Confidence Intervals for a Proportion in Finite Population Sampling

Seung-Chun Lee[1,a]

[a]Department of Statistics, Hanshin University

Abstract

Recently the interval estimation of binomial proportions is revisited in various literatures. This is mainly due to the erratic behavior of the coverage probability of the well-known Wald confidence interval. Various alternatives have been proposed. Among them, the Agresti-Coull confidence interval, the Wilson confidence interval and the Bayes confidence interval resulting from the noninformative Jefferys prior were recommended by Brown *et al.* (2001). However, unlike the binomial distribution case, little is known about the properties of the confidence intervals in finite population sampling. In this note, the property of confidence intervals is investigated in finite population sampling.

Keywords: Wald interval, Agresti-Coull interval, Wilson interval, Weighted Polya posterior interval, exact interval, combined interval, hypergeometric distribution.

1. Introduction

The Wald confidence interval has been considered as a standard method for the interval estimation of a binomial proportion. However, the erratic behavior of the coverage probability of the Wald interval has been recognized in various literatures. See for example, Blyth and Still (1983), Staner (1998), Agresti and Coull (1998), Brown *et al.* (2001, 2002) and Lee (2005). In particular Brown *et al.* (2001) investigated the unsatisfactory coverage properties of the Wald interval in details. However, it seems that some adjustments of the Wald interval are very useful. For example, Agresti and Coull (1998) showed that an improved interval for p can be obtained by adding two successes and two failures to the observed counts and then uses the standard method. They claimed that the modification performs well even for small samples. Another useful Wald-like interval is the Wilson interval discussed by Wilson (1927). It is shown by Agresti and Coull (1998) that the Wilson interval performs much better than the Clopper-Pearson exact interval or the Wald interval in terms of having coverage probabilities close to the nominal confidence level. The Wilson interval was also recommended by Brown *et al.* (2001) for use with nearly all sample sizes and parameter values. Lee (2005) developed another Wald-like confidence interval based on weighted Polya posterior. It was showed that the interval outperforms the Agresti-Coull interval and is essentially equivalent to the Wilson interval with simplicity.

There are many other confidence intervals for the binomial proportion such as the arcsine interval, the logit interval, the Jefferys prior interval (Berger, 1985) and the likelihood ratio interval. Among them, the Wilson interval and the Jefferys prior interval for small n, say less then, or equal to 40, and the Agresti-Coull interval for larger n were recommended by Brown *et al.* (2001) in terms of their

coverage properties and parsimony as measured by expected length. They claimed that for $n \geq 40$, the Wilson, the Jefferys and the Agresti-Coull intervals were all very similar but Agresti-Coull interval was recommended because of its simplicity.

Unlike the case of the binomial proportion in infinite population, little is investigated in finite population. Most literatures introduce only the Wald interval. Although Lee (2005) gave a brief summary for the problem, it is hard to find literatures describing the property of well-known confidence intervals. Since many confidence intervals such as the Wilson interval can be modified easily for the interval estimation problem in finite population sampling, it may be surprising. Also, Cai and Krishnamoorthy (2005) recently developed a simple but effective confidence interval. A summary of various interval estimators in infinite population is seemed to be necessary. In this note, I investigate the property of those intervals in terms of coverage probability and expected width in finite population sampling.

2. Confidence Intervals for a Proportion in Finite Population Sampling

Instead of the binomial distribution, the hypergeometric distribution is related to the estimation of a proportion in finite population sampling. In what follows, $hyper(x; N, n, p)$ will denote the probability mass function with N, n and p representing population size, sample size and proportion of success, respectively. That is,

$$hyper(x; N, n, p) = \begin{cases} \dfrac{\binom{n}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}, & \max(0, n - N + Np) \leq x \leq \min(n, Np), \\ 0, & \text{otherwise.} \end{cases}$$

In this scenario, N and n are known constants, and p is unknown parameter with parameter space $\mathcal{P} = \{0, 1/N, \ldots, (N-1)/N, 1\}$.

