Research Report
Department of Statistics

Stockholms universitet

No. 2020:1

Approximate Confidence Intervals for a Binomial p - Once Again

Per Gösta Andersson

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Approximate Confidence Intervals for a Binomial p - Once Again

Per Gösta Andersson
Department of Statistics, Stockholm University

Abstract

The problem of constructing a reasonably simple yet well behaved confidence interval for a binomial parameter p is old but still fascinating and surprisingly complex. During the last century many alternatives to the poorly behaved standard Wald interval have been suggested. This paper builds on a special case of a general technique for adjusted intervals primarily based on Wald type statistics. The main idea is to construct an approximate pivot with uncorrelated, or nearly uncorrelated, components. The resulting ANNE interval, as well as a modification thereof, is compared with the well renowned Wilson and Agresti-Coull intervals and the subsequent discussion will in itself hopefully shed some new light on this seemingly elementary interval estimation situation. Generally, an alternative to the Wald interval is to be judged not only by performance, its expression should also indicate *why* we will obtain a better behaved interval.

KEY WORDS: Correlation; Coverage probability; Score statistic; Skewness; Wald statistic.

1. Introduction

The Wald (standard) interval

$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{1}$$

which can be traced as far back as Laplace (1812) is surely the most well-known (and still most used?) confidence interval for a binomial p, where $\hat{p} = x/n$, x is the observed value of $X \sim Bin(n,p)$ and z is the $1-\alpha/2$ quantile of the standard normal distribution. The actual coverage probability is however often far from the nominal level $1-\alpha$, a fact which is highlighted in e.g. Agresti and Coull (1998), Newcombe (1998) and Brown et al. (2001, 2002). Papers dealing with this interval estimation problem also include, in chronological order, Crow (1956), Ghosh (1979), Blyth and Still (1983), Blyth (1986), Santner (1998), Agresti and Caffo (2000), Schilling and Doi (2014), Easterling (2015) and Jin et al. (2017).
Over the years many alternative intervals have been proposed, among them the "exact" Clopper-Pearson (1934), the arcsine (constructed from a variance stabilizing transformation), the Jeffreys (a Bayesian approach using the noninformative Jeffreys prior), the logit and the likelihood ratio interval.
However, we will as an initial alternative to the Wald interval consider the Wilson (1927) interval, which is constructed from the score statistic

$$\frac{\hat{p}-p}{\sqrt{V(\hat{p})}}, \tag{2}$$

where $V(\hat{p}) = p(1-p)/n = 1/I_n(p)$, where $I_n(p)$ is the Fisher information number of X, whereas (1) is based on the Wald statistic

$$\frac{\hat{p}-p}{\sqrt{\hat{V}(\hat{p})}}, \tag{3}$$

where $\hat{V}(\hat{p}) = \hat{p}(1-\hat{p})/n = 1/\hat{I}_n(\hat{p})$. Another interval with, as it has turned out, favorable properties is the Agresti-Coull (1998) interval. This is constructed simply by replacing $\hat{p} = x/n$ in (1) by the midpoint of the interval resulting from the Wilson procedure where n is modified accordingly. The Wilson and Agresti-Coull intervals are among the most recommended in Brown et al. (2001), which makes them good benchmarks for comparison with the alternative intervals discussed in this paper.
Before moving on, let us have a look at Chapter 4 in DasGupta (2008) concerning transformations in a more general setting. There it is assumed

that for an estimator $\hat{\theta}$ of some parameter θ the following convergence in distribution holds:

$$\sqrt{n}(\hat{\theta} - \theta) \overset{\mathrm{D}}{\to} N(0, \sigma^2(\theta))$$

For a fixed n one can therefore argue that

$$\sqrt{n}\,\frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})} \overset{\mathrm{appr}}{\sim} N(0, 1), \tag{4}$$

using a consistent plug-in variance estimator $\hat{\sigma}^2(\hat{\theta})$ of $\sigma^2(\hat{\theta})$ and Slutsky's Theorem.
However, there are now concerns about the following issues.
1) $\hat{\theta}$ may be biased though asymptotically unbiased
2) $\hat{\sigma}(\hat{\theta})$ is stochastic
3) $\hat{\theta}$ may have a substantially skewed distribution
For the binomial case 1) is not an issue but regarding 2) this should affect e.g. the variance of (3) and we can also note that $E(\hat{p}(1 - \hat{p})/n) = p(1 - p)/n \cdot ((n-1)/n)$, implying a slight negative bias. We could of course instead work with the unbiased variance estimator $\hat{p}(1 - \hat{p})/(n - 1)$, but, for n not too small, the gain will be numerically negligible in terms of e.g. coverage probabilities. Finally, concerning 3) it is known that for $X \sim Bin(n, p)$ the skewness coefficient is

