

BETTER SADDLEPOINT CONFIDENCE INTERVALS VIA BOOTSTRAP CALIBRATION

XIAODONG ZHENG

(Communicated by Wei-Yin Loh)

ABSTRACT. Confidence interval construction for parameters of lattice distributions is considered. By using saddlepoint formulas and bootstrap calibration, we obtain relatively short intervals and bounds with $O(n^{-3/2})$ coverage errors, in contrast with $O(n^{-1})$ and $O(n^{-1/2})$ coverage errors for normal theory intervals and bounds when the population distribution is absolutely continuous. Closed form solutions are also provided for the cases of binomial and Poisson distributions. The method is illustrated by some simulation results.

1. INTRODUCTION

The topic of confidence interval construction is ubiquitous in statistical inference. The major concern here is how to achieve good coverage accuracy for small sample size or asymptotically. Recent advances in Edgeworth expansion and bootstrap have made available a variety of confidence intervals with much improved coverages over traditional normal theory intervals (see, e.g. Hall, 1988). These results however only apply to the cases when the underlying distributions are continuous. The more difficult discrete distribution counterpart is quite underdeveloped.

The obstacle lies in the fact that the $O(n^{-1/2})$ term of the Edgeworth expansion for the normalized sample mean of, for example, a lattice distribution consists of two parts, one of which is continuous representing the skewness of the population, the other being discontinuous. The usual bootstrap approximation of the distribution of the normalized mean consistently estimates the skewness part but not the discontinuous part (Singh 1981, Theorem 1E). As a result, various bootstrap confidence intervals have coverage errors $O(n^{-1/2})$, rather than $O(n^{-1})$ for the continuous distribution counterpart. Similarly, the approach of inverting an Edgeworth expansion can only correct the skewness (Hall 1982). See also Hall (1987) where an improved bootstrap is used to reduce (but not completely eliminate) the error in the $n^{-1/2}$ order term.

In this paper, we try to narrow the gap by providing some new accurate and short intervals for parameters of lattice distributions through saddlepoint formulas and bootstrap calibration. Unlike Edgeworth expansions, the saddlepoint formulas estimate the tail probabilities of a sample mean through local approximations of the density via the steepest descent method. The results are much more accurate

Received by the editors October 2, 1996 and, in revised form, February 28, 1997.
1991 *Mathematics Subject Classification*. Primary 62F25; Secondary 62E20.

Key words and phrases. Bootstrap, hypothesis testing, lattice random variables.

and relatively smooth density-like approximations, which can be inverted to derive a confidence interval with the same order error via the duality between hypothesis testing and confidence intervals. Thus our approach is similar in spirit to that of Blyth (1986). As far as we are aware, saddlepoint formulas have been used for confidence interval construction in Tingley and Field (1990), but only in the cases when the sampling distributions are continuous. Their results do not apply to discrete distributions. Another related article is Zheng and Loh (1995), which employs smoothing technique. Our approach here seems more natural.

2. EXACT SADDLEPOINT CONFIDENCE INTERVALS

The lattice version of the Lugannani-Rice saddlepoint formula (Lugannani and Rice, 1980) provides very accurate approximation to the tail probability of the sample mean of a sequence of lattice random variables. It is also very compact for practical use compared to other saddlepoint formulas. See Daniels (1987) for some numerical examples.

Suppose X, X_1, \dots, X_n are a sequence of iid lattice random variables with mean μ and sample mean \bar{X} . Because any lattice variable X taking values on lattice points $\{a + jh; j \text{ an integer}\}$ with span $h > 0$ can be linearly transformed into a variable $Y = (X - a)/h$ with span 1, we therefore will assume that the span of X is 1. Denote the cumulant generating function (CGF) of X by $K(t) = \log E \exp(tX)$ which is assumed to exist. Let Φ and ϕ be the standard normal distribution and density respectively. Then the Lugannani-Rice formula takes the form:

$$(1) \quad P(\bar{X} \geq d) = 1 - \Phi(\zeta) + \phi(\zeta)\{\eta^{-1} - \zeta^{-1}\} + O(n^{-3/2})$$

where d is a possible value of \bar{X} , the saddlepoint W is defined by $K^{(1)}(W) = d$ and

$$(2) \quad \begin{aligned} \eta &= \{1 - \exp(-W)\}[nK^{(2)}(W)]^{1/2}, \\ \zeta &= \{2n[Wd - K(W)]\}^{1/2} \text{sgn}(W). \end{aligned}$$

Frequently, instead of tail probability approximations, we are interested in estimating an unknown parameter θ of the distribution of X by confidence intervals. Typically, the inference is based on the sample mean and the duality between hypothesis testing and confidence intervals. Formula (1) is used to approximate the significance level of a test for θ or more essentially the distribution of the test statistic.

