Improved Confidence Intervals for the Bernoulli Parameter

Wheyming Tina Song, Chia-Jung Chang
Department of Industrial Engineering National Tsing Hua University, Hsinchu, Taiwan, Republic of China

Abstract

Despite the simplicity of the Bernoulli process, developing good confidence-interval procedures for its parameter—the probability of success p—is deceptively difficult. The binary data yield a discrete number of successes from a discrete number of trials, n. This discreteness results in actual coverage probabilities that oscillate with the n for fixed values of p (and with p for fixed n). We suggest ideas designed to reduce the effect of discreteness on coverage probability while having minimal effect on properties of the interval width. The ideas are of four types: ways to improve the standard-error estimator, ways to improve choice of the Student-t distribution's degrees of freedom, ways to randomly perturb the interval, and ways to combine intervals. After reviewing two confidence-interval procedures and a family of point estimators for p, we discuss four specific new procedures. We illustrate improved performance with Monte Carlo experiments. An example procedure that performs well is to randomly mix intervals from E. B. Wilson's (1927) procedure with intervals from an adjusted standard interval.

Keywords: Confidence Interval, Bernoulli Parameter, Monte-Carlo Simulation.

1 Introduction

Consider a Bernoulli process $X_1, X_2, \ldots, X_n$ with parameter p. The probability mass function of each X_i has the form

$$f_X(x) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Also X_i and X_j are independent for all $i \neq j$. The confidence interval (CI) procedure for p with confidence level $1-\alpha$ is a procedure to determine L^* and U^* such that the following probability statement is satisfied:

$$\mathrm{P}(L^* \leq p \leq U^*) = 1 - \alpha, \tag{1}$$

where $0 < \alpha < 1$. Both L^* and U^* are functions of $X_1, X_2, \ldots, X_n$; hence, they are random variables. Once we select a sample of size n (say $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$), and compute the values of L^* and U^* using the sample values, we obtain a confidence interval (l^*, u^*) for p. Both l^* and u^* are real values.

In practice, it is difficult or impossible to find a CI procedure to satisfy Equation(1) for any given α, n and p. Our objective is to construct a CI procedure to form (L, U) for p at a nominal confidence level $1-\alpha$ such that (i) its actual coverage probability (CP), $\mathrm{P}(L \leq p \leq U)$, is close to the nominal level $1-\alpha$, (ii) its expected interval width, $\mathrm{E}(U-L)$, is narrow enough to be informative, and (iii) its standard deviation of the interval width, $\mathrm{Std}(U-L)$, is small enough to be stable. We aim to meet the three criteria simultaneously because it is easy to construct a CI procedure satisfying just one criterion.

We define some notation here. First we make a distinction between "CI procedure" and "CI". We use CI procedure to refer to a procedure to form (L, U) which is a random interval, and use CI to refer to a real-value interval (l, u). Let $\hat{P}$ and $\hat{p}$ be the general point estimator (random variable) and the corresponding point estimate (real number) of p, respectively. To refer to a specific CI procedure, we sometimes add a subscript to CI, $\hat{P}$ and $\hat{p}$. For example, to refer to a standard CI procedure, we then use notation $\mathrm{CI_S}$, $\hat{P}_\mathrm{S}$ and $\hat{p}_\mathrm{S}$. The notation $\mathrm{CP}(n, p)$ is refer to the coverage probability measured at values n and p.

The organization of this paper is as follows. In Section 2, we review two CI procedures, and expand upon point estimators for p. In Section 3, we discuss four types of ideas for improving CI procedure performance. In Section 4, we propose four new procedures. Section 5 is the summary, conclusions, and future research.

2 Literature Review

We first review the standard CI procedure and show empirically that a commonly used rule of thumb "$n\ \min\{p, 1-p\} > 5$" for the standard CI procedure does not guarantee a good approximation of CP when n increases or p is close to 0.5. Then, we review Wilson's CI procedure. Finally, we review and expand upon estimators of p in that we derive the statistical properties including bias, variance, and mean-squared-error for Bayesian posterior-mean estimators of p. For a review of CI procedures of Bernoulli parameter p, see Clopper and Pearson (1934), Schader and Schmid (1989), Agresti and Coulli (1998), Henderson and Meyer (2001), Chew (1971), and Brow, *et al.* (2001, 2002).

2.1 The Standard Confidence Interval

The confidence interval that is often found in introductory textbooks of probability and statistics (for example, Montgomery and Runger 2002, p.266) is called the standard confidence interval,

$$\mathrm{CI_S} = (\hat{p}_\mathrm{S} - z_{\alpha/2}\hat{\sigma}_\mathrm{S},\ \hat{p}_\mathrm{S} + z_{\alpha/2}\hat{\sigma}_\mathrm{S}), \tag{2}$$

where

- the estimate of p is

$$\hat{p}_S = k/n, \tag{3}$$

where $k = \sum_{i=1}^{n} x_i$ is the number of successes;

- the estimated standard error of the point estimator of p is

$$\hat{\sigma}_S = \sqrt{\frac{\hat{p}_S(1-\hat{p}_S)}{n}}; \text{and} \tag{4}$$

- $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2) = P(Z \le 1-\alpha/2)$, where $Z \sim$ Standard Normal distribution with mean 0 and variance 1.

Many introductory text books (for example, Montgomery and Runger 2002, p.119) suggest a rule of thumb "$n\ \min\{p, 1-p\} > 5$" for CI_S to ensure that the actual coverage probability is a good approximation of the nominal coverage probability. We use three examples to show that when $n\ \min\{p, 1-p\} > 5$,

- there is no guarantee that the true coverage probabilities are close to the nominal confidence level,
- the larger sample sizes do not guarantee a larger coverage probability, and
- the coverage probabilities are not close to the nominal confidence level when p is close to 0.5.