$$\gamma_n = (1 - 2p)/\sqrt{np(1 - p)},$$

where $1 - 2p$ turns out to be a key component for the procedures in this paper. Going back to the list in DasGupta (2008) we would like to add:
4) $\hat{\theta}$ and $\hat{\sigma}(\hat{\theta})$ are most likely correlated.
This will affect the skewness of the approximate pivot in (4).
What could then be the consequence(s) of 4) for the resulting Wald interval? From e.g. Andersson (2009) we get that a second order Taylor approximation of $(\hat{\theta} - \theta)/\sqrt{\hat{V}(\hat{\theta})}$ around $(\theta, V(\hat{\theta}))$ is

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{V}(\hat{\theta})}} \approx \frac{\hat{\theta} - \theta}{\sqrt{V(\hat{\theta})}} - \frac{(\hat{\theta} - \theta)(\hat{V}(\hat{\theta}) - V(\hat{\theta}))}{2V(\hat{\theta})^{3/2}}$$

Investigating the bias, assuming that $E(\hat{\theta}) = \theta$ and $E(\hat{V}(\hat{\theta})) = V(\hat{\theta})$:

$$E\Big(\frac{\hat{\theta} - \theta}{\sqrt{\hat{V}(\hat{\theta})}}\Big) \approx -\frac{Cov(\hat{\theta}, \hat{V}(\hat{\theta}))}{2V(\hat{\theta})^{3/2}} \tag{5}$$

So, here is a direct effect of the correlation between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$, which will cause the subsequent Wald interval to be centered around the "wrong" value. It is also to be observed that in our present binomial situation 3) and 4) are directly connected, since one can show that

$$Cov\Big(\hat{p}, \frac{\hat{p}(1-\hat{p})}{n-1}\Big) = \frac{p(1-p)(1-2p)}{n^2} = \gamma_n\Big(\frac{p(1-p)}{n}\Big)^{3/2} \tag{6}$$

Returning to the binomial case it is perhaps tempting to draw the conclusion that the approximate bias of (3) is $-\gamma_n/2 = O(1/\sqrt{n})$, but formally (3) has a degenerate distribution with no moments defined, since there is a positive probability that $X = 0$ or $X = n$.
In this paper we will bring forward a special case of an adjusted confidence interval for a mean (Andersson (2009), based on Andersson and Nerman (2000)). This is constructed from the general idea that for a statistic (approximate pivot) such as (3) the numerator should be uncorrelated (or less correlated) with the denominator. (This is obviously fulfilled for the score statistic.) The main reason for this is to avoid induced skewness of the resulting pivot, since we will argue that its distribution is approximately normal. This interval together with a modified version will be presented together with the Wald, Wilson and Agresti-Coull intervals in Section 2 in the case of $n = 40$ and the nominal confidence level 95%. The paper concludes with a summary and some final remarks.

2. A Presentation and Comparison of Intervals for $n = 40$ and Nominal Confidence Level 95%

We will present and illustrate the properties of the Wald, Wilson, Agresti-Coull and ANNE intervals, where the emphasis is put on coverage and non-coverage properties, but we will also consider interval lengths and disributions of pivots.

2.1 The Wald Interval

As mentioned, the coverage probability of the Wald interval (1) is usually lower than the chosen confidence level. A less commented (one excellent exception being Cai (2005)) but nevertheless serious drawback for this supposedly symmetric interval is the lack of balance between the lower and upper non-coverage probabilities for values of p not close to 0.5. Evidently, this has a serious negative impact on the coverage probability of one-sided confidence intervals. As a graphical illustration Figure 1 (left) displays coverage probabilities for varying values of p when $n = 40$ and $\alpha = 0.05$, whereas in Figure 1 (right) we find the corresponding lower non-coverage probabilities (lower non-coverage probability signifying the probability that the upper endpoint of the interval gets a lower value than p, i.e. a "miss to the left of p").