2.1. Approximate confidence intervals

2.1.1. Wald interval

For the interval estimation problem, the Wald method may be the most popular, but it seems that there are two versions of the Wald method. In elementary level textbook, the 95% Wald interval is given as:

$$\hat{p} \pm 2\sqrt{(1 - f)}\sqrt{\hat{p}\hat{q}/(n - 1)} \tag{2.1}$$

where $\hat{p} = X/n, \hat{q} = 1 - \hat{p}$ and $f = n/N$. See Yamane (1967). However, more advanced textbooks, see for example Cochran (1977), gives

$$\hat{p} \pm \left[z_{\alpha/2}\sqrt{1 - f}\sqrt{\hat{p}\hat{q}/(n - 1)} + \frac{1}{2n} \right]. \tag{2.2}$$

By taking 2 instead of $z_{0.025} \approx 1.96$ and the continuity correction, both (2.1) and (2.2) are wider and provide more proper coverage probabilities than usual Wald interval when sample size is small. Although by taking wider width, the coverage property is improved, the main problem of the Wald interval is not its width. In fact, the Wald interval tends to have wider width than its rival intervals, but to give poor coverage probability.

Among the two Wald intervals, the coverage probability of (2.2) is close to the nominal level than (2.1). Thus, in what follows, (2.2) will represent the Wald interval.

2.1.2. Wilson interval

The Wilson interval is the set of p satisfying

$$\left| (\hat{p} - p) / \sqrt{\frac{N-n}{N-1} \frac{p(1-p)}{n}} \right| \leq z_{\alpha/2}.$$

Writing $f^* = (n-1)/(N-1)$, the confidence interval can be written as:

$$\tilde{p}^* \pm z_{\alpha/2} \frac{\sqrt{1-f^*}}{\tilde{n}^*} \sqrt{n\hat{p}\hat{q} + (1-f^*)z_{\alpha/2}^2/4} \tag{2.3}$$

where $\tilde{n}^* = n + (1-f^*)z_{\alpha/2}^2$ and $\tilde{p}^* = (X + (1-f)z_{\alpha/2}^2/2)/\tilde{n}^*$.

Note that, putting $f^* = 0$, (2.3) is the Wilson confidence interval for p under the binomial distribution. It was numerically shown by Brown $et\ al.$ (2001) that the Wilson interval behaves quiet well in terms of the coverage probability and the expected width. However, the first term in the last square root is getting to zero as $\hat{p}$ goes to 0 or 1. Thus, the interval width is getting smaller. Although $z_{\alpha/2}^2/4$ prevent the interval width from being zero as the continuity correction in (2.2), it tends to have short width than necessary when p is near 0 or 1, and as a result, there are big down-spikes in its coverage probability plots. See also Figure 1. You may refer Lee (2005) to the same phenomenon in the binomial distribution case.

2.1.3. Agresti-Coull interval

Agresti and Coull (1998) developed a new confidence interval for a binomial proportion by so-called "adding two successes and two failures" to the observed count. Presumably the strategy will work for the finite population sampling as well.

Under our circumstance, the Agresti-Coull interval would be

$$\tilde{p} \pm z_{\alpha/2} \sqrt{1-\tilde{f}} \sqrt{\tilde{p}\tilde{q}/(\tilde{n}-1)} \tag{2.4}$$

where $\tilde{n} = n + 4, \tilde{p} = (X+2)/\tilde{n}$ and $\tilde{f} = \tilde{n}/N$. Thus the formula is exactly same as (2.2) except the continuity correction. Since my previous numerical studies showed that the Agresti-Coull interval with the continuity correction is too conservative, it is omitted.

Four artificial observations (two successes and two failures) prevent $\tilde{p}\tilde{q}/(\tilde{n}-1)$ from being zero, and there is no down-spike in its coverage probability plots. In fact, the Agresti-Coull interval is little bit conservative. Note that, Agresti and Coull (1998) justified the Agresti-Coull interval by Bayesian prospect, but the variance of $\tilde{p}$ is slightly overestimated in a Bayesian paradigm. The conservatism may be due to the overestimation. However, the Agresti-Coull interval was recommended by Brown $et\ al.$ (2001) for large sample sizes, say $n \geq 40$, for estimating a proportion of the binomial distribution.

2.1.4. Weighted Polya posterior interval

The weighted Polya posterior interval was developed by Lee (2005) using weighted Polya posterior. Originally the weighted Polya posterior was developed by Meeden (1999) to solve statistical problems in skewed population and small sample size. Since in the interval estimation of a proportion, most problems occur when sample size is small and p is near 0 or 1, his approach is applicable.

