Matching pseudocounts for interval estimation of binomial and Poisson parameters

Haruhiko Ogasawara*
Otaru University of Commerce, Otaru, Japan – hogasa@res.otaru-uc.ac.jp

Abstract

For interval estimation of a binomial proportion and a Poisson mean, matching pseudocounts are derived, which give the one-sided Wald confidence intervals with second-order accuracy. The confidence intervals remove the bias of coverage probabilities given by the score confidence intervals. Partial poor behavior of the confidence intervals by the matching pseudocounts is corrected by hybrid methods using the score confidence interval depending on sample values.

Keywords: one-sided confidence interval; Cornish-Fisher expansion; score confidence interval; Wald confidence interval.

1. Introduction

Interval estimation of a binomial proportion is a basic statistical problem, which can be typically performed by using the studentized sample proportion $\hat{p}$ following an approximate standard normal distribution. This approximation gives the two-sided confidence interval (CI) with an asymptotic confidence level $1-\alpha$:

$$\hat{p} \pm n^{-1/2} z_{\alpha/2} (\hat{p}\hat{q})^{1/2}, \tag{1.1}$$

where n is the sample size, $z_{\alpha/2}$ is the $(1-(\alpha/2))$th quantile of the standard normal, and $\hat{q} = 1-\hat{p}$. Unfortunately, it is known that the Wald CI of (1.1) behaves poorly especially when the population proportion p_0 is close to 0 or 1 with moderate sample sizes (Ghosh, 1979; Blyth & Still, 1983; Newcombe, 1998, 2011; Santner, 1998; Agresti & Coull, 1998; Brown, Cai & DasGupta, 2001, 2002; Pires & Amado, 2005; Cousins, Hymes & Tucker, 2010).

A CI for a Poisson mean corresponding to (1.1) is

$$\bar{X} \pm n^{-1/2} z_{\alpha/2} \bar{X}^{1/2}, \tag{1.2}$$

where $\bar{X}$ is a sample mean, which also shows a poor behavior (Agresti & Coull, 1998; Brown, Cai & DasGupta, 2002; for one-sided CIs corresponding to (1.1) and (1.2), see Cai, 2005).

Practical CIs with excellent behavior even with a small sample size are given by Wilson's (1927) score CIs corresponding to (1.1) and (1.2), which use the sample proportion and mean standardized by the population standard errors: $n^{1/2}(\hat{p}-p_0)/(p_0 q_0)^{1/2}$ and $n^{1/2}(\bar{X}-\lambda_0)/\lambda_0^{1/2}$, where $\mathrm{E}(\hat{p}) = p_0$ with $q_0 = 1-p_0$ and $\mathrm{E}(\bar{X}) = \lambda_0$. Setting these standardized values, rather than studentized ones, equal to $\pm z_{\alpha/2}$ and solving for p_0 and λ_0, we have the CIs of p_0 and λ_0:

$$\frac{1}{1+n^{-1}z_{\alpha/2}^2}\left\{\hat{p} + \frac{n^{-1}}{2} z_{\alpha/2}^2 \pm n^{-1/2} z_{\alpha/2} \left(\hat{p}\hat{q} + \frac{n^{-1}}{4} z_{\alpha/2}^2\right)^{1/2}\right\} \tag{1.3}$$

and

$$\bar{X} + \frac{n^{-1}}{2} z_{\alpha/2}^2 \pm n^{-1/2} z_{\alpha/2} \left(\bar{X} + \frac{n^{-1}}{4} z_{\alpha/2}^2\right)^{1/2}, \tag{1.4}$$

respectively.

While the two-sided score CIs of (1.3) and (1.4) generally show excellent performances, the corresponding one-sided score CIs have non-negligible biases of coverage with opposite directions to the corresponding Wald CIs though less severe than the latter CIs (Cai, 2005). The purpose of this paper is to show that the systematic biases of the one-sided score CIs can be removed by using pseudocounts based on an optimal power of the Jeffreys (1946; 1961, Section 3.6) prior for the canonical or natural parameters i.e., logit and log λ for the binomial and Poisson distributions, respectively. When the pseudocounts, which can be negative, are added, the one-sided Wald CIs are shown to be asymptotically more accurate than the usual Wald and score CIs when lattice properties of the discrete distributions are neglected.

