LEAST COST FORMULATIONS, LTD.
824 Timberlake Drive, Virginia Beach, VA 23464-3239
Tel: (757) 467-0954 Fax: (757) 467-2947
E-mail: office@lcftld.com URL: http://lcfltd.com/

TECHNICAL REPORT

NUMBER: TR297 DATE: 09 June 7

TITLE: Coverage accuracy for binomial proportion 95% confidence intervals for 12 to 100 replicates.

AUTHOR: R. A. LaBudde

ABSTRACT: Four methods were considered for determining confidence intervals for a binary proportion resulting from N trial replicates. The methods were: 1) Agresti-Coull ("AC"); 2) Wilson ("W"); 3) modified Wilson ("WM"); 4) Clopper-Pearson ("CP"); and 5) Blaker ("BK"). The CP method was strongly conservative, with coverage accuracy for a 95% confidence interval ranging from 96-100% for N = 8, 12, 16, 20, 30, and 100 and true proportions ρ from 0.01 to 0.50. The AC method gave near identical results to the BK method. Of the five methods considered, the WM method generally most accurate, although the BK method is recommended for N more than 100, and the W method for N of 20 or more. The BK method is recommended over the CP method. For general work, any of the AC, W or WM methods should give acceptable results, although the W method has a tendency to low coverage for small N and ρ close to 0 or 1, and the AC method is somewhat conservative.

KEYWORDS: 1) AGRESTI 2) WILSON 3) BLAKER 4) CLOPPER 5) BINOMIAL 6) CONFIDENCE

REL.DOC.:

REVISED:

INTRODUCTION

There are a variety of methods available for specifying a confidence interval for a binomial proportion found in an experiment with N replications and X successes.

The simple Wald confidence interval exhibited in most introductory texts on statistics is highly inaccurate (coverage accuracy well below nominal confidence level) unless N is large (100 or more).

A variety of methods have been proposed to give improved confidence intervals for proportions. There is some controversy surrounding the question of which method actually provides the best coverage accuracy, as the accuracy is a discontinuous function of N and the true proportion ρ [1-7].

In what follows, we investigate the coverage accuracy of several commonly used or recommended methods for determining 95% confidence intervals on a single binomial proportion, with N between 8 and 100 (small to medium sample size).

METHODS INVESTIGATED

Let $z = z_{1-\alpha/2}$ denote the 1-α/2 quantile of the standard normal distribution, where α = 1 – Confidence Level. Here the confidence level of interest is 95%, α = 0.05 and and $z_{1-\alpha/2} = 1.9600$.

In pseudo-BASIC, the algorithms for the simpler methods are:

AGRESTI-COULL (“AC”):

```
w = (x + z * z * 0.5) / (n + z * z)
d = z * sqrt(w * (1 - w) / (n + z * z))
LCL = max(0,w - d)
UCL = min(1,w + d)
```

WILSON (“W”): With no continuity correction.

```
if  (x = 0) then
   LCL = 0
   UCL = z*z / (n + z*z)
elseif (x = n) then
   LCL = n / (n + z*z)
   UCL = 1
else
   d = z * sqrt(x - x*x/n + 0.25*z*z)
   LCL = (x + 0.5*z*z - d)/(n + z*z)
   UCL = (x + 0.5*z*z + d)/(n + z*z)
end if
```

WILSON-MODIFIED (“WM”): Wilson score interval above, adjusted at x = 1, N-1 values.

```
if (x = 0) then
  LCL = 0
  UCL =z*z / (n + z*z)
elseif (x = n) then
  LCL = n / (n + z*z)
  UCL = 1
else
  d = z * sqrt(x - x*x/n + 0.25*z*z)
  LCL = (x + 0.5*z*z - d)/(n + z*z)
  UCL = (x + 0.5*z*z + d)/(n + z*z)
end if
if (x = n - 1) then UCL = 1
if (x = 1) then LCL = 0
```

CLOPPER-PEARSON (“CP”): Based on inversion of the exact binomial distribution.

```
LCL = 0
UCL = 1
if (x  < > 0)  then LCL = qbeta(α/2, x, n - x + 1)
if (x  < > n)  then UCL = qbeta(1- α/2, x + 1, n - x)
```

BLAKER (“BK”): Improved Clopper-Pearson interval.

The AC and W methods are approximations based upon large-sample properties. The WM method is introduced here (not previously published) as an attempted correction to the problems the W method has for low coverage accuracy when ρ is near 0 or 1. The CP and BK methods are “exact” in the sense that coverage accuracy is never less than nominal (here, 95%).

Numerous other possible intervals of good quality exist, but only those listed above are considered here. They exemplify the issues involved with all methods.

The BK method was calculated using the ‘binGroup’ package of R, and the others were programmed directly in R.