We consider here a general class of one-parameter exponential families of lattice distributions, where the canonical form of the density of X with respect to the counting measure is

$$(3) \quad f_{\theta}(x) = \exp\{\beta T(x) - A(\beta)\}h(x)$$

with $\beta = \beta(\theta)$ a strictly monotone function of θ . Without loss of generality, assume that $\beta(\theta)$ is strictly increasing. Then the joint density of X_1, \dots, X_n admits monotone likelihood ratio in $\bar{T}_n = n^{-1} \sum_1^n T(X_i)$. By inverting the uniformly most powerful test of $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$, an exact $100(1 - \alpha)\%$ upper confidence bound for θ would be the solution $\theta = L(\bar{T}_n)$ to $F_{\theta}(\bar{T}_n) = \alpha$ if the distribution function $F_{\theta}(t)$ of \bar{T}_n were continuous in both t and θ (the latter is satisfied since \bar{T}_n admits an exponential family of distributions); cf. Lehmann (1986, Section 3.5).

To overcome the discrete nature of the distribution of \bar{T}_n and avoid randomization, we use the Lugannani-Rice formula as a smooth estimate \widehat{F}_θ of F_θ :

$$\widehat{F}_\theta(d) = \Phi(\zeta) - \phi(\zeta)\{\eta^{-1} - \zeta^{-1}\},$$

where η and ζ are defined by (2) and the CGF of $T(X)$ is $K(t) = A(t + \beta(\theta)) - A(\beta(\theta))$. Note that $A^{(2)}(\beta) = \text{Var}(T(X)) > 0$ and hence $A^{(1)}$ is a strictly increasing function. The saddle point W , which satisfies $A^{(1)}(W + \beta(\theta)) = d$, has the explicit expression $W = [A^{(1)}]^{-1}(d) - \beta(\theta)$. A nominal $100(1 - \alpha)\%$ saddlepoint based upper confidence bound U_1 for θ is the solution to $\widehat{F}_\theta(\bar{T}_n) = \alpha$. The coverage error of U_1 is $O(n^{-3/2})$ because of the $O(n^{-3/2})$ error in (1). Similarly, a nominal $100(1 - 2\alpha)\%$ two-sided confidence intervals for θ achieving $O(n^{-3/2})$ coverage error is $[L_1, U_1]$ with L_1 the solution in θ to $\widehat{F}_\theta(\bar{T}_n) = 1 - \alpha$.

We now give specific examples on the two most important lattice distributions.

Example 1. $X \sim \text{Bernoulli}(p)$ where $p \in (0, 1)$ and $q = 1 - p$. Then $\bar{T}_n = \widehat{p} = \sum_{i=1}^n X_i/n$. Since $K(t) = \log(q + p \exp[t])$, the saddlepoint is

$$W = \log[qdp^{-1}(1 - d)^{-1}].$$

Also

$$\begin{aligned} \eta &= (qd)^{-1}(d - p)[nd(1 - d)]^{1/2}, \\ (4) \quad \zeta &= \sqrt{2n}\{d \log[qdp^{-1}(1 - d)^{-1}] - \log[q(1 - d)^{-1}]\}^{1/2} \text{sgn}(d - p). \end{aligned}$$

Example 2. $X \sim \text{Poisson}(\lambda)$ where $\lambda = E(X) > 0$. This distribution has monotone likelihood ratio in \bar{T}_n , the sample mean. Furthermore $K(t) = \lambda[\exp(t) - 1]$ and $W = \log(d\lambda^{-1})$. The formulas for η and ζ are

$$\begin{aligned} \eta &= \sqrt{nd}(1 - \lambda/d), \\ (5) \quad \zeta &= \sqrt{2n}\{d \log(d\lambda^{-1}) - (d - \lambda)\}^{1/2} \text{sgn}(d - \lambda). \end{aligned}$$

3. APPROXIMATE SADDLEPOINT CONFIDENCE INTERVALS

The saddlepoint-based confidence interval $[L_1, U_1]$ does not have a convenient closed form solution. In this section, we attempt to fill this gap by giving two confidence intervals with explicit formulas for the Bernoulli (or binomial) and Poisson parameters. Because their functional forms are as simple as the normal theory intervals, the computation can be easily carried out by a hand calculator. Furthermore, they achieve the same coverage accuracy as those in the previous section, i.e., the coverage errors are $O(n^{-3/2})$.