Example 1. Table 1 lists the smallest coverage probability (SCP) when the sample size n is greater than n_0, where $\lfloor n_0 \min\{p, 1-p\}\rfloor = 5$ ($\lfloor a \rfloor$ is the smallest integer that is greater than a or equal to a). The last column is $SCP(n_0, p) = \min\{CP(n,p), n > n_0\}$. For instance, $SCP(n_0 = 10, p = 0.5) = \min\ \{CP(n, p = 0.5), n > 10\} = 0.86$, listed in the right-most column and last row. The values in Table 1 show that all SCPs are between 0.86 and 0.90, which are below the nominal confidence level 0.95.

Table 1: The smallest CP for the sample size which is greater than n_0, where "$\lfloor n_0 \min\{p, 1-p\}\rfloor = 5$". $\alpha = 0.05$.

p	0.01	0.1	0.2	0.3	0.4	0.5
n_0	500	50	25	17	13	10
$SCP(n_0, p)$	0.87	0.88	0.88	0.89	0.90	0.86

Example 2. Figure 1 plots the CPs for $p = 0.1$. The sample sizes considered are $50 \le n \le 150$ such that $n\ \min\{p, 1-p\} \ge 5$. The numbers shown in Figure 1 are the sample sizes which yield a local maximum and minimum CPs. The oscillation phenomenon is clear and this shows that the CPs do not steadily get closer to the nominal CP as n increases. For instance, the $CP(n = 147, p = 0.1)$, which is the bottom and right-most local minimum, is smaller than $CP(n = 63, p = 0.1)$, which is the top and left-most local optimum.

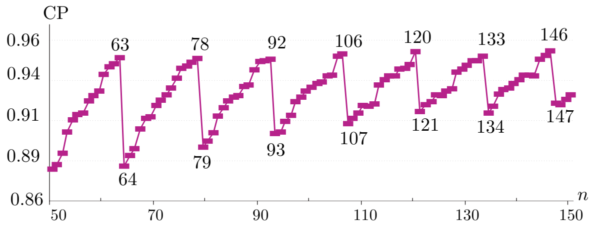

Figure 1: The coverage probabilities of CI_S, $p = 0.1$

Example 3. Figure 2 plots the CPs for $n = 100$. The probability of success considered are $0.01 \le p \le 0.5$ such that $n\ \min\{p, 1-p\} \ge 5$. The numerical numbers shown in Figure 2 are values of p which yield a local maximum and minimum CPs. Again, there is an oscillation and this suggests that CP does not steadily get closer to the nominal CP as p increases to 0.5. For instance, the CPs at $p = 0.09, 0.12, 0.17, 0.23, 0.30, 0.40$ (which are the local maximum) are larger than $CP(n = 100, p = 0.5)$, which is the bottom and right-most local minimum.

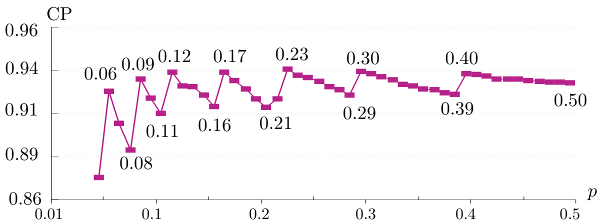

Figure 2: The coverage probabilities of CI_S, $n = 100$

The probabilities reported in Table 1 and Figures 1 and 2 and all other tables and figures throughout this paper are simulation results. In all cases their true value are within + or - two units of the last digit reported.

More examples regarding the performance of the standard CI procedure can be found in Henderson and Meyer (2001) and Brown, *et al.* (2001, 2002).

2.2 The Wilson Confidence Interval

E. B. Wilson's confidence-interval (Wilson, 1927), denoted by CI_W, has the form

$$CI_W = (\hat{p}_W - z^2_{\alpha/2}\hat{\sigma}_W,\ \hat{p}_W + z^2_{\alpha/2}\hat{\sigma}_W), \tag{5}$$

where

- $\hat{p}_W = \dfrac{k + (z^2_{\alpha/2}/2)}{n + z^2_{\alpha/2}}$ and
-

$$\hat{\sigma}_W = \sqrt{\frac{n\hat{p}_S(1-\hat{p}_S) + (z^2_{\alpha/2}/4)}{(n + z^2_{\alpha_2})^2}}, \tag{6}$$

where $\hat{p}_{\mathrm{S}} = k/n$ defined in Equation (3).

The lower and upper bounds in Wilson's interval are two roots of p for the following equation

$$\frac{\hat{p}_S - p}{\sqrt{\frac{p(1-p)}{n}}} = \pm z_{\alpha/2}. \tag{7}$$

Brown, *et al.* (2001 and 2002) surveys many existing CI procedures and conclude that Wilson's CI procedure works better for small n (say $n \le 40$) than many other CI procedures and performs as good as other CI procedures for larger n in terms of the coverage probability.

2.3 Point Estimators of p

In this section, we review, expand upon, and evaluate a number of point estimators for estimating the Bernoulli parameter. A generalization of the estimator of the Bernoulli parameter p used in many well-known CI procedures has the form

$$\hat{P} = \frac{K + r}{n + 2r}, \tag{8}$$

where $r \ge 0$. Five examples are well known.

1. The standard CI procedure uses $r = 0$. The corresponding point estimator of p is defined by $\hat{P}_{\mathrm{S}} = K/n$, already shown in Section 2.1. The K/n is the maximum likelihood estimator of p.

2. Jeffrey's CI procedure uses $r = 0.5$. See Brown *et al.*(2001).
   The corresponding point estimator of p is defined by $\hat{P}_{\mathrm{J}} = \dfrac{K + 0.5}{n + 1}$.

3. Laplace estimator uses $r = 1$. See Chew (1971).
   The corresponding point estimator of p is defined by

   $$\hat{P}_{\mathrm{L}} = \frac{K + 1}{n + 2}. \tag{9}$$

   Laplace estimate is a Bayesian estimate using Uniform(0, 1) as the prior distribution.