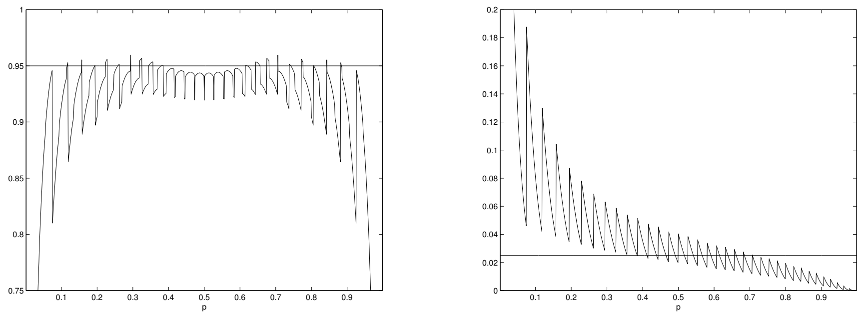

Figure 1. Coverage probability (left) and lower non-coverage probability (right) for the Wald interval when $n = 40$ and $\alpha = 0.05$.

As seen from Figure 1 (left) the coverage probability very rarely reaches (or exceeds) the nominal level 0.95 and for values of p close to 0 or 1 there is a steep drop towards 0. Correspondingly, Figure 1 (right) clearly shows that the lower non-coverage probability is almost solely close to 0.025 around $p = 0.5$. For values of $p < 0.5$ (> 0.5) the non-coverage probability is often $>> 0.025$ ($<< 0.025$). Due to symmetry the opposite holds for the upper non-coverage probability and, which was observed in Figure 1 (left), the sum of the non-coverage probabilities are often considerably larger than 0.05. Brown et al. (2001) coined the terms "lucky" and "unlucky" n regarding combinations of n and $p = 0.5$ yielding high (≥ 0.95) or low (≤ 0.92) coverage probabilities, respectively. As a matter of fact, according to Table 2 in Brown et al. (2001) and Figure 1 (left), one unlucky n is $n = 40$ (coverage probability 0.912).
We note that the confidence coefficient equals 0 for all n (the infimum of coverage probabilities over the parameter space $(0, 1)$ for p). Another deficiency of the Wald interval is the inability of providing an interval for the extreme, yet important, cases of $x = 0$ and $x = n$. However, for these values of X one may without normal approximation straightforwardly utilize the binomial distribution, see e.g. Bickel and Doksum (1977) and Louis (1981).
Generally, problems arise whenever a discrete distribution is approximated by a continuous one, in this case a binomial to a normal. One quite popular method to improve on this approximation is the the use of the continuity

correction. An alternative to (3) would then be

$$\frac{\hat{p} + \frac{1}{2n} - p}{\sqrt{\hat{V}(\hat{p})}}$$

leading to the confidence interval

$$\hat{p} + \frac{1}{2n} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

However, for $n = 40$ and $\alpha = 0.05$ the mean coverage probability is 0.890 with and 0.894 without the continuity correction. (The mean is taken over values of p: $0.001, 0.002, \ldots, 0.999$). It is well known that continuity corrections do not always work well, especially in the tails of distributions and this technique will be considered further in this paper.

2.2 The Wilson Interval

One of the first suggested alternatives to the Wald interval is due to Wilson (1927), also named the score interval. Instead of utilizing that

$$\frac{\hat{p} - p}{\sqrt{\hat{V}(\hat{p})}} \overset{\text{appr}}{\sim} N(0,1), \tag{7}$$

leading to the Wald interval, the Wilson interval is based on, as it turns out, the more accurate approximation

$$\frac{\hat{p} - p}{\sqrt{V(\hat{p})}} \overset{\text{appr}}{\sim} N(0,1), \tag{8}$$

(In (7) and (8) it is understood that $\hat{p} = X/n$ and in this paper this holds whenever $\hat{p}$ should be interpreted as stochastic.) As previously remarked the approximate Wald pivot in (7) is not defined for $x = 0$ and $x = n$, a fact which in itself violates the normal distribution approximation. As opposed to the Wald pivot the pivot in (8) is defined for the entire support of X with zero mean, unit variance and skewness coefficient $(1-2p)/\sqrt{np(1-p)} = O(1/\sqrt{n})$ and yields the Wilson interval

$$\frac{\hat{p} + \frac{z^2}{2n} \pm z\sqrt{\hat{V}(\hat{p}) + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \tag{9}$$

