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 \( 1 - \alpha \):

\[
\hat{p} \pm n^{-1/2} z_{\alpha/2} (\hat{p}\hat{q})^{1/2},
\]

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

\[
\bar{X} \pm n^{-1/2} z_{\alpha/2} \bar{X}^{1/2},
\]

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: \( n^{1/2} \left( \hat{p} - p_0 \right) / (p_0 q_0)^{1/2} \) and \( n^{1/2} \left( \bar{X} - \lambda_0 \right) / \lambda_0^{1/2} \), where \( \text{E}(\hat{p}) = p_0 \) with \( q_0 = 1 - p_0 \) and \( \text{E}(\bar{X}) = \lambda_0 \). Setting these standardized values, rather than studentized ones, equal to \( \pm z_{\alpha/2} \) and solving for \( p_0 \) and \( \lambda_0 \), we have the CIs of \( p_0 \) and \( \lambda_0 \):

\[
\frac{1}{1 + n^{-1} z_{\alpha/2}^2} \left\{ \hat{p} + n^{-1} z_{\alpha/2}^2 \pm n^{-1/2} z_{\alpha/2} \left( \hat{p}\hat{q} + n^{-1} z_{\alpha/2}^2 \right)^{1/2} \right\}
\]

and

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

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 \lambda 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 \( \theta_0 \) with \( \Pr\{a \leq \theta_0 \geq L(\theta_0, 1-\alpha, n^{-i/2})\} = 1-\alpha + O(n^{-i/2}) \) \((i = 1, 2, \ldots)\). 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 \( \theta \), where estimation equation is

\[
\frac{\partial T}{\partial \theta} + n^{-1} q^*(\theta) = 0,
\]

where \( T \) is the log likelihood averaged over \( n \) observations and \( q^* \) becomes the log prior derivative when Bayesian estimation is used. The estimator \( \hat{\theta}_W \) by the weighted score is a posterior mode when the weight is given by a prior.

Define

\[ t^*_W = n^{1/2} \hat{\theta}_W - \theta_0, \]

where \( \hat{\theta}_W \) is the estimated Fisher information per observation given by the maximum likelihood estimator (MLE) \( \hat{\theta}_{ML} \). When \( \theta \) is a canonical parameter in the exponential family, it is known that the population Fisher information \( \hat{T}_0 \) is equal to \( \var{X} \) where \( X \) is the associated observable variable.

For instance, \( \hat{T}_0 = p_0 q_0 \) and \( \hat{T}_0 = \lambda_0 \) for the binomial and Poisson distributions, respectively. Then, asymptotically, \( t^*_W \rightarrow N(0,1) \) as \( t_{ML} \equiv n^{1/2} \hat{\theta}_{ML} - \theta_0 \rightarrow N(0,1) \).

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 \( \beta_{ML,j} \) and the skewness and excess kurtosis of \( X \) by \( \text{sk}(X) \) and \( \text{kt}(X) \), respectively, where \( \beta_{ML2} \) 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

\[
\kappa_1(t_{ML}) = n^{-1/2} \beta_{ML1} + O(n^{-3/2}) = O(n^{-3/2}) \quad (\beta_{ML1} = 0),
\]

\[
\kappa_2(t_{ML}) = 1 + n^{-1} \left\{-\frac{3}{4} \text{sk}(X)^2 + \frac{\text{kt}(X)}{2}\right\} + O(n^{-2})
\equiv 1 + n^{-1} \beta_{ML2} + O(n^{-2}) \quad (\beta_{ML2} = 1),
\]

\[
\kappa_3(t_{ML}) = n^{-1/2} \text{sk}(X) + O(n^{-3/2}) \equiv n^{-1/2} \beta_{ML3} + O(n^{-3/2}),
\]

\[
\kappa_4(t_{ML}) = n^{-1}(-3\text{sk}(X)^2 + 3\text{kt}(X)) + O(n^{-2}) \equiv n^{-1} \beta_{ML4} + O(n^{-2}),
\]

