PHP 2515 Final Project: Interval Estimation for a Binomial Proportion

Peirong Hao, Cici Huang, William Qian

July 29, 2024

1 Introduction

In statistical analysis, estimating the interval of a binomial proportion has great importance, particularly in fields ranging from clinical trials to market research. In practice, we may encounter different estimation methods. We compare two different interval estimators, the Wald interval and the Agresti-Coull interval, for an unknown binomial proportion in this report. By doing simulations, we obtained the performance of those two estimators by comparing them to the nominal/stated coverage of 95% and found the Agresti-Coull interval is better than the Wald interval. We analyzed the reasons for that result from a Bayesian interpretation.

2 Theoretical Background

2.1 Confidence Interval

Confidence intervals are a critical concept in statistics, offering a range of values, often calculated from sample data, that are likely to contain an unknown population parameter. Unlike a single point estimate, a confidence interval provides a range that reflects the uncertainty inherent in sample data. Typically expressed at a 95% confidence level, it implies that if we repeatedly sample from the same population, approximately 95% of the confidence intervals calculated from those samples will contain the true parameter value. This tool is invaluable in hypothesis testing and data analysis, allowing researchers to assess the reliability and precision of their estimates.

Consider observing n iid realizations from the following data generating process (DGP)

$$X_1, X_2, ..., X_n \overset{iid}{\sim} Ber(p)$$

Consider estimator $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. To from an interval estimator, recall that $\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{Var[X]}} \overset{A}{\sim} N(0,1)$ by the CLT. Thus, asymptotically $P\{Z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{Var[X]}} \leq Z_{1-1/\alpha}\} = 1-\alpha$. Here, z_v is the $v-$th percentile of the standard normal. Solving the inner inequality for p yields a $(1-\alpha)100\%$ confidence interval p.

$$\begin{aligned}
P(Z_{0.025} \leq \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{Var(X)}} \leq Z_{0.975}) &= P(Z_{0.025}\sqrt{Var(X)} \leq \sqrt{n}(\hat{p}-p) \leq Z_{0.975}\sqrt{Var(X)}) \\
&= P(Z_{0.025}\sqrt{\frac{Var(X)}{n}} \leq \hat{p}-p \leq Z_{0.975}\sqrt{\frac{Var(X)}{n}} \\
&= P(-\hat{p}+Z_{0.025}\sqrt{\frac{Var(X)}{n}} \leq -p \leq -\hat{p}+Z_{0.975}\sqrt{\frac{Var(X)}{n}} \\
&= P(\hat{p}-Z_{0.975}\sqrt{\frac{Var(X)}{n}} \leq p \leq \hat{p}-Z_{0.025}\sqrt{\frac{Var(X)}{n}} \\
&= P(\hat{p}-Z_{0.975}\sqrt{\frac{Var(X)}{n}} \leq p \leq \hat{p}+Z_{0.975}\sqrt{\frac{Var(X)}{n}}
\end{aligned}$$

Method	Coverage
Wald Interval	0.8922
Agresti-Coull Interval	0.9780

Table 1: True p coverage comparison between methods.

The 95% confidence interval for p is $(\hat{p} - Z_{0.975}\sqrt{\frac{Var(X)}{n}}, \hat{p} + Z_{0.975}\sqrt{\frac{Var(X)}{n}})$, which has the form $\hat{p} \pm Z_{0.975}\sqrt{\frac{Var(X)}{n}}$.

2.2 The Wald Interval

The Wald Confidence Interval is a specific type of interval used for estimating the true value of a parameter from a statistical sample. It's most commonly used for estimating proportions in binomial distributions. The Wald interval estimator does this inversion and estimates $Var[X]$ by plugging in p wherever p appears in the expression for $Var[X]$. The CI plugged in with the Wald Interval is:

$$\hat{p} \pm Z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

2.3 The Agresti-Coull Interval

The Agresti-Coull Interval is an adjusted form of the Wald Confidence Interval, specifically designed to improve its performance, especially for small sample sizes or when the sample proportion is near 0 or 1. The adjustment involves adding "pseudocounts" to both the number of successes and failures in the sample. The Agresti-Coull interval "adds 2 success in 4 trials" and computes an interval using $\tilde{p} = \frac{1}{n}(2 + \sum_{i=1}^{n} X_i)$, where $\tilde{n} = n + 4$. The Agresti-Coull interval is given by

$$\tilde{p} \pm Z_{0.975}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

3 Simulation

To better understand and compare the output of those two interval methods, we conducted several simulations. In each of the simulations, we simulate 10000 data sets from the DGP defined above with $p = 0.5$

3.1 Definite Sample Size Simulation

First, we set the sample size, $n = 10$. For each simulated dataset, compute both the Wald interval and the Agresti-Coull interval and record whether the intervals include the true $p = 0.5$. We get the following result in Table 1. Compared to the nominal/stated coverage of 95%, the Agresti-Coull Interval seems to cover a larger region.

3.2 Varied Sample Size Simulation

In the next step, we repeat the above procedures for $n \in \{10, 12, 13, 15, 17, 18, 20, 23, 25, 28, 30, 33, 35, 37, 40, 42, 44, 49\}$ to see how do the two methods perform when sample size gets bigger. The result is shown in Table 2, and the data is plotted in Figure 1. We can see it very clearly that when the sample size (n) gets larger, the coverage of the Wald Interval increases, while the coverage of the Agresti-Coull interval gets slightly smaller and they all eventually converge to the nominal/stated coverage of 95%. Considering their performance, the Agresti-Coull interval is a better choice because its coverage is closer to the nominal/state coverage when the sample size is small. When the sample size gets larger, it can also perform well.