2. Pseudocounts based on asymptotic expansions

For clarity, we use the standard definition of the asymptotic order of accuracy for a CI in the following way. Assume that a lower endpoint denoted by $L(\theta_0, 1-\alpha, n^{-i/2})$ of a one-sided CI has nominal upper $1-\alpha$ coverage probability for a population quantity θ_0 with $\Pr\{\theta_0 \geq L(\theta_0, 1-\alpha, n^{-i/2})\} = 1-\alpha + O(n^{-i/2})$ $(i = 1, 2, \ldots)$. Then, the CI with $L(\theta_0, 1-\alpha, n^{-i/2})$ is said to be i-th order accurate.

While Cai's (2005) third-order accurate CI is for a mean parameter in the exponential family, a similar CI is obtained for a corresponding canonical parameter. Let $q^* = q^*(\theta)$ be a weight in the weighted score or penalized likelihood for a scalar parameter θ, whose estimation equation is

$$\frac{\partial \bar{l}}{\partial \theta} + n^{-1} q^*(\theta) = 0, \tag{2.1}$$

where $\bar{l}$ is the log likelihood averaged over n observations and q^* becomes the log prior derivative when Bayesian estimation is used. The estimator $\hat{\theta}_{\mathrm{W}}$ by the weighted score is a posterior mode when the weight is given by a prior.

Define

$$t_{\mathrm{W}}^* = n^{1/2} \hat{\bar{i}}_{\mathrm{ML}}^{1/2} (\hat{\theta}_{\mathrm{W}} - \theta_0), \tag{2.2}$$

where $\hat{\bar{i}}_{\mathrm{ML}}$ is the estimated Fisher information per observation given by the maximum likelihood estimator (MLE) $\hat{\theta}_{\mathrm{ML}}$. When θ is a canonical parameter in the exponential family, it is known that the population Fisher information $\bar{i}_0$ is equal to var(X) where X is the associated observable variable. For instance, $\bar{i}_0 = p_0 q_0$ and $\bar{i}_0 = \lambda_0$ for the binomial and Poisson distributions, respectively. Then, asymptotically, $t_{\mathrm{W}}^* \to \mathrm{N}(0,1)$ as $t_{\mathrm{ML}} \equiv n^{1/2} \hat{\bar{i}}_{\mathrm{ML}}^{1/2} (\hat{\theta}_{\mathrm{ML}} - \theta_0) \to \mathrm{N}(0,1)$.

Denote the j-th $(j = 1, 2, \Delta 2, 3, 4)$ order asymptotic cumulants, independent of n, for t_{ML} of a canonical scalar parameter by $\beta_{\mathrm{ML}\,j}^{(t)}$ and the skewness and excess kurtosis of X by sk(X) and kt(X), respectively, where $\beta_{\mathrm{ML}\Delta 2}^{(t)}$ is the added higher-order asymptotic variance of t_{ML}. Then, Ogasawara (2013, Equation (4.4)) showed that the cumulants of t_{ML} up to the fourth-order are

$$\begin{aligned}
\kappa_1(t_{\mathrm{ML}}) &= n^{-1/2} \beta_{\mathrm{ML}1}^{(t)} + O(n^{-3/2}) = O(n^{-3/2}) \ (\beta_{\mathrm{ML}1}^{(t)} = 0), \\
\kappa_2(t_{\mathrm{ML}}) &= 1 + n^{-1} \left\{ -\frac{3}{4} \mathrm{sk}(X)^2 + \frac{\mathrm{kt}(X)}{2} \right\} + O(n^{-2}) \\
&\equiv 1 + n^{-1} \beta_{\mathrm{ML}\Delta 2}^{(t)} + O(n^{-2}) \ (\beta_{\mathrm{ML}2}^{(t)} = 1), \\
\kappa_3(t_{\mathrm{ML}}) &= n^{-1/2} \mathrm{sk}(X) + O(n^{-3/2}) \equiv n^{-1/2} \beta_{\mathrm{ML}3}^{(t)} + O(n^{-3/2}), \\
\kappa_4(t_{\mathrm{ML}}) &= n^{-1} \{-3\mathrm{sk}(X)^2 + 3\mathrm{kt}(X)\} + O(n^{-2}) \equiv n^{-1} \beta_{\mathrm{ML}4}^{(t)} + O(n^{-2}),
\end{aligned} \tag{2.3}$$

where note that the asymptotic bias of order $O(n^{-1/2})$ is $0\,(\beta_{\mathrm{ML}1}^{(t)} = 0)$. For t_{W}^*, noting