4. Wilson's CI procedure (1927) uses $r = z_{\alpha/2}^2\ /2$.
   The corresponding point estimator of p is defined by $\hat{P}_{\mathrm{W}} = \dfrac{K + z_{\alpha/2}^2\ /2}{n + z_{\alpha/2}^2}$.

5. Henderson and Meyers's CI procedure (2001, p. 339) uses $r = 2$.
   The corresponding point estimator of p is defined by $\hat{P}_{\mathrm{H}} = \dfrac{K + 2}{n + 4}$.

The parameter r in Equation (8) has a special meaning in Bayesian estimation. If $r > 0$, $\hat{P} = \dfrac{K + r}{n + 2r}$ is shown (for example, in Chew 1971) to be a posterior-mean Bayesian estimator if the prior probability distribution function for the parameter p has the form

$$\frac{\Gamma(2r) p^{r-1} (1-p)^{r-1}}{\Gamma(r)\Gamma(r)}. \tag{10}$$

Steinhaus (1957) shows that when $\hat{P}_{\mathrm{S}} = K/n$ which corresponding to $r = 0$ is not a posterior-mean Bayesian estimator (i.e., there does not exist a prior distribution of p such that the posterior mean reduces to K/n).

We list some statistical properties of $\hat{P}$ below.

Result 1. *The statistical properties of $\hat{P}$ in terms of the bias, variance, and mse are*

1. $\mathrm{bias}(\hat{P}) = \dfrac{r(1-2p)}{n+2r}$,
2. $\mathrm{var}(\hat{P}) = \dfrac{np(1-p)}{(n+2r)^2}$,
3. $\mathrm{mse}(\hat{P}) = \dfrac{r^2 + (n - 4r^2)p(1-p)}{(n+2r)^2}$, and
4. $\mathrm{mse}(\hat{P}) \le \mathrm{mse}(\hat{P}_{\mathrm{S}})$ if and only if

$$p \in [1/2 - w/2, 1/2 + w/2], \tag{11}$$

where $w = \sqrt{\dfrac{n+r}{n+r+nr}}$.

Consider one example for $r = 1$ and $n = 100$, Equation (11) shows that $\mathrm{mse}(\hat{P}_{\mathrm{A}}) \le \mathrm{mse}(\hat{P}_{\mathrm{S}})$ if and only if $p \in [0.145, 0.854]$.

3 Four Types of Ideas for Improving Performance

3.1 Determining Student-t Degrees of Freedom

The coverage probabilities of $\mathrm{CI_S}$ shown in Table 1, Figures 1, and 2 are all below the nominal probability 0.95. One way to improve the coverage probability for $\mathrm{CI_S}$ is to replace the $z_{\alpha/2}$ in Equation (2) by $t_{\alpha/2}(\upsilon)$ because the value of $t_{\alpha/2}(\upsilon)$ is larger than $z_{\alpha/2}$ for any value of υ. The question now is the selection of the value for υ. A natural thought of a d.f. is to use $n-1$. Define

$$\upsilon_{\mathrm{S}} = n - 1. \tag{12}$$

The value of υ_{S} defined in Equation (12) does not depend on the value of p. Because the amplitudes of oscillations (see examples shown in Figures 1 and 2) depend on both n and p, a reasonable υ_{S} that works for many n and p should also depend on both n and p. In this section, we propose a new df, which is called "the adjusted degrees of freedom", as a function of both n and p.

Let $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli (p). The adjusted degrees of freedom is defined by

$$\upsilon_{\mathrm{A}} = \frac{2n}{\frac{1}{p(1-p)} - 3 - \frac{n-3}{n-1}}, \tag{13}$$

and the corresponding t-value is defined by

$$t(v_{\mathrm{A}}) = \begin{cases} t(\lceil v_{\mathrm{A}} \rceil), & v_{\mathrm{A}} \le 1 \\ \beta\, t(\lfloor v_{\mathrm{A}} \rfloor) + (1-\beta)\, t(\lceil v_{\mathrm{A}} \rceil), & v_{\mathrm{A}} > 1 \end{cases}, \tag{14}$$

where $\beta = \lceil v_{\mathrm{A}} \rceil - v_{\mathrm{A}}$, $\lceil v_{\mathrm{A}} \rceil$ is the smallest integer greater than or equal to v_{A}, and $\lfloor v_{\mathrm{A}} \rfloor$ is the largest integer less than or equal to v_{A}. The reason of using v_{A} is given in Result 2 and the following explanation.

Result 2. *Assuming that $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli (p), and $Y_1, Y_2, \ldots, Y_{n_1} \sim$ iid Normal (μ, σ^2). Let $\hat{\sigma}^2_{\bar{X}} = S^2_X/n$ and $\hat{\sigma}^2_{\bar{Y}} = S^2_Y/n_1$ be the unbiased estimators of the variance of $\bar{X}$ and $\bar{Y}$, respectively, where S^2_X and S^2_Y are sample variance of X and Y, respectively. If the sample sizes n and n_1 satisfies the following equation*

$$n_1 = 1 + \frac{2n}{\dfrac{1}{p(1-p)} - 3 - \dfrac{n-3}{n-1}}, \tag{15}$$

then the "coefficients of variation" of $\hat{\sigma}^2_{\bar{X}}$ and $\hat{\sigma}^2_{\bar{Y}}$ are equal. That is,

$$\frac{\mathrm{std}(\hat{\sigma}^2_{\bar{X}})}{\mathrm{E}(\hat{\sigma}^2_{\bar{X}})} = \frac{\mathrm{std}(\hat{\sigma}^2_{\bar{Y}})}{\mathrm{E}(\hat{\sigma}^2_{\bar{Y}})}. \tag{16}$$

Result 2 shows that the sample size n requested for a Bernoulli distribution is equivalent to the sample size n_1 requested for a normal distribution to ensure that the same coefficients of variation of the variance of the sample mean for both distributions. This leads us to use $n_1 - 1$ which is v_{A} defined in Equation (13) as the corresponding degrees of freedom for a Bernoulli distribution.