As noted by Agresti and Coull (1998) the midpoint of (9) is

$$\hat{p}(\frac{n}{n+z^2}) + \frac{1}{2}(\frac{z^2}{n+z^2}), \tag{10}$$

implying that for $\hat{p} < 0.5\,(> 0.5)$ the midpoint of the Wilson interval is, compared with the Wald interval, moved to the right (left). The expression (10) can be viewed as a shrinkage estimate towards 0.5, but, as we shall see, it can also be interpreted as an effect of the correlation between $\hat{p}$ and its standard error $\sqrt{\hat{V}(\hat{p})}$ (or its variance estimator $\hat{V}(\hat{p})$). Finally, observe that for e.g. $x = 0$ the Wilson procedure actually leads to an interval, which is $(0, z^2/(z^2+n))$, where the upper bound is a decent approximation of the "exact" $1 - \alpha^{1/n}$.
The general increase of coverage probability for the Wilson interval vis-á-vis the Wald interval is evident when comparing Figure 1 (left) with Figure 2 (left). The confidence coefficient is around 0.852 for the Wilson interval and the improvement carries on to the non-coverage probabilities as seen from comparing Figure 1 (right) with Figure 2 (right). Furthermore, note that the Wald and Wilson non-coverage probabilities also differ in terms of a trend: Increasing values of p means (on average) decreasing values of the lower non-coverage probabilities for the Wald interval and the opposite holds for the Wilson interval. (This was also observed by Cai (2005).) We will return to this phenomenom.

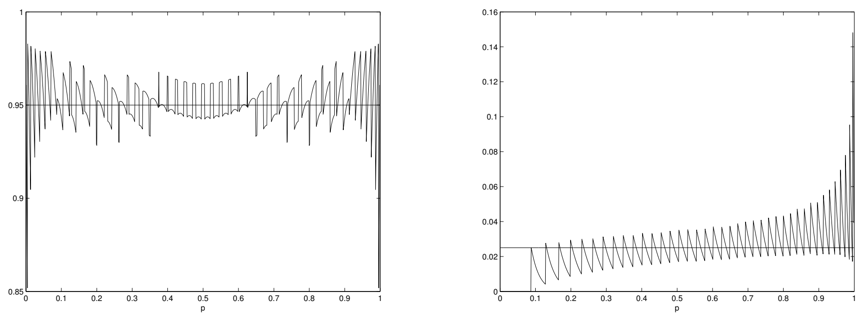

Figure 2. Coverage probability (left) and lower non-coverage probability (right) for the Wilson interval when $n = 40$ and $\alpha = 0.05$.

2.3 The Agresti-Coull Interval

This interval was proposed in order to be able to utilize the Wald interval expression with $\hat{p}$ substituted by

$$\tilde{p} = \frac{x + z^2/2}{n + z^2}$$

and n by $\tilde{n} = n + z^2$. (When $\alpha = 0.05$ we can let $\tilde{p} = (x+2)/(n+4)$, thus obtaining the appealingly simple rule "add two successes and two failures".) As noted by Agresti and Coull (1998) this midpoint value agrees with the Bayes estimate of p assuming a prior which is $Beta$(2,2). (The previously mentioned Jeffreys prior is $Beta$(1/2,1/2).)

Now, from the assumption that

$$\frac{p - \tilde{p}}{\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}} \tag{11}$$

we thus arrive at the confidence interval

$$\tilde{p} \pm z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{12}$$

The results for $n = 40$ are shown in Figure 3 and from these we observe that the behaviour in terms of both coverage and non-coverage probabilities are quite different from the previous intervals. Here we actually get the most conservative coverage probabilites for the lowest and highest values of p and the non-coverage probabilities do not exhibit a monotone trend over the whole range of p. We can note however that there is a quite high similarity between the Wilson and Agresti-Coull intervals for $0.2 \leq p \leq 0.8$. The confidence coefficient is 0.933 for the Agresti.Coull interval.