where note that the asymptotic bias of order \( O(n^{-1/2}) \) is 0 \((\beta_{ML1} = 0)\). For \( t^*_W \), noting
\( \hat{\theta}_w = \hat{\theta}_{\text{ML}} + n^{-1/2} \hat{t}_{\text{ML}}^* q^*_{\text{ML}} + O_p(n^{-2}) \) (Ogasawara, 2014, Equation (5.8)) and using Ogasawara (2014, Theorem 3 and Corollary 1) and (2.3), it can be shown that

\[
\kappa_j(t_w^*) = n^{-1/2} T_0^{-1/2} q_0^* + O(n^{-3/2}) \equiv n^{-1/2} \beta^{(i)}_w + O(n^{-3/2}),
\]

\[
\kappa_j(t_w^*) = 1 + n^{-1} \{ \beta^{(i)}_{ML2} + 2 T_0^{-1/2} n \text{acov}(\hat{\theta}_{ML}, \hat{t}_{ML}^*, q_{ML}^*) \} + O(n^{-2})
\]

\[
= 1 + n^{-1} \{ \beta^{(i)}_{ML2} - T_0^{-2} q_0 T_0^{-1}(D1) + 2n \text{acov}(\hat{\theta}_{ML}, \hat{t}_{ML}^*, q_{ML}^*) \} + O(n^{-2})
\]

\[
\equiv 1 + n^{-1} \beta^{(i)}_{wa2} + O(n^{-2}) \ (\beta^{(i)}_{wa2} = 1),
\]

\[
\kappa_j(t_w^*) = n^{-1/2} \beta^{(i)}_{ML3} + O(n^{-3/2}) \ (\beta^{(i)}_{W3} = \beta^{(i)}_{ML3}),
\]

\[
\kappa_j(t_w^*) = n^{-1/2} \beta^{(i)}_{ML4} + O(n^{-2}) \ (\beta^{(i)}_{W4} = \beta^{(i)}_{ML4}),
\]

where \( q_0^* = q^*(\theta_0), \hat{q}_{ML}^* = q^*(\hat{\theta}_{ML}), \text{acov}(\cdot) \) is the asymptotic covariance (variance) of order \( O(n^{-1}) \) for variables in parentheses and \( \hat{t}_0^{(D1)} \equiv \partial T / \partial \theta \big|_{\theta = \theta_0} \).

Denote \( t \) as a generic studentized estimator, whose special cases are \( t_{ML} \) and \( t_w^* \). Let \( \beta^{(i)}_j (j = 1, 2, \Delta 2, 3, 4) \) be the generic asymptotic cumulants, independent of \( n \), for \( t \), whose special cases are \( \beta^{(i)}_{ML} \) and \( \beta^{(i)}_w \). 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 \( \theta_0 \), where \( \theta \) can be a parameter other than a canonical one:

\[
L(\theta_0, 1 - \alpha, n^{-3/2}) = \hat{\theta} - \hat{\beta}_2^{1/2} z_\alpha - n^{-1/2} \hat{\beta}_2^{1/2} \left[ \hat{\beta}_1^{(i)} + \frac{\hat{\beta}^{(i)}_3}{6} \left( z_\alpha^2 - 1 \right) \right]
\]

\[
- n^{-1/2} \hat{\beta}_2^{1/2} \left[ \frac{1}{2} \left( \hat{\beta}^{(i)}_{a2} - 2 \hat{\beta}_2^{1/2} n \text{acov}(\hat{\theta}, \hat{t}_2^{(i)} + \frac{\hat{\beta}^{(i)}_4}{6} (z_\alpha^2 - 1) ) \right) \right] z_\alpha
\]