Table 2: v_{S} and v_{A}. $\alpha = 0.05$.

n	2	5	30	50	100	300
$v_{\mathrm{S}} = n-1$	1	4	29	49	99	299
$v_{\mathrm{A}}(p=0.01)$	0.04	0.1	0.6	1.0	2.1	6.2
$v_{\mathrm{A}}(p=0.1)$	0.4	1.3	8.4	14.0	28.0	84.3
$t(v_{\mathrm{S}})$	12.7	2.8	2.1	2.0	1.98	1.96
$t(v_{\mathrm{A}})(p=0.01)$	12.7	12.7	12.7	12.7	4.2	2.4
$t(v_{\mathrm{A}})(p=0.1)$	12.7	10.2	2.3	2.2	2.1	2.0

The comparison of v_{S} and v_{A} are listed in Table 2. The value of v_{A} increases as p increases. When $n = 2$, the $t(v_{\mathrm{S}})$ and $t(v_{\mathrm{A}})$ are identical, which is 12.7. For all cases of $n > 2$ and p, v_{S} is larger than v_{A}. When $n = 5$, the $t(v_{\mathrm{S}}) = 2.8$, while $t(v_{\mathrm{A}})$ is 12.7 and 10.2 for $p = 0.01$ and $p = 0.1$, respectively. The difference of $t(v_{\mathrm{S}})$ and $t(v_{\mathrm{A}})$ becomes smaller when the sample sizes get closer to 300. For instance, when $n = 100$, $t(v_{\mathrm{S}}) = 1.98$ and $t(v_{\mathrm{A}})(p = 0.01) = 4.2$; and when $n = 300$, $t(v_{\mathrm{S}}) = 1.96$ and $t(v_{\mathrm{A}})(p = 0.01) = 2.4$.

3.2 Estimating the Point Estimator's Standard Error

A desirable property of an estimator is unbiasedness. In Result 3 we propose an unbiased estimator $\hat{\mathrm{V}}(\hat{P})$ for $\mathrm{var}(\hat{P})$, where $\hat{P} = \dfrac{K+r}{n+2r}$ defined in Equation (8). The proof of Result 3 is straightforward and therefore is skipped here.

Result 3. *Assume that $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli (p) with the number of successes $K = \sum_{i=1}^n X_i$. Let $\hat{P} = \dfrac{K+r}{n+2r}$. Then,*

$$\hat{\mathrm{V}}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{(n-1) + \dfrac{r(n+r)}{n\hat{P}(1-\hat{P})}} \tag{17}$$

is an unbiased estimator of $\mathrm{V}(\hat{P})$ (i.e., $\mathrm{E}(\hat{\mathrm{V}}(\hat{P})) = \mathrm{V}(\hat{P})$.)

Based on Result 3, the corresponding three unbiased estimates of $\mathrm{V}(\hat{P})$ for $r = 0, r = 0.5,$ and $r = 1$ are

- $\dfrac{\hat{p}_{\mathrm{S}}(1-\hat{p}_{\mathrm{S}})}{n-1}$,
- $\dfrac{\hat{p}_{\mathrm{W}}(1-\hat{p}_{\mathrm{W}})}{(n-1) + \dfrac{0.5z^2_{\alpha/2}(n+0.5z_{\alpha/2})}{n\hat{p}_{\mathrm{W}}(1-\hat{p}_{\mathrm{W}})}}$, and
- $$\frac{\hat{p}(1-\hat{p}_{\mathrm{L}})}{(n-1) + \dfrac{n+1}{n\hat{p}_{\mathrm{L}}(1-\hat{p}_{\mathrm{L}})}}. \tag{18}$$

The estimate of $\hat{P}_{\mathrm{S}}$ defined in Equation (4) for the standard CI procedure having a denominator n instead of $n-1$ is not an unbiased estimate of $\mathrm{V}(\hat{P}_{\mathrm{S}})$. Also, the estimate of $\hat{P}_{\mathrm{W}}$ defined in Equation (6) for the Wilson's CI procedure is not an unbiased estimate of $\mathrm{V}(\hat{P}_{\mathrm{W}})$.

3.3 Randomly Perturbing Intervals

As shown earlier, the actual coverage probability for the standard CI procedure contains oscillations in the sample size n for fixed Bernoulli parameter p (and in p for fixed n.) In fact, all existing CI procedures have an oscillation phenomenon. The amplitude of the oscillation for CP can be reduced by replacing some constant parameters in a CI procedure by random variables. For example, let $\mathrm{CI_R}$ be a randomly perturbed interval of $\mathrm{CI_S}$ by replacing the sample size n in $\hat{\sigma}_{\mathrm{S}}$, defined in Equation (4) by a random variable N. That is,

$$\mathrm{CI_R} = (\hat{p}_{\mathrm{S}} - z_{\alpha/2}\hat{\sigma}_{\mathrm{S}}(N),\ \hat{p}_{\mathrm{S}} + z_{\alpha/2}\hat{\sigma}_{\mathrm{S}}(N)), \tag{19}$$

where $\hat{\sigma}_{\mathrm{S}}(N) = \sqrt{\hat{p}_s(1-\hat{p}_s)/N}$, and the random variable N follows some probability distribution, say, $N \sim$ discrete uniform$(n-d, n+d)$, where d is some constant.

3.4 Combining Multiple Intervals

If the oscillation patterns of the coverage probabilities of two confidence intervals are out of phase, then the amplitude of the oscillation can be reduced by using mixtures or linear combinations of these two intervals. In this section, we define mixtures or linear combinations of finite multiple intervals. The idea of combining intervals can be extended to an infinite numbers of intervals.