3.1. Binomial parameter. We use the notation of Example 1. Consider the construction of a nominal $100(1 - \alpha)\%$ ($\alpha < 1/2$) upper confidence bound for p . The basic idea is to find an approximate d_n such that $\widehat{F}_p(d_n) = \alpha + O(n^{-3/2})$ and set d_n equal to the observed value \widehat{p} . According to arguments in Section 2, the solution to this equation in p is then the desired upper bound.

Let d_n satisfy $P(\widehat{p} \geq d_n) \rightarrow 1 - \alpha$ and $e_n = \sqrt{n}(d_n - p)/\sqrt{pq}$. Then $e_n \rightarrow z_\alpha < 0$ and $d_n - p = n^{-1/2}\sqrt{pq}e_n = O(n^{-1/2})$ is negative, where $\Phi(z_\alpha) = \alpha$. Hence, for

large n , the saddlepoint $W = \log\{qd_n[p(1-d_n)]^{-1}\}$ is negative. Expand

$$\begin{aligned}\zeta &= -\sqrt{2n}\{d_n \log[qd_n p^{-1}(1-d_n)^{-1}] - \log[q(1-d_n)^{-1}]\}^{1/2} \\ &= -\{e_n^2 + (1/3)n^{-1/2}e_n^3(2p-1)(pq)^{-1/2} \\ &\quad + (1/6)n^{-1}(1-3p+3p^2)e_n^4(pq)^{-1} + O(n^{-3/2})\}^{1/2} \\ &= e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n^3(14p^2-14p+5)(pq)^{-1} + O(n^{-3/2})\end{aligned}$$

and

$$\begin{aligned}\eta &= (qd_n)^{-1}(d_n-p)[nd_n(1-d_n)]^{1/2} \\ &= e_n - (1/2)n^{-1/2}e_n^2(pq)^{-1/2} + (3/8)n^{-1}e_n^3(1-2p)(pq)^{-1} + O(n^{-3/2}).\end{aligned}$$

We have

$$\begin{aligned}\eta^{-1} - \zeta^{-1} &= (1/3)n^{-1/2}(1+p)(pq)^{-1/2} \\ &\quad + (1/12)n^{-1}e_n(p^2+8p-1)(pq)^{-1} + O(n^{-3/2}).\end{aligned}$$

Then (1) becomes

$$\begin{aligned}P\{\hat{p} \geq p + n^{-1/2}e_n(pq)^{1/2}\} &= 1 - \Phi\{e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n^3[14p^2-14p+5](pq)^{-1}\} \\ &\quad + \phi\{e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n^3[14p^2-14p+5](pq)^{-1}\}[(1/3)n^{-1/2}(1+p)(pq)^{-1/2} \\ &\quad + (1/12)n^{-1}e_n(p^2+8p-1)(pq)^{-1}] + O(n^{-3/2}) \\ &= 1 - \Phi\{e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n^3[14p^2-14p+5](pq)^{-1} \\ &\quad - (1/3)n^{-1/2}(1+p)(pq)^{-1/2} - (1/12)n^{-1}e_n(p^2+8p-1)(pq)^{-1}\} \\ &\quad - (1/2)[(1/3)n^{-1/2}(1+p)(pq)^{-1/2} \\ &\quad \quad + (1/12)n^{-1}e_n(p^2+8p-1)(pq)^{-1}]^2 \\ &\quad \times [e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad \quad + (1/72)n^{-1}e_n^3[14p^2-14p+5](pq)^{-1}] \\ &\quad \times \phi\{e_n + (1/6)n^{-1/2}e_n^2(2p-1)(pq)^{-1/2} \\ &\quad \quad + (1/72)n^{-1}e_n^3[14p^2-14p+5](pq)^{-1}\} \\ (6) \quad &+ O(n^{-3/2}) \\ &= 1 - \Phi\{e_n + (1/6)n^{-1/2}[(2p-1)e_n^2 - 2(1+p)](pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n[e_n^2(14p^2-14p+5) - 6(p^2+8p-1)](pq)^{-1}\} \\ (7) \quad &- (1/18)n^{-1}e_n(1+p)^2(pq)^{-1}\phi(e_n) + O(n^{-3/2}) \\ &= 1 - \Phi\{e_n + (1/6)n^{-1/2}[(2p-1)e_n^2 - 2(1+p)](pq)^{-1/2} \\ &\quad + (1/72)n^{-1}e_n[e_n^2(14p^2-14p+5) - 2(p^2+20p-5)](pq)^{-1}\} \\ (8) \quad &+ O(n^{-3/2})\end{aligned}$$

where (6) follows from $\Phi(x) - y\phi(x) = \Phi(x - y) + (1/2)y^2x\phi(x) + O(y^3)$ with

$$x = e_n + (1/6)n^{-1/2}e_n^2(2p - 1)(pq)^{-1/2} + (1/72)n^{-1}e_n^3[14p^2 - 14p + 5](pq)^{-1}$$

and

$$y = (1/3)n^{-1/2}(1 + p)(pq)^{-1/2} + (1/12)n^{-1}e_n(p^2 + 8p - 1)(pq)^{-1}.$$

