Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

Niuman M Comas

Departamento de Matemática, Universidad de Holguín, Ave XX Aniversario, 82100, Holguín, Cuba.

Abstract

This paperexplains a methodology to obtain confidence intervals for proportion without any normal approximation. It is based on the use ofbinomial distribution in order to calculate the confidence bounds using polynomial interpolation techniques to estimate "artificial" probabilities associated to fractional distribution arguments.

The methodology includes the estimation of cumulative probabilities with continuity simulation and the respective inverse root. Hypothesis test is also incorporated. Abrief review of other common methods is previously shown.

The implementation of the "Direct" method wascarried outunder open source by means of JavaScript programming and it is available in the on-line statistical laboratory DynStats.com.

The results were compared with other procedures like the Wald, Clopper-Pearson and Mid-P methods. Most similarities were found for the Wald and Mid-p methods according to the proportion's region.

Keywords: Confidence interval, binomial proportion, binomial distribution, polynomial interpolation.

Introduction

The estimation of confidence intervals for variance, mean, and proportion are typical procedures of inferential statistics. The mean is associated directly with the normal distribution and the Student's t distribution, while the confidence interval for variance is based in its adjustment to the chi-squared distribution. (Stuart & Kendall, 1973) (Larson, 1974).

The estimation of confidence intervals for proportion may be done by different methods, they are classified in approximate and "exact", like the Wald's method, Mid-P and Clopper-Pearson which are among the better known ones. Their advantages and limitations have been object of frequent discussion in recent times, as much for their accuracy as for the computational cost. (Brown, Cai, & DasGupta, 2001) (Agresti & Gottard, 2005) (Newcombe & Nurminen, 2011) (Thulin, 2013).

The probability distribution directly associated with the proportion is the binomial distribution. However this it is a discrete distribution and it doesn't allow to evaluate fractional values of the argument

directly. That's why it is necessary to use indirect techniques to estimate confidence intervals.Many of these techniques are based on the approach to the normal distribution. To improve this approach several continuity corrections are introduced. An example of exact method is the Mid-p, it is based on the use of the binomial distribution, but without moving away from its intrinsic discrete character.

This paper describes an alternative procedure to estimate confidence intervals for proportion using methods of polynomial interpolation for the continuity simulation of the cumulative binomial distribution. The procedure may seem not orthodox, but the results are reliable.

Common Methods of Confidence Interval Estimation for Binomial Proportion

Wald method: It is based on the approach to the normal distribution according to the theorem of the central limit. This method has the handicap that it only offers good results for relatively big samples (bigger than 100) with the condition that n·p·q > 5. It is the simplest method and the first one that it is studied in statistic courses. (Koroliuk, 1981) (Snedecor & Cochran, 1989) (Brown, Cai, & DasGupta, 2001) (Wallis S., 2013).

The main formula is:

$$\theta = p \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

(1)

Where:

θ: population proportion

p: sample proportion

n: sample size

Z: statistic from normal distribution

α: significance level

Clopper-Pearson: It is sometimes qualified as an exact method, exact binomial intervals were originally obtained by inversion of the equal-tail binomial test using conventional P-values (Clopper & Pearson, 1934)

It is considered a conservative method because it returns confidence bounds in an interval that frequently surpasses the requirements of the significance level. (Agresti & Gottard, 2005) (Thulin, 2013)

Alternatively there are formulaic methods derived from a mathematical relationship between the binomial distribution and various continuous distributions including the beta-binomial and the *F*-distribution. It is usual the use of the following formula to estimate the confidence interval: (Agresti & Coull, 1998) (Habtzghi, Midha, & Das, 2014).

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

$$\left(1+\frac{n-x+1}{x\cdot F_{INV}\left(1-\frac{\alpha}{2},2x,2(n-x+1)\right)}\right)^{-1}<\theta<\left(1+\frac{n-x}{(x+1)\cdot F_{INV}\left(\frac{\alpha}{2},2(x+1),2(n-x)\right)}\right)^{-1} \quad (2)$$

Where:

x: success count (n·p)

n: sample size

F_{INV}: Inverse F distribution

α: significance level

This alternative is one of those more used in statistical packages, among them Statgraphics and Minitab. (StatPoint Technologies, Inc., 2015) (Minitab, Inc., 2016).