When underlying distribution is binomial, the weighted Polya posterior can be essentially obtained by a conjugate beta prior distribution with mean $1/2$ and variance $1/4(2\xi+1)$ where ξ is the weight. In
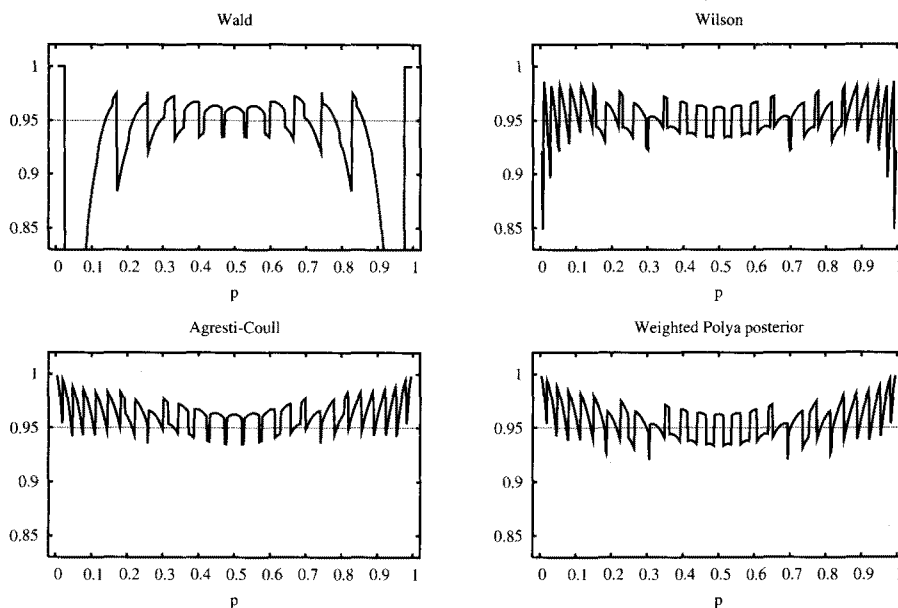
Figure 1: *Coverage probabilities of 4 approximate 95% confidence intervals when N = 500 and n = 20.*

finite population sampling, the weighted Polya posterior is not an known distribution, but the posterior mean and variance of p are known as (Lee, 2005)

$$\bar{p} = \hat{p} + 2(1 - f)\frac{\xi}{(n + 2\xi)}\left(\frac{1}{2} - \hat{p}\right) \quad \text{and} \quad (1 - f)\frac{N + 2\xi}{N}\frac{1}{n + 2\xi + 1}\frac{X + \xi}{n + 2\xi}\left(1 - \frac{X + \xi}{n + 2\xi}\right).$$

Note that when $\xi = 0$, the conditional mean is just $\hat{p}$ and variance is $(1 - f)\hat{p}(1 - \hat{p})/(n - 1)[(n - 1)/(n + 1)]$. Thus, the variance is the same as the usual frequentist formulation of variance except the factor $(n - 1)/(n + 1)$. If we construct a confidence interval using Wald procedure with $\hat{p}$ and $(1 - f)\hat{p}(1 - \hat{p})/(n - 1)$, it would be wider than a proper interval in the Bayesian point of view except that the sample size is large enough. Thus it is desirable to set ξ as an appropriate nonzero value. For the binomial distribution case, $\xi = z_{\alpha/2}^2/2$ was chosen. Taking the same value of ξ, and considering $(N + z_{\alpha/2}^2)/N \approx 1$, the weighted Polya posterior interval is constructed as:

$$\bar{p} \pm z_{\alpha/2}\sqrt{1 - f}\sqrt{\frac{X + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2}\left(1 - \frac{X + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2}\right)\bigg/(n + z_{\alpha/2}^2 + 1)}. \tag{2.5}$$

Let $CI_A(X)$ be a confidence interval due to method A. An usual criterion for judging an approximate method A is its coverage probability which is defined as

$$C_A(p) = \sum_{x \in X} I\left(p \in CI_A(x)\right) hyper(x; N, n, p) \tag{2.6}$$