Any e_n satisfying

$$\begin{aligned} & e_n + (1/6)n^{-1/2}[(2p - 1)e_n^2 - 2(1 + p)](pq)^{-1/2} \\ & + (1/72)n^{-1}e_n[e_n^2(14p^2 - 14p + 5) - 2(p^2 + 20p - 5)](pq)^{-1} \\ (9) \quad & = z_\alpha + O(n^{-3/2}) \end{aligned}$$

will fulfill the requirement that

$$(10) \quad P[\hat{p} \geq p + n^{-1/2}e_n(pq)^{1/2}] = 1 - \alpha + O(n^{-3/2}).$$

In particular,

$$\begin{aligned} e_n &= z_\alpha \{1 + (36npq)^{-1}[z_\alpha^2(2p^2 - 2p - 1) - 2(7p^2 - 16p + 1)]\}^{1/2} \\ (11) \quad & - (36npq)^{-1/2}[(2p - 1)z_\alpha^2 - 2(1 + p)] \end{aligned}$$

satisfies (10). Now set $d_n = p + n^{-1/2}e_n(pq)^{1/2} = \hat{p}$ and rewrite it as

$$\begin{aligned} p - \hat{p} - (6n)^{-1}[(2p - 1)z_\alpha^2 - 2(1 + p)] \\ = -n^{-1/2}z_\alpha\{pq + (36n)^{-1}[z_\alpha^2(2p^2 - 2p - 1) - 2(7p^2 - 16p + 1)]\}^{1/2}. \end{aligned}$$

The latter may be reduced to a quadratic equation in p by taking square of both sides. The solution can be easily verified to be

$$(12) \quad p = U_2 = \{\delta_1 + (\delta_1^2 - \delta_2\delta_3)^{1/2}\}/\delta_2,$$

where

$$\begin{aligned} \delta_1 &= \hat{p} + (3n)^{-1}\hat{q}(z_\alpha^2 - 1) + (6n)^{-2}(z_\alpha^4 + 18z_\alpha^2 - 4), \\ \delta_2 &= 1 + (3n)^{-1}(2 + z_\alpha^2) + 2(6n)^{-2}(z_\alpha^4 + 3z_\alpha^2 + 2), \\ \delta_3 &= \hat{p}^2 - (3n)^{-1}\hat{p}(z_\alpha^2 + 2) + 2(6n)^{-2}(z_\alpha^4 + 3z_\alpha^2 + 2). \end{aligned}$$

Note that U_2 is well-defined because asymptotically $\delta_1^2 - \delta_2\delta_3 = n^{-1}pqz_\alpha^2 + O_p(n^{-2})$ is nonnegative. The derivation of a nominal two-sided interval for p is similar. We thus have

Theorem 1. *Let $\alpha < 1/2$. Then*

- (a) *The nominal $100(1 - \alpha)\%$ upper confidence bound U_2 has coverage error $O(n^{-3/2})$.*

(b) The nominal $100(1 - 2\alpha)\%$ two-sided confidence interval $[L_2, U_2]$ with

$$L_2 = \{\delta_1 - (\delta_1^2 - \delta_2\delta_3)^{1/2}\}/\delta_2$$

also has coverage error $O(n^{-3/2})$.

3.2. Poisson parameter. We only give an outline of the construction of a nominal $100(1 - \alpha)\%$ upper confidence bound U_2 for the Poisson parameter $\lambda = E(X)$ based on \bar{X} . Since $\text{Var}(X) = \lambda$, any d_n such that $P(\bar{X} \geq d_n) \rightarrow 1 - \alpha$ must satisfy $d_n = \lambda + \sqrt{\lambda}e_n n^{-1/2} = E(X) + O(n^{-1/2})$ with $e_n \rightarrow z_\alpha < 0$. Thus the saddlepoint $W = \log(d_n/\lambda) < 0$ for all large n . Following Example 2 and section 3.1, we have

$$\begin{aligned}\zeta &= e_n\{1 - (1/6)e_n(n\lambda)^{-1/2} + (5/72)e_n^2(n\lambda)^{-1}\} + O(n^{-3/2}), \\ \eta &= e_n\{1 - (1/2)e_n(n\lambda)^{-1/2} + (3/8)e_n^2(n\lambda)^{-1}\} + O(n^{-3/2}), \\ \eta^{-1} - \zeta^{-1} &= (1/3)(n\lambda)^{-1/2} - (1/12)e_n(n\lambda)^{-1} + O(n^{-3/2}).\end{aligned}$$

