frac{1}{3}\kappa^2 + \frac{1}{6})(1+2\mu), \tag{51}$$
$$w_2(\mu) = \left\{(\frac{13}{36}\kappa^3 + \frac{17}{36}\kappa)(\mu+\mu^2) + (\frac{1}{36}\kappa^3 + \frac{1}{18}\kappa)\right\}(\mu+\mu^2)^{-1/2}. \tag{52}$$

Rewriting the approximate limits (50) in terms of $\mu = p/(1-p)$, n, $\hat\mu = x/n$, after some algebra one has

$$(\mu_l,\ \mu_u) = (\hat\mu + w_1(\hat\mu)n^{-1}) \pm \{\kappa(\hat\mu+\hat\mu^2)^{1/2}n^{-1/2} + w_2(\hat\mu)n^{-3/2}\} + O(n^{-2}) \tag{53}$$

with the $+$ sign going with μ_u and the $-$ sign with μ_l.

Step 5. Now we expand the coverage probability by using (33). In order to use (33) we invert the inequalities $\mu_l \le \mu \le \mu_u$ into the form $\ell_J \le n^{1/2}(\hat\mu-\mu)/(\mu+\mu^2)^{1/2} \le u_J$. We need the following lemma. The proof, which we omit here, is straightforward.

Lemma 1. *Let w_1 and w_2 be two functions with a continuous first derivative. Then the roots x_* of the equations*

$$x \pm \kappa[x(1+x)]^{1/2}n^{-1/2} + w_1(x)n^{-1} + w_2(x)n^{-3/2} - \mu = 0 \tag{54}$$

can be expressed as

$$x_* = \mu \mp (\mu+\mu^2)^{1/2}\kappa n^{-1/2} + [(\frac{1}{2}+\mu)\kappa^2 - w_1(\mu \mp (\mu+\mu^2)^{1/2}\kappa n^{-1/2})]n^{-1}$$
$$-w_2(\mu)n^{-3/2} \mp \{[\frac{1}{8}(\mu+\mu^2)^{-1/2} + (\mu+\mu^2)^{1/2}]\kappa^3 - (\frac{1}{2}+\mu)(\mu+\mu^2)^{-1/2}w_1(\mu)\kappa\}n^{-3/2}$$
$$+O(n^{-2}). \tag{55}$$

All the $-$ $(+)$ *signs in* $\mp$ *in* (55) *go with the* $+$ $(-)$ *sign in* $\pm$ *in* (54).

Applying Lemma 1 to (53), we obtain $P(p \in CI_J) = P(\ell_J \le n^{1/2}(\hat{\mu}-\mu)/(\mu+\mu^2)^{1/2} \le u_J)$ with

$$\begin{aligned}(\ell_J, u_J) &= \pm\kappa + [(\frac{1}{2}+\mu)\kappa^2 - w_1(\mu \pm (\mu+\mu^2)^{1/2}\kappa n^{-1/2})](\mu+\mu^2)^{-1/2}n^{-1/2} \\ &\quad \pm\{[\kappa^3(\frac{1}{8}+\mu+\mu^2)-\kappa(\frac{1}{2}+\mu)w_1(p)](\mu+\mu^2)^{-1/2}+w_2(\mu)\}(\mu+\mu^2)^{-1/2}n^{-1} \\ &\quad +O(n^{-3/2}) \\ &= \frac{1}{6}(\kappa^2-1)(1+2\mu)(\mu+\mu^2)^{-1/2}n^{-1/2} \\ &\quad \pm\{\kappa + [(\frac{1}{36}\kappa^3 - \frac{7}{36}\kappa) - (\frac{1}{72}\kappa^3 + \frac{1}{36}\kappa)(\mu+\mu^2)^{-1}]n^{-1}\} + O(n^{-3/2}), \quad (56)\end{aligned}$$

with all $+$ signs going with u_J and all $-$ signs with ℓ_J. Now the expansion (21) for the negative binomial case follows from (33).