\[
+ \{ \hat{\beta}_3^{(i)} \} \left[ \frac{z_\alpha^3}{18} + \frac{5}{36} z_\alpha \right] + \hat{\beta}_4^{(i)} \left( \frac{z_\alpha^3}{24} - \frac{z_\alpha}{8} \right),
\]

where \( \hat{\beta}_2 \) is the estimated \( \beta_2 \), \( n^{-1} \beta_2 = \text{avar}(\hat{\theta}) \), \( \hat{\beta}_j^{(i)} \) is a sample version of \( \beta_j^{(i)} (j = 1, 2, \Delta 2, 3, 4) \), and oscillation terms are neglected. The upper bound \( U(\theta_0, 1 - \alpha, n^{-3/2}) \) corresponding to (2.5) is given by replacing \( z_\alpha \) 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 \( \theta_0 \) is a generic population mean (binomial proportion or Poisson mean).

Proof. Define tentatively \( \theta \) as a single canonical parameter. Let \( C^* \) be a common pseudocount for generic cell(s). Then, \( q^* = C^* \partial \log T / \partial \theta = C^* T^{(D1)} / T \) (recall \( T_0 = p_0 q_0 \) and \( T_0 = \lambda_0 \) for the canonical parameters in the binomial and Poisson distributions), which gives from (2.3) and (2.4)

\[
\beta^{(i)}_w = T_0^{-1/2} q_0^* = C^* \text{sk}(X) \quad \text{and} \quad \beta^{(i)}_{W3} = \beta^{(i)}_{ML3} = \text{sk}(X).
\]

The third term \( -n^{-1/2} \hat{\beta}_2^{1/2} z_\alpha \{ \hat{\beta}_1^{(i)} + (\hat{\beta}_3^{(i)} / 6)(z_\alpha^2 - 1) \} \) on the right-hand side of (2.5) in the case of \( t_w^* \), a case of \( t \), vanishes when \( C^* = C^*_M \equiv -(z_\alpha^2 - 1) / 6 \), which indicates that the one-sided Wald
CIs for $\theta_0$ 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 $\theta_0$ for these means, are second-order accurate except for oscillation terms. Q.E.D.

Noting that the Jeffreys prior is given by $\frac{1}{2}i$, $C^*$ 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

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

(2.7)

where

$$L(\theta_0, 1-\alpha, n^{-3/2}) = \log \frac{\hat{\theta} + n^{-1}C^*_M}{\hat{\theta} + n^{-1}C^*_M} - n^{-1/2}(\hat{\theta} - \hat{z}_a^{-1/2})z_a$$

$$- n^{-3/2}(\hat{\theta} - \hat{z}_a^{-1/2})\left\{\left(\frac{z_a^3}{36} - \frac{5}{72}z_a\right)(\hat{\theta} - \hat{z}_a^{-1/2}) - \frac{z_a^3}{36} + \frac{7}{36}z_a\right\}. \quad (2.8)$$

For a Poisson mean,

$$L(\lambda_0, 1-\alpha, n^{-3/2}) = (\hat{X} + n^{-1}C^*_M)\exp\left\{-n^{-1/2}\hat{X}z_a - n^{-3/2}\hat{X}z_a^{-1/2}\left(\frac{z_a^3}{36} - \frac{5}{72}z_a\right)\right\},$$

(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 $z_a$ in the lower endpoints with $-z_a$.

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

$$\beta_w^{(i)} = \beta_{ML2}^{(i)} - 3T_0^{-1}\hat{q}_{ML} + 2n\text{acov}(\hat{\theta}_{ML}, \hat{q}_{ML})$$

$$= \beta_{ML2}^{(i)} + C^*_M\left\{-3T_0^{-1}(T_0^{(1)})^2 + 2T_0^{-2}T_0^{(2)}\right\}$$

$$= -\frac{3}{4}\text{sk}(X)^2 + k_t(X) - \frac{z_a^2 - 1}{6}\left\{-3\text{sk}(X)^2 + 2k_t(X)\right\},$$