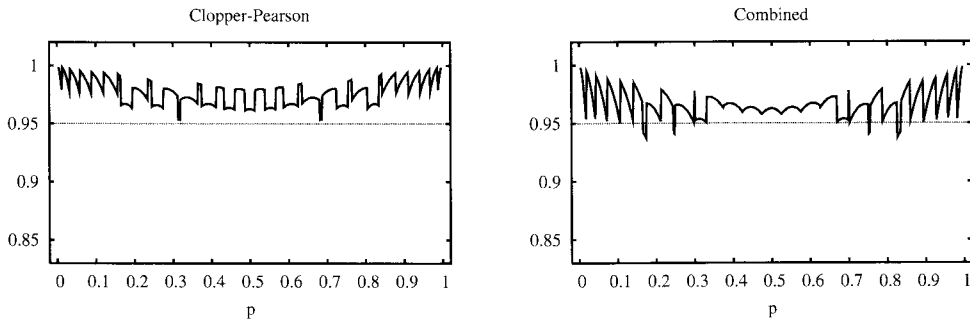where I is the usual indicator function.

Figure 2: *Coverage probabilities of 2 exact 95% confidence intervals when $N = 500$ and $n = 20$.*

The general patterns of coverage probability of 4 approximate confidence intervals are shown in Figure 1. Except the Wald interval, the nominal level is well approximated by the coverage probabilities. Also comparing Figure 1 of Lee (2005), it can be observed that the confidence intervals behave very similarly to the corresponding intervals in binomial distribution case. For instance, the Agresti-Coull interval is more conservative than the Wilson and the weighted Polya posterior, and the Wald interval should be used with great care when sample size is small. I examined the coverage patterns of various population sizes N up to 2000 but N did not play much role in those patterns. See Figure 3.

2.2. Exact confidence intervals

Recently Cai and Krishnamoorthy (2005) developed a confidence interval by combining an exact test and a test considered in Kulkarni and Shah (1995) and Krishnamoorthy et al. (2002). Using the notation of Cai and Krishnamoorthy (2005), an obvious exact confidence interval, due to Clopper and Pearson (1936), can be obtained by considering testing hypotheses $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$. That is, for a given α, and observed value x, an exact test reject H_0 when the p-value

$$P_e(p_0) = 2 \min \{ \Pr[X \leq x \mid p_0], \Pr[X \geq x \mid p_0] \}$$

is less or equal to α, while an exact confidence interval is the set $\{p : Pe(p) > \alpha\}$. It is well-known that the Clopper-Pearson exact confidence interval is too conservative when underlying distribution is discrete. To adjust the conservatism, Cai and Krishnamoorthy (2005) applied another test which reject H_0 when

$$P_a(p_0) = \Pr \left[(X - np_0)^2 > (x - np_0)^2 \mid p_0 \right].$$

is less than or equal to α. A new combined test rejects the null hypothesis when both tests reject. Thus, a new combined interval can be constructed as a set of p satisfying

$$\min \left(P_e(p), P_a(p) \right) > \alpha. \tag{2.7}$$

Cai and Krishnamoorthy (2005) noted the combined one is not exact. However, based on my numerical study, the combined one was near exact, and hence it is classified as an exact confidence interval in this paper. The performance of combined interval is even better than the separated ones.

The coverage patterns of the Clopper-Pearson interval and the combined interval are shown in Figure 2. As one can expect, both intervals are conservative, but the combined interval behaves reasonably compared with the approximate intervals.

Table 1: Six 95% confidence intervals for p when $N = 500, n = 100, x = 37$.

Method	Lower limit	Upper limit	Interval width
Wald(with continuity correction)	0.280	0.460	0.180
Agresti-Coull	0.292	0.458	0.166
Weighted Polya posterior	0.292	0.456	0.164
Wilson	0.292	0.456	0.164
Clopper-Pearson	0.286	0.462	0.176
Combined	0.290	0.460	0.170

Example 1. Cochran (1977) gave a simple example for the estimation of a proportion in finite population sampling. In a random sample of size 100, from a population size 500, 37 units were classified into a certain class. In this example, 95% confidence limits are calculated by the 5 methods as shown in Table 1. It can be observed that the Wald interval is widest. It is even wider than the Clopper-Pearson exact confidence interval. The weighted Polya posterior and the Wilson intervals are narrower than any other intervals.