3.4.1 Mixtures of intervals

Let CI_{M} be a finite mixture of g confidence intervals,

$$\mathrm{CI}_{\mathrm{M}} = \begin{cases} \mathrm{CI}_1, & \text{with probability } c_1 \\ \mathrm{CI}_2, & \text{with probability } c_2 \\ \vdots & \\ \mathrm{CI}_g, & \text{with probability } c_g, \end{cases} \tag{20}$$

where $c_1, c_2, \ldots, c_g$ are the mixture probabilities and $\sum_{i=1}^{g} c_i = 1$.

Some mixture CIs are also special cases of randomly perturbed intervals discussed in Section 3.3. Consider an example of CI_{M} in Equation (20) by choosing

$$\mathrm{CI}_i = (\hat{p}_{\mathrm{S}} - z_{\alpha/2}\hat{\sigma}_{\mathrm{S}}(n_i),\ \hat{p}_{\mathrm{S}} + z_{\alpha/2}\hat{\sigma}_{\mathrm{S}}(n_i)),$$

where $\hat{\sigma}_{\mathrm{S}}(n_i) = \sqrt{\dfrac{\hat{p}_{\mathrm{s}}(1-\hat{p}_{\mathrm{s}})}{n_i}}$, $i = 1, 2, \ldots, g$. The corresponding mixture interval can be written as the form in Equation (19), where the random variable N follows a discrete uniform $(n_1, n_2, \ldots, n_g)$.

3.4.2 Linear combinations of multiple intervals

Let $\mathrm{CI}_{\mathrm{L}} = (l_{\mathrm{L}}, u_{\mathrm{L}})$ be a linear combination of g confidence intervals $\mathrm{CI}_i = (l_i, u_i), i = 1, 2, \ldots, g$,

$$(l_{\mathrm{L}}, u_{\mathrm{L}}) = (\sum_{i=1}^{g} c_i l_i, \sum_{i=1}^{g} c_i u_i), \tag{21}$$

where $c_1, c_2, \ldots, c_g$ are the combination coefficients.

4 Four Example Procedures

4.1 Procedure Definitions

We propose four example CI procedures in this section. They are called Procedures A, R, M, and L; their corresponding confidence intervals are denoted by CI_{A}, CI_{R}, CI_{M}, and CI_{L}. Each example procedure utilizes some or all ideas proposed in Section 3. After we define these procedures in Section 4.1.1 to 4.1.4, we illustrate their performance with Monte Carlo experiments in terms of the three criteria discussed in Section 4.2.

4.1.1 Procedure A: adjusted standard

The adjusted standard CI procedure is defined to be

$$\mathrm{CI}_{\mathrm{A}} = (\hat{p}_{\mathrm{L}} - t(\upsilon_{\mathrm{A}})\hat{\sigma}_{\mathrm{A}},\ \hat{p}_{\mathrm{L}} + t(\upsilon_{\mathrm{A}})\hat{\sigma}_{\mathrm{A}}),$$

where

- $\hat{p}_{\mathrm{L}} = (k+1)/(n+2)$, which is Laplace estimator defined in Equation (9),
- $\upsilon_{\mathrm{A}} = \dfrac{2n}{\dfrac{1}{\hat{p}_{\mathrm{S}}(1-\hat{p}_{\mathrm{S}})} - 3 - \dfrac{n-3}{n-1}}$, defined in Equation(13), and
- $\hat{\sigma}_{\mathrm{A}} = \sqrt{\dfrac{\hat{p}_{\mathrm{L}}(1-\hat{p}_{\mathrm{L}})}{(n-1) + \dfrac{n+1}{n\hat{p}_{\mathrm{L}}(1-\hat{p}_{\mathrm{L}})}}}$, defined in Equation (18).

4.1.2 Procedure R: randomly perturbed adjusted standard

The randomly perturbed adjusted standard CI procedure is defined to be

$$\mathrm{CI}_{\mathrm{R}} = (\hat{p}_{\mathrm{A}} - t(\upsilon_{\mathrm{A}})\hat{\sigma}_{\mathrm{R}},\ \hat{p}_{\mathrm{A}} + t(\upsilon_{\mathrm{A}})\hat{\sigma}_{\mathrm{R}})$$

where

- $\hat{\sigma}_{\mathrm{R}} = \sqrt{\dfrac{\hat{p}_{\mathrm{A}}(1-\hat{p}_{\mathrm{A}})}{(N-1) + \frac{N+1}{N\hat{p}_{\mathrm{A}}(1-\hat{p}_{\mathrm{A}})}}}$, where $N \sim$ discrete uniform $[n - 0.2n, n + 0.2n]$.

4.1.3 Procedure M: a mixture of W and A

Figure 3 shows that the oscillation patters of CI_{W} and CI_{A} is out of phase. Define the mixture interval CI_{M} of CI_{A} and CI_{W} to be

$$\mathrm{CI}_{\mathrm{M}} = \begin{cases} \mathrm{CI}_{\mathrm{A}}, & \text{with probability } 0.5 \\ \mathrm{CI}_{\mathrm{W}}, & \text{with probability } 0.5 \end{cases},$$

where CI_{A} and CI_{W} are defined in Sections 2 and 5, respectively.