Therefore

$$\begin{aligned}P(\bar{X} \geq \lambda + \sqrt{\lambda}e_n n^{-1/2}) &= 1 - \Phi\left\{e_n - (1/2)e_n^2(n\lambda)^{-1/2} + (3/8)e_n^3(n\lambda)^{-1}\right\} \\ &\quad + \phi\left\{e_n - (1/2)e_n^2(n\lambda)^{-1/2} + (3/8)e_n^3(n\lambda)^{-1}\right\} \\ &\quad \times [(1/3)(n\lambda)^{-1/2} - (1/12)e_n(n\lambda)^{-1}] + O(n^{-3/2}) \\ &= 1 - \Phi\left\{e_n - (1/2)e_n^2(n\lambda)^{-1/2} + (3/8)e_n^3(n\lambda)^{-1} - (1/3)(n\lambda)^{-1/2}\right. \\ &\quad \left.+ (1/12)e_n(n\lambda)^{-1}\right\} - (1/18)(n\lambda)^{-1}e_n\phi(e_n) + O(n^{-3/2}) \\ &= 1 - \Phi\left\{e_n - (1/6)(e_n^2 + 2)(n\lambda)^{-1/2} + (5/72)(e_n^2 + 2)e_n(n\lambda)^{-1}\right\} \\ (13) \quad &\quad + O(n^{-3/2}).\end{aligned}$$

Since

$$(14) \quad e_n = z_\alpha \left\{1 - (1/36)(z_\alpha^2 + 2)(n\lambda)^{-1}\right\}^{1/2} + (1/6)(z_\alpha^2 + 2)(n\lambda)^{-1/2}$$

satisfies

$$e_n - (1/6)(e_n^2 + 2)(n\lambda)^{-1/2} + (5/72)(e_n^2 + 2)e_n(n\lambda)^{-1} = z_\alpha + O(n^{-3/2}),$$

(13) implies that the event

$$(15) \quad \bar{X} \geq \lambda + n^{-1/2}z_\alpha \left\{\lambda - (1/36)(z_\alpha^2 + 2)n^{-1}\right\}^{1/2} + (1/6)(z_\alpha^2 + 2)n^{-1}$$

has probability $1 - \alpha + O(n^{-3/2})$. Now set the right-hand side of (15) to the observed value $\bar{X} = \bar{x}$ and get

$$\bar{x} - \lambda - (1/6)(z_\alpha^2 + 2)n^{-1} = n^{-1/2}z_\alpha \left\{\lambda - (1/36)(z_\alpha^2 + 2)n^{-1}\right\}^{1/2}$$

and square both sides to reduce it to a quadratic inequality in λ . The resulting solution is

$$(16) \quad \lambda \leq U_2 = \bar{x} - z_\alpha n^{-1/2} \left\{\bar{x} + (18n)^{-1}(z_\alpha^2 - 7)\right\}^{1/2} - (1/3)n^{-1}(1 - z_\alpha^2).$$

We have the following

Theorem 2. Let $\alpha < 1/2$.

(a) The nominal $100(1 - \alpha)\%$ upper confidence bound U_2 for λ has coverage error $O(n^{-3/2})$.

(b) *The nominal $100(1 - 2\alpha)\%$ two-sided confidence interval $[L_2, U_2]$ with*

$$L_2 = \bar{x} + z_\alpha n^{-1/2} \{ \bar{x} + (18n)^{-1} (z_\alpha^2 - 7) \}^{1/2} - (1/3)n^{-1}(1 - z_\alpha^2)$$

has coverage $1 - 2\alpha + O(n^{-3/2})$.

4. BOOTSTRAP CALIBRATION

Although the above saddlepoint-based intervals achieve $O(n^{-3/2})$ coverage errors, our simulation results in the next section indicate that, in finite sample situations, they usually have higher coverages than the nominal level and shorter lengths than normal theory intervals. This phenomenon suggests that the value α used in computing the nominal $100(1 - \alpha)\%$ upper bounds U_1 and U_2 , and $100(1 - 2\alpha)\%$ two-sided intervals $[L_1, U_1]$ and $[L_2, U_2]$ is often too conservative or small. A natural adjustment is to use an $\alpha^* > \alpha$ instead of α in all the above formulas. Intuitively, this adjustment has another appealing effect of reducing the length of the interval. The question is how much adjustment is needed, or how to determine the corrected value α^* . It is also reasonable to ask that the new corrected intervals should theoretically maintain the original $O(n^{-3/2})$ coverage accuracy.