3. Comparison of Confidence Intervals

The coverage probabilities of 95% confidence intervals with various population and sample sizes are demonstrated in Figure 3 by the Box-Percentile plot proposed by Esty and Banfield (2003). It seems that the weighted Polya posterior and the Wilson intervals approximate well the nominal level than other intervals regardless of the population and sample sizes, and the Wald interval behaves very poorly. We may exclude it from our consideration when sample size is small.

Brown *et al.* (2001) argued that "most statistical methods are only felt to be approximately valid as representations of the true situation." Hence the resulting coverage properties from those models are best only approximately accurate. The coverage should be close to the nominal confidence level rather than to guarantee it is at least $(1 - \alpha) \times 100\%$. In this prospective, a confidence interval is good, if the coverage probability is as close to the nominal level as possible. Thus, the Mean Absolute Error(MAE) or the Mean Squared Error(MSE) of the coverage probabilities could be excellent criterions for comparing confidence intervals. The Wilson interval is the best in these criterions. Note however, the weighted Polya posterior interval has competitive power in that the difference in the MAE or the MSE not large enough, and the Wilson interval has certain weaknesses. As shown in Figure 1, it has big down-spikes at each edge.

The expected width of a confidence interval A, which is defined as

$$\text{EW}_A(p) = \text{E}_p\left(W_A(X)\right) = \sum_{x \in \mathcal{X}} W_A(x) \, hyper(x; N, n, p) \tag{3.1}$$

where $W_A(x)$ is the width of confidence interval when x is observed, is another important factor for evaluating the performance of confidence intervals. Since a wide interval is noninformative, a confidence interval with small expected width is preferable. Figure 4 demonstrates the expected width of 95% confidence intervals when $N = 500$ and n up to 50. It can be observed that the Wald interval tends to be wide when p is around 0.5, even wider than the exact Copper-Pearson interval. Thus, the poor coverage probability of the Wald in that area is not due to its width. Rather, it is mainly due to the wrong center of interval. Note that the other three approximate intervals adjust the center $\hat{p}$ toward 1/2. In view of Figure 1, this kind of adjustment seems to be good. Thus, this might justify the adjustment of the Wilson, the Agresti-Coull and the weighted Polya posterior intervals. Among the three, the Agresti-Coull interval has the largest expected width except each extreme tail areas.
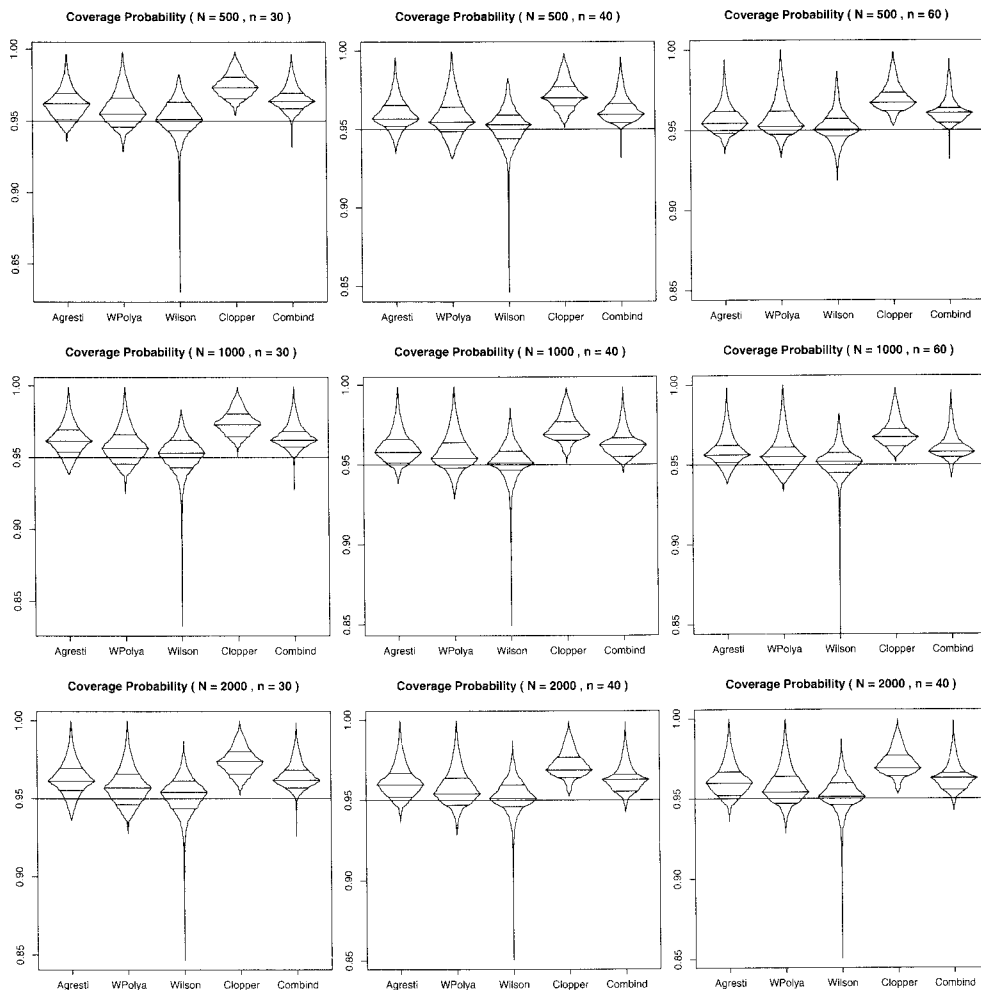
Figure 3: *The Box-percentile plot of the coverage probabilities of 95% confidence intervals.*