$\hat{\theta}_{\mathrm{W}} = \hat{\theta}_{\mathrm{ML}} + n^{-1}\hat{\bar{i}}_{\mathrm{ML}}^{-1}\hat{q}_{\mathrm{ML}}^{*} + O_p(n^{-2})$ (Ogasawara, 2014, Equation (5.8)) and using Ogasawara (2014, Theorem 3 and Corollary 1) and (2.3), it can be shown that

$$
\begin{aligned}
\kappa_1(t_{\mathrm{W}}^{*}) &= n^{-1/2}\bar{i}_0^{-1/2}q_0^{*} + O(n^{-3/2}) \equiv n^{-1/2}\beta_{\mathrm{W}1}^{(t)} + O(n^{-3/2}),\\
\kappa_2(t_{\mathrm{W}}^{*}) &= 1 + n^{-1}\{\beta_{\mathrm{ML}\Delta 2}^{(t)} + 2\bar{i}_0^{1/2} n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{\bar{i}}_{\mathrm{ML}}^{-1/2}\hat{q}_{\mathrm{ML}}^{*})\} + O(n^{-2})\\
&= 1 + n^{-1}\{\beta_{\mathrm{ML}\Delta 2}^{(t)} - \bar{i}_0^{-2}q_0^{*}\bar{i}_0^{(\mathrm{D}1)} + 2n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{q}_{\mathrm{ML}}^{*})\} + O(n^{-2})\\
&\equiv 1 + n^{-1}\beta_{\mathrm{W}\Delta 2}^{(t)} + O(n^{-2})\ (\beta_{\mathrm{W}2}^{(t)} = 1),\\
\kappa_3(t_{\mathrm{W}}^{*}) &= n^{-1/2}\beta_{\mathrm{ML}3}^{(t)} + O(n^{-3/2})\ (\beta_{\mathrm{W}3}^{(t)} = \beta_{\mathrm{ML}3}^{(t)}),\\
\kappa_4(t_{\mathrm{W}}^{*}) &= n^{-1}\beta_{\mathrm{ML}4}^{(t)} + O(n^{-2})\ (\beta_{\mathrm{W}4}^{(t)} = \beta_{\mathrm{ML}4}^{(t)}),
\end{aligned}
\tag{2.4}
$$

where $q_0^{*} = q^{*}(\theta_0)$, $\hat{q}_{\mathrm{ML}}^{*} = q^{*}(\hat{\theta}_{\mathrm{ML}})$, $\mathrm{acov}(\cdot)$ $(\mathrm{avar}(\cdot))$ is the asymptotic covariance (variance) of order $O(n^{-1})$ for variables in parentheses and $\bar{i}_0^{(\mathrm{D}1)} \equiv \partial\bar{i}/\partial\theta\,|_{\theta=\theta_0}$.

Denote *t* as a generic studentized estimator, whose special cases are t_{ML} and t_{W}^{*}. Let $\beta_j^{(t)}(j = 1, 2, \Delta 2, 3, 4)$ be the generic asymptotic cumulants, independent of *n*, for *t*, whose special cases are $\beta_{\mathrm{ML}\,j}^{(t)}$ and $\beta_{\mathrm{W}j}^{(t)}$. Then, from the Cornish-Fisher expansion (Cornish & Fisher, 1937; Hall, 1992), Ogasawara (2012, Equation (2.5)) gave the following lower endpoint for a one-sided third-order accurate CI for θ_0, where θ can be a parameter other than a canonical one:

$$
\begin{aligned}
L(\theta_0, 1-\alpha, n^{-3/2}) &= \hat{\theta} - n^{-1/2}\hat{\beta}_2^{1/2}z_\alpha - n^{-1}\hat{\beta}_2^{1/2}\left\{\hat{\beta}_1^{(t)} + \frac{\hat{\beta}_3^{(t)}}{6}(z_\alpha^2 - 1)\right\}\\
&\quad - n^{-3/2}\hat{\beta}_2^{1/2}\left[\frac{1}{2}\left\{\hat{\beta}_{\Delta 2}^{(t)} - 2\hat{\beta}_2^{-1/2} n\,\widehat{\mathrm{acov}}\left(\hat{\theta}, \hat{\beta}_1^{(t)} + \frac{\hat{\beta}_3^{(t)}}{6}(z_\alpha^2 - 1)\right)\right\}z_\alpha\right.\\
&\quad \left. + \{\hat{\beta}_3^{(t)}\}^2\left(-\frac{z_\alpha^3}{18} + \frac{5}{36}z_\alpha\right) + \hat{\beta}_4^{(t)}\left(\frac{z_\alpha^3}{24} - \frac{z_\alpha}{8}\right)\right],
\end{aligned}
\tag{2.5}
$$

where $\hat{\beta}_2$ is the estimated β_2, $n^{-1}\beta_2 = \mathrm{avar}(\hat{\theta})$, $\hat{\beta}_j^{(t)}$ is a sample version of $\beta_j^{(t)}(j = 1, 2, \Delta 2, 3, 4)$, and oscillation terms are neglected. The upper bound $U(\theta_0, 1-\alpha, n^{-3/2})$ corresponding to (2.5) is given by replacing z_α with $z_{1-\alpha} = -z_\alpha$.