$$\beta_{w3}^{(i)} = \beta_{ML3}^{(i)} = \text{sk}(X),$$

$$\beta_{w4}^{(i)} = \beta_{ML4}^{(i)} = -3\text{sk}(X)^2 + 3k_t(X),$$

where $T_0^{(2)} \equiv \frac{\partial^2}{\partial \theta^2} |_{\theta = 0}$. When the matching pseudocount $C^*_M$ is used, noting that

$$\text{avar}(\hat{\theta}_{ML}) = \text{avar}(\hat{\theta}_w) = n^{-1}\beta_2 = n^{-1}T_0^{-1}, \quad (2.5)$$

based on $t_{ML}^*$ becomes

$$L(\theta_0, 1-\alpha, n^{-3/2}) = \hat{\theta}_w - n^{-1/2}\hat{z}_a$$

$$- n^{-3/2}(\hat{\theta}_w - \hat{z}_a)\left[\frac{1}{2}\beta_{w2}^{(i)}z_a + \left(\beta_{ML3}^{(i)}\right)^2 - \frac{z_a^3}{18} + \frac{5}{36}z_a\right] + \hat{\beta}_{ML4}^{(i)}\left(\frac{z_a^3}{24} - \frac{z_a}{8}\right). \quad (2.11)$$
\[ \hat{W} - n^{-1/2} \tilde{t}_{ML} z_a \]
\[ -n^{-3/2} \tilde{t}_{ML} \left[ -\frac{3}{8} \text{sk}(X)^2 z_a + \frac{kt(X)}{4} z_a - \frac{z_a^3 - z_a}{12} \left\{ -3 \text{sk}(X)^2 + 2kt(X) \right\} \right] \]
\[ + \text{sk}(X)^2 \left( \frac{z_a^3}{18} + \frac{5}{36} z_a \right) + \left\{ -3 \text{sk}(X)^2 + 3kt(X) \right\} \left( \frac{z_a^3}{24} - \frac{z_a}{8} \right) \]
\[ = \hat{W} - n^{-1/2} \tilde{p}^{1/2} z_a - n^{-3/2} \tilde{t}_{ML} \left\{ \right\}
\[ = \frac{5}{72} z_a^3 - \frac{z_a}{9} \text{sk}(X)^2 + \left( -\frac{z_a^3}{24} + \frac{z_a}{24} \right) kt(X) \right\}. \]

For the binomial distribution with canonical parametrization, it is known that
\[ 2 \text{sk}(X)^2 = \frac{1 - 2 p_0}{p_0q_0} = (p_0q_0)^{-1} - 4 \]
and
\[ \text{kt}(X) = \frac{1 - 6 p_0 + 6 p_0^2}{p_0q_0} = (p_0q_0)^{-1} - 6. \]

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
\[ \tilde{X}_0 = \lambda_0, \text{sk}(X)^2 = (\lambda_0 / \lambda_0^{3/2})^2 = \lambda_0^{-1} \text{ and } \text{kt}(X) = \lambda_0 / \lambda_0^2 = \lambda_0^{-1}. \]

Substituting sample versions of (2.13) for (2.11),
\[ L(\theta_0, 1 - \alpha, n^{-3/2}) \]
\[ = \log(X + n^{-1} C_{\alpha}^r) - n^{-1/2} X^{-1/2} z_a - n^{-3/2} X^{-3/2} \left( \frac{z_a^3}{36} - \frac{5}{72} z_a \right) \]
follows. Transformation of (2.14) gives (2.9). Q.E.D.

In (2.8) for a binomial proportion, \( \hat{p} + n^{-1} C_{\alpha}^r \) or \( \hat{q} + n^{-1} C_{\alpha}^r \) 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 \( \hat{p} + n^{-1} C_{\alpha}^r \) or \( \hat{q} + n^{-1} C_{\alpha}^r \) 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 \( X + n^{-1} C_{\alpha}^r \) 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