and hence is the most conservative. In those areas, the expected width of the Agresti-Coull interval is slightly smaller than the weighted Polya posterior, but the difference between two is negligible. However, the difference with the Wilson seems to be meaningful, because the down-spikes of the Wilson occur in those areas. It seems that the Wilson interval is too narrow in each tail area. It also seems that the Agresti-Coull and the weighted Polya posterior intervals would not have proper width in those areas when sample size small, because the expected widths of the two are larger than the near exact combined interval, and hence more conservative. On the other hand, the combined interval is much wider than the other two. Thus, none can be best, but the weighted Polya posterior and the Wilson intervals are recommendable in view of the approximation and the expected width criterions when sample size is small, say $n \leq 40$.

We have chosen two approximate confidence intervals by the two criterions. Note however, the exact interval is conservative to guarantee the minimum level of coverage probability at sacrifice of the expected width, while the approximate confidence interval oscillates the nominal level with rela-
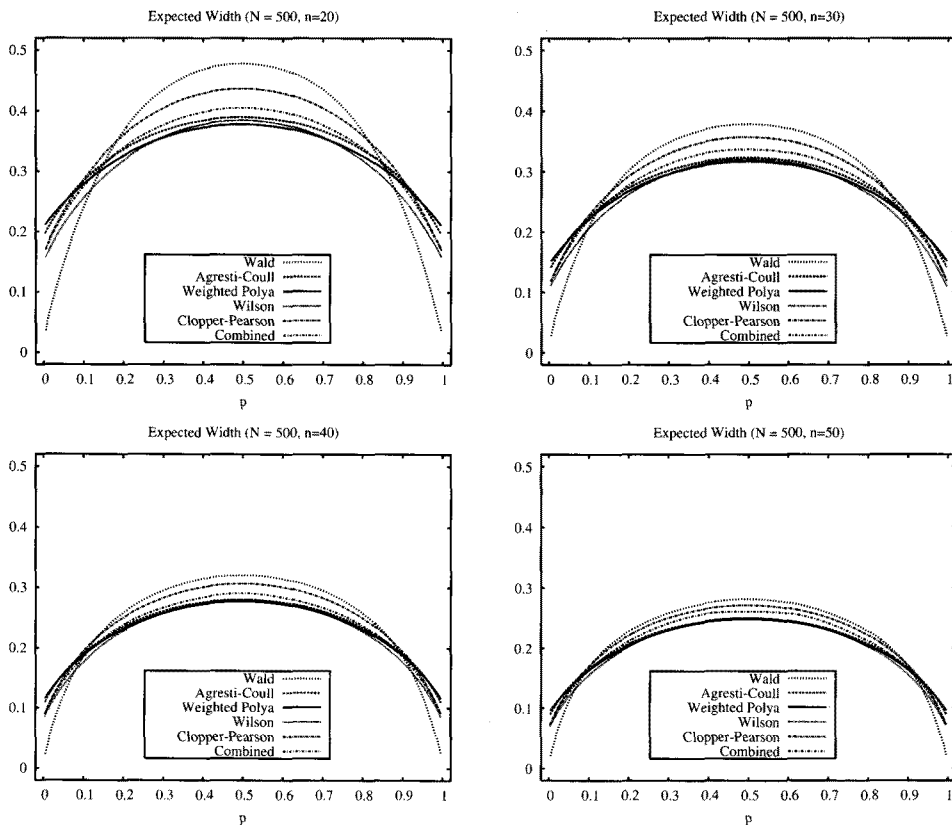
Figure 4: *Expected widths of 95% confidence intervals when N = 500 and n = 20, 30, 40 and 50.*