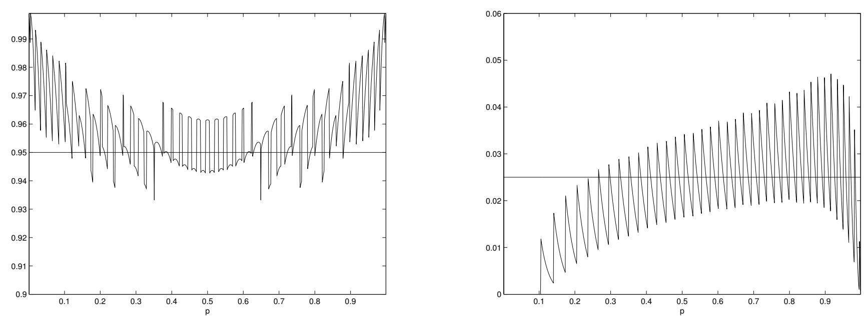

Figure 3. Coverage probability (left) and lower non-coverage probability (right) for the Agresti-Coull interval when $n = 40$ and $\alpha = 0.05$.

2.4 The ANNE Intervals

A different approach to both the Wald and the Wilson procedure is outlined in Andersson and Nerman (2000) and Andersson (2009), where in the latter paper the interval estimation problem of a probability (proportion) is a special case of the confidence interval construction of a mean based on an iid sample. In short, the idea (transferred to the binomial situation) in the first step is to replace $\hat{V}(\hat{p}) = \hat{p}(1-\hat{p})/n$ in the denominator of (3) by a variance estimator of $\hat{p}$ which is uncorrelated with $\hat{p}$ and which has reduced variance. The proposed candidate is $\hat{V}(\hat{p}) - K(\hat{p}-p)$, where $K = Cov(\hat{p}, \hat{V}(\hat{p}))/V(\hat{p}) \approx (1-2p)/n$. (To get equality we would have to use $\hat{V}(\hat{p}) = \hat{p}(1-\hat{p})/(n-1)$.) Estimating K by $\hat{K} = (1-2\hat{p})/n$ leads to

$$\frac{\hat{p}-p}{\sqrt{\hat{V}(\hat{p}) - \frac{1-2\hat{p}}{n}(\hat{p}-p)}} \overset{\text{appr}}{\sim} N(0,1) \tag{13}$$

and the resulting ANNE (Andersson-Nerman) confidence interval is readily seen to be

$$\hat{p} + \frac{z^2}{2n}(1-2\hat{p}) \pm z\sqrt{\hat{V}(\hat{p}) + z^2\frac{(1-2\hat{p})^2}{4n^2}} \tag{14}$$

(Since K is a function of p only one could alternatively consider using the approximate pivot

$$\frac{\hat{p}-p}{\sqrt{\hat{V}(\hat{p}) - \frac{1-2p}{n}(\hat{p}-p)}}, \tag{15}$$

without first estimating K by $\hat{K}$. As it turns out, the corresponding confidence region will not be an interval though.)
The factor $1-2\hat{p}$ determines if the interval is shifted to the left or right and by how much, which, as been shown previously, is directly linked to the estimated correlation between the numerator and denominator of the Wald statistic (3). At first expression (14) seems quite different from the Wilson interval, but observe that the midpoint in (9) can be written as

$$\hat{p} + \frac{z^2}{2n}\frac{1-2\hat{p}}{1+\frac{z^2}{n}}$$

On the other hand, the midpoint in (14) can be written as

$$\hat{p}(1-\frac{z^2}{n}) + \frac{1}{2}\frac{z^2}{n},$$

the shrinkage towards 0.5 being slightly more pronounced than for the Wilson interval.
Even though the Wilson and ANNE intervals have rather closely related expressions, the numerical results for $n = 40$ are, at least partially, substantially different, as seen when comparing Figure 2 with Figure 4. The coverage probability is much the same for small (and large) values of p (both have the

confidence coefficient 0.852), but in between the ANNE interval is with a few exceptions conservative, whereas the Wilson interval has quite often coverage probabilities below 0.95 as seen in Figure 2 (left). The noncoverage probabilities have similar patterns for these intervals (Figure 2 (right) and Figure 4 (right)), the ANNE interval displaying slightly smaller values of the lower non-coverage probabilities for most values of p.