Poisson case: The posterior distribution of λ given $X = x$ is Gamma$(x + 1/2, 1/n)$. Denote by $F(z; m_1, m_2)$ the cdf of the Gamma(m_1, m_2) distribution and denote by $G(\alpha; m_1, m_2)$ the inverse of the cdf. Then

$$\begin{aligned}P_\lambda(\lambda \in CI_J) &= P(G(\alpha/2; X+1/2, 1/n) \le \lambda \le G(1-\alpha/2; X+1/2, 1/n)) \\ &= P(\alpha/2 \le F(\lambda; X+1/2, 1/n) \le 1-\alpha/2).\end{aligned}$$

Holding other parameters fixed, the function $F(\lambda; X+1/2, 1/n)$ is strictly decreasing in X. So there exist unique $X_l = \rho_1(1-\alpha/2, \lambda)$ and $X_u = \rho_2(\alpha/2, \lambda)$ satisfying

$$\begin{aligned}F(\lambda; X_l+1/2, 1/n) \le 1-\alpha/2 \;\text{ and }\; F(\lambda; X_l-1/2, 1/n) > 1-\alpha/2, \\ F(\lambda; X_u+1/2, 1/n) \ge \alpha/2 \;\text{ and }\; F(\lambda; X_u+3/2, 1/n) < \alpha/2.\end{aligned}$$

Therefore $P_\lambda(\lambda \in CI_J) = P(\ell_J \le \frac{n^{1/2}(\bar{X}-\lambda)}{\lambda^{1/2}} \le u_J)$ with

$$\ell_J = [\rho_1(1-\alpha/2, \lambda) - n\lambda]/(n\lambda)^{1/2} \;\text{ and }\; u_J = [\rho_2(\alpha/2, \lambda) - n\lambda]/(n\lambda)^{1/2}. \quad (57)$$

The quantities ℓ_J and u_J are defined implicitly in (57) through ρ_1 and ρ_2. Again, the proof of (21) for the Poisson case requires an expansion for both ℓ_J and u_J. We do this below.

Step 1. Denote $x_1 = x - 1/2$, $\lambda_1 = x_1/n$, $s = x_1^{1/2}n^{-1} = \lambda_1^{1/2}n^{-1/2}$, and $\gamma = n^{x+1/2}/\Gamma(x_1+1)$. Here λ_1 is the mode of the posterior distribution. Let $Y = (\lambda - \lambda_1)/s$. Then the conditional density of Y given $X = x$ is $\psi(y) = \gamma \cdot s(\lambda_1 + sy)^{x_1}e^{-n(\lambda_1+sy)}$.

Step 2. Let $L(y) = \log\psi(y)$. Then it is easy to see that $L'(0) = 0,\ L''(0) = -1$, $L^{(3)}(0) = 2\lambda_1^{-1/2}n^{-1/2}$, and $L^{(4)}(0) = -6\lambda_1^{-1}n^{-1}$. Applying Stirling's formula to the Gamma functions in $L(0)$, one gets, after some algebra $L(0) = -\frac{1}{2}\log(2\pi) - \frac{1}{12}\lambda_1^{-1}n^{-1} + O(n^{-3/2})$. Expanding $L(y)$ at 0, one has

$$L(y) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}y^2 + \frac{1}{3}\lambda_1^{-1/2}y^3n^{-1/2} - (\frac{1}{12} + \frac{1}{4}y^4)\lambda_1^{-1}n^{-1} + O(n^{-3/2}), \tag{58}$$

$$\psi(y) = e^{L(y)} = \phi(y)[1 + \frac{1}{3}\lambda_1^{-1/2}y^3n^{-1/2} + (-\frac{1}{12} - \frac{1}{4}y^4 + \frac{1}{18}y^6)\lambda_1^{-1}n^{-1}] + O(n^{-3/2}). \tag{59}$$

Step 3. Integrating both sides of (59) from $-\infty$ to z, we have

$$H(z) \equiv \int_{-\infty}^{z}\psi(y)dy = \Phi(z) + v_1(z)\phi(z)n^{-1/2} + v_2(z)\phi(z)n^{-1} + O(n^{-3/2}), \tag{60}$$

where $v_1(z) = -\frac{1}{3}\lambda_1^{-1/2}(z^2+2)$ and $v_2(z) = -\lambda_1^{-1}(\frac{1}{18}z^5 + \frac{1}{36}z^3 + \frac{1}{12}z)$.

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 $\xi_{\alpha,n} = z_\alpha + \tau_1 n^{-1/2} + \tau_2 n^{-1} + o(n^{-1})$. Plugging in (60) and solving for τ_1 and τ_2, after some algebra we get $\tau_1 = \frac{1}{3}(z_\alpha^2+2)\lambda_1^{-1/2}$ and $\tau_2 = (\frac{1}{36}z_\alpha^3 + \frac{11}{36}z_\alpha)\lambda_1^{-1}$.