tively small expected width. The criterions favors the approximate confidence interval too much. One might think that the positive difference between coverage probability and nominal level should be treated differently from the negative one. Since, a wide interval tends to be conservative, the coverage probability and the expected width have a trade-off relationship. Thus, we should compare the approximate and the exact confidence intervals separately, or need another criterion for fair comparison. One criterion for doing this can be found in Lee (2009) or Casella et al. (1994). However, I do not want to use their approaches, because they require an appropriate value of weight between the coverage probability and the expected length. I believe that the weight is a matter of preference. If one's prime interest is the minimum level of coverage, then an exact confidence interval is one's choice. In this case, the combined interval rather than the Clopper-Pearson could be excellent one, because it is almost exact with much smaller expected width.

4. Conclusion

The approximate confidence intervals considered in this paper behave quite similarly to the corresponding intervals in binomial distribution case. Thus similar conclusions can be extracted among the approximate confidence intervals. That is, the Wald interval should be not be considered for a serious statistical problem specially when sample size n is small, say $n \leq 40$. In this case, the weighted

Polya posterior and the Wilson intervals are recommendable. Since they give quite reasonable coverage probabilities for almost all p, and the combined interval tends to be too wide, I believe that the recommendation is still valid, even if we consider exact intervals together.

For large sample, the Wilson, the weighted Polya posterior, the Agresti-Coull and the combined intervals are all equally good. However, the Agresti-Coull interval is recommendable because of the simplicity emphasized by Brown et al. (2001). Unlike the binomial distribution case, the weighted Polya posterior interval is not as simple as the Agresti-Coull.

References

Agresti, A. and Coull, B. A. (1998). Approximation is better than "exact" for interval estimation of binomial proportions, *American Statistician*, **52**, 119–126.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.

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

Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion, *Statistical Science*, **16**, 101–133.

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

Casella, G. T., Hwang, T. G. and Robert, C. P. (1994). Loss functions for set estimation, *Statistical Decision Theory and Related Topics V* (Edited by S. S. Gupta and J. O. Berger), Springer-Verlag.

Cai, Y. and Krishnamoorthy, K. (2005). A simple improved inference methods for some discrete distributions, *Computational Statistics & Data Analysis*, **48**, 605–621.

Clopper, C. J. and Pearson, E. S. (1936). The use of confidence or fiducial limits illustrated in the case of binomial, *Biometriks*, **26**, 404–413.

Chocran, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York.

Esty, W. E. and Banfield, J. D. (2003). The box-percentile plot, *Journal of Statistical Software*, **8**, Issue 17.

Krishnamoorthy, K., Thomson, J. and Cai, Y. (2004). An exact method of testing equality of several binomial proportions to a specified standard, *Computational Statistics & Data Analysis*, **45**, 697–707.

Kulkarni, P. M. and Shah, A. K. (1995). Testing the equality of several binomial proportions to prespecified standard, *Statistic & Probability Letter*, **25**, 213–219.

Lee, S.-C. (2005). Interval estimation of binomial proportions based on weighted Polya posterior, *Computational Statistics & Data Analysis*, **51**, 1012–1021.

Lee, S.-C. (2009). Choosing between the exact and the asymptotic confidence intervals: For the difference of two independent binomial proportions, *Communications of the Korean Statistical Society*, **16**, 363–372.

Meeden, G. D. (1999). Interval estimators for the population mean for skewed distributions with a small sample size, *Journal of Applied Statistics*, **26**, 81–96.

Staner, 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.

Yamane, T. (1967). *Elementary Sampling Theory*, Prentice-Hall Inc., Englewood Cliffs, New York.