# INTERVAL ESTIMATION IN DISCRETE EXPONENTIAL FAMILY

by

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

Anirban DasGupta
Purdue University

Technical Report #00-01

Department of Statistics
Purdue University
West Lafayette, IN USA

January 2000

# Interval Estimation in Discrete Exponential Family*

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

January 4, 2000

**Abstract**

The erratic and poor performance of the popular Wald confidence interval for a Binomial proportion was demonstrated in Brown, Cai and DasGupta (1999a,b) through extensive computation and analytical calculations. It was also shown that the equal tailed Jeffreys prior interval, the score interval obtained from inversion of Rao's score test, and another specific interval suggested in Agresti and Coull (1998) provide significant improvements over the Wald interval in the Binomial problem.

We show in this article that across a class of practically important discrete distributions, most of the key phenomena evidenced in the Binomial case are almost exactly replicated. In the discrete Exponential family with a quadratic variance function, we derive two term Edgeworth expansions for the coverage and two term expansions for the expected length for the Wald and four other natural intervals. They are the Rao score interval, the likelihood ratio interval, the equal tailed Jeffreys interval, and an interval similar to the Agresti-Coull interval for the Binomial case.

These calculations reveal a great amount of common structure. For instance, the calculations show that the Jeffreys and the likelihood ratio interval nearly annihilate the systematic bias term in the Edgeworth expansion in all cases, and these same two intervals are also the best in expected length in all the cases.

Simplicity of computation aside, our theoretical calculations show that the likelihood ratio and the Jeffreys interval are the two best all rounded alternatives in all these lattice problems.

The results are complemented by an array of illustrative examples.

**Keywords:** 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.

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

# 1 Introduction

Confidence intervals for a binomial proportion have been studied by many researchers for more than fifty years, and it had been generally known that the popular Wald confidence interval $\hat{p} \pm z_{\alpha/2}\, n^{-1/2}(\hat{p}(1-\hat{p}))^{1/2}$ was deficient for $p$ near 0 or 1 in the sense of poor coverage probability. In two recent interesting articles, Santner (1998), and Agresti and Coull (1998) initiated a more serious discussion of the severely poor performance of the Wald interval. Agresti and Coull (1998) also did a comparison of a number of intervals for the nominal 95% case for selected values of $n$ and suggested explicitly a new confidence interval. Brown, Cai and DasGupta (1999a, b) then addressed the problem on a more comprehensive basis, and based on extensive numerical evaluations and analytical calculations, concluded that the Wald interval indeed is far too poor and unreliable, and recommended the Jeffreys prior interval or the score interval for small $n$, and the interval previously suggested in Agresti and Coull (1998) for larger $n$. Among other things, Brown, Cai and DasGupta (1999a, b) showed that the Wald interval suffers from a systematic negative bias as well as a chaotic oscillation in its coverage probability, and contrary to common perception, these problems are not just for $p$ near 0 or 1, and not just for small or moderate $n$. Appropriate Edgeworth expansions in Brown, Cai and DasGupta (1999b) identified the source of these problems in the lattice nature of the Binomial distribution, and an incorrect centering of the popular Wald interval.

The principal purpose of this article is to present a unified and coherent description of the interval estimation problem across a class of important discrete distributions. The satisfactory part of this work is that the unification is not merely in the form of a technical extension. The unification is many faceted. It will be seen in the constancy of the unpredictable performance of the popular Wald interval, in wrong centering being a key reason for this problem, in commonality of the mathematical results, and a sound uniformity in the ultimate resolutions of the problem. For example, the results show that across a class of discrete distributions, the equal tailed Jeffreys prior interval and the interval obtained from inversion of the popular likelihood ratio test are perhaps the best all rounded alternatives to the Wald interval. In addition, with a mystical conformity, these two intervals founded on altogether different methods have virtually identical coverage and length properties. In fact, the nearly identical Edgeworth and length expansions for these two intervals are quite intriguing. The results also show that the interval produced by inversion of Rao's score test also always provides major improvements in coverage, but suffers in comparison in parsimony with respect to length.

It turns out that for the purpose of the analytical calculations we are interested in, the correct theoretical framework is the discrete Exponential family with a quadratic variance function. Morris (1982, 1983) gives a very comprehensive account of the Exponential family with quadratic variance functions; he shows that among the discrete ones, the Binomial, Negative Binomial, and the Poisson have the quadratic variance property. Fortunately, these are exactly the most important lattice cases in practical applications. The practical importance of the Binomial case is obvious; Santner and Duffy (1989) give very interesting examples of the practical importance of the interval estimation problem in the Negative Binomial and the Poisson case. The applications range from oil exploration (Clevenson and

Zidek (1975)), safety of nuclear plants (Kaplan (1983)), to epidemiology (Lui (1995)), and numerous others. These problems are certainly practically important.

In Section 2, the discrete 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, with a brief motivation and background, alternative intervals. In contrast to Brown, Cai and DasGupta (1999a, b), 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 their 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.

In Section 7, 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 case. This is certainly an exceptionally positive feature of the likelihood interval. 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 9, a technical appendix, contains the proofs. The results are illustrated by various examples and computation throughout the article.

## 2 Discrete Natural Exponential Family

We consider interval estimation of the mean in the discrete natural exponential family (DNEF) with quadratic variance functions (QVF) (the variance is at most a quadratic function of the mean). DNEFs with a QVF consist of three important discrete distributions: binomial, negative binomial, and Poisson (see, e.g., Morris (1982) and Brown (1986)). First we state some basic facts about the discrete exponential family for use in the rest of this article.

The distributions in a natural exponential family have the form

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

$\xi$ is called the natural parameter. The mean, variance and cumulant generating function are

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

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 $\xi$ only through the mean $\mu$, and indeed,

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

for constants $a_0$, $a_1$, and $a_2$.

DNEFs with a QVF consist of the Binomial, Negative Binomial, and the Poisson distribution. Let us list the important facts for the three distributions separately.

- Binomial, Bin$(1,p)$: 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 in this case $a_0 = 0$, $a_1 = 1$ and $a_2 = -1$.

- Negative binomial, NBin$(1,p)$, the number of successes before the first failure; let $p$ = probability of success. Now $\xi = \log p$, $\psi(\xi) = -\log(1-e^{\xi})$, and $h(x) = 1$. And $\mu = p/q$, $V(\mu) = p/q^2 = \mu + \mu^2$. Thus in this case, $a_0 = 0$, $a_1 = 1$ and $a_2 = 1$.

- Poisson, Poi$(\lambda)$: In this case, $\xi = \log\lambda$, $\psi(\xi) = e^{\xi}$, and $h(x) = 1/x!$. And $\mu = \lambda$, $V(\mu) = \mu$. Thus here $a_0 = 0$, $a_1 = 1$ and $a_2 = 0$.

Note that in all three cases $a_0 = 0$ and $a_1 = 1$. Hereafter we will drop $a_0$ and $a_1$ in (1) and write $V(\mu) = \mu + a_2\mu^2$.

The common setup throughout the rest of the article is that we have iid observations $X_1, X_2, ..., X_n \sim f(x|\xi)$ with $f$ as one of the three cases above, and we want to estimate $\mu$. Estimation of monotone functions of $\mu$ is certainly a relevant and important problem, but is not considered here mainly due to space consideration.

# 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 (1999a,b) described that the problems persist for even very large $n$ and even when $p$ is near or even exactly equal to .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 will show by two quick examples that indeed this is the case. The examples are for the Poisson case.

**Example 1.** Suppose we want to estimate a Poisson mean $\lambda$ on the basis of $n$ iid 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. Figure 1 plots the coverage probability of the Wald interval for variable $\lambda$. The most striking aspect of the plot is that the coverage never reaches .99; we see the clear systematic negative bias. What was previously observed in the binomial proportion problem resurfaces in the Poisson problem. The Wald interval has a systematic bias problem in discrete cases in some generality.

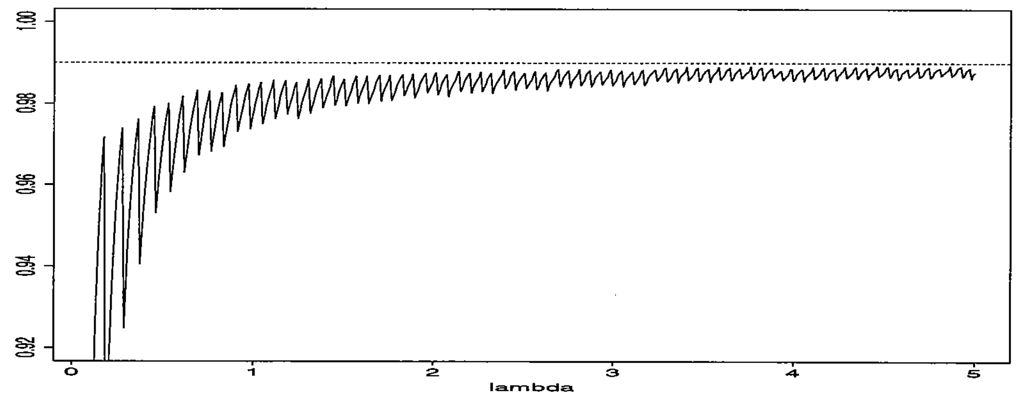


Figure 1: Coverage probability of the Wald interval for $n = 20$ and $.1 \leq \lambda \leq 5$.

**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 $\lambda$, say, $\lambda = .5$. Naively, one may expect that the coverage gets systematically closer to the nominal level as the sample size $n$ 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 .936, when $n = 16$, the coverage is only .892, when $n = 18$, the coverage is .940, and yet when $n = 72$, the coverage is .933, actually smaller than the coverage for $n = 9$ ! Exactly as it was seen in Brown, Cai and DasGupta (1999a) in the Binomial case, the phenomenon of unpredictable arrival of large unlucky values of $n$ reappears in the Poisson problem. The Wald interval is simply too erratic and unreliable in these lattice problems, and alternative intervals with better properties are earnestly needed.

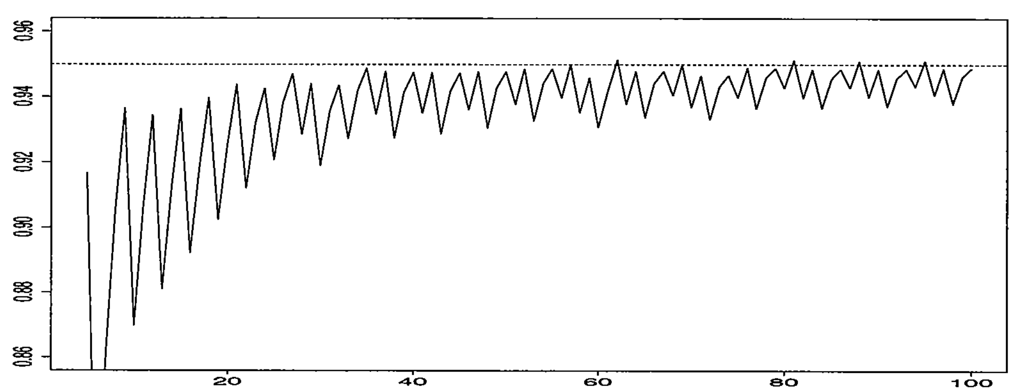


Figure 2: Coverage probability of the Wald interval for fixed $\lambda = .5$ and variable $n$ from 5 to 100.