Mid-P: This method uses the discrete binomial distribution directly to calculate the confidence interval. (Neyman, 1935) (Stevens, 1950). The bounds (θ_L, θ_U) are determined as solution of the equations:

$$\sum_{k=x}^{n}\binom{n}{k}\theta_L^k(1-\theta_L)^{n-k}=\alpha/2$$

$$\sum_{k=0}^{x}\binom{n}{k}\theta_U^k(1-\theta_U)^{n-k}=\alpha/2 \quad (3)$$

Usually the method uses a procedure of successive approaches by bisection, where the indicator

$$MidP=1+\frac{B(x,n,p)}{2}-BC(x,n,p) \quad (4)$$

is used as decision criteria to approach the searched root.

The functions B() and BC() correspond to the binomial and cumulative binomial distribution respectively.

Recently the Mid-P method has received bigger attention by several investigators that recommend its election at first term. (Newcombe R., 1998) (Brown, Cai, & DasGupta, 2001) (Sullivan & Soe, 2006) (Thulin, 2013).

Other methods have been developed starting from basic normal approximation, among them the Score (Wilson, 1927), Modiffied Wald (Agresti & Coull, 1998) and Score with Continuity Correction (Fleiss Quadratic) (Fleiss, 1981). A reference example of its implementation is the one carried out by OpenEpi (Sullivan & Soe, 2006).

Continuity Simulation for Binomial Distribution

It is well-known that the binomial distribution formula is:

$$B(x,n,p) \; = \; \frac{n!}{x!(n\text{-}x)\,!}p^{x}(1-p)^{n-x} \tag{5}$$

Where n is the sample size and x the success count, integer values.

p: probability to obtain an independent success into Bernoulli experiment (expected proportion) (Walck, 1996)(NIST/SEMATECH, 2012).

The cumulative probability is determined as the sum of the exact probabilities for discrete values from zero until k.

$$BC(x \leq k,n,p) \; = \; \sum_{0}^{k}\frac{n!}{x!(n\text{-}x)\,!}p^{x}(1-p)^{n-x} \tag{6}$$

Since this distribution doesn't allow to calculate probabilities for fractional values its direct use is hindered in the determination of confidence intervals for the proportion.

To introduce fractional arguments lead to an inconsistent fact. This is equivalent to situations like: the probability that in a sample of 7 people, the half is exactly of feminine sex? This is impossible, 3.5 women can't exist, and therefore the probability is zero.

It would be interesting to calculate the hypothetical probability that would correspond to the previous situation if the binomial distribution was continuous.

The above-mentioned situation can be eased by means of a simple transformation of the binomial distribution: substituting the factorial operators for the function gamma, $\Gamma(\;)$.

$$B(x, n, p) = \frac{\Gamma(n+1)}{\Gamma(x+1)\,\Gamma(n\text{ - }x+1)}p^{x}(1-p)^{n-x} \tag{7}$$

This formula bears some resemblance with the beta distribution. It facilitates to plot the binomial distribution as a continuous curve and to obtain hypothetical probabilities for non-integer arguments (Figure 1), but it is not useful to sum them and they are not probability densities. Cumulative probabilities are interesting to the effects of the confidence bounds estimation, they require a different treatment.

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

Figure 1. Binomial distribution with continuous simulation. (n = 5, p = 0.40)
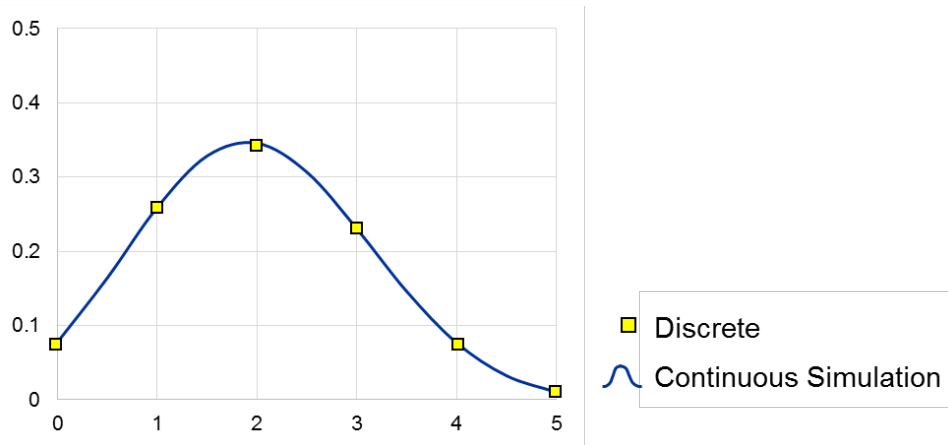
Interpolation methods can be used to obtain approximate cumulative probabilities for fractional arguments. A good approximation can be obtained using polynomial interpolation, Lagrange's method is very efficient to achieve it. (Atkinson, 1989).