A particularly useful tool for this purpose is the bootstrap calibration technique, which was introduced by Loh (1987, 1988); see also Loh (1991). The idea of the bootstrap calibration is to first use bootstrap to estimate the true coverage $\pi(\alpha)$ of a confidence interval. The interval is then adjusted by comparing $\pi(\alpha)$ with the target nominal level.

In our context, parametric bootstrap is more convenient and reasonable than the nonparametric bootstrap. Under the one-parameter exponential family distributions (3), one first obtains the \sqrt{n} -consistent maximum likelihood estimator (MLE) for β : $\hat{\beta} = [A^{(1)}]^{-1}(\bar{T}_n)$ (Lehmann, 1983, Example 6.2.4). Since $\beta(\theta)$ is a monotone increasing function of θ , the MLE for θ is $\hat{\theta} = \beta^{-1}(\hat{\beta})$. The parametric bootstrap distribution is exactly (3) with β and θ replaced by $\hat{\beta}$ and $\hat{\theta}$. Denote the bootstrap measure by $P^*(\cdot)$.

To bootstrap the coverage $\pi(\alpha)$ of a saddlepoint-based nominal $100(1 - \alpha)\%$ upper bound $U(\bar{T}_n)$ for θ , let X_1^*, \dots, X_n^* be an iid sample from P^* . Compute the bootstrap version $\bar{T}_n^* = n^{-1} \sum_{i=1}^n T(X_i^*)$ of \bar{T}_n . Then $\pi(\alpha) = P[\theta \leq U(\bar{T}_n)]$ is estimated by $\hat{\pi}(\alpha) = P^*[\hat{\theta} \leq U(\bar{T}_n^*)]$. Define the excess of the estimated coverage over the target coverage by

$$\xi = \psi(\hat{\pi}(\alpha)) - \psi(1 - \alpha)$$

where ψ is a strictly increasing unbounded function on $(0, 1)$ with continuous first derivative. Then the adjusted level α^* is the solution to

$$\psi(1 - \alpha^*) = \psi(1 - \alpha) - \xi.$$

That is, the calibrated version U_1^* of U_1 is the solution in θ to $\hat{F}_\theta(\hat{T}_n) = \alpha^*$ and similarly, by replacing z_α in U_2 we get U_2^* . Note that if an upper bound has conservative coverage, the estimated excess ξ will detect this and bring α down to α^* . Otherwise, it will increase α . It is not difficult to see that since $\pi(\alpha) = 1 - \alpha + O(n^{-3/2})$ and the bootstrap distribution has exactly the same form as (3) except that θ is replaced by $\hat{\theta}$, we have $\hat{\pi}(\alpha) = 1 - \alpha + O_p(n^{-3/2})$. Hence $\alpha^* = \alpha + O_p(n^{-3/2})$ and the calibrated upper bounds retain the low coverage errors of $O(n^{-3/2})$.

The calibration for a nominal $100(1 - \alpha)\%$ lower bound $L(\bar{T}_n)$ is similar with obvious changes of definitions of $\pi(\alpha)$ and $\hat{\pi}(\alpha)$. The calibrated lower bounds L_1^* and L_2^* together with U_1^* and U_2^* give nominal $100(1 - 2\alpha)\%$ two-sided intervals $[L_1^*, U_1^*]$ and $[L_2^*, U_2^*]$ with $O(n^{-3/2})$ coverage errors.

5. SIMULATION RESULTS

We conducted a Monte Carlo experiment to examine the finite sample properties of the above intervals. The normal theory interval and an interval given by Blyth (1986) were also included for comparison. The latter also has $O(n^{-3/2})$ coverage error and its $100(1 - \alpha)\%$ upper bound is given by $(a - z_\alpha\sqrt{b})/c$ where

$$\begin{aligned}
 a &= \hat{p}[1 + (1 - z_\alpha^2)(3n)^{-1}] + (2 + z_\alpha^2)(3n)^{-1}, \\
 b &= n^{-1}(\hat{p} + n^{-1})\hat{q}[1 + (7 - z_\alpha^2)/(18n)] - (1 + n^{-1})(7 - z_\alpha^2)/(18n^2),
 \end{aligned}$$