THEORETICAL COVERAGE ACCURACIES

Coverage accuracy for N = 8					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	0.99731	*0.92274*	0.99731	0.99731	0.99731
0.0200	*0.98966*	0.85076	*0.98966*	*0.98966*	*0.98966*
0.0500	*0.94276*	*0.94276*	*0.94276*	0.99421	0.99421
0.1000	*0.96191*	*0.96191*	*0.96191*	0.99498	*0.96191*
0.2000	*0.94372*	*0.94372*	*0.94372*	0.98959	0.98959
0.5000	*0.92969*	*0.92969*	*0.92969*	0.99219	0.99219
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

Coverage accuracy for N = 12					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	*0.99383*	0.88638	*0.99383*	*0.99383*	*0.99383*
0.0200	*0.97689*	*0.97689*	*0.97689*	*0.97689*	*0.97689*
0.0500	*0.98043*	*0.98043*	*0.98043*	*0.98043*	*0.98043*
0.1000	*0.97436*	*0.97436*	*0.97436*	0.99567	*0.97436*
0.2000	*0.98059*	*0.98059*	*0.98059*	*0.98059*	*0.98059*
0.5000	*0.96143*	*0.96143*	*0.96143*	*0.96143*	*0.96143*
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

Coverage accuracy for N = 16					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	*0.98907*	0.85146	*0.98907*	*0.98907*	*0.98907*
0.0200	*0.96014*	*0.96014*	*0.96014*	0.99631	*0.96014*
0.0500	*0.95706*	*0.95706*	*0.95706*	0.99300	*0.95706*
0.1000	0.98300	*0.93159*	*0.93159*	0.98300	0.98300
0.2000	0.97334	*0.94520*	*0.94520*	0.99300	0.97334
0.5000	*0.92319*	*0.92319*	*0.92319*	0.97873	0.97873
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

Coverage accuracy for N = 20					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	*0.98314*	*0.98314*	*0.98314*	*0.98314*	*0.98314*
0.0200	0.99293	*0.94010*	*0.94010*	0.99293	0.99293
0.0500	0.98410	*0.92452*	*0.92452*	0.98410	0.98410
0.1000	*0.95683*	*0.95683*	*0.95683*	0.98875	*0.95683*
0.2000	*0.95633*	*0.95633*	*0.95633*	0.97849	*0.95633*
0.5000	*0.95861*	*0.95861*	*0.95861*	*0.95861*	*0.95861*
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

Coverage accuracy for N = 30					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	0.99668	*0.96385*	*0.96385*	0.99668	*0.96385*
0.0200	*0.97828*	*0.97828*	*0.97828*	*0.97828*	*0.97828*
0.0500	0.98436	*0.93923*	*0.93923*	0.98436	0.98436
0.1000	*0.97417*	*0.97417*	*0.97417*	0.99222	*0.97417*
0.2000	*0.96386*	*0.96386*	*0.96386*	0.97998	*0.96386*
0.5000	*0.95723*	*0.95723*	*0.95723*	*0.95723*	*0.95723*
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

Coverage accuracy for N = 100					
ρ	***AC***	***W***	***WM***	***CP***	***BK***
0.0100	0.98163	*0.92063*	*0.92063*	0.98163	0.98163
0.0200	0.98452	*0.94917*	*0.94917*	0.98452	0.98452
0.0500	*0.96589*	*0.96589*	*0.96589*	0.98261	*0.96589*
0.1000	0.97156	0.93640	0.93640	*0.95569*	*0.95569*
0.2000	0.94052	0.94052	0.94052	0.96740	*0.95465*
0.5000	*0.94311*	*0.94311*	*0.94311*	0.96480	0.96480
NOTE: Results for 1 – ρ are equal to those for ρ. Minimum error results in italics for each row.					

CONCLUSIONS

1. Coverage is sometimes too low (< 90%) for the Wilson intervals for ρ close to 0 or 1 for N < 20.
2. The Clopper-Pearson method is very conservative, having coverage 96-100%.
3. The Blaker method is better than Clopper-Pearson, still "exact" and conservative, but involves the most complex algorithm. It is the probably the best of the methods shown for N > 100.
4. The Agresti-Coull method is typically identical to the Blaker method, except for an occasional difference with lower coverage (e.g., $\rho = 0.50$ for $N = 16$).
5. Only specific values for coverage are possible, based upon sums of the N binomial probabilities involved.
6. Of the four methods considered, the modified Wilson method is most accurate for N < 100.
7. The Wilson method is equally accurate for N of 20 or more.

REFERENCES

1. Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. Stat. Sci. 16(2): 101-133.
2. Agresti, A. and Coull, B.A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. Amer. Stat. 52: 119-126.
3. 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.
4. Newcombe, R.G. (1998). Two-sided confidence intervals for the single proportion: comparison of several methods. Stat. in Med. 17: 857-872.
5. Vollset, S.E. (1993). Confidence intervals for a binomial proportion. Stat. in Med. 12: 809-824.
6. Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. J. Amer. Stat. Assoc. 22: 209-212.
7. Bolboaca, S. and Cadariu, A.A. (2003). Binomial distribution sample confidence intervals estimation. 2. Proportion-like medical key paramaters. Leonardo Elec. J. Pract. & Technol. 3: 75-110.