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

$$(\lambda_l,\ \lambda_u) = \lambda_1 + \frac{1}{3}(\kappa^2+2)n^{-1} \pm \{\kappa\lambda_1^{1/2}n^{-1/2} + (\frac{1}{36}\kappa^3 + \frac{11}{36}\kappa)\lambda_1^{-1/2}n^{-3/2}\} + O(n^{-2}). \tag{61}$$

Rewriting the approximate limits (61) in terms of $\hat\lambda = x/n$, one has, after some algebra,

$$(\lambda_l,\ \lambda_u) = \hat\lambda + (\frac{1}{3}\kappa^2 + \frac{1}{6})n^{-1} \pm \{\kappa\hat\lambda^{1/2}n^{-1/2} + (\frac{1}{36}\kappa^3 + \frac{1}{18}\kappa)\hat\lambda^{-1/2}n^{-3/2}\} + O(n^{-2}). \tag{62}$$

Step 5. Now we expand the coverage probability by using (33). In order to use (33) we invert the inequalities $\lambda_l \le \lambda \le \lambda_u$ into the form $\ell_J \le n^{1/2}(\hat\lambda - \lambda)/\lambda^{1/2} \le u_J$. We need the following lemma. The proof, which we omit here, is straightforward.

Lemma 2. *Let w be a function with continuous first derivative. Then the roots x_* of the equations*

$$x \pm x^{1/2}\kappa n^{-1/2} + (\frac{1}{3}\kappa^2 + \frac{1}{6})n^{-1} + w(x)n^{-3/2} - \lambda = 0 \tag{63}$$

can be expressed as

$$x_* = \lambda + \frac{1}{6}(\kappa^2 - 1)n^{-1} - w(\lambda)n^{-3/2} \mp \{\lambda^{1/2}\kappa n^{-1/2} - (\frac{1}{24}\kappa^3 + \frac{1}{12}\kappa)\lambda^{-1/2}n^{-3/2}\} + O(n^{-2}). \quad (64)$$

The $-$ $(+)$ *sign in* $\mp$ *in* (*64*) *goes with the* $+$ $(-)$ *sign in* $\pm$ *in* (*63*).

Applying Lemma 2 to (62), we obtain $P(\lambda \in CI_J) = P(\ell_J \leq n^{1/2}(\hat{\lambda} - \lambda)/\lambda^{1/2} \leq u_J)$ with

$$(\ell_J, u_J) = \frac{1}{6}(\kappa^2 - 1)\lambda^{-1/2}n^{-1/2} \pm \{\kappa - (\frac{1}{72}\kappa^3 + \frac{1}{36}\kappa)\lambda^{-1}n^{-1}\} + O(n^{-3/2}). \quad (65)$$

The expansion (21) for the Poisson case now follows from (33).

Expansions for Expected Length

We now prove Theorem 5. The derivation of the expected length expansions in equations (27)$-$(31) is algebraically intense. We report the main steps below and skip the many intermediate algebraic simplifications. We denote below $Z_n = (\bar{X} - \mu)(\mu + a_2\mu^2)^{-1/2}n^{1/2}$.

The interval CI_J*.* The limits of CI_J admit the general representation

$$\bar{X} + w_1(\bar{X})n^{-1} \pm \{\kappa(\bar{X} + a_2\bar{X}^2)^{1/2}n^{-1/2} + w_2(\bar{X})n^{-3/2}\} + R_J(n),$$

where the remainder $R_J(n)$ satisfies $E(|R_J(n)|) = O(n^{-2})$, and the function $w_2(\cdot)$ is defined as $w_2(\mu) = \frac{1}{36}(\mu + a_2\mu^2)^{-1/2}\{(\kappa^3 + 3\kappa) + a_2(\mu + a_2\mu^2)(13\kappa^3 + 17\kappa)\}$. Thus, directly, the length L_J of CI_J satisfies

$$\begin{aligned} E(L_J) &= E[2\kappa(\bar{X} + a_2\bar{X}^2)^{1/2}n^{-1/2} + 2w_2(\bar{X})n^{-3/2}] + O(n^{-2}) \\ &= 2\kappa(\mu + a_2\mu^2)^{1/2}n^{-1/2}[1 - \frac{1}{8}(\mu + a_2\mu^2)^{-1}n^{-1}] + 2w_2(\mu)n^{-3/2} + O(n^{-2}), \end{aligned}$$

which simplifies to (31) after some algebra.