The method consists on calculating the cumulative probabilities for consecutive integer values of the argument x (from 0 until n). When it is necessary to estimate the probability for a fractional value, the two immediate previous values and two immediate later are taken. These values are used to approximate the required probability by means of polynomial interpolation.

Polynomial interpolation guarantees a high approach to the probability obtained if the function was continuous and inside the interval defined by the cumulative probabilities corresponding to two consecutive integers.

Then the values x_0, x_1, x_2, x_3 and its ordinates y_0, y_1, y_2, y_3 are obtained.

Where the probability to estimate corresponds to the pair x, y, and $x_1 < x < x_2$.

The corresponding interpolator polynomial parameters are:

$$L_0 = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)}$$

$$L_1 = \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)}$$

$$L_2 = \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)}$$

$$L_3 = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} \tag{8}$$

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

The value of the ordinate (y) is determined as follows:

$$y = y_0 L_0 + y_1 L_1 + y_2 L_2 + y_3 L_3 \tag{9}$$

It is possible to define a new function of cumulative binomial distribution that returns the discrete cumulative probability for integer arguments and it carries out the polynomial interpolation for non-integer arguments. To differentiate it from the discrete distribution it will be denoted with asterisk.

$$BC^*(x \le k, n, p) = \begin{cases} \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } x \in Z \\ \text{Interpolate}(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3, x) & \text{if } x \notin Z \end{cases} \tag{10}$$

Polynomial interpolation can be also useful to obtain the inverse root of cumulative binomial distribution for a given area. The procedure requires the successive evaluation of discrete binomial distribution until to obtain a cumulative probability greater than given area. Thenthe values x_0, x_1, x_2, x_3 and its ordinates y_0, y_1, y_2, y_3 are taken, these are used to determine the expected argument valueby polynomial interpolation. It defines the «continuous» inverse function:

$$BCInv^*(Area, n, p) = \begin{cases} 1. \text{ Successive evaluation of } y = B(x, n, p) & (\text{until } y > Area) \\ 2. \text{ Obtain } x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3 \text{ value} & (y_1 < Area < y_2) \\ 3. \text{ Interpolate}(y_0, y_1, y_2, y_3, x_0, x_1, x_2, x_3, Area) & (\text{return the x value}) \end{cases} \tag{11}$$

Figure 2. Comparison among the cumulative discrete binomial distribution and the continuous simulation with polynomial interpolation (p = 0.50)
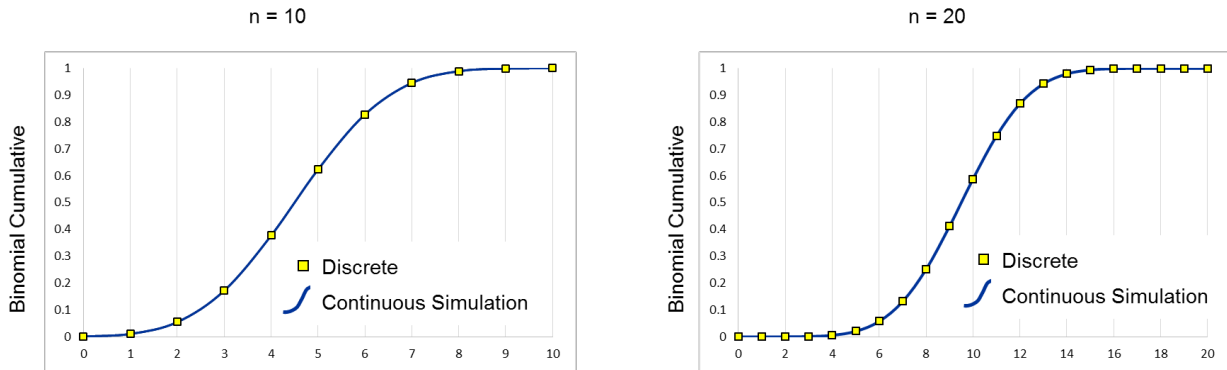
To be exact, the graphics of the binomial distribution never cut the coordinates origin, for x = 0 it always returns a probability greater than zero (see figure 1), this is remarkable mainly for small samples and small values of the parameter p (n·p < 5).