### 3.1 Explaining the Bias

The standard interval is based on the fact that

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

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 helpful in understanding part of the reason for the very poor performance of the Wald interval in these lattice problems.

As in (9), denote

$$Z_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\mu + a_2 \mu^2}}.$$

Then simple algebra shows

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

A standard Taylor expansion and calculations using the central moments of $f(x|\xi)$ yield an approximation to the bias:

$$\begin{aligned} EW_n &= -(\frac{1}{2} + a_2\mu)\sigma^{-1}n^{-1/2} + (\frac{3}{8} + a_2\mu + a_2^2\mu^2)\sigma^{-5}K_3 n^{-3/2} \\ &\quad - (\frac{1}{2} + a_2\mu)(\frac{15}{8} + 3a_2\mu + 3a_2^2\mu^2)\sigma^{-3}n^{-3/2} + O(n^{-2}), \end{aligned} \tag{2}$$

where $K_3$ is the third cumulant of $f$. Denote $\rho = V(\mu) = \psi''(\xi)$. Then by chain rule,

$$K_3 = \frac{d\rho}{d\xi} = \frac{d\rho}{d\mu} \cdot \frac{d\mu}{d\xi} = (1 + 2a_2\mu) \cdot V(\mu) = (1 + 2a_2\mu)\sigma^2. \tag{3}$$

Now (2) yields

$$EW_n = -(\frac{1}{2} + a_2\mu)\,[1 + (\frac{9}{8} + a_2\mu + a_2^2\mu^2)\sigma^{-2}n^{-1}]\,\sigma^{-1}n^{-1/2} + O(n^{-2}). \tag{4}$$

Again, consider the three special cases separately.

- Binomial (Bin$(1,p)$): In this case, $a_2 = -1$, $\mu = p$, $V(\mu) = pq$, and $K_3 = pq(q-p)$. After some algebra, we have from (4)

$$EW_n = \frac{p - 1/2}{\sqrt{npq}}(1 + \frac{7}{2n} + \frac{9(p-1/2)^2}{2npq}) + O(n^{-2}). \tag{5}$$

- Negative Binomial (NBin$(1,p)$): Now, $a_2 = 1$, $\mu = p/q$, $V(\mu) = p/q^2$, and $K_3 = (p+p^2)/q^3$. It follows from (2)

$$EW_n = -\frac{1+p}{2\sqrt{np}}(1 + \frac{1}{n} + \frac{9q^2}{8np}) + O(n^{-2}). \tag{6}$$

- Poisson (Poi$(\lambda)$): In this case, $a_2 = 0$, and $\mu = V(\mu) = K_3 = \lambda$. We have

$$EW_n = -\frac{1}{2\sqrt{n\lambda}}(1 + \frac{9}{8n\lambda}) + O(n^{-2}). \tag{7}$$

### 3.2 Discussion

These bias expressions give us useful information. From equation (5), 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 (1999a, b) show that 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 equation (7), we see that $W_n$ appears to have a negative bias for all $\lambda$. So the bias problem persists, but now in contrast to the Binomial case, the center of the Wald interval for $\lambda$ should always be moved up. And indeed, our subsequent calculations confirm that by moving up the center of the Wald interval, we can significantly curtail the systematic negative bias in the coverage of the Wald interval (see Figure 6).

Let us also see some actual numerical evidence of a serious bias problem. Figure 3 below plots the bias (i.e. $E(W_n)$) for $\lambda = .5$ and $\lambda = 1$ and for $n = 1$ to 100. The clearly significant bias even when $n\lambda$ is 40 or so is certainly disconcerting.

Similar disturbing bias is also present in the Negative Binomial problem, and again examination of equation (6) would suggest that here too the center of the Wald interval needs to be moved up to address a potential bias problem.

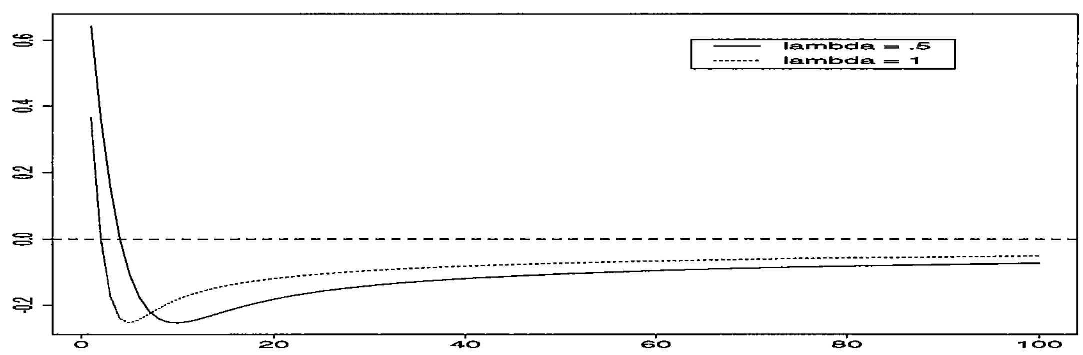


Figure 3: Bias in Poisson case for $n = 1$ to 100 with $\lambda = .5$ and $\lambda = 1$.

# 4 The Confidence Intervals

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

$$W_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma}} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\hat{\mu} + a_2\hat{\mu}^2}} \xrightarrow{\mathcal{L}} N(0, 1), \tag{8}$$

and

$$Z_n = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\mu + a_2\mu^2}} \xrightarrow{\mathcal{L}} N(0, 1), \tag{9}$$

We can construct confidence intervals for $\mu$ based on the approximations (8) and (9).

1. The *Wald interval* is based on (8):

$$CI_s = \hat{\mu} \pm \kappa\hat{\sigma}n^{-1/2} = \hat{\mu} \pm \kappa(\hat{\mu} + a_2\hat{\mu}^2)^{1/2}n^{-1/2} \tag{10}$$

2. Define the *recentered interval* as

$$CI_{rs} = \tilde{\mu} \pm \kappa\hat{\sigma}n^{-1/2} = \tilde{\mu} \pm \kappa(\hat{\mu} + a_2\hat{\mu}^2)^{1/2}n^{-1/2}. \tag{11}$$

3. The *score interval* is based on (9). This interval is formed by inverting Rao's score test of equal-tailed tests of $H_0 : \mu = \mu_0$. Hence, one accepts $H_0$ based on Rao's score test if and only if $\mu_0$ 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

$$CI_R = \tilde{\mu} \pm \frac{\kappa n^{1/2}}{n - \kappa^2 a_2}(\hat{\mu} + a_2\hat{\mu}^2 + \frac{\kappa^2}{4n})^{1/2}. \tag{12}$$

4. The "Agresti-Coull" ($AC$) *interval* has the same simple form as the standard interval $CI_s$, but with a different $\hat{\mu}$ and a modified value for $n$. Let $\tilde{X} = X+\kappa^2/2$, $\tilde{n} = n-\kappa^2 a_2$ and $\tilde{\mu} = \tilde{X}/\tilde{n}$. Then the AC interval is defined as

$$CI_{AC} = \tilde{\mu} \pm \kappa(V(\tilde{\mu}))^{1/2}\tilde{n}^{-1/2} = \tilde{\mu} \pm \kappa(\tilde{\mu} + a_2\tilde{\mu}^2)^{1/2}\tilde{n}^{-1/2}. \tag{13}$$

   Note that here $X = \sum_{i=1}^{n} X_i$.

## 4.1 Likelihood Ratio Intervals

In Section 4 we introduced the Wald and the score interval which are obtained by inversion of the acceptance regions of the Wald and the Rao's score test, respectively. Another method suggests itself. This is to construct an interval 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 $\Lambda_n$ is the likelihood ratio

$$\Lambda_n = L(\mu_0)/\sup_{\mu} L(\mu),$$

with $L$ as the likelihood function based on $n$ iid observations from the underlying density. 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 may further convince us that the interval deserves very serious consideration.

Figure 4 plots the exact coverage probability of the likelihood ratio interval and four other intervals we discussed earlier for $\lambda = .5$ and $n$ from 5 to 100 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 in Figure 4, we see that the likelihood ratio interval has substantially smaller oscillation. Much of our subsequent technical calculations will confirm this impressive performance of the likelihood ratio interval.

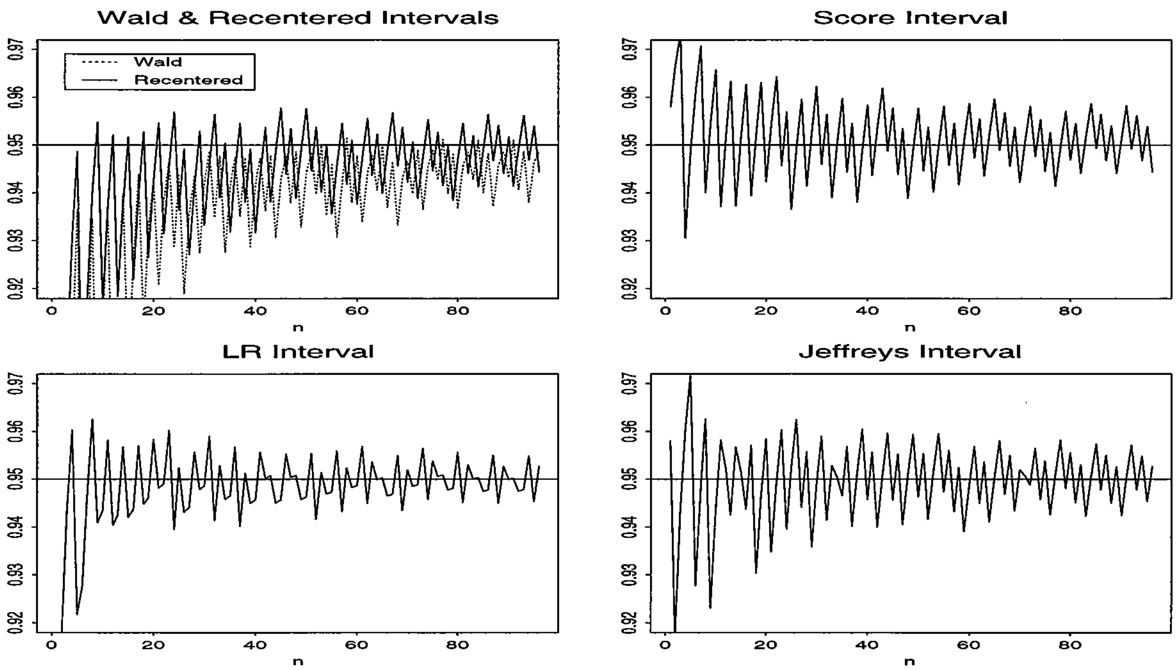


Figure 4: Coverage probability of five confidence intervals for a Poisson mean with $\lambda = .5$, $\alpha = .05$ and $n$ from 5 to 100.