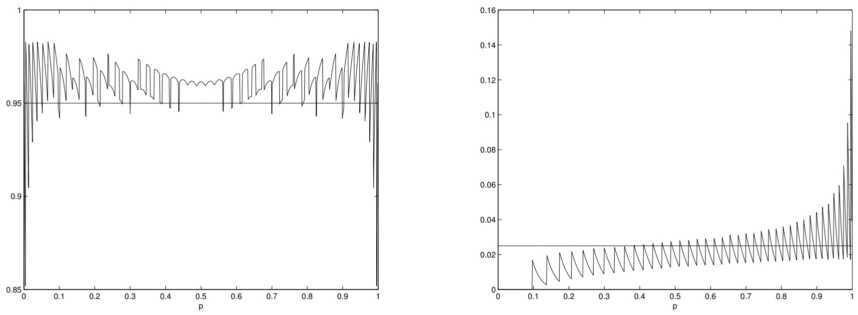

Figure 4. Coverage probability (left) and lower non-coverage probability (right) for the ANNE interval when $n = 40$ and $\alpha = 0.05$.

Actually, as observed in Kott et al. (2001), the Wilson and ANNE intervals are asymptotically identical in the sense that if we drop terms of order $O_p(n^{-3/2})$, both equal

$$\hat{p} + \frac{z^2}{2n}(1 - 2\hat{p}) \pm z\sqrt{\hat{V}(\hat{p}) + \frac{z^2}{4n^2}} \tag{16}$$

This interval, the Simplified ANNE, and its coverage/non-coverage properites are shown in Figure 5. It is evident from expression (16) that we simply get a longer ANNE interval. This should improve upon the coverage probability, but by how much? The answer is: Quite a lot, as we can see from Figure 5. In fact, we now have unnessecarily conservative coverage for all but the most extreme values of p. Therefore, this interval will not be considered further in this paper.

2.5 A Comparison of expected lengths of confidence intervals

Not only coverage and non-coverage probabilities should be used as measures of quality for a confidence interval, but also e.g. the expected length for a given p. There is often a price to pay in terms of a longer interval when the coverage probability increases. Figure 6 shows the expected lengths for the Wilson, Agresti-Coull and ANNE interval. (The Wald and Simplified ANNE intervals are now disregarded.) In this respect the Wilson interval is

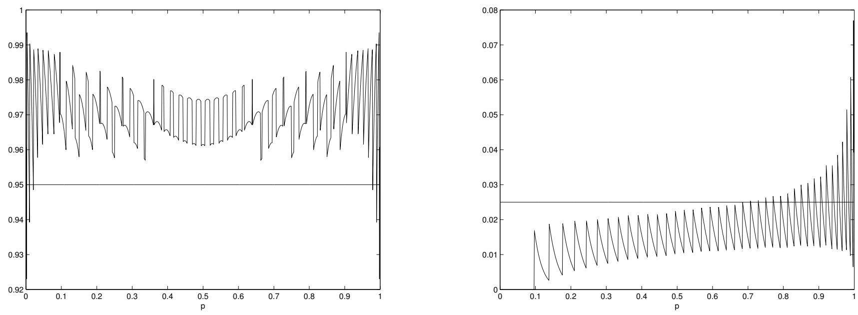

Figure 5. Coverage probability (left) and lower non-coverage probability (right) for the Simplified ANNE interval when $n = 40$ and $\alpha = 0.05$.

the overall shortest and the ANNE interval si the longest, except for very small/large values of p for which the Agresti-Coull interval is longer than the other two. This last observation should come as no surprise, since the Agresti-Coull interval is highly conservative for such values of p.

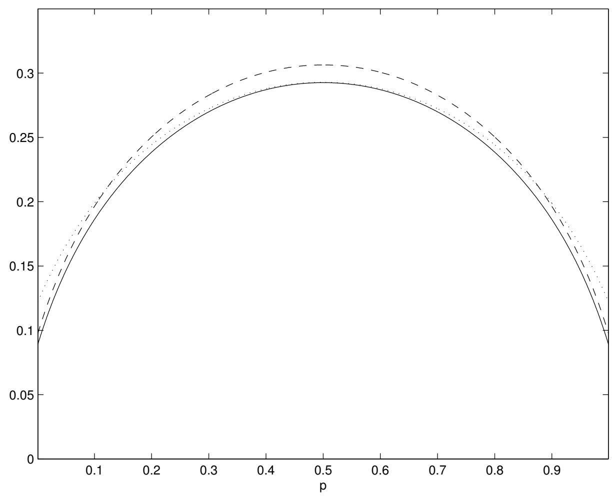

Figure 6. Expected length for the Wilson (sold), Agresti-Coull (dotted) and ANNE interval (dashed) when $n = 40$ and $\alpha = 0.05$.