Theorem 1. *The pseudocount for each cell of two binomial categories and a single cell of a Poisson count, which yields second-order accurate* $L(\theta_0, 1-\alpha, n^{-1})$ *and* $U(\theta_0, 1-\alpha, n^{-1})$ *using the Wald confidence interval based on* t_{W}^{*} *in(2.2), is given by* $-(z_\alpha^2 - 1)/6$, *where* θ_0 *is a generic population mean (binomial proportion or Poisson mean).*

Proof. Define tentatively θ as a single canonical parameter. Let C^{*} be a common pseudocount for generic cell(s). Then, $q^{*} = C^{*}\partial\log\bar{i}/\partial\theta = C^{*}\bar{i}^{(\mathrm{D}1)}/\bar{i}$ (recall $\bar{i}_0 = p_0 q_0$ and $\bar{i}_0 = \lambda_0$ for the canonical parameters in the binomial and Poisson distributions), which gives from (2.3) and (2.4)

$$
\beta_{\mathrm{W}1}^{(t)} = \bar{i}_0^{-1/2}q_0^{*} = C^{*}\mathrm{sk}(X) \text{ and } \beta_{\mathrm{W}3}^{(t)} = \beta_{\mathrm{ML}3}^{(t)} = \mathrm{sk}(X). \tag{2.6}
$$

The third term $-n^{-1/2}\hat{\beta}_2^{1/2}z_\alpha\{\hat{\beta}_1^{(t)} + (\hat{\beta}_3^{(t)}/6)(z_\alpha^2 - 1)\}$ on the right-hand side of (2.5) in the case of t_{W}^{*}, a case of *t*, vanishes when $C^{*} = C_{\mathrm{M}}^{*} \equiv -(z_\alpha^2 - 1)/6$, which indicates that the one-sided Wald

CIs for θ_0 are second-order accurate. Since the mean parameters in the two distributions are monotone functions of the corresponding canonical parameters, the monotonically transformed CIs for the original population mean parameters, restoring notation θ_0 for these means, are second-order accurate except for oscillation terms. Q.E.D.

Noting that the Jeffreys prior is given by $\bar{i}^{1/2}$, C^* gives the $2C^*$th power of the Jeffreys prior.

Corollary 1. *The lower endpoint for the third-order accurate one-sided CI using the pseudocount C^*_{M} in Theorem 1 for a binomial proportion is given by*

$$L(p_0, 1-\alpha, n^{-3/2}) = \frac{1}{1+\exp\{-L(\theta_0, 1-\alpha, n^{-3/2})\}}, \tag{2.7}$$

where

$$\begin{aligned} L(\theta_0, 1-\alpha, n^{-3/2}) &= \log\frac{\hat{p}+n^{-1}C^*_{\mathrm{M}}}{\hat{q}+n^{-1}C^*_{\mathrm{M}}} - n^{-1/2}(\hat{p}\hat{q})^{-1/2} z_\alpha \\ &\quad - n^{-3/2}(\hat{p}\hat{q})^{-1/2}\left\{\left(\frac{z_\alpha^3}{36}-\frac{5}{72}z_\alpha\right)(\hat{p}\hat{q})^{-1} - \frac{z_\alpha^3}{36}+\frac{7}{36}z_\alpha\right\}. \end{aligned} \tag{2.8}$$

For a Poisson mean,

$$\begin{aligned} &L(\lambda_0, 1-\alpha, n^{-3/2}) \\ &= (\bar{X}+n^{-1}C^*_{\mathrm{M}})\exp\left\{-n^{-1/2}\bar{X}^{-1/2}z_\alpha - n^{-3/2}\bar{X}^{-3/2}\left(\frac{z_\alpha^3}{36}-\frac{5}{72}z_\alpha\right)\right\}. \end{aligned} \tag{2.9}$$

The corresponding upper endpoints $U(p_0, 1-\alpha, n^{-3/2})$ and $U(\lambda_0, 1-\alpha, n^{-3/2})$ are given by replacing z_α in the lower endpoints with $-z_\alpha$.