and $c = 1 + (2 + z_\alpha^2)/(3n)$ and the lower bound is obtained by replacing z_α by $-z_\alpha$ in the above formulas. Three sample sizes ($n = 10, 50, 500$) were tested and the nominal level was 95%. For the binomial two-sided intervals, $p = 0.05, 0.1, 0.3$ and 0.5 were used. For upper bounds, we tested $p = 0.1, 0.3, 0.5, 0.7$ and 0.9 . For Poisson intervals, values of λ were set to 1, 3 and 10 respectively. The number of simulation trials was 1,000. The bootstrap iteration was 500 with calibration function the inverse standard normal cdf. The exact saddlepoint intervals are undefined for $\hat{p} = 0, 1$ or $\hat{\lambda} = 0$. In these cases, we used the formulas of the approximate intervals instead. The binomial lower and upper bounds were truncated at 0 and 1 respectively. The Poisson lower bounds were also truncated at 0. The results are reported in Tables 1–4, except for the exact saddlepoint intervals for the binomial parameter which are given in Zheng and Loh (1995) and not duplicated here.

TABLE 1. Monte Carlo estimates of coverages of 95% two-sided intervals for binomial p ; average interval lengths in parentheses; ψ -function is inverse normal cdf.

Interval type		Normal Theory	Blyth	Approx. Saddlept.	Calibrated Approx. Saddlept.
n	p				
10	.05	.339(.110)	.918(.303)	.973(.204)	.965(.186)
50	.05	.731(.109)	.971(.120)	.935(.081)	.945(.092)
500	.05	.940(.038)	.942(.032)	.954(.033)	.954(.034)
10	.1	.641(.255)	.941(.390)	.930(.377)	.941(.377)
50	.1	.877(.159)	.938(.165)	.961(.150)	.955(.146)
500	.1	.926(.052)	.948(.050)	.952(.049)	.952(.046)
10	.3	.857(.507)	.940(.468)	.923(.470)	.918(.472)
50	.3	.939(.251)	.922(.241)	.947(.241)	.948(.238)
500	.3	.939(.080)	.951(.078)	.953(.067)	.948(.065)
10	.5	.868(.590)	.922(.511)	.941(.517)	.944(.513)
50	.5	.939(.274)	.953(.260)	.958(.261)	.954(.245)
500	.5	.937(.088)	.951(.082)	.949(.082)	.945(.070)
Max. SE		.015(.009)	.009(.005)	.009(.005)	.009(.002)

TABLE 2. Monte Carlo estimates of coverages of 95% upper bounds for binomial p ; average interval lengths in parentheses; ψ -function is inverse normal cdf.

Interval type		Normal Theory	Blyth	Approx. Saddlept.	Calibrated Approx. Saddlept.
n	p				
10	.1	.650(.228)	.945(.291)	.951(.214)	.942(.180)
50	.1	.889(.172)	.925(.169)	.927(.172)	.941(.176)
500	.1	.910(.122)	.958(.124)	.934(.111)	.942(.104)
10	.3	.864(.525)	.928(.514)	.920(.495)	.924(.505)
50	.3	.934(.408)	.942(.414)	.936(.363)	.944(.402)
500	.3	.950(.359)	.951(.335)	.950(.359)	.967(.400)
10	.5	.938(.749)	.943(.770)	.940(.699)	.939(.741)
50	.5	.940(.619)	.938(.601)	.941(.620)	.946(.605)
500	.5	.955(.538)	.954(.537)	.946(.534)	.943(.526)
10	.7	.938(.907)	.957(.859)	.946(.845)	.946(.795)
50	.7	.962(.809)	.942(.804)	.946(.791)	.947(.783)
500	.7	.959(.734)	.955(.732)	.957(.722)	.953(.716)
10	.9	.984(.994)	.970(.968)	.925(.971)	.942(.937)
50	.9	.977(.968)	.959(.947)	.944(.947)	.947(.920)
500	.9	.956(.922)	.949(.921)	.945(.918)	.945(.912)
Max. SE		.015(.011)	.009(.006)	.009(.006)	.008(.003)

TABLE 3. Monte Carlo estimates of coverages of 95% two-sided intervals for Poisson λ ; average interval lengths in parentheses; ψ -function is inverse normal cdf.