The interval CI_{LR}*.* This is the most complex case and the expansions for the expected length have to be first derived separately for each of the three cases. The three separate expansions can then be unified into the general expression (30) stated in Theorem 5.

The limits of the likelihood ratio interval are the roots of the equation $-\log\Lambda_n = \kappa^2/2$, where Λ_n is the likelihood ratio statistic for testing a simple null on the relevant parameter. The general method followed in each of the three cases is to first find asymptotic expansions for these roots up to the order $n^{-3/2}$ and then find expansions for the expected difference of the roots. The asymptotic expansions for the roots are found in each case by the method of Theorem 3, as described in equations (40)−(45). We will now describe the main steps in the Poisson and the negative binomial cases. (The binomial case was derived in Brown, Cai, and DasGupta (2002)).

1. *The Poisson case*:

Step 1. The likelihood ratio Λ_n is given by

$$\Lambda_n = \frac{\lambda^{n\bar{X}} e^{-n\lambda}}{\bar{X}^{n\bar{X}} e^{-n\bar{X}}}. \tag{66}$$

For an expansion of the expected length up to the order $n^{-3/2}$, the case $\bar{X} = 0$ does not matter. If $\bar{X} > 0$, by a unimodality argument, the equation $-\log \Lambda_n = \kappa^2/2$ has two roots in λ; these are the limits of the interval CI_{LR}. Writing $t = \lambda/\bar{X} - 1$, the roots of $-\log \Lambda_n = \kappa^2/2$ satisfy $t - \log(1+t) = \kappa^2/(2n\bar{X})$.
Step 2. By the same steps as in equations (40)−(45), the roots, say $\underline{t}$ and $\bar{t}$, satisfy

$$\underline{t} = -\kappa(n\bar{X})^{-1/2} + \frac{1}{3}\kappa^2(n\bar{X})^{-1} - \frac{1}{36}\kappa^3(n\bar{X})^{-3/2} + R_{1,n}, \tag{67}$$

$$\bar{t} = \kappa(n\bar{X})^{-1/2} + \frac{1}{3}\kappa^2(n\bar{X})^{-1} + \frac{1}{36}\kappa^3(n\bar{X})^{-3/2} + R_{2,n}, \tag{68}$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1,\ 2$. From (67) and (68), the length L_{LR} of CI_{LR} satisfies

$$E(L_{LR}) = 2\kappa E(\bar{X}^{1/2})n^{-1/2} + \frac{1}{18}\kappa^3\lambda^{1/2}n^{-3/2} + O(n^{-2}). \tag{69}$$

Step 3. Writing $Z_n = n^{1/2}(\bar{X} - \lambda)/\lambda^{1/2}$, a straightforward calculation has $E(\bar{X}^{1/2}) = \lambda^{1/2}[1-(8n\lambda)^{-1}] + O(n^{-3/2})$, and so from (69) one obtains $E(L_{LR}) = 2\kappa\lambda^{1/2}[1-(8n\lambda)^{-1}]n^{-1/2} + \frac{1}{18}\kappa^3\lambda^{-1/2}n^{-3/2} + O(n^{-2})$, which easily simplifies to $E(L_{LR}) = 2\kappa\lambda^{1/2}\ n^{-1/2}(1-(9-2\kappa^2)/(72n\lambda)) + O(n^{-2})$.

2. *The negative binomial case*:

Step 1. In our parametrization (See Section 2), the mean $\mu = p/q$, and so $p = \mu/(1+\mu)$. The likelihood ratio is given by

$$\Lambda_n = \left(\frac{\mu}{\bar{X}}\right)^{n\bar{X}} \left(\frac{1+\bar{X}}{1+\mu}\right)^{n(1+\bar{X})}. \tag{70}$$

Assume $\bar{X} > 0$ and write $t = \mu/\bar{X} - 1$. The equation $-\log \Lambda_n = \kappa^2/2$ is equivalent to $(1+\bar{X})\log(1+\bar{X}t/(1+\bar{X})) - \bar{X}\log(1+t) = \kappa^2/2n$.