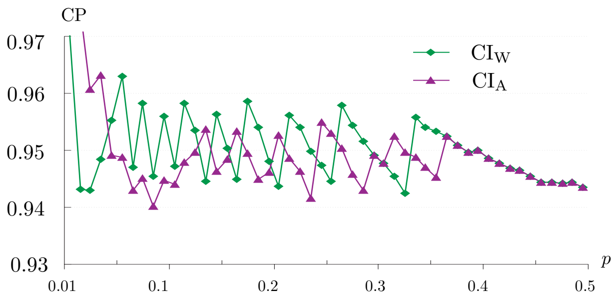

Figure 3: Oscillation patterns, $n = 300$

4.1.4 Procedure L: linear combination of W and A

We linearly combine $\mathrm{CI}_{\mathrm{W}} = (l_{\mathrm{W}}, u_{\mathrm{W}})$ and $\mathrm{CI}_{\mathrm{A}} = (l_{\mathrm{A}}, u_{\mathrm{A}})$ to be $\mathrm{CI}_{\mathrm{L}} = (l_{\mathrm{L}}, u_{\mathrm{L}})$, where

$$(l_{\mathrm{L}}, u_{\mathrm{L}}) = (0.5 l_{\mathrm{A}} + 0.5 l_{\mathrm{W}},\ 0.5 u_{\mathrm{A}} + 0.5 u_{\mathrm{W}}). \tag{22}$$

4.2 Monte Carlo Results

We adopt three criteria for comparison of different CI procedures: (i) coverage probability (CP), (ii) the standardized interval width, which is the expected interval width divided by $\sqrt{p(1-p)/n}$, and (iii) the estimated standard deviation of the interval width, which is the estimated deviation of the interval width divided by $\sqrt{p(1-p)/n}$.

4.2.1 Coverage probabilities

In this section, we illustrate the improved performance of the four example procedures proposed in Section 4 with simulation experiments. In all illustrations, the nominal coverage probability is 0.95.

We compare the six Procedures S, W, A, R, M, and L in terms of the smallest sample size n for which the coverage probability stays above 0.92. Let $n_{\rm S}, n_{\rm W}, n_{\rm A}, n_{\rm R}, n_{\rm M}$, and $n_{\rm L}$ denote the required samples sizes for Procedures S, W, A, R, M, and L, respectively, to guarantee 92% coverage probability. Table 3 shows that the values of $n_{\rm S}$ are the largest among all required sample sizes for the six procedures studied here. When $p = 0.01$, it takes 1503 samples for Procedure S to ensure that the coverage probability is at least 0.92. The value of $n_{\rm W}$, shown in the third row, is 103 when $p = 0.01$; and is smaller or equal to 16 when $0.1 \le p \le 0.5$. The required samples sizes for the proposed Procedures A, R, M, and L are less than 41 (occured at $p = 0.5$) to guarantee 92% coverage probability for all p. The values of $n_{\rm A}, n_{\rm L}, n_{\rm M}$, and $n_{\rm R}$ are the same when $p = 0.01, 0.1$, and 0.2. Procedure R is the obvious choice of interval for all cases except when $p = 0.5$.

Table 3: The smallest n for which the coverage probability stays above 0.92 ($\alpha = 0.05$)

p	0.01	0.10	0.20	0.30	0.40	0.45	0.50
$n_{\rm S}$	1503	136	80	45	30	35	41
$n_{\rm W}$	103	6	2	16	7	10	5
$n_{\rm A}$	2	2	2	16	11	24	41
$n_{\rm R}$	2	2	2	2	2	2	16
$n_{\rm M}$	2	2	2	16	2	10	2
$n_{\rm L}$	2	2	2	16	2	10	14

Figure 4 plots the CP for the six procedures when $n = 300$ and $0.01 \le p \le 0.5$. Figure 5 plots the CP for the six Procedures when $p = 0.2$ and $2 \le n \le 400$. We have the following observations from both figures:

- Most of the coverage probabilities for Procedure S are underestimated (i.e., smaller than 0.95).
- The coverage probabilities for all six procedures except Procedure S oscillate around the nominal probability 0.95.
- The proposed Procedures R, M, and L have smaller amplitudes of oscillation than those for the existing Procedures S and W.
- Among the six CI procedures, Procedure R has the smallest amplitudes of oscillation.

4.2.2 The expected half width

We use the estimated standardized expected half width, $\hat{\rm E}(\text{width})/\sqrt{p(1-p)/n}$, instead of $\hat{\rm E}(\text{width})$ as an criterion because the expected half width of CI is proportional to $\sqrt{p(1-p)/n}$. Figure 6 shows $\hat{\rm E}(\text{width})/\sqrt{p(1-p)/n}$ for $n = 300$ and $0.01 \le p \le 0.5$ and Figure 7 shows $\hat{\rm E}(\text{width})/\sqrt{p(1-p)/n}$ for $p = 0.2$ and $2 \le n \le 400$. Both Figures 6 and 7 show that $\text{CI}_{\rm S}$ has shorter expected half width when $p \le 0.05$ or $n \le 50$. The differences of the estimated standardized expected half width for the rest of five procedures are negligible.

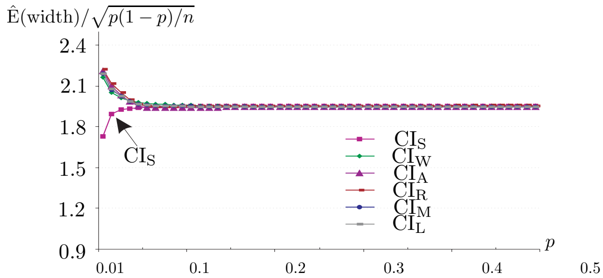

Figure 6: The estimated standardized interval width,$n = 300$

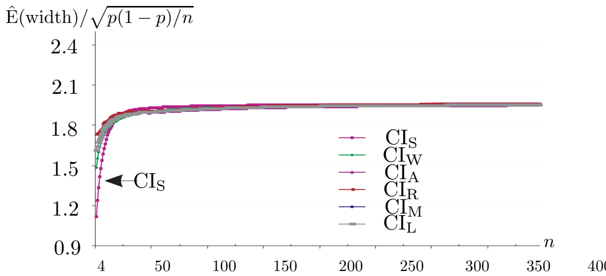

Figure 7: The estimated standardized interval width, $p = 0.2$

4.2.3 The standard deviation of the half width

The standardized standard deviation of the half width is denoted by

$$\widehat{\rm std}(\text{width})/\sqrt{p(1-p)/n}$$

, shown in Figure 8 ($n = 300$ and $0.01 \le p \le 0.5$) and 9 ($p = 0.2$ and $2 \le n \le 400$). In Figure 8, we see that the $\text{CI}_{\rm S}$ has a larger standard deviation of the half width than other CIs when $p \le 0.05$, while the $\text{CI}_{\rm R}$ has a larger standard deviation of the half width than other CIs when $p \ge 0.05$. In Figure 9, we see that the $\text{CI}_{\rm S}$ has a larger standard deviation of the half width than other CIs when