## 4.2 The Jeffreys Interval

Now consider the Jeffreys interval. Denote $b(\cdot) = (\psi')^{-1}(\cdot)$. Then $b$ is a strictly increasing function and $\xi = b(\mu)$. The Fisher information about $\mu$ 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 $\psi''(b(\mu)) = \psi''(\xi) = \mu + a_2\mu^2$ and $b'(\mu) = 1/\psi''(\xi) = (\mu + a_2\mu^2)^{-1}$, we have

$$I(\mu) = n(\mu + a_2\mu^2)^{-1}.$$

Thus the Jeffreys prior is proportional to $I^{1/2}(\mu) = n^{1/2}(\mu + a_2\mu^2)^{-1/2}$ and the posterior

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

5. The *Jeffreys equal-tailed interval* for $\mu$ is given by

$$CI_J = [J_{\alpha/2},\ J_{1-\alpha/2}] \tag{14}$$

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

Consider the three special distributions for illustration separately.

- Binomial: here $\psi(\xi) = \log(1+e^{\xi})$ and $b(\mu) = \log(\mu/(1-\mu))$. 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

$$CI_J = [p_l,\ p_u] = [B_{\alpha/2,X+1/2,\,n-X+1/2},\ B_{1-\alpha/2,X+1/2,\,n-X+1/2}]. \tag{15}$$

- Negative Binomial: here $\psi(\xi) = -\log(1-e^{\xi})$ and $b(\mu) = \log(\mu/(1+\mu))$. The Jeffreys prior for $\mu$ is proportional to $\mu^{-1/2}(1+\mu)^{-1/2}$ and the posterior is a beta-prime distribution (see Johnson, *et al.* (1995)).

  The Jeffreys interval is transformation-invariant. We can obtain the Jeffreys prior interval for $\mu$ through the Jeffreys prior 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

$$CI_J^p = [p_l,\ p_u] = [B_{\alpha/2,X+1/2,\,n},\ B_{1-\alpha/2,X+1/2,\,n}]. \tag{16}$$

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

$$CI_J = [p_l/(1-p_l),\ p_u/(1-p_u)]. \tag{17}$$

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

$$CI_J = [\lambda_l,\ \lambda_u] = [G_{\alpha/2,X+1/2,\,1/n},\ G_{1-\alpha/2,X+1/2,\,1/n}]. \tag{18}$$

**Example 3**. We have introduced a number of different confidence intervals above as alternatives to the Wald interval. To the practitioner, a key question would be how do the actual limits of these various intervals differ among themselves. If two different intervals have very similar limits, a practitioner is likely to consider them as practically equivalent. For the theoretician, examination of the limits is instructive in that two intervals with similar limits are anticipated to have similar coverage and length properties.

In Figure 5 below, we have plotted the limits of some of the intervals in a Binomial and a Poisson case. In the Binomial case, the limits are for the Wald, Jeffreys and the likelihood ratio interval, and $n = 20$. In the Poisson case, the limits are for the Wald, score, Agreti-Coull, Jeffreys and the likelihood ratio interval, and the plot is truncated at $x = 20$.

First, in the Poisson plot, we see a clear clustering; the score and the Agresti-Coull intervals have close limits and these are visibly further out than the limits of the Jeffreys and the likelihood ratio intervals, which are in a separate cluster. The Wald interval, on the other hand, is all by itself, markedly separated from the other four 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 and the Agresti-Coull interval are a bit further out once again, although less so than in the Poisson case. We do not show them on the plot to keep the plot less clumsy.

It would be reasonable to expect that the Jeffreys and the likelihood ratio interval have comparable coverage and length properties. Later in our detailed theoretical calculations in Sections 5 and 7, these visual conjectures would in fact be vindicated.

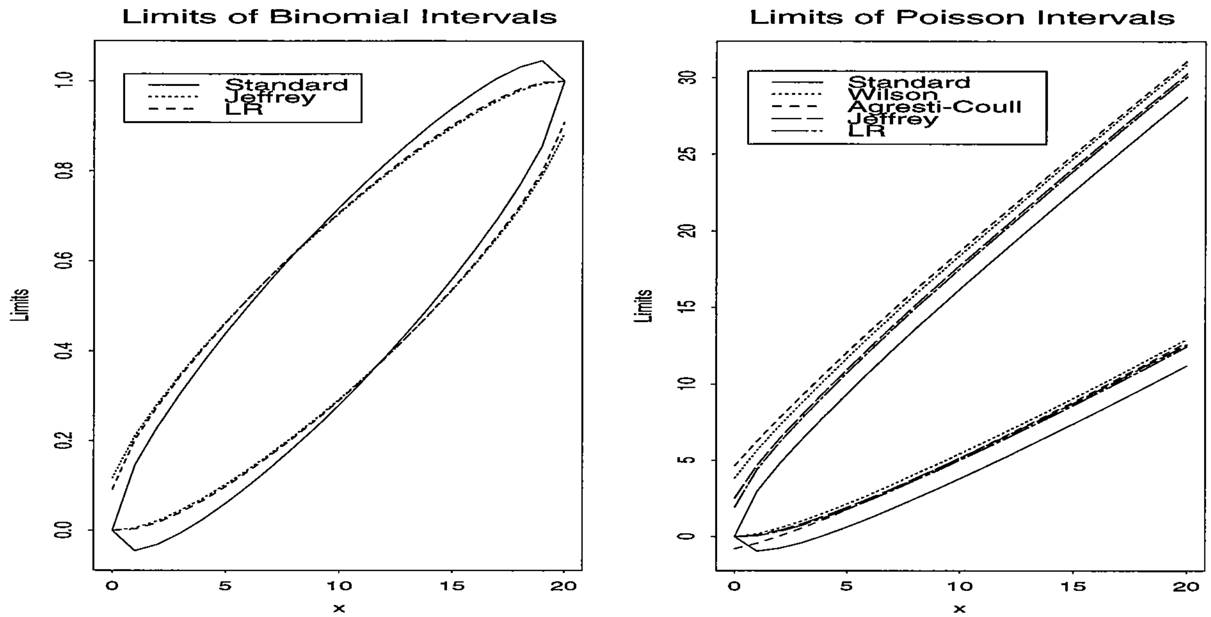


Figure 5: Comparison of the limits of the intervals with $n = 20$ and $\alpha = .05$.

# 5 The Edgeworth Expansions

Define

$$h(x) = x - x_- \tag{19}$$

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

$$g(\mu, z) = g(\mu, z, n) = h(n\mu + n^{1/2}\sigma z) \tag{20}$$

We suppress in (20) and later the dependence of $g$ on $n$ and denote

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

**Theorem 1** *Let $\mu \in \mu$ and $0 < \alpha < 1$. 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 (10) satisfies*

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

*where the quantities $\ell_s$ and $u_s$ are defined in (46) in the appendix.*

**Theorem 2** *Let $\mu \in \mu$ and $0 < \alpha < 1$. Suppose $n\mu - n^{1/2}\sigma\ell_{rs}$ is not an integer. Then the coverage probability of the confidence interval $CI_{rs}$ defined in (11) satisfies*

$$\begin{aligned} P_{rs} &= P_\mu(\mu \in CI_{rs}) = (1-\alpha) + \sigma^{-1}\{g(\mu,\ell_{rs}) - g(\mu,u_{rs})\} \cdot \phi(\kappa)n^{-1/2} \\ &+ \{-\frac{a_2}{18}(2\kappa^5 + 25\kappa^3 + 3\kappa) - \frac{1}{36\sigma^2}(\kappa^5 + 2\kappa^3 + 6\kappa)\} \cdot \phi(\kappa)n^{-1} \\ &+ \{(1+2a_2\mu)(\frac{1}{6}\kappa^2 - \frac{1}{2})Q_{21}(\ell_{rs},u_{rs}) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{22}$$

*where the quantities $\ell_{rs}$ and $u_{rs}$ are defined in (46) in the appendix.*

**Theorem 3** *Let $\mu \in \mu$ and $0 < \alpha < 1$. Suppose $n\mu - n^{1/2}\sigma\kappa$ is not an integer. Then the coverage probability of the confidence interval $CI_R$ defined in (12) satisfies*

$$\begin{aligned} P_R &= P_\mu(\mu \in CI_R) = (1-\alpha) + \sigma^{-1}\{g(\mu,-\kappa) - g(\mu,\kappa)\} \cdot \phi(\kappa)n^{-1/2} \\ &+ \{-\frac{a_2}{18}(2\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{1}{36\sigma^2}(\kappa^5 - 7\kappa^3 + 6\kappa)\} \cdot \phi(\kappa)n^{-1} \\ &+ \{(1+2a_2\mu)(\frac{1}{6}\kappa^2 - \frac{1}{2})Q_{21}(-\kappa,\kappa) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{23}$$

**Theorem 4** *Let $\mu \in \mu$ and $0 < \alpha < 1$. Suppose $n\mu - n^{1/2}\sigma\ell_{AC}$ is not an integer. Then the coverage probability of the confidence interval $CI_{AC}$ defined in (13) satisfies*

$$\begin{aligned} P_{AC} &= P_\mu(\mu \in CI_{AC}) = (1-\alpha) + \sigma^{-1}\{g(\mu,\ell_{AC}) - g(\mu,u_{AC})\} \cdot \phi(\kappa)n^{-1/2} \\ &+ \{-\frac{a_2}{18}(2\kappa^5 - 29\kappa^3 + 3\kappa) - \frac{1}{36\sigma^2}(\kappa^5 - 16\kappa^3 + 6\kappa)\} \cdot \phi(\kappa)n^{-1} \\ &+ \{(1+2a_2\mu)(\frac{1}{6}\kappa^2 - \frac{1}{2})Q_{21}(\ell_{AC},u_{AC}) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{24}$$

*where the quantities $\ell_{AC}$ and $u_{AC}$ are defined in (54) in the appendix.*

Contrary to the all at one stroke derivations for the other intervals in the entire DNEF with a quadratic variance function, for the likelihood ratio interval and the Jeffreys interval a general Edgeworth expansion of the coverage probability seems to be very difficult, if not impossible. So we will be forced to consider the specific negative binomial and Poisson cases separately (The binomial case was derived in Brown, Cai, and DasGupta (1999b)). The expansions themselves, however, can then be written in a general form.

The following theorem gives a unified expression for the two-term Edgeworth expansion of the coverage probability of the likelihood ratio interval.