2.6 A Comparison of Approximate Pivots

One way to explain the performance of the approximate confidence intervals is to study the distributions of the corresponding approximate pivots with

respect to expectation (bias), variance and skewness. (In the following, when referring to a pivot, it will be understood in an approximate sense.)
In Figure 7 we have dispayed bar charts for the probability distributions of the Wilson, Agresti-Coull and ANNE pivotd for $p = 0.01$ and $p = 0.05$. Clearly the Agresti-Coull pivot has a substantial positive bias and, as foreseen, the fit to the no rmal diatribution is improved for all three pivots when $p = 0.5$ as compared with $p = 0.01$. A summary of moments is also telling:

		Expectation	Variance	Skewness
	Wilson	0	1	1.50
$p = 0.01$	Agresti-Coull	1.25	0.0589	1.30
	ANNE	-0.0162	0.898	1.33
	Wilson	0	1	0.629
$p = 0.05$	Agresti-Coull	0.857	0.356	0.0105
	ANNE	-0.0069	0.927	0.546

Concerning mean and variance, at least, the Wilson and ANNE pivots are "well-behaved". The Agresti-Coull is not. Yet, as we have sen, the coverage probability for small (and large) values of p is conservative, which is an effect of longer intervals than generated by the other two procedures. In fact, the Agresti-Coull interval suffers from the same deficiency as the Wald interval, namely that for small/high values of p the interval is not necessarily within $(0, 1)$.

3. Concluding remarks

We have studied properties of some alternative confidence intervals to the Wald interval and to conclude we ill make some comments about their algebraic expressions. The ANNE interval

$$\hat{p} + \frac{z^2}{2n}(1 - 2\hat{p}) \pm z\sqrt{\hat{V}(\hat{p}) + z^2\frac{(1 - 2\hat{p})^2}{4n^2}}$$

appears in this respect more informative than the Wilson interval. In the former both the change of mid-point and the length depends on $1 - 2\hat{p}$.

$$\tilde{p} \pm z\sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}},$$

The sign of the shift is determined by the size of $\hat{p}$: negative for $\hat{p} > 0.5$ and postive for $\hat{p} > 0.5$. The factor $1 - 2\hat{p}$ is furthermore a key component when estimating the correlation between $\hat{p}$ and $\hat{p}(1 - \hat{p})$, which in its turn affects

the distribution of the Wald pivot. Also, for e.g. $\hat{p} = 0$ the interval is simply $(0, z^2/n)$.
Looking instead at the Agresti-Coull interval

$$\tilde{p} \pm z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}},$$

the modification of the Wald interval is based on a change of point estime from $\hat{p}$ to $\tilde{p}$, which can be argued within a Bayesian framework, by, as mentioned previously, considering a prior which has a $Beta(2, 2)$ distribution. However, this does not imply that it is problematic to interpret the interval from a frequentist's point of view, merely that the interval is not symmetric around an unbiased point estimate of p. Also, for $\alpha = 0.05$ we have that $\tilde{p} \approx (x + 2)/(n + 4)$, so when e.g. $x = 0$, we get a Wald interval around the point $2/(n + 4)$. This is then interpreted as having obtained two successes out of $n + 4$ repetitions and the obstacle of the Wald interval for $x = 0$ is conveniently avoided.
And the winner for best motivated and best modified Wald interval is? Well, that is for the reader to decide...

References

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

Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and difference of proportions result from adding two successes and two failures. *The American Statistician*, **54(4)**, 280-288.

Andersson, P.G. (2009). A simple correlation adjustment procedure applied to confidence interval construction. *The American Statistician*, **63**, 258-262.

Andersson, P.G. and Nerman, O. (2000). A balanced adjusted confidence interval procedure applied to finite population sampling. *Proceedings of ICES II*, Buffalo, New York, USA (CD-ROM) American Statistical Association, Alexandria, Virginia, USA.

Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day.

Blyth, C.R. (1986). Approximate binomial confidence limits. *Journal of the American Statistical Association*, **81**, 843-855.