Sample Size(n)	Wald Interval	Agresti-Coull Interval
10	0.8949	0.9789
12	0.859	0.9667
13	0.9114	0.9788
15	0.8854	0.967
17	0.9506	0.9506
18	0.9052	0.9676
20	0.9595	0.9595
23	0.9045	0.9614
25	0.9567	0.9567
28	0.9139	0.9631
30	0.9573	0.9573
33	0.9214	0.9655
35	0.9574	0.9574
37	0.952	0.952
40	0.9178	0.9595
42	0.958	0.958
44	0.9527	0.9527
49	0.9598	0.9598

Table 2: Converge of true p under different sample size

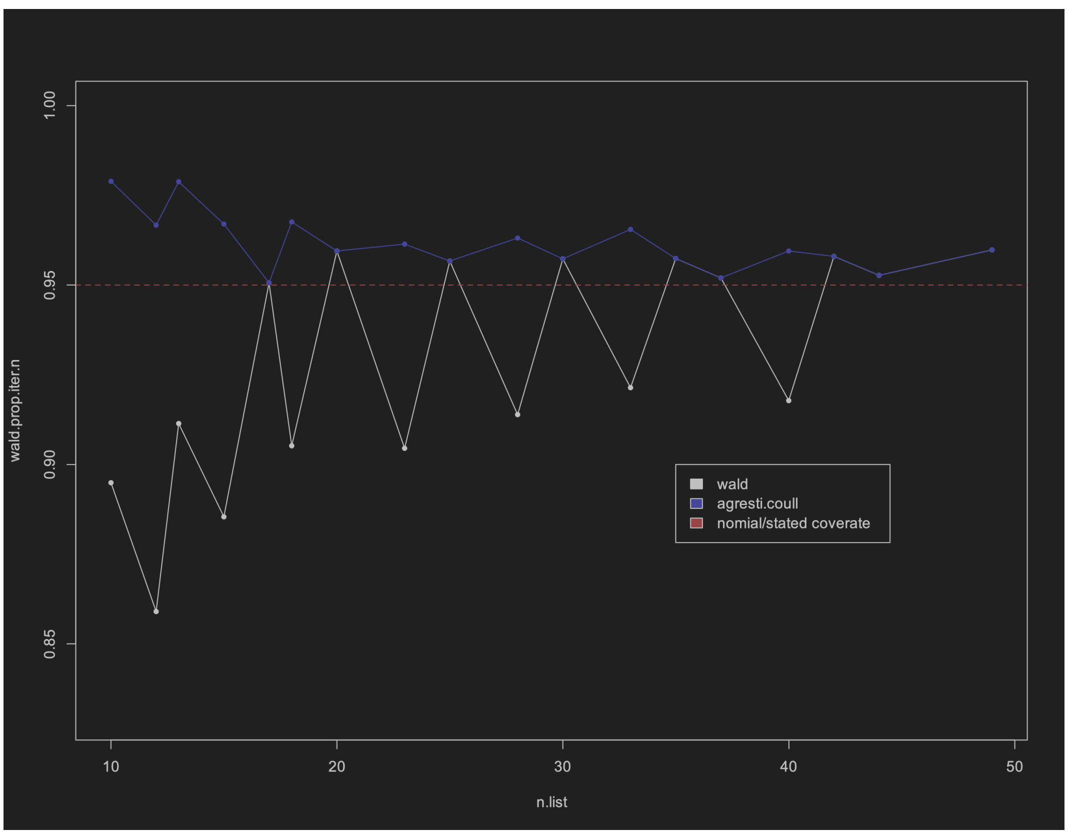

Figure 1: Coverage trend over sample size

4 Bayesian Interpretation of Agresti-Coull Interval

Adding 2 successes and 2 failures is equivalent to using a $Beta(2,2)$ prior. This $Beta(2,2)$ prior is a symmetric distribution centered at 0.5, indicating all values of p are equally likely, but with a tendency towards the middle value of 0.5. By incorporating this prior, the MLE of p is adjusted. This can be seen as "shrinking" the MLE towards the prior mean of 0.5. This shrinkage is more pronounced when n is small and becomes negligible for large n.

5 Conclusion

In our analysis, we focused on examining the performance of two statistical methods - the Wald Interval and the Agresti-Coull Interval.

For smaller sample sizes, our findings revealed a notable divergence in the performance of the two intervals. The Wald Interval's coverage was consistently below the standard 95% confidence level. This underperformance suggests a lack of reliability in this interval when dealing with small sample sizes, potentially leading to less accurate estimations or predictions.

In contrast, the Agresti-Coull Interval demonstrated a more robust performance in the same scenario. It consistently maintained coverage above the 95% threshold, indicating a higher reliability and accuracy in its estimates even with limited data. This characteristic makes the Agresti-Coull Interval a preferable choice in situations where sample sizes are small.

As we expanded our analysis to a broader range of sample sizes (various 'n'), the behavior of the Wald test displayed an erratic pattern. Its performance varied significantly with different sample sizes; in some instances (the 'lucky n'), it achieved the desired 95% coverage, but in many others (the 'not so lucky n'), it failed to do so. This inconsistency underscores the limitation of the Wald Interval in providing reliable confidence levels across different sample sizes.

On the other hand, the Agresti-Coull Interval maintained a more consistent and conservative approach across various sample sizes. It tended to produce higher confidence intervals, which, although possibly more conservative than necessary at times, ensured a more consistent adherence to the 95% confidence threshold. This consistent performance reiterates its suitability for a wider range of sample sizes, offering a dependable tool for statistical analysis that requires confidence interval estimation.

In summary, our comparative study highlights the strengths and weaknesses of the Wald and Agresti-Coull intervals in different statistical scenarios. The Agresti-Coull Interval, with its consistent performance across various sample sizes, emerges as a more reliable method for calculating confidence intervals, especially in cases where sample size varies or is limited.