**Theorem 5** *Denote by $CI_{LR}$ a generic LR interval. Then the coverage probability of $CI_{LR}$ satisfies the general representation*

$$\begin{aligned} P_{LR} &= P_\mu(\mu \in CI_{LR}) = (1-\alpha) + \sigma^{-1}\{g(\mu, \ell_{LR}) - g(\mu, u_{LR})\} \cdot \phi(\kappa) n^{-1/2} \\ &+ \{-\frac{a_2}{6}\kappa - \frac{1}{6\sigma^2}\kappa\} \cdot \phi(\kappa) n^{-1} \\ &+ \{-\frac{1}{2}(1+2a_2\mu)Q_{21}(\ell_{LR}, u_{LR}) + Q_{22}(-\kappa, \kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{25}$$

*where the quantities $\ell_{LR}$ and $u_{LR}$ are defined in (57) 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 6** *Denote by $CI_J$ a generic equal-tailed Jeffreys prior interval as defined in (15) in the binomial case, (17) in the negative binomial case, and (18) in the Poisson case. Then for any fixed $0 < p < 1$ (binomial and negative binomial cases) or any fixed $\lambda > 0$ (Poisson case) and any $0 < \alpha < 1$, the coverage probability of $CI_J$ satisfies*

$$\begin{aligned} P_\mu(\mu \in CI_J) &= (1-\alpha) + [g(\lambda,\ \ell_J) - g(\lambda,\ u_J)]\sigma^{-1}\phi(\kappa)\cdot n^{-1/2} - \frac{1}{12\sigma^2}\kappa\phi(\kappa)n^{-1} \\ &+ [-\frac{1}{3}(1+2a_2\mu)Q_{21}(\ell_J,\ u_J) + Q_{22}(-\kappa,\ \kappa)]\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &+ O(n^{-3/2}) \end{aligned} \tag{26}$$

The Edgeworth expansions for the three specific distributions, binomial, negative binomial, and Poisson, can be obtained easily from Theorems 1 - 6 by plugging in the corresponding $a_2$, $\mu$ and $\sigma$.

- Binomial: $a_2 = -1$, $\mu = p$ and $\sigma^2 = pq$.
- Negative binomial: $a_2 = 1$ , $\mu = p/q$ and $\sigma^2 = p/q^2$.
- Poisson: $a_2 = 0$ and $\mu = \sigma^2 = \lambda$.

# 6 Comparison of Coverage Probability

We will now use the two term Edgeworth expansions to compare the coverage properties of the standard interval $CI_s$ and the various alternative intervals, for all three distributions simultaneously. The encouraging part is that we can reach general conclusions for all the three distributions. The recommendations therefore carry a unifying character. First we will show how the non-oscillatory part of the second order term can be used to explain the deficiency of the standard procedure and the much better performance of competing ones such as 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 the three distributions. The score interval $CI_R$ does significantly better than the standard interval $CI_s$, and especially so near the boundaries. On the other hand, the Agresti-Coull interval $CI_{AC}$ has higher coverage probability than $CI_R$ (and likewise the others), and again, the difference is the most noticeable 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. Although the likelihood ratio interval has, strictly speaking, a slightly larger negative bias than the Jeffreys interval, the difference is entirely academic and is of no practical relevance at all. The Edgeworth expansions thus 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.

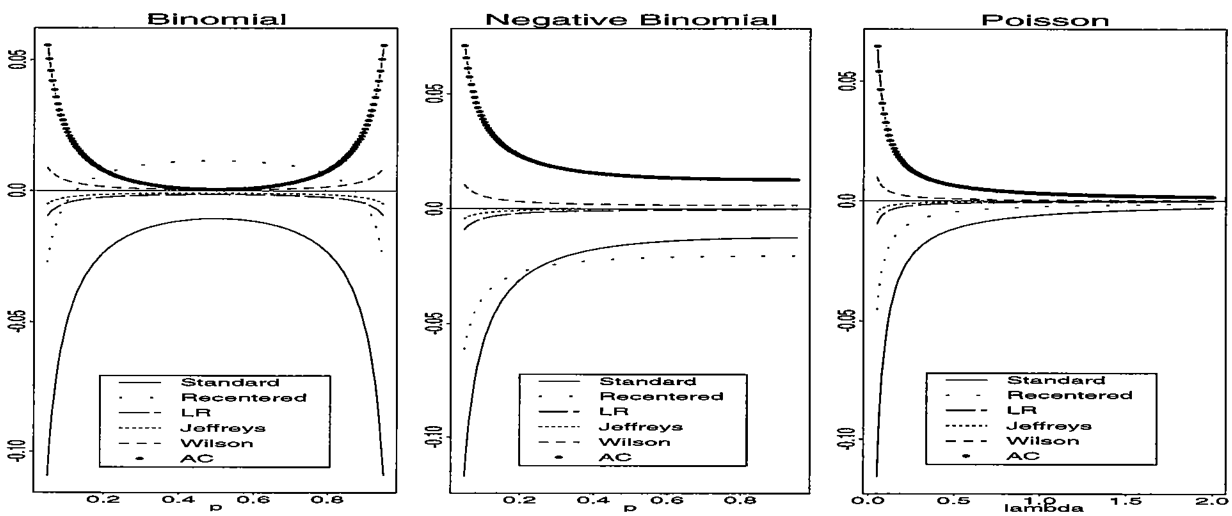


Figure 6: Comparison of the nonoscillating terms with $n = 40$ and $\alpha = .05$.

Directly from equations (21) - (26), we have:

$$P_{rs} - P_s = \{a_2(\frac{1}{3}\kappa^5 - 2\kappa^3) + \frac{1}{12\sigma^2}\kappa^5\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (27)$$

$$P_R - P_s = \{\frac{1}{3}a_2\kappa^5 + \frac{1}{\sigma^2}(\frac{1}{12}\kappa^5 + \frac{1}{4}\kappa^3)\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (28)$$

$$P_{AC} - P_s = \{a_2(\frac{1}{3}\kappa^5 + \kappa^3) + \frac{1}{\sigma^2}(\frac{1}{12}\kappa^5 + \frac{1}{2}\kappa^3)\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (29)$$

$$P_{LR} - P_s = \{\frac{a_2}{18}(8\kappa^5 - 11\kappa^3) + \frac{1}{18\sigma^2}(2\kappa^5 + \kappa^3)\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (30)$$

$$P_J - P_s = \{\frac{a_2}{18}(8\kappa^5 - 11\kappa^3 + 3\kappa) + \frac{1}{36\sigma^2}(4\kappa^5 + 2\kappa^3 + 3\kappa)\}\phi(\kappa) \cdot n^{-1} + O(n^{-3/2}) + \text{oscillations} \quad (31)$$

## 6.1 Further Discussion

By consideration of the coefficients of the $n^{-1}$ terms in (27) - (31), 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 (27) - (31) 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 bias term in $P_R - P_s$, $P_{AC} - P_s$, $P_{AC} - P_R$, and $P_{AC} - P_{rs}$ is positive for all $\kappa$ and all values of the underlying parameter. From this we can expect that each of the score and the Agresti-Coull interval will improve on the standard interval as regards the systematic negative bias phenomenon for all three distributions; we can further expect that of all these different intervals under consideration, the Agresti-Coull interval provides the maximum improvement, and it may even be that the improvement is so much that the Agresti-Coull interval is somewhat conservative. Some of these can in fact be readily seen in Figure 6 above. 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 would also provide relief to the systematic bias problem.

(b). In the particular Poisson case, the coefficient $a_2 = 0$. Comparison of equations (30) and (31) 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 enlightening phenomenon that the two intervals, constructed by using totally different methods, have nearly identical coverage properties, in terms of their Edgeworth expansions.

(c). In Figure 6 above, we see something interesting when we compare the coefficients of the bias term for the recentered interval and the standard interval. In the Poisson case, the curves do not cross, but in the negative binomial case they do. On the other hand, in Brown, Cai, and DasGupta (1999b), it was seen that in the Binomial case the two curves do not cross either.

There is a unifying explanation of all of these. From the expressions in (27), (28), and (29), we see that the coefficient of the $n^{-1}$ term in $P_{rs} - P_s$ is

$$-(\frac{1}{3}\kappa^5 - 2\kappa^3) + \frac{1}{12pq}\kappa^5$$

in the Binomial case, which is positive for all $\kappa$ and all $p$. The coefficient is $\kappa^5/12\lambda$ in the Poisson case and so is again positive for all $\kappa$ and all $\lambda$. So in these two cases, recentering the interval to $\tilde{\mu}$ always reduces the bias problem. However, in the negative binomial case, that coefficient of the $n^{-1}$ term is

$$\kappa^5/3 - 2\kappa^3 + \kappa^5 q^2/(12p),$$

which is positive only if $q^2/p \geq 24\kappa^{-2} - 4$. If $1 - \alpha$ is .99, then $\kappa = 2.575$, and in this case $24\kappa^{-2} - 4 = -.38$, and so vacuously, the required inequality on $q^2/p$ holds for all $p$. However, if $1-\alpha$ is .95, then $\kappa = 1.96$, and in this case $24\kappa^{-2} - 4 = 2.2474$, and the required inequality on $q^2/p$ holds only if $p \leq .25$. This is exactly what we are seeing in Figure 6. The conclusion is that in the negative binomial case recentering to $\tilde{\mu}$ always reduces the bias problem in the nominal 99% case, but not necessarily in the nominal 95% case. In the latter case, recentering helps in reducing the bias only for relatively small $p$.

(d). 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 likelihood ratio interval.

# 7 Expansions for Expected Length

The two term Edgeworth expansions presented in Section 5 compare the coverage property of the various intervals. However, in mutual comparison of different confidence intervals, in addition to coverage, parsimony in length is also an important issue. Therefore, for the intervals we discussed in Section 4 , we will now provide an expansion for their expected lengths correct up to the order $O(n^{-3/2})$. The expansions unfold a lot of common structure.

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 technical calculations. Below we present such a short preview.

## 7.1 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 five intervals in expected length pointwise for every value of the parameter. The intervals are ranked as $CI_s$, $CI_{LR}$, $CI_J$, $CI_R$ and $CI_{AC}$ from the shortest to the longest. Thus, among the four 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 $CI_J$, $CI_{LR}$, $CI_R = CI_s$ and $CI_{AC}$ 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 (1999b) 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, and the Agresti-Coull type interval $CI_{AC}$ is always the longest. 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.

## 7.2 Expansions and Comparisons

The expansion for length differs qualitatively from the two term Edgeworth expansion for coverage probability in that the Edgeworth expansion includes terms involving $n^{-1/2}$ and $n^{-1}$, whereas the expansion for length includes terms $n^{-1/2}$ and $n^{-3/2}$. The coefficient of the $n^{-1/2}$ term is the same for all the intervals, but the coefficient for the $n^{-3/2}$ term differs. So, naturally, the coefficients of the $n^{-3/2}$ term will be used as a basis for comparison of their length.