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(2)**, 101-117.

Brown, L.D., Cai, T.T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, **30(1)**, 160-201.

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

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

Crow, E.L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423-435.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer.

Easterling, R. (2015). There's nothing wrong with Clopper-Pearson binomial confidence limits. *The American Statistician*, **69(2)**, 154-155.

Ghosh, B.K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, **74**, 894-900.

Jin, S., Thulin, M. and Larsson, R. (2017). Approximate Bayesianity of frequentist confidence intervals for a binomial proportion. *The American Statistician*, **71(2)**, 106-111.

Kott, P., Andersson, P.G. and Nerman, O. (2001). Two-sided coverage intervals for small proportions based on survey data. *Proceedings of the FCSM Research Conference*, Arlington, Virginia, USA.

Laplace, P.S. (1812). Théorie analytique des probabilités. Paris: *Courcier*.

Louis, T.A. (1981). Confidence intervals for a binomial parameter after observing no successes. *The American Statistician*, **35(3)**, 154.

Newcombe, R.G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, **17**, 857-872.

Santner, T.J. (1998). Teaching large-sample binomial confidence intervals. *Teaching Statistics*, **20**, 20-23.

Schilling, M.F. and Doi, J.A. (2014). A coverage probability approach to finding an optimal binomial confidence procedure. *The American Statistician*, **68(3)**, 133-145.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209-212.

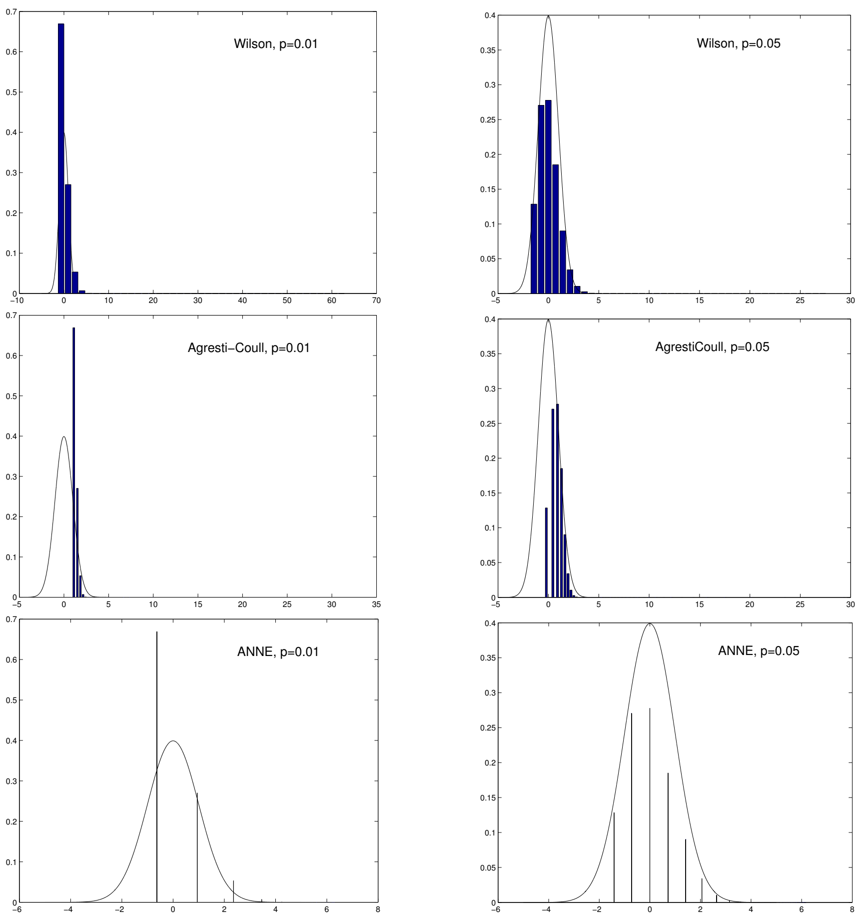

Figure 7. Bar charts displaying, together with the density function of the standard normal distribution, the probability distributions of the Wald, Wilson, Agresti-Coull and ANNE pivots for $n = 40$ and $\alpha = 0.05$. (For the Wald pivot the cases of $x = 0$ and $x = 40$ are omitted.)