Proof. In (2.5), when using C^*_{M} for a canonical parameter, from (2.3), (2.4) and $\hat{q}^*_{\mathrm{ML}} = C^*_{\mathrm{M}}\hat{\bar{i}}^{(\mathrm{D1})}_{\mathrm{ML}}\hat{\bar{i}}^{-1}_{\mathrm{ML}}$, we have

$$\begin{aligned} \beta^{(t)}_{\mathrm{W}\Delta 2} &= \beta^{(t)}_{\mathrm{ML}\Delta 2} - \bar{i}_0^{-2} q_0^* \bar{i}_0^{(\mathrm{D1})} + 2n\,\mathrm{acov}(\hat{\theta}_{\mathrm{ML}}, \hat{q}^*_{\mathrm{ML}}) \\ &= \beta^{(t)}_{\mathrm{ML}\Delta 2} + C^*_{\mathrm{M}}\{-3\bar{i}_0^{-3}(\bar{i}_0^{(\mathrm{D1})})^2 + 2\bar{i}_0^{-2}\bar{i}_0^{(\mathrm{D2})}\} \\ &= -\frac{3}{4}\mathrm{sk}(X)^2 + \frac{\mathrm{kt}(X)}{2} - \frac{z_\alpha^2 - 1}{6}\{-3\mathrm{sk}(X)^2 + 2\mathrm{kt}(X)\}, \\ \beta^{(t)}_{\mathrm{W}3} &= \beta^{(t)}_{\mathrm{ML}3} = \mathrm{sk}(X), \\ \beta^{(t)}_{\mathrm{W}4} &= \beta^{(t)}_{\mathrm{ML}4} = -3\mathrm{sk}(X)^2 + 3\mathrm{kt}(X), \end{aligned} \tag{2.10}$$

where $\bar{i}_0^{(\mathrm{D2})} \equiv \partial^2\bar{i}/\partial\theta^2 |_{\theta=\theta_0}$. When the matching pseudocount C^*_{M} is used, noting that $\mathrm{avar}(\hat{\theta}_{\mathrm{ML}}) = \mathrm{avar}(\hat{\theta}_{\mathrm{W}}) = n^{-1}\beta_2 = n^{-1}\bar{i}_0^{-1}$, (2.5) based on t^*_{W} becomes

$$\begin{aligned} L(\theta_0, 1-\alpha, n^{-3/2}) &= \hat{\theta}_{\mathrm{W}} - n^{-1/2}\hat{\beta}_2^{1/2} z_\alpha \\ &\quad - n^{-3/2}\hat{\beta}_2^{1/2}\left[\frac{1}{2}\hat{\beta}^{(t)}_{\mathrm{W}\Delta 2} z_\alpha + \{\hat{\beta}^{(t)}_{\mathrm{ML}3}\}^2\left(-\frac{z_\alpha^3}{18}+\frac{5}{36}z_\alpha\right) + \hat{\beta}^{(t)}_{\mathrm{ML}4}\left(\frac{z_\alpha^3}{24}-\frac{z_\alpha}{8}\right)\right] \end{aligned} \tag{2.11}$$

$$= \hat{\theta}_{\mathrm{W}} - n^{-1/2} i_{\mathrm{ML}}^{-1/2} z_\alpha$$

$$-n^{-3/2} i_{\mathrm{ML}}^{-1/2} \Bigg[-\frac{3}{8} \widehat{\mathrm{sk}(X)}^2 z_\alpha + \frac{\widehat{\mathrm{kt}(X)}}{4} z_\alpha - \frac{z_\alpha^3 - z_\alpha}{12} \{-3\widehat{\mathrm{sk}(X)}^2 + 2\widehat{\mathrm{kt}(X)}\}$$

$$+\widehat{\mathrm{sk}(X)}^2 \left(-\frac{z_\alpha^3}{18} + \frac{5}{36} z_\alpha \right) + \{-3\widehat{\mathrm{sk}(X)}^2 + 3\widehat{\mathrm{kt}(X)}\} \left(\frac{z_\alpha^3}{24} - \frac{z_\alpha}{8} \right) \Bigg]$$

$$= \hat{\theta}_{\mathrm{W}} - n^{-1/2} \hat{\beta}_2^{1/2} z_\alpha - n^{-3/2} i_{\mathrm{ML}}^{-1/2} \left\{ \left(\frac{5}{72} z_\alpha^3 - \frac{z_\alpha}{9} \right) \widehat{\mathrm{sk}(X)}^2 + \left(-\frac{z_\alpha^3}{24} + \frac{z_\alpha}{24} \right) \widehat{\mathrm{kt}(X)} \right\}.$$