**Theorem 7** *Let $CI$ be a generic notation for any of the five intervals, $CI_s$, $CI_R$, $CI_{AC}$, $CI_{LR}$ and $CI_J$, for estimating the mean $\mu$, as defined in equations (10) – (14). Then,*

$$L(n,\ \mu) = E(\textit{length of } CI) = 2\kappa(\mu + a_2\mu^2)^{1/2}n^{-1/2}\left(1 - \frac{\delta(\kappa,\ \mu)}{72n(\mu + a_2\mu^2)}\right) + O(n^{-2}) \quad (32)$$

*where*

$$\begin{aligned} \delta(\kappa,\ \mu) &= 9 \textit{ for } CI_s; & (33)\\ &= 9(1-\kappa^2) - 72\kappa^2 a_2(\mu + a_2\mu^2) \textit{ for } CI_R; & (34)\\ &= 9(1-2\kappa^2) - 108\kappa^2 a_2(\mu + a_2\mu^2) \textit{ for } CI_{AC}; & (35)\\ &= 9 - 2\kappa^2 - 26\kappa^2 a_2(\mu + a_2\mu^2) \textit{ for } CI_{LR}; & (36)\\ &= 9 - 2(\kappa^2+2) - 2(13\kappa^2+17)a_2(\mu + a_2\mu^2) \textit{ for } CI_J; & (37) \end{aligned}$$

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

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

**Remark:** Hence, up to the error $n^{-2}$, very interestingly, for every $\lambda > 0$, the ranking of the intervals is $CI_s$, $CI_{LR}$, $CI_J$, $CI_R$ and $CI_{AC}$ from the shortest to the longest, as long as $\kappa > 2/\sqrt{7} = .76$. In practice, $\kappa$ will certainly be larger than .76 and so, we have the quite remarkable fact that pointwise in $\lambda$, 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 beats the Jeffreys interval $CI_J$ at every $\lambda$.

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 $CI_s$, $CI_{LR}$, $CI_J$, $CI_R$ and $CI_{AC}$ admit the expansions*

$$
\begin{aligned}
E(L_s) &= 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2}{72np}] + O(n^{-2}) \\
E(L_{LR}) &= 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 2\kappa^2(1+11p+p^2)}{72np}] + O(n^{-2}) \\
E(L_J) &= 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 2\kappa^2(1+11p+p^2) - 2(2+13p+2p^2)}{72np}] + O(n^{-2}) \\
E(L_R) &= 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 9\kappa^2(1+6p+p^2)}{72np}] + O(n^{-2}) \\
E(L_{AC}) &= 2\kappa p^{1/2}q^{-1}n^{-1/2}[1 - \frac{9q^2 - 18\kappa^2(1+4p+p^2)}{72np}] + O(n^{-2})
\end{aligned}
$$

**Remark:** From the above expressions in Corollary 2, one can verify that up to an error of $O(n^{-2})$, pointwise at every $p > 0$, the ranking of the intervals is $CI_s$, $CI_{LR}$, $CI_J$, $CI_R$ and $CI_{AC}$, from the shortest to the longest, provided $\kappa > \sqrt{17/23} = .86$. Note that this ranking exactly coincides with the ranking we previously derived in the Poisson case. Again we see the quite impressive performance of the likelihood ratio interval.

Finally, now the Binomial case is addressed. Unlike the Poisson and the Negative Binomial cases, a uniform ranking in length pointwise for all $p$ is not valid in the Binomial case. However, if the expansions are integrated over $p$, then a clear ranking still emerges.

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

$$
\begin{aligned}
\int_0^1 E(L_J)dp &= \frac{\kappa\pi}{4}n^{-1/2} - (\frac{37}{36} + \frac{5\kappa^2}{36})\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}) \\
\int_0^1 E(L_{LR})dp &= \frac{\kappa\pi}{4}n^{-1/2} - (1 + \frac{5\kappa^2}{36})\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}) \\
\int_0^1 E(L_R)dp &= \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}); \\
\int_0^1 E(L_s)dp &= \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}); \\
\int_0^1 E(L_{AC})dp &= \frac{\kappa\pi}{4}n^{-1/2} + (\frac{\kappa^2}{2} - 1)\frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2});
\end{aligned}
$$

**Remark:** Hence, up to the error $n^{-2}$, the ranking of the intervals is $CI_J$, $CI_{LR}$, $CI_R = CI_s$ and $CI_{AC}$ 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.

# 8 Summary and Conclusions

The examples and theoretical results we have presented demonstrate that the popular Wald interval is uniformly poor in a number of important lattice distributions, and better alternatives are urgently needed. Our comprehensive comparisons show that fortunately, in each of these cases, a number of alternative intervals provide significant improve ments 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 and the Agresti-Coull type interval are easier to present and compute in an informal environment. But in the absence of an overriding need for easy computation, the likelihood ratio and the Jeffreys interval can be resolutely recommended.

# 9 Proofs

All of the distributions in the discrete natural exponential families under consideration are lattice distributions with the maximal span of one. The two-term Edgeworth expansion for a lattice distribution with the maximal span of one is

$$\begin{aligned} F_n(z) &= \Phi(z) + p_1(z)\phi(z)n^{-1/2} + \sigma^{-1}[g(\mu,z) + \frac{1}{2}]\phi(z)n^{-1/2} + p_2(z)\phi(z)n^{-1} \\ &\quad + \{[g(\mu,z) + \frac{1}{2}]\sigma p_3(z) - [\frac{1}{2}g^2(\mu,z) - \frac{1}{2}g(\mu,z) + \frac{1}{12}]\}\sigma^{-2}z\phi(z)n^{-1} \\ &\quad + O(n^{-3/2}) \end{aligned} \tag{38}$$

where

$$\begin{aligned} p_1(z) &= \frac{1}{6}\beta_3(1-z^2) \\ p_2(z) &= -\frac{1}{24}\beta_4(z^3-3z) - \frac{1}{72}\beta_3^2(z^5 - 10z^3 + 15z) \\ p_3(z) &= \frac{1}{6}\beta_3(z^2-3) \end{aligned}$$

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

$$\begin{aligned} F_n(z) &= \Phi(z_0) + \tilde{p}_1(z)\phi(z_0)n^{-1/2} + \sigma^{-1}[g(\mu,z) + \frac{1}{2}]\phi(z)n^{-1/2} + \tilde{p}_2(z)\phi(z_0)n^{-1} \\ &\quad + \{\sigma[g(\mu,z) + \frac{1}{2}]\tilde{p}_3(z_0) - [\frac{1}{2}g^2(\mu,z) - \frac{1}{2}g(\mu,z) + \frac{1}{12}]\}\sigma^{-2}z_0\phi(z_0)n^{-1} \\ &\quad + O(n^{-3/2}) \end{aligned} \tag{39}$$

where

$$\tilde{p}_1(z) = c_1 + \frac{1}{6}\beta_3(1 - z_0^2) \tag{40}$$

$$\tilde{p}_2(z) = c_2 - \frac{1}{2}z_0c_1^2 + \frac{1}{6}(z_0^3 - 3z_0)\beta_3c_1 - \frac{1}{24}\beta_4(z_0^3 - 3z_0) - \frac{1}{72}\beta_3^2(z_0^5 - 10z_0^3 + 15z_0) \tag{41}$$

$$\tilde{p}_3(z) = -c_1 + \frac{1}{6}\beta_3(z_0^2 - 3). \tag{42}$$

*Proof:* The expansion (38) 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 $z_0$:

$$\Phi(z) = \Phi(z_0) + c_1\phi(z_0)n^{-1/2} + (c_2 - \frac{1}{2}z_0c_1^2)\phi(z_0)n^{-1} + O(n^{-3/2}) \tag{43}$$

$$\phi(z) = \phi(z_0) - z_0c_1\phi(z_0)n^{-1/2} + O(n^{-1}) \tag{44}$$

$$z^2 = z_0^2 + 2z_0c_1n^{-1/2} \tag{45}$$

Now plugging (43) - (45) into (38) and noting that $g^2(\mu, z) - g(\mu, z)$ admits a one-term Taylor expansion, we obtain (39). ∎

**Remark:** In (39), the second $O(n^{-1/2})$ and the second $O(n^{-1})$ terms are oscillation terms.

*Proof of Theorems 1 and 2*: We consider the standard interval and the recentered interval together. Let $\tilde{\mu} = (1 + \delta_1)\hat{\mu} + \delta_2$ be the center of the confidence interval for $\mu$. In the case of the standard interval, we set $\delta_1 = \delta_2 = 0$ and in the recentered interval case we set $\delta_1 = \kappa^2a_2/(n - \kappa^2a_2)$ and $\delta_2 = \frac{1}{2}k^2/(n - \kappa^2a_2)$. Denote

$$\begin{aligned} A &= n(1+\delta_1)^2 - \kappa^2a_2 \\ B &= 2n(1+\delta_1)(\mu - \delta_2) + \kappa^2 \\ C &= n(\mu - \delta_2)^2 \end{aligned}$$

By solving a quadratic equation, we have

$$\alpha(\mu,\ n) = P_\mu(\mu \in CI_*) = P(\ell_* \le n^{1/2}(\hat{\mu} - \mu)/\sigma \le u_*)$$

where

$$(\ell_*,\ u_*) = \left(\frac{B \pm \sqrt{B^2 - 4AC}}{2A} - \mu\right)\sigma^{-1}n^{1/2}. \tag{46}$$

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

$$\begin{aligned} (\ell_*,\ u_*) &= \{(\frac{1}{2} + a_2\mu)\kappa^2 - (\mu\delta_1 + \delta_2)n\}\sigma^{-1}n^{-1/2} \\ &\quad \pm (\kappa + \frac{1}{2}\kappa\Delta\sigma^{-2} + a_2\kappa^3n^{-1} - 2\delta_1\kappa) + O(n^{-3/2}) \end{aligned}$$

where $\Delta = (\delta_1 - 2a_2\delta_2)\mu - \delta_2 + \kappa^2/(4n)$.