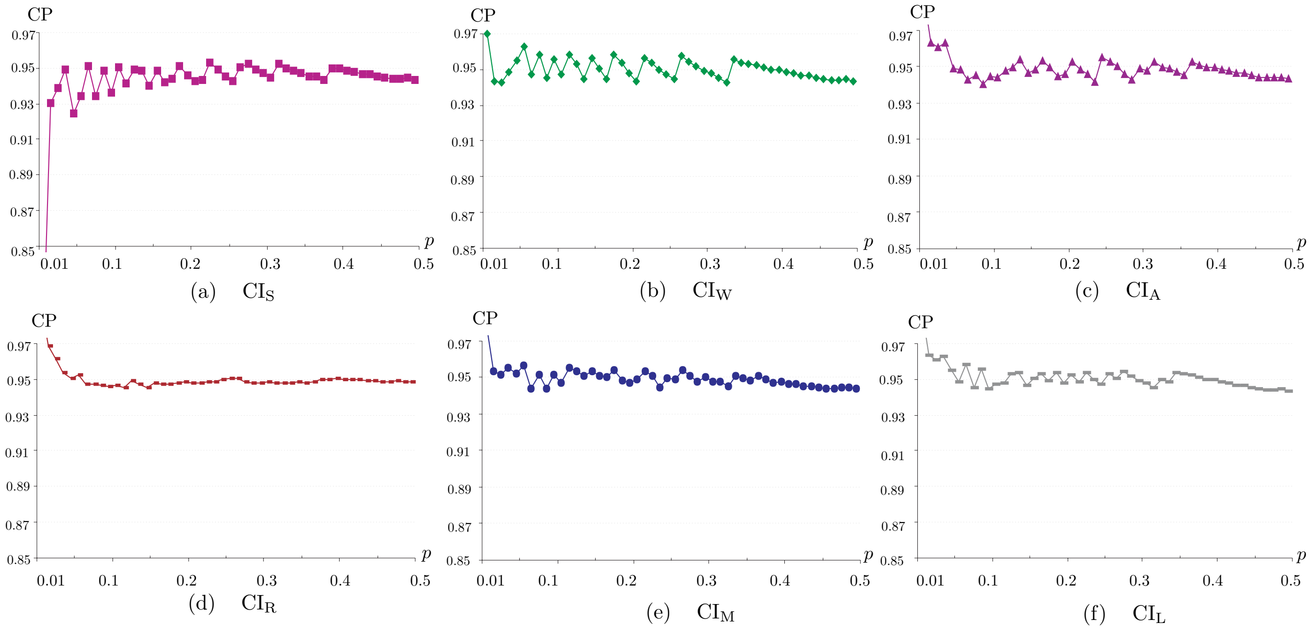

Figure 4: The coverage probabilities, $n = 300$

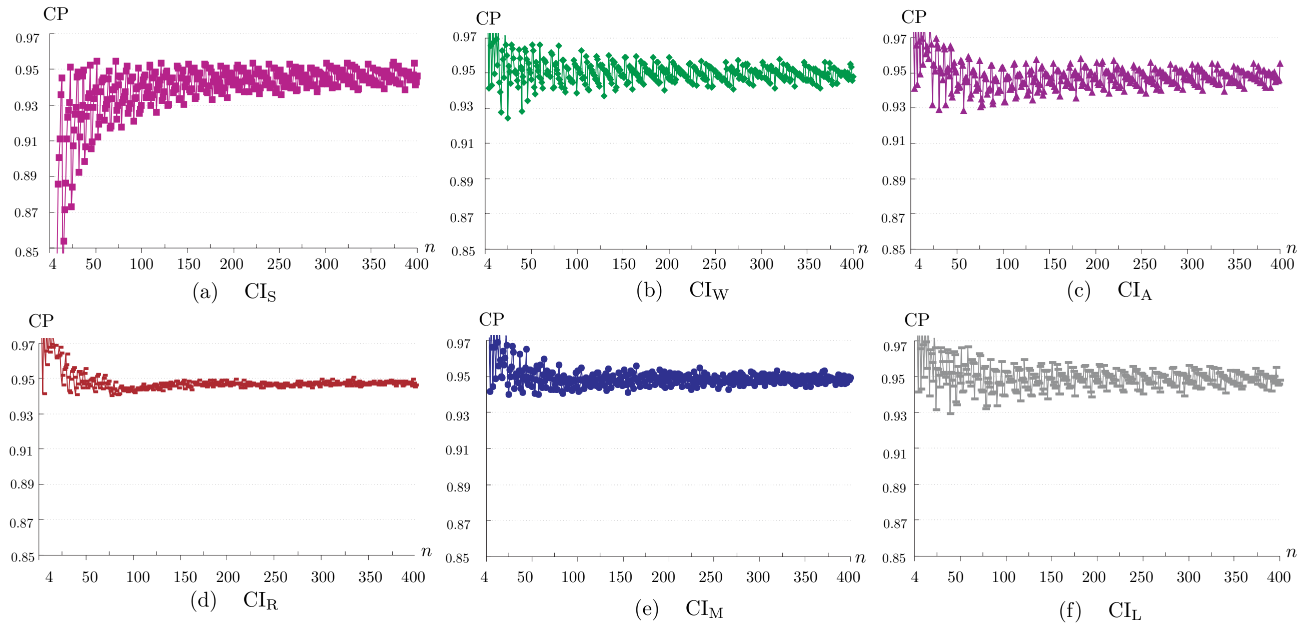

Figure 5: The coverage probabilities, $p = 0.2$

$n \leq 50$, while the $\mathrm{CI_R}$ has a larger standard deviation of the half width than other CIs when $n \geq 50$. The differences of the estimated standardized deviation of the half width for $\mathrm{CI_A}, \mathrm{CI_W}, \mathrm{CI_M}$, and $\mathrm{CI_L}$ are negligible.