For the binomial distribution with canonical parametrization, it is known that

$$\bar{i}_0 = p_0 q_0,\ \mathrm{sk}(X)^2 = \frac{(1-2p_0)^2}{p_0 q_0} = (p_0 q_0)^{-1} - 4 \tag{2.12}$$

$$\text{and } \mathrm{kt}(X) = \frac{1 - 6p_0 + 6p_0^2}{p_0 q_0} = (p_0 q_0)^{-1} - 6\,.$$

Inserting sample versions of (2.12) into (2.11), (2.8) and consequently (2.7) follow.

For the Poisson distribution with a canonical parameter, we have

$$\bar{i}_0 = \lambda_0,\ \mathrm{sk}(X)^2 = (\lambda_0 / \lambda_0^{3/2})^2 = \lambda_0^{-1} \text{ and } \mathrm{kt}(X) = \lambda_0 / \lambda_0^2 = \lambda_0^{-1}. \tag{2.13}$$

Substituting sample versions of (2.13) for (2.11),

$$L(\theta_0, 1-\alpha, n^{-3/2})$$
$$= \log(\bar{X} + n^{-1} C_{\mathrm{M}}^*) - n^{-1/2} \bar{X}^{-1/2} z_\alpha - n^{-3/2} \bar{X}^{-3/2} \left(\frac{z_\alpha^3}{36} - \frac{5}{72} z_\alpha \right) \tag{2.14}$$

follows. Transformation of (2.14) gives (2.9). Q.E.D.

In (2.8) for a binomial proportion, $\hat{p} + n^{-1} C_{\mathrm{M}}^*$ or $\hat{q} + n^{-1} C_{\mathrm{M}}^*$ can be non-positive with a finite sample size. In these cases, $L(p_0, 1-\alpha, n^{-3/2})$ is defined to be 0 or 1 when $\hat{p} + n^{-1} C_{\mathrm{M}}^*$ or $\hat{q} + n^{-1} C_{\mathrm{M}}^*$ is non-positive, respectively. In such cases, similar definitions are also used for $U(p_0, 1-\alpha, n^{-3/2})$. In (2.9) for a Poisson mean, when $\bar{X} + n^{-1} C_{\mathrm{M}}^*$ is negative $L(\lambda_0, 1-\alpha, n^{-3/2})$ and $U(\lambda_0, 1-\alpha, n^{-3/2})$ are defined to be 0.

References

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

Blyth, C. R., & Still, H. A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association, 78*, 108-116.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science, 16*, 101-123.

Brown, L. D., Cai, T. T., & DasGupta, A. (2002). Confidence interval for a binomial proportion and asymptotic expansions. *The Annals of Statistics, 30*, 160-201.

Cai, T. T. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference, 131*, 63-88.

Cornish, E. A., & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Revue de l’Institut International de Statistique, 5*, 307-320.

Cousins, R. D., Hymes, K. E., & Tucker, J. (2010). Frequentist evaluation of intervals estimated for a binomial parameter and for the ratio of Poisson means. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrrameters, Detectors and Associated*

Equipment, *612*, 388-398.
Ghosh, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, *74*, 894-900.
Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer. Corrected printing, 1997.
Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*, *Series A*, *186*, 453-461.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon.
Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, *17*, 857-872.
Newcombe, R. G. (2011). Measures of location for confidence intervals for proportions. *Communications in Statistics - Theory and Methods, 40*, 1743-1767.
Ogasawara, H. (2012). Cornish-Fisher expansions using sample cumulants and monotonic transformations. *Journal of Multivariate Analysis*, *103*, 1-18.
Ogasawara, H. (2013). Asymptotic cumulants of the estimator of the canonical parameter in the exponential family. *Journal of Statistical Planning and Inference*, *143*, 2142-2150.
Ogasawara, H. (2014). Bias adjustment minimizing the asymptotic mean square error. *Communications in Statistics – Theory and Methods*. DOI:10.1080/03610926.2013.786788.
Pires, A. M., & Amado, C. (2008). Interval estimators for a binomial proportion: Comparison of twenty methods. *REVSTAT – Statistical Journal*, *6*, 165-197.
Santner, T. J. (1998). A note on teaching binomial confidence intervals. *Teaching Statistics*, *20*, 20-23.
Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*, 209-212.