Step 2. The roots $\underline{t}$ and $\bar{t}$ of this equation satisfy

$$\underline{t} = -\kappa[(1+\bar{X})/\bar{X}])^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1+2\bar{X})(n\bar{X})^{-1} - \frac{1}{36}\kappa^3(1+13\bar{X}+13\bar{X}^2)(1+\bar{X})^{-1/2}(n\bar{X})^{-3/2} + R_{1,n}, \quad (71)$$

$$\bar{t} = \kappa[(1+\bar{X})/\bar{X}])^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1+2\bar{X})(n\bar{X})^{-1} + \frac{1}{36}\kappa^3(1+13\bar{X}+13\bar{X}^2)(1+\bar{X})^{-1/2}(n\bar{X})^{-3/2} + R_{2,n}, \quad (72)$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1,\ 2$. From (71) and (72),

$$E(L_{LR}) = 2\kappa E[(\bar{X}+\bar{X}^2)^{1/2}]n^{-1/2} + \frac{1}{18}\kappa^3(1+13\mu+13\mu^2)(\mu+\mu^2)^{-1/2}n^{-3/2} + O(n^{-2}). \quad (73)$$

Step 3. Let $Z_n = n^{1/2}(\bar{X}-\mu)/(\mu+\mu^2)^{1/2}$. From (73) by a straightforward expansion,

$$\begin{aligned} E(L_{LR}) &= 2\kappa(\mu+\mu^2)^{1/2}\{1-[8n(\mu+\mu^2)]^{-1}\}n^{-1/2} \\ &\quad + \frac{1}{18}\kappa^3(1+13\mu+13\mu^2)(\mu+\mu^2)^{-1/2}n^{-3/2} + O(n^{-2}) \\ &= 2\kappa(\mu+\mu^2)^{1/2}n^{-1/2}[1-\frac{9-2\kappa^2(1+13\mu+13\mu^2)}{72n(\mu+\mu^2)}] + O(n^{-2}). \end{aligned}$$

Remark. The unified expression

$$E(L_{LR}) = 2\kappa(\mu+a_2\mu^2)^{1/2}n^{-1/2}[1-\frac{9-2\kappa^2-26a_2\kappa^2(\mu+a_2\mu^2)}{72n(\mu+\mu^2)}] + O(n^{-2}).$$

Follows from the specific expressions for the three cases.

Acknowledgement

This research was supported in part by NSF Grants DMS-9971751, DMS-0072578, and NSA Grant MDA 904-97-1-0031.

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 edition. Springer-Verlag, New York.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation And Asymptotic Expansions.* Wiley, New York.

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

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., Cai, T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and Edgeworth expansions. *Ann. Statist.* **30**, in press.

Brown, L. D., Cai, T. and DasGupta, A. (2000). Interval estimation in exponential families. Technical Report, Department of Statistics, University of Pennsylvania.

Casella, G. and Robert, C. (1989). Refining Poisson confidence intervals. *Canad. J. Statist.* **17**, 45-57.

Clevenson, M. and Zidek, J. (1975). Simultaneous estimation of the means of independent Poisson laws, *J. Amer. Stat. Assoc.* **70**, 698-705.

Crow, E. L. and Gardner, R. S. (1959). Confidence intervals for the expectation of a Poisson variable. *Biometrika* **46**, 441-453.

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.

Hall, P. (1992). *The Bootstrap And Edgeworth Expansion.* Springer-Verlag, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Second edition, Vol. 2. Wiley, New York.

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

Kaplan, S. (1983). On a "two stage" Bayesian procedure for determining failure rates from experimental data, *IEEE Trans. on Power App. and Sys.* **102**, 195-202.

Kabaila, P. and Byrne, J. (2001). Exact short Poisson confidence intervals. *Canad. J. Statist.* **29**, 99-106.

Lui, K.-J. (1995). Confidence limits for the population prevalence rate based on the negative binomial distribution. *Statist. Medicine* **14**, 1471-1477.

Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10**, 65-80.

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.* **11**, 515-529.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Santner, T. J. and Duffy, D. E. (1989). *The Statistical Analysis of Discrete Data.* Springer-Verlag, Berlin.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Department of Statistics, The Wharton School, University of Pennsylvania, 3000 Steinberg Hall-Dietrich Hall, 3620 Locust Walk, Philadelphia, PA 19104-6302, U.S.A.

E-mail: lbrown@wharton.upenn.edu

Department of Statistics, The Wharton School, University of Pennsylvania, 3000 Steinberg Hall-Dietrich Hall, 3620 Locust Walk, Philadelphia, PA 19104-6302, U.S.A.

E-mail: tcai@wharton.upenn.edu

Purdue University, 150 N. University Street, West Lafayette, IN 47907-2068, U.S.A.

E-mail: dasgupta@stat.purdue.edu

(Received July 2001; accepted July 2002)