- For the standard interval, letting $\delta_1 = \delta_2 = 0$, we have

$$(\ell_s,\ u_s) = (\frac{1}{2} + a_2\mu)\kappa^2\sigma^{-1}n^{-1/2} \pm \{\kappa + (a_2 + \frac{1}{8}\sigma^{-2})\kappa^3 n^{-1}\} + O(n^{-3/2}) \qquad (47)$$

Now, $P_\mu(\mu \in CI_s) = F_n(u_s) - F_n(\ell_s)$. Then (39) yields

$$\begin{aligned} P_\mu(\mu \in CI_s) &= (1-\alpha) + \sigma^{-1}\{g(\mu,\ell_s) - g(\mu,u_s)\} \cdot \phi(\kappa)n^{-1/2} + 2\tilde{p}_2(u_s)\phi(\kappa)n^{-1} \\ &\quad + \{\sigma\tilde{p}_3(u_s)Q_{21}(\ell_s,u_s) + Q_{22}(-\kappa,\kappa)\}\sigma^{-2}\kappa\phi(\kappa)n^{-1} \\ &\quad + O(n^{-3/2}) \qquad (48) \end{aligned}$$

where $\tilde{p}_2(\cdot)$ and $\tilde{p}_2(\cdot)$ are given as in (41) and (42) with $z_0 = \kappa$, $c_1 = (1/2 + a_2\mu)\kappa^2\sigma^{-1}$ and $c_2 = (a_2 + 1/(8\sigma^2))\kappa^3$. Similar as in (3), we have

$$K_4 = \psi^{(4)}(\xi) = \frac{dK_3}{d\mu} \cdot \frac{d\mu}{d\xi} = V(\mu) + 6a_2(V(\mu))^2 = \sigma^2 + 6a_2\sigma^4. \qquad (49)$$

Hence,

$$\beta_3 = K_3/\sigma^3 = (1 + 2a_2\mu)\sigma^{-1} \quad \text{and} \quad \beta_4 = K_4/\sigma^4 = \sigma^{-2} + 6a_2 \qquad (50)$$

Using (50), after some algebra, we have

$$2\tilde{p}_2(u_s) = -\frac{a_2}{18}(8\kappa^5 - 11\kappa^3 + 3\kappa) - \frac{1}{18\sigma^2}(2\kappa^5 + \kappa^3 + 3\kappa) \qquad (51)$$

and

$$\sigma\tilde{p}_3(u_s) = -(1 + 2a_2\mu)(\frac{1}{3}\kappa^2 + \frac{1}{2}) \qquad (52)$$

We obtain (21) by putting (51) and (52) in (48).

- For the recentered interval, by letting $\delta_1 = \kappa^2 a_2/(n - \kappa^2 a_2) = \kappa^2 a_2 n^{-1} + O(n^{-2})$ and $\delta_2 = \frac{1}{2}k^2/(n - \kappa^2 a_2) = \frac{1}{2}\kappa^2 n^{-1} + O(n^{-2})$, we have

$$(\ell_{rs},\ u_{rs}) = \pm\{\kappa - (a_2 + \frac{1}{8}\sigma^{-2})\kappa^3 n^{-1}\} + O(n^{-3/2}) \qquad (53)$$

Similarly, $P_\mu(\mu \in CI_{rs}) = F_n(u_{rs}) - F_n(\ell_{rs})$, and (22) follows from (39) with the same derivation as that for the standard interval. We omit the detailed algebra here.

*Proof of Theorem 3*: The Edgeworth expansion for $P_\mu(\mu \in CI_R)$ is slightly simpler because

$$P_R = P_\mu(\mu \in CI_R) = P(-\kappa \le n^{1/2}(\hat{\mu} - \mu)/\sigma \le \kappa)$$

And now (23) follows from (39). ∎

*Proof of Theorem 4*: The Edgeworth expansion for $P_\mu(\mu \in CI_{AC})$ can be derived in a similar way. Denote

$$\begin{aligned} A &= (n - 2\kappa^2 a_2)n^2 \\ B &= 2\mu(n - \kappa^2 a_2)^2 n + \kappa^4 a_2 n \\ C &= \mu^2(n - \kappa^2 a_2)^3 - \kappa^2\mu(n - \kappa^2 a_2)^2 - \frac{1}{4}\kappa^4 n \end{aligned}$$

We have, after some straightforward algebra,

$$P_\mu(\mu \in CI_{AC}) = P(\ell_{AC} \le n^{1/2}(\hat{\mu} - \mu)/\sigma \le u_{AC})$$

where

$$(\ell_{AC},\ u_{AC}) = \left(\frac{B \pm \sqrt{B^2 - 4AC}}{2A} - \mu\right)\sigma^{-1}n^{1/2} \tag{54}$$

The + sign goes with $u_{AC}$ and the − sign with $\ell_{AC}$. Expanding $\ell_{AC}$ and $u_{AC}$, one has

$$(\ell_{AC},\ u_{AC}) = \pm\{\kappa + (\frac{1}{2}a_2 + \frac{1}{8}\sigma^{-2})\kappa^3 n^{-1}\} + O(n^{-3/2}) \tag{55}$$

with the + sign going with $u_{AC}$ and the − sign with $\ell_{AC}$. Now $P_{AC} = F_n(u_{AC}) - F_n(\ell_{AC})$, and (24) follows from (39). ∎

## Expansion for the Likelihood Ratio Interval

We now prove Theorem 5. The proofs of the three cases are similar. We will give the proof of the negative binomial case in detail, the proof for the binomial and Poisson two cases are slightly simpler and will be omitted here.

Let $X_1, \cdots, X_n \overset{iid}{\sim} \mathrm{NBin}(1, p)$. Then the MLE for $p$ is $\hat{p} = \bar{X}/(1+\bar{X})$ and

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

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

$$p(1 + (pn)^{-1/2}z)\log(1 + (pn)^{-1/2}z) - (1 + p^{1/2}n^{-1/2}z)\log(1 + p^{1/2}n^{-1/2}z) - \frac{q\kappa^2}{2n} \le 0. \tag{56}$$

Denote the RHS of (56) by $d(z)$. It is easy to verify that $d(\cdot)$ is a convex function and so has at most two roots. Denote by $\ell_{LR}$ and $u_{LR}$ the roots of the equation $d(z) = 0$. So

$$d(\ell_{LR}) = d(u_{LR}) = 0 \tag{57}$$

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

$$b(t) = t + \frac{1}{2}t^2 - \frac{1}{6}t^3 + \frac{1}{12}t^4 + O(t^5) \tag{58}$$

Now applying (58) to (56), we have, after some simplification, that $d(z) = 0$ is equivalent to

$$z^2 - \frac{1}{3}(1+p)p^{-1/2}n^{-1/2}z^3 + \frac{1}{6}(1+p+p^2)p^{-1}n^{-1}z^4 - \kappa^2 = O(n^{-3/2}) \tag{59}$$

Now let $z = \pm\kappa + b_1 n^{-1/2} + b_2 n^{-1}$. Plugging into (59) and solving for $b_1$ and $b_2$, we have

$$\begin{aligned} b_1 &= \frac{1}{6}(1+p)p^{-1/2}\kappa^2 = \frac{1}{6}(1+2\mu)\sigma^{-1}\kappa^2 \\ b_2 &= \mp\frac{1}{72}(1-4p+p^2)p^{-1}\kappa^3 = \mp\frac{1}{72}(\sigma^{-2}-2)\kappa^3. \end{aligned}$$

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

$$(\ell_{LR},\ u_{LR}) = \frac{1}{6}(1+p)p^{-1/2}\kappa^2 n^{-1/2} \pm \{\kappa - \frac{1}{72}(1-4p+p^2)p^{-1}\kappa^3 n^{-1}\} + O(n^{-3/2}). \quad (60)$$

The $+$ sign goes with $u_{LR}$ and the $-$ sign with $\ell_{LR}$. 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

$$(\ell_{LR},\ u_{LR}) = \frac{1}{6}(1+2a_2\mu)\sigma^{-1}\kappa^2 n^{-1/2} \pm \{\kappa - \frac{1}{72}(\sigma^{-2} - 2a_2)\kappa^3 n^{-1}\} + O(n^{-3/2}). \quad (61)$$

Now $P_{LR} = F_n(u_{LR}) - F_n(\ell_{LR})$, and the Edgeworth expansion (25) follows from (39). ∎

## Expansion for Jeffreys Prior Intervals

We now prove Theorem 6. We will use the direct expansion method to derive (26) (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 DNEF with a quadratic variance function, for the Jeffreys interval a general Edgeworth expansion of the coverage probability seems to be basically impossible. So we will be forced to consider the specific negative binomial and Poisson cases separately (The binomial case was derived in Brown, Cai, and DasGupta (1999b)). Theses specific cases are already very complex, as shall be 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, *et al.* (1995)). So there exist unique $X_l = \rho_1(1-\alpha/2, p)$ and $X_u = \rho_2(\alpha/2, p)$ satisfying

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

Therefore

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

with

$$\begin{aligned} \ell_J &= [\rho_1(1-\alpha/2, p) - np/q]/(np/q^2)^{1/2} \\ u_J &= [\rho_2(\alpha/2, p) - np/q]/(np/q^2)^{1/2} \end{aligned} \tag{62}$$

The quantities $\ell_J$ and $u_J$ are defined implicitly in (62) through $\rho_1$ and $\rho_2$. The proof of (26) for the negative binomial case requires an asymptotic expansion for both $\ell_J$ and $u_J$. We do this below.

**Step 1.** Denote

$$\begin{aligned} x_1 &= x - 1/2, \;\; n_1 = n + x - 3/2, \;\; p_1 = x_1/n_1, \;\; q_1 = 1 - p_1 \\ s &= (p_1 q_1)^{1/2} n_1^{-1/2}, \;\; \gamma = \frac{\Gamma(n_1+2)}{\Gamma(x_1+1)\Gamma(n_1-x_1+1)} \end{aligned}$$

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

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

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

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

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

$$L(y) = -\frac{1}{2}\log(2\pi) + c_0 n_1^{-1} - \frac{1}{2}y^2 + c_1 n_1^{-1/2} y^3 + c_2 n_1^{-1} y^4 + O(n_1^{-3/2}) \tag{63}$$

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

$$\psi(y) = e^{L(y)} = \phi(y)[1 + c_1 n_1^{-1/2} y^3 + (c_0 + c_2 y^4 + \frac{1}{2}c_1^2 y^6) n_1^{-1}] + O(n_1^{-3/2}) \tag{64}$$

**Step 3.** Integrating both sides of (64) from $-\infty$ to $z$, we have

$$H(z) \equiv \int_{-\infty}^{z} \psi(y)dy = \Phi(z) - v_1(z)\phi(z)n_1^{-1/2} + v_2(z)\phi(z)n_1^{-1} + O(n_1^{-3/2}) \qquad (65)$$

where $v_1(z) = -c_1(z^2+2)$ and $v_2(z) = -[\frac{1}{2}c_1^2(z^5+5z^3+15z) + c_2(z^3+3z)]$. (Because the $O(n_1^{-3/2})$ term in (64) 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 (65) and solving for $\tau_1$ and $\tau_2$, after some algebra, we get

$$\begin{aligned}\tau_1 &= \frac{1}{3}(1-2p_1)(z_\alpha^2+2)(p_1q_1)^{-1/2}\\ \tau_2 &= (\frac{1}{36}z_\alpha^3 + \frac{11}{36}z_\alpha)(p_1q_1)^{-1} - (\frac{13}{36}z_\alpha^3 + \frac{71}{36}z_\alpha)\end{aligned}$$

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

$$\begin{aligned}(p_l,\, p_u) &= p_1 + \frac{1}{3}(1-2p_1)(\kappa^2+2)n_1^{-1} \pm \Big\{\kappa(p_1q_1)^{1/2}n_1^{-1/2}\\ &\quad + \kappa(p_1q_1)^{1/2}n_1^{-3/2}[(\frac{1}{36}\kappa^2 + \frac{11}{36})(p_1q_1)^{-1} - (\frac{13}{36}\kappa^2 + \frac{71}{36})]\Big\} + O(n_1^{-2}) \qquad (66)\end{aligned}$$

Let

$$w_1(\mu) = (\frac{2}{3}\kappa^2 + \frac{1}{3})(\mu + \frac{1}{2}) \qquad (67)$$

$$w_2(\mu) = \left\{(\frac{13}{36}\kappa^3 + \frac{17}{36}\kappa)(\mu+\mu^2) + (\frac{1}{36}\kappa^3 + \frac{1}{18}\kappa)\right\}(\mu+\mu^2)^{-1/2}. \qquad (68)$$

Rewriting the approximate limits (66) in terms of $\mu = p/(1-p)$, $n$, $\hat{\mu} = x/n$, after some algebra, one has

$$(\mu_l,\, \mu_u) = (\hat{\mu} + w_1(\hat{\mu})n^{-1}) \pm \{\kappa(\hat{\mu}+\hat{\mu}^2)^{1/2}n^{-1/2} + w_2(\hat{\mu})n^{-3/2}\} + O(n^{-2}) \qquad (69)$$

with the $+$ sign going with $\mu_u$ and the $-$ sign with $\mu_l$..