Interval type		Normal Theory	Approx. Saddlept.	Exact Saddlept.	Calibrated Approx. Saddlept.	Calibrated Exact Saddlept.
n	λ					
10	1	.922(1.228)	.957(1.216)	.957(1.216)	.953(1.139)	.954(1.141)
50	1	.937(0.553)	.929(.551)	.940(.550)	.940(.553)	.953(.554)
500	1	.935(.176)	.937(.175)	.945(.169)	.943(.177)	.944(.169)
10	3	.942(2.140)	.957(2.123)	.970(1.838)	.954(1.974)	.975(1.843)
50	3	.949(.960)	.944(.958)	.953(.919)	.951(.950)	.947(.876)
500	3	.946(.304)	.946(.303)	.955(.304)	.946(.300)	.949(.297)
10	10	.942(3.914)	.949(3.909)	.975(3.573)	.954(3.909)	.965(3.422)
50	10	.951(1.752)	.956(1.752)	.968(1.688)	.943(1.194)	.957(1.198)
500	10	.954(.558)	.957(.554)	.972(.537)	.952(.531)	.957(.534)
Max. SE		.008(.014)	.008(.010)	.008(.009)	.008(.009)	.007(.008)

In summary, saddlepoint-based intervals and the interval given by Blyth (1986) perform substantially better than the normal theory interval. The improvements are more salient when parameters assume boundary values (p close to 0 or 1, λ close to 0) or when the sample size is small. The interval by Blyth is comparable to the approximate saddlepoint interval in coverage and sometimes is slightly longer.

TABLE 4. Monte Carlo estimates of coverages of 95% upper bounds for Poisson λ ; average interval lengths in parentheses; ψ -function is inverse normal cdf.

Interval type		Normal Theory	Approx. Saddlept.	Exact Saddlept.	Calibrated Approx. Saddlept.	Calibrated Exact Saddlept.
n	λ					
10	1	.926(1.524)	.926(1.574)	.972(1.509)	.944(1.580)	.963(1.502)
50	1	.933(1.233)	.933(1.231)	.970(1.195)	.932(1.228)	.966(1.173)
500	1	.945(1.074)	.949(1.073)	.985(1.070)	.954(1.102)	.969(1.005)
10	3	.921(3.869)	.953(3.912)	.987(3.913)	.947(3.867)	.982(3.875)
50	3	.932(3.406)	.942(3.405)	.979(3.386)	.939(3.409)	.967(3.277)
500	3	.943(3.126)	.943(3.127)	.980(3.056)	.952(3.200)	.974(3.005)
10	10	.937(11.693)	.938(11.690)	.996(10.173)	.941(10.346)	.998(10.874)
50	10	.948(10.739)	.948(10.741)	.989(10.047)	.939(10.023)	.977(10.004)
500	10	.946(10.231)	.947(10.230)	.992(10.019)	.954(10.003)	.982(9.763)
Max. SE		.009(.012)	.008(.009)	.005(.009)	.008(.009)	.008(.009)

Bootstrap calibration reduces interval length while retaining or improving coverage accuracy. In conclusion, the intervals having closed form solutions are recommended for routine use while calibration can be combined if high precision and short length are desirable.

ACKNOWLEDGEMENTS

The author is grateful to Professor Wei-Yin Loh for suggestions leading to this research. The revision also benefits substantially from a referee's detailed comments, one of which points out an error in Theorem 1 of the earlier version.

REFERENCES

- Blyth, C.R. (1986). Approximate binomial confidence limits. *J. Amer. Statist. Assoc.* **81** 843–855; Corridenda (1989), *84*, 636. CMP 19:02
- Daniels, H.E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55** 37–48. MR **90e**:62029
- Hall, P. (1982). Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika* **69** 647–652. MR **85i**:62025
- Hall, P. (1987). On the bootstrap and continuity correction. *J. R. Statist. Soc. B* **49** 82–89.
- Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. MR **89h**:62085
- Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York. MR **85a**:62001
- Lehmann, E.L. (1986). *Testing Statistical Hypotheses*. 2nd. ed., Wiley, New York. MR **87j**:62001
- Loh, W.-Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82** 155–162. MR **88e**:62082
- Loh, W.-Y. (1988). Discussion of “Theoretical comparison of bootstrap confidence intervals” by P. Hall. *Ann. Statist.* **16** 972–976.
- Loh, W.-Y. (1991). Bootstrap calibration for confidence interval construction and selection. *Statistica Sinica* **1** 477–491. CMP 92:02
- Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12** 475–490. MR **81f**:60034

12. Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195. MR **83c**:62047
13. Tingley, M. and Field, C. (1990). Small-sample confidence intervals. *J. Amer. Statist. Assoc.* **85** 427–434. MR **92k**:62059
14. Zheng, X. and Loh, W.-Y. (1995). Bootstrapping binomial confidence intervals. *J. Statist. Planning and Inference.* **43** 355–380. MR **96f**:62075

DEPARTMENT OF MATHEMATICS AND STATISTICS, UTAH STATE UNIVERSITY, LOGAN, UTAH
84322-3900