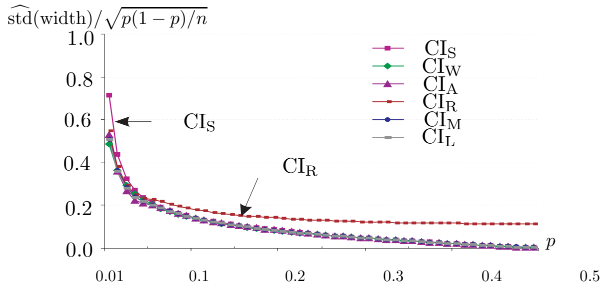

Figure 8: The standardized standard deviation of the interval width, $n = 300$

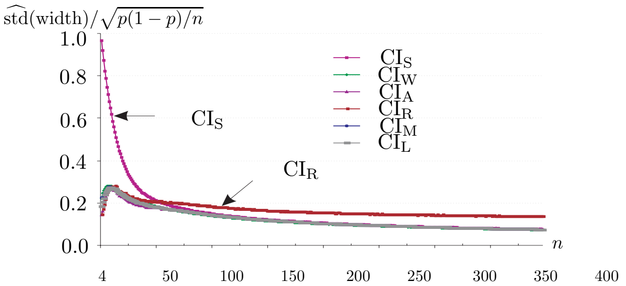

Figure 9: The standardized standard deviation of the interval width, $p = 0.2$

5 Summary, Conclusions, and Future Research

We address the classic problem of constructing confidence intervals of the Bernoulli parameter. The goal is to have the actual coverage probability close to the nominal level $1-\alpha$, a narrow expected interval width, and a small standard deviation of the interval width.

The standard confidence interval, introduced in most probability and statistics introductory text books tends to have much smaller coverage probabilities than the nominal coverage probability. Many CI procedures such as Wilson's interval overcome this under-coverage shortcoming of the standard interval procedure. Such CI procedures, however, do not overcome the oscillation phenomenon on coverage probability, which is due to the effect of the discreteness of the Bernoulli data. By reducing the amplitudes of certain oscillation, the quality of the coverage probability is improved.

We suggest four ideas: (i) improve the standard-error estimator, (ii) improve the choice of the Student-t, (iii) randomly perturb the interval, and (iv) combine intervals. We propose four example Procedures, A, R, M, and L, to implement the respective ideas. In comparison with two well-known CI procedures, the standard S and Wilson's CI procedures, the proposed Procedures M and L have smaller amplitudes of oscillation of coverage probability than Procedures S and W, while having negligible effects on the standard deviation of the interval width. Procedure R seems to perform the best among the four proposed procedures in the sense that it achieves the nominal coverage probability; and this has the smallest amplitudes of oscillation among the six CI procedures (at the price of a small increase in the variance of the interval width).

Whether Procedures R, L, and M achieve similar results for other α, p, and n remains to be investigated. All that can be claimed with certainty is that Procedures R, L, and M have demonstrably large potential to perform better than the standard confidence interval and Wilson's interval in terms of the three criteria. This study lays the groundwork for more-general methods of combining confidence intervals.

Acknowledgments

This paper is based upon work supported by the National Science Council under Grant No. NSC-93-2213-E-007-060. The authors deeply thank Professor Bruce Schmeiser for spending a significant amount of time in discussion of the technical content of this paper. We also thank Professors Dennis Engi, David Goldsman and Raghu Pasupathy for their perspective comments.

References

Agresti, A. and Coulli, Brent A. (1998), "pproximate is better than "exact" for interval estimation of binomial proportions," *The American Statistician*, **52**, 119-126.

Brown, L. D., Cai, T. and Dasgupta, A. (2001),"Interval estimation for a binomial proportion (with discussion)," *Statist. Sci*, **16**, 101-133.

Brown, L. D., Cai, T. and Dasgupta, A. (2002), "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions," *The Annals of Statistics* ,**30**, 160-201.

Chew, V. (1971),"Point estimation of the Parameter of the Binomial Distribution," *The American Statistician*, **25**, 47-50.

Clopper, C. J. and Pearson, E. S. (1934), "The use of confidence of fiducial limits illustration in the case of the binomial,"*Biometrika*, **26**, 404-413.

Henderson, M. and Meyer, M. C. (2001), "Exploring the confidence interval for a binomial parameter in a first course in statistical computing," *The American Statistician*, **55**, 337-344.

Montgomery, D. C. and Runger, George C. (2002), *Applied statistics and probability for engineers*, third eds.

Schader, M. and Schmid, F. (1989), "Two rules of thumb for the approximation of the binomial distribution by the Normal distribution," *The American Statistician*, **43**, 23-24.

Steinhaus, H. (1957), "The problem of estimation," *The Annals of Mathematical Statistics*, **28**, 633-648.

Wilks, S. S. (1962) *Mathematical Statistics*, Princeton University Press, Princeton, N. J, 200.

Wilson, E. B. (1927), "Probable inference, the law of succession, and statistical inference,"*J. Ame. Statist. Assoc.*, **22**, 209-212.