**Step 5.** Now we expand the coverage probability by using (39). In order to use (39) we invert the inequalities $\mu_l \le \mu \le \mu_u$ into the form of

$$\ell_J \le n^{1/2}(\hat{\mu} - \mu)/(\mu+\mu^2)^{1/2} \le u_J.$$

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 continuous first derivative. Then the roots $x_*$ of the equations*

$$x \pm \kappa[x(1+x)]^{1/2}n^{-1/2} + w_1(x)n^{-1} + w_2(x)n^{-3/2} - \mu = 0 \tag{70}$$

*can be expressed as*

$$\begin{aligned} x_* &= \mu \mp (\mu+\mu^2)^{1/2}\kappa n^{-1/2} + [(\frac{1}{2}+\mu)\kappa^2 - w_1(\mu \mp (\mu+\mu^2)^{1/2}\kappa n^{-1/2})]n^{-1} - w_2(\mu)n^{-3/2} \\ &\mp\{[\frac{1}{8}(\mu+\mu^2)^{-1/2} + (\mu+\mu^2)^{1/2}]\kappa^3 - (\frac{1}{2}+\mu)(\mu+\mu^2)^{-1/2}w_1(\mu)\kappa\}n^{-3/2} + O(n^{-2}) \end{aligned} \tag{71}$$

*All the $-$ $(+)$ signs in $\mp$ in (71) go with the $+$ $(-)$ sign in $\pm$ in (70).*

Applying Lemma 1 to (69), 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

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

with all $+$ signs go with $u_J$ and all $-$ signs with $\ell_J$. Now the expansion (26) for the negative binomial case follows from (39).

<u>Poisson case</u>: The posterior distribution of $\lambda$ 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

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

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

$$\ell_J = [\rho_1(1-\alpha/2,\lambda) - n\lambda]/(n\lambda)^{1/2} \text{ and } u_J = [\rho_2(\alpha/2,\lambda) - n\lambda]/(n\lambda)^{1/2}. \qquad (73)$$

The quantities $\ell_J$ and $u_J$ are defined implicitly in (73) through $\rho_1$ and $\rho_2$. Again, the proof of (26) for the Poisson case requires an asymptotic expansion for both $\ell_J$ and $u_J$. We do this below.

**Step 1.** Denote

$$x_1 \;=\; x - 1/2,\;\; \lambda_1 = x_1/n,\;\; s = x_1^{1/2}n^{-1} = \lambda_1^{1/2}n^{-1/2},\;\; \gamma = \frac{n^{x+1/2}}{\Gamma(x_1+1)}$$

Here $\lambda_1$ 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

$$L(y) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}y^2 + \frac{1}{3}\lambda_1^{-1/2}y^3n^{-1/2} - (\frac{1}{12} + \frac{1}{4}y^4)\lambda_1^{-1}n^{-1} + O(n^{-3/2}). \qquad (74)$$

Therefore

$$\psi(y) = e^{L(y)} = \phi(y)[1 + \frac{1}{3}\lambda_1^{-1/2}y^3n^{-1/2} + (-\frac{1}{12} - \frac{1}{4}y^4 + \frac{1}{18}y^6)\lambda_1^{-1}n^{-1}] + O(n^{-3/2}). \quad (75)$$

**Step 3.** Integrating both sides of (75) from $-\infty$ to $z$, we have

$$H(z) \equiv \int_{-\infty}^{z} \psi(y)dy = \Phi(z) + v_1(z)\phi(z)n^{-1/2} + v_2(z)\phi(z)n^{-1} + O(n^{-3/2}) \qquad (76)$$

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 (76) and solving for $\tau_1$ and $\tau_2$, after some algebra, we get

$$\tau_1 = \frac{1}{3}(z_\alpha^2 + 2)\lambda_1^{-1/2} \text{ and } \tau_2 = (\frac{1}{36}z_\alpha^3 + \frac{11}{36}z_\alpha)\lambda_1^{-1}.$$

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

$$(\lambda_l,\ \lambda_u) = \lambda_1 + \frac{1}{3}(\kappa^2+2)n^{-1} \pm \{\kappa\lambda_1^{1/2}n^{-1/2} + (\frac{1}{36}\kappa^3 + \frac{11}{36}\kappa)\lambda_1^{-1/2}n^{-3/2}\} + O(n^{-2}) \tag{77}$$

Rewriting the approximate limits (77) in terms of $\hat{\lambda} = x/n$, one has, after some algebra,

$$(\lambda_l,\ \lambda_u) = \hat{\lambda} + (\frac{1}{3}\kappa^2 + \frac{1}{6})n^{-1} \pm \{\kappa\hat{\lambda}^{1/2}n^{-1/2} + (\frac{1}{36}\kappa^3 + \frac{1}{18}\kappa)\hat{\lambda}^{-1/2}n^{-3/2}\} + O(n^{-2}) \tag{78}$$

**Step 5.** Now we expand the coverage probability by using (39). In order to use (39) we invert the inequalities $\lambda_l \le \lambda \le \lambda_u$ into the form of

$$\ell_J \le n^{1/2}(\hat{\lambda}-\lambda)/\lambda^{1/2} \le u_J.$$

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*

$$x \pm x^{1/2}\kappa n^{-1/2} + (\frac{1}{3}\kappa^2 + \frac{1}{6})n^{-1} + w(x)n^{-3/2} - \lambda = 0 \tag{79}$$

*can be expressed as*

$$x_* = \lambda + \frac{1}{6}(\kappa^2-1)n^{-1} - w(\lambda)n^{-3/2} \mp \{\lambda^{1/2}\kappa n^{-1/2} - (\frac{1}{24}\kappa^3 + \frac{1}{12}\kappa)\lambda^{-1/2}n^{-3/2}\} + O(n^{-2}) \tag{80}$$

*The $-$ $(+)$ sign in $\mp$ in (80) goes with the $+$ $(-)$ sign in $\pm$ in (79).*

Applying Lemma 2 to (78), we obtain

$$P(\lambda \in CI_J) = P(\ell_J \le n^{1/2}(\hat{\lambda}-\lambda)/\lambda^{1/2} \le u_J)$$

with

$$(\ell_J, u_J) = \frac{1}{6}(\kappa^2-1)\lambda^{-1/2}n^{-1/2} \pm \{\kappa - (\frac{1}{72}\kappa^3 + \frac{1}{36}\kappa)\lambda^{-1}n^{-1}\} + O(n^{-3/2}) \tag{81}$$

The expansion (26) for the Poisson case now follows from (39). ∎

## Expansions for Expected Length

We now prove Theorem 7. The derivation of the expected length expansions in equations (32) - (37) is algebraically intense. We will report the main steps below and skip the many intermediate algebraic simplifications. We denote below $Z_n = \frac{\bar{X}-\mu}{\sqrt{\mu+a_2\mu^2}}$.

*The interval $CI_s$.* The length of the Wald interval $CI_s$ is

$$\begin{aligned} L_s &= 2\kappa(\bar{X}+a_2\bar{X}^2)^{1/2}n^{-1/2} \\ &= 2\kappa(\mu+a_2\mu^2)^{1/2}n^{-1/2}(1+Z_n(1+2a_2\mu)(\mu+a_2\mu^2)^{-1/2}n^{-1/2}+a_2Z_n^2n^{-1})^{-1/2}, \end{aligned}$$

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

$$\begin{aligned} L_s &= 2\kappa(\mu + a_2\mu^2)^{1/2}n^{-1/2}\Big\{1 + \frac{1}{2}Z_n(1+2a_2\mu)(\mu+a_2\mu^2)^{-1/2}n^{-1/2} \\ &+ [\frac{a_2}{2}Z_n^2 - \frac{1}{8}Z_n^2(1+2a_2\mu)^2(\mu+a_2\mu^2)^{-1}]n^{-1} + R_s(Z_n)\Big\}, \end{aligned} \tag{82}$$

where $E(|R_s(Z_n)|) = O(n^{-3/2})$. Note that this statement on $E(|R_s(Z_n)|)$ has to be verified separately for each of the Binomial, Poisson, and Negative Binomial case, by using the specific central moment formulas in the three cases. The first six central moments are used; see Brown, Cai and DasGupta (19999b) for further details of this step. Now from (82), on using $E(Z_n) = 0$ and $E(Z_n^2) = 1$, we have

$$E(L_s) = 2\kappa(\mu + a_2\mu^2)^{1/2}n^{-1/2}[1 + \frac{1}{2}a_2n^{-1} - -\frac{1}{8}(1+2a_2\mu)^2(\mu+a_2\mu^2)^{-1}n^{-1}] + O(n^{-2}),$$

which simplifies to expression (33) on some algebra.

*The interval $CI_R$.* For the Rao score interval $CI_R$, the length is

$$\begin{aligned} L_R &= 2\kappa(1 - a_2\kappa^2n^{-1})^{-1}(\bar{X} + a_2\bar{X}^2 + \frac{1}{4}\kappa^2n^{-1})^{1/2}n^{-1/2} \\ &= 2\kappa(\mu+a_2\mu^2)^{1/2}(1-a_2\kappa^2n^{-1})^{-1}n^{-1/2}\Big\{1 + Z_n(1+2a_2\mu)(\mu+a_2\mu^2)^{-1/2}n^{-1/2} \\ &+ (a_2Z_n^2 + \frac{1}{4}\kappa^2(\mu+a_2\mu^2)^{-1})n^{-1}\Big\}^{-1/2}, \end{aligned}$$

again on some intermediate algebra by using the definition of $Z_n$. So

$$\begin{aligned} L_R &= 2\kappa(\mu+a_2\mu^2)^{1/2}(1-a_2\kappa^2n^{-1})^{-1}n^{-1/2}\Big\{1 + \frac{1}{2}Z_n(1+2a_2\mu)(\mu+a_2\mu^2)^{-1/2}n^{-1/2} \\ &+ [\frac{1}{2}a_2 + \frac{1}{8}(\mu+a_2\mu^2)^{-1} - \frac{1}{8}(1+2a_2\mu)^2(\mu+a_2\mu^2)^{-1}]Z_n^2n^{-1} + R_R(Z_n)\Big\}, \end{aligned}$$

where exactly as in (82) above, $E(|R_R(Z_n)|) = O(n^{-3/2})$. Thus

$$\begin{aligned} E(L_R) &= 2\kappa(\mu+a_2\mu^2)^{1/2}(1-a_2\kappa^2n^{-1})^{-1}n^{-1/2}\Big\{1 + \frac{a_2}{2}n^{-1} + \frac{1}{8}(\mu+a_2\mu^2)^{-1}n^{-1} \\ &- \frac{1}{8}(1+2a_2\mu)^2(\mu+a_2\mu^2)^{-1}n^{-1}\Big\} + O(n^{-2}) \\ &= 2\kappa(\mu+a_2\mu^2)^{1/2}n^{-1/2}\{1 + (a_2\kappa^2 + \frac{1}{8}(\kappa^2-1)(\mu+a_2\mu^2)^{-1})n^{-1}\} + O(n^{-2}) \end{aligned}$$

which simplifies to expression (34).

*The interval $CI_{AC}$.* For the interval $CI_{AC}$, the length is $L_{AC} = 2\kappa(\tilde{\mu} + a_2\tilde{\mu}^2)^{1/2}\tilde{n}^{-1/2}$, where $\tilde{n} = n - a_2\kappa^2$ and $\tilde{\mu} = (X + \kappa^2/2)/\tilde{n}$. By using the definition of $Z_n$,

$$\tilde{\mu} = (\mu + Z_n(\mu+a_2\mu^2)^{1/2}n^{-1/2} + \frac{1}{2}\kappa^2n^{-1})(1-a_2\kappa^2n^{-1})^{-1},$$

and hence, after some algebra,

$$\begin{aligned} L_{AC} &= 2\kappa(\mu + a_2\mu^2)^{1/2}(n - a_2\kappa^2)^{-1/2}\Big\{1 + [a_2 + \frac{1}{4}(\mu + a_2\mu^2)^{-1}]\kappa^2 n^{-1} \\ &+ [\frac{1}{2}a_2 - \frac{1}{8}(1 + 2a_2\mu)^2(\mu + a_2\mu^2)^{-1}]Z_n^2 n^{-1} + R_{AC}(Z_n)\Big\}, \end{aligned}$$

where $E(|R_{AC}(Z_n)|) = O(n^{-3/2})$. Thus, finally, from (83),

$$\begin{aligned} E(L_{AC}) &= 2\kappa(\mu + a_2\mu^2)^{1/2}(1 + \frac{1}{2}a_2\kappa^2 n^{-1})n^{-1/2}\Big\{1 + a_2\kappa^2 n^{-1} + \frac{1}{2}a_2 n^{-1} \\ &+ \frac{1}{4}\kappa^2(\mu + a_2\mu^2)^{-1}n^{-1} - \frac{1}{8}(1 + 2a_2\mu)^2(\mu + a_2\mu^2)^{-1}n^{-1}\Big\} + O(n^{-2}) \end{aligned}$$

which simplifies to equation (35) stated in Theorem 7 on a few steps of algebra.

*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

$$w_2(\mu) = \frac{1}{36}(\mu + a_2\mu^2)^{-1/2}\{(\kappa^3 + 3\kappa) + a_2(\mu + a_2\mu^2)(13\kappa^3 + 17\kappa)\} \tag{83}$$

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 the expression (37) 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 the Binomial, Poisson, and the Negative Binomial case. The three separate expansions can then be unified into the general expression (36) stated in Theorem 7.

The limits of the likelihood ratio interval are the roots of the equation $-\log\Lambda_n = \kappa^2/2$, where $\Lambda_n$ 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 5, as described in equations (56) - (61). We will now describe the main steps in each of the Binomial, Poisson, and the Negative binomial case.

*1. The Poisson case*:

**Step 1.** The likelihood ratio $\Lambda_n$ is given by

$$\Lambda_n = \frac{\lambda^{n\bar{X}}e^{-n\lambda}}{\bar{X}^{n\bar{X}}e^{-n\bar{X}}} \tag{84}$$

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 $\lambda$; these are the limits of the interval $CI_{LR}$.
**Step 2**. 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 3**. By the same steps as in equations (56) - (61), the roots, say $\underline{t}$ and $\bar{t}$, satisfy

$$\underline{t} = -\kappa(n\bar{X})^{-1/2} + \frac{1}{3}\kappa^2(n\bar{X})^{-1} - \frac{1}{36}\kappa^3(n\bar{X})^{-3/2} + R_{1,n} \quad (85)$$

$$\text{and } \bar{t} = \kappa(n\bar{X})^{-1/2} + \frac{1}{3}\kappa^2(n\bar{X})^{-1} + \frac{1}{36}\kappa^3(n\bar{X})^{-3/2} + R_{2,n}, \quad (86)$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1, 2$. From (85) and (86), the length $L_{LR}$ of $CI_{LR}$ satisfies

$$E(L_{LR}) = 2\kappa E(\bar{X}^{1/2})n^{-1/2} + \frac{1}{18}\kappa^3\lambda^{1/2}n^{-3/2} + O(n^{-2}) \quad (87)$$

**Step 4**. Writing now $Z_n = n^{1/2}(\bar{X} - \lambda)/\lambda^{1/2}$, by a straightforward calculation,

$$E(\bar{X}^{1/2}) = \lambda^{1/2}[1 - (8n\lambda)^{-1}] + O(n^{-3/2}),$$

and so from (87) 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

$$E(L_{LR}) = 2\kappa\lambda^{1/2}n^{-1/2}(1 - \frac{9 - 2\kappa^2}{72n\lambda}) + O(n^{-2}).$$

*2. The Binomial case*:

**Step 1**. With $X \sim Bin(n, p)$, the likelihood ratio is given by

$$\Lambda_n = \frac{p^X(1-p)^{n-X}}{(X/n)^X(1 - X/n)^{n-X}} \quad (88)$$

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

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

**Step 3**. The roots $\underline{t}$ and $\bar{t}$ satisfy

$$\underline{t} = -\kappa(\hat{q}/\hat{p})^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1 - 2\hat{p})(n\hat{p})^{-1} - \frac{1}{36}\kappa^3(1 - 13\hat{p}\hat{q})\hat{q}^{-1/2}(n\hat{p})^{-3/2} + R_{1,n} \quad (89)$$

and

$$\bar{t} = \kappa(\hat{q}/\hat{p})^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1 - 2\hat{p})(n\hat{p})^{-1} + \frac{1}{36}\kappa^3(1 - 13\hat{p}\hat{q})\hat{q}^{-1/2}(n\hat{p})^{-3/2} + R_{2,n} \quad (90)$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1,\ 2$. From (89) and (90),

$$E(L_{LR}) = 2\kappa E[(\hat{p}\hat{q})^{1/2}]n^{-1/2} + \frac{1}{18}\kappa^3(1 - 13pq)(pq)^{-1/2}n^{-3/2} + O(n^{-2}) \tag{91}$$

**Step 4**. Writing $Z_n = n^{1/2}(\hat{p} - p)/(pq)^{1/2}$, by a straightforward expansion,

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

and so from (91) one obtains

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

which simplifies to

$$E(L_{LR}) = 2\kappa(pq)^{1/2}n^{-1/2}[1 - \frac{9 - 2\kappa^2(1 - 13pq)}{72npq}] + O(n^{-2}).$$

*3. The Negative Binomial case*: In our parametrization (See Section 2), the mean $\mu = p/q$, and so $p = \mu/(1 + \mu)$.

**Step 1**. The likelihood ratio is given by

$$\Lambda_n = \left(\frac{\mu}{\bar{X}}\right)^{n\bar{X}} \left(\frac{1 + \bar{X}}{1 + \mu}\right)^{n(1+\bar{X})}. \tag{92}$$

**Step 2**. Assuming $\bar{X} > 0$, if we write $t = \mu/\bar{X} - 1$, then the equation $-\log \Lambda_n = \kappa^2/2$ is equivalent to

$$(1 + \bar{X})\log(1 + \frac{\bar{X}}{1 + \bar{X}}t) - \bar{X}\log(1 + t) = \frac{\kappa^2}{2n}.$$

**Step 3**. The roots $\underline{t}$ and $\bar{t}$ of this equation satisfy

$$\begin{aligned} \underline{t} &= -\kappa[(1 + \bar{X})/\bar{X}])^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1 + 2\bar{X})(n\bar{X})^{-1} \\ &\quad - \frac{1}{36}\kappa^3(1 + 13\bar{X} + 13\bar{X}^2)(1 + \bar{X})^{-1/2}(n\bar{X})^{-3/2} + R_{1,n} \end{aligned} \tag{93}$$

and

$$\begin{aligned} \bar{t} &= \kappa[(1 + \bar{X})/\bar{X}])^{1/2}n^{-1/2} + \frac{1}{3}\kappa^2(1 + 2\bar{X})(n\bar{X})^{-1} \\ &\quad + \frac{1}{36}\kappa^3(1 + 13\bar{X} + 13\bar{X}^2)(1 + \bar{X})^{-1/2}(n\bar{X})^{-3/2} + R_{2,n} \end{aligned} \tag{94}$$

where $E(|R_{i,n}|) = O(n^{-2})$, $i = 1,\ 2$. From (93) and (94),

$$E(L_{LR}) = 2\kappa E[(\bar{X} + \bar{X}^2)^{1/2}]n^{-1/2} + \frac{1}{18}\kappa^3(1 + 13\mu + 13\mu^2)(\mu + \mu^2)^{-1/2}n^{-3/2} + O(n^{-2}) \tag{95}$$

**Step 4**. As usual, writing $Z_n = n^{1/2}(\bar{X} - \mu)/(\mu + \mu^2)^{1/2}$, from (95) by a straightforward expansion,

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

**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 Poisson, Binomial, and the Negative Binomial cases.

# References


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

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

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

[4] Bhattacharya, R.N. & Rao, R.R. (1976). *Normal Approximation And Asymptotic Expansions.* Wiley, New York.

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

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

[7] Brown, L.D., Cai, T. & DasGupta, A. (1999b). Confidence intervals for a binomial proportion and Edgeworth expansions. Technical Report, Department of Statistics, Purdue University.

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

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

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

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

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

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

[14] 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.

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

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

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

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

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

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