Figure 3 shows a graph about the binomial probabilities obtained for x = 0 and various n, p levels. The points represent the minimum probability available to obtain for each n, p combination, smaller probabilities are lost. The existence of positive probabilities for x = 0 grows in importance when they are

greater than α/2. It can influence the estimation of the lower confidence bound and causes that it will be truncated to zero.

Figure 3. Influence of success probability (proportion, p) and trial number (sample size, n) in absolute lower bound of binomial distribution (when x = 0).
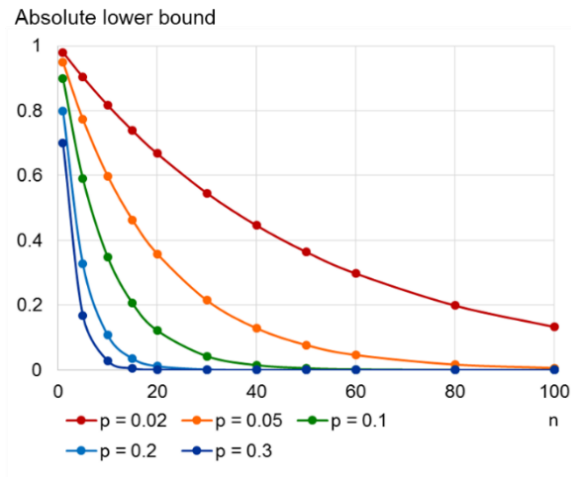
The computational cost related with the factorial calculation for the evaluation of binomial distribution for big values can be diminished using sums of logarithms or applying the Stirling's formula for x greater than 100. (Hazewinkel, 2001)

$$\log(x!) = \sum_{k=2}^{x} \log(k) \tag{12}$$

$$x! \approx \sqrt{2\pi x}\left(\frac{x}{e}\right)^x e^{\frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} - \frac{1}{1680x^7} + \cdots} \tag{13}$$

$$\log(x!) \approx \frac{\log(2\pi x)}{2} + x\log\left(\frac{x}{e}\right) + \frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} - \cdots \tag{14}$$

Obtaining of Confidence Intervals for Proportion by Continuous Simulation of Binomialdistribution

The method that is developed next should not be confused with the Score with Continuity Correction, also known as Fleiss Quadratic method (Fleiss, 1981), because it follows a different idea. Instead of introducing corrections to the formulaeto calculate the confidence interval based on the normal approach, it is carried out the simulation of continuity directly in binomial distribution.

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

Confidence Bounds

Let a sample with size n and x successes, where the proportion is determined as $p = \dfrac{x}{n}$.

Assuming this as population proportion, is it needed to determine the confidence interval with two tails for a significance level alpha (α).

The lower and upper bounds of the confidence interval will be calculated directly by means of the binomial distribution. The inverse cumulative distribution will be used for $\alpha/2$ and $1 - \alpha/2$.

As these bounds may coincide with non-integer values of the x argument, it becomes necessary to artificially extend the domain of the binomial distribution for fractional arguments in order to simulate a continuous behavior. The needful treatment of binomial distribution was treated previously.

The abbreviated procedure to obtain a confidence interval is represented in the following algorithm:

1. Read sample data: x, n, (p=x/n)
2. Settle down the significance level: α
3. Calculate the lower confidence bound (x_1) for $\alpha/2$
   $x_1 = $ BCInv* ($\alpha/2$, n, p)
4. Calculate the upper confidence bound (x_2) for $(1-\alpha/2)$
   $x_2 = $ BCInv* ($1-\alpha/2$, n, p)
5. Calculate the proportions for the confidence interval:
   $p_1 = x_1/n$ $p_2 = x_2/n$

Hypothesis Test

To contrast the sample proportion with a null hypothesis value (H_0: $p_{Ho} = \ldots$), the P-value is determined according to the following procedure:

$x_0 = p_{Ho} \cdot n$

For the two-sided test: (H_1: $p_{Ho} \neq \ldots$)

  P-value = CB* (x_0, n, p)

  if(P-value > 0.5) P-value = 1 – P-value

  P-value = 2·P-value

For the left tail test (H_1: $p_{Ho} < \ldots$)

  P-value = 1 – CB* (x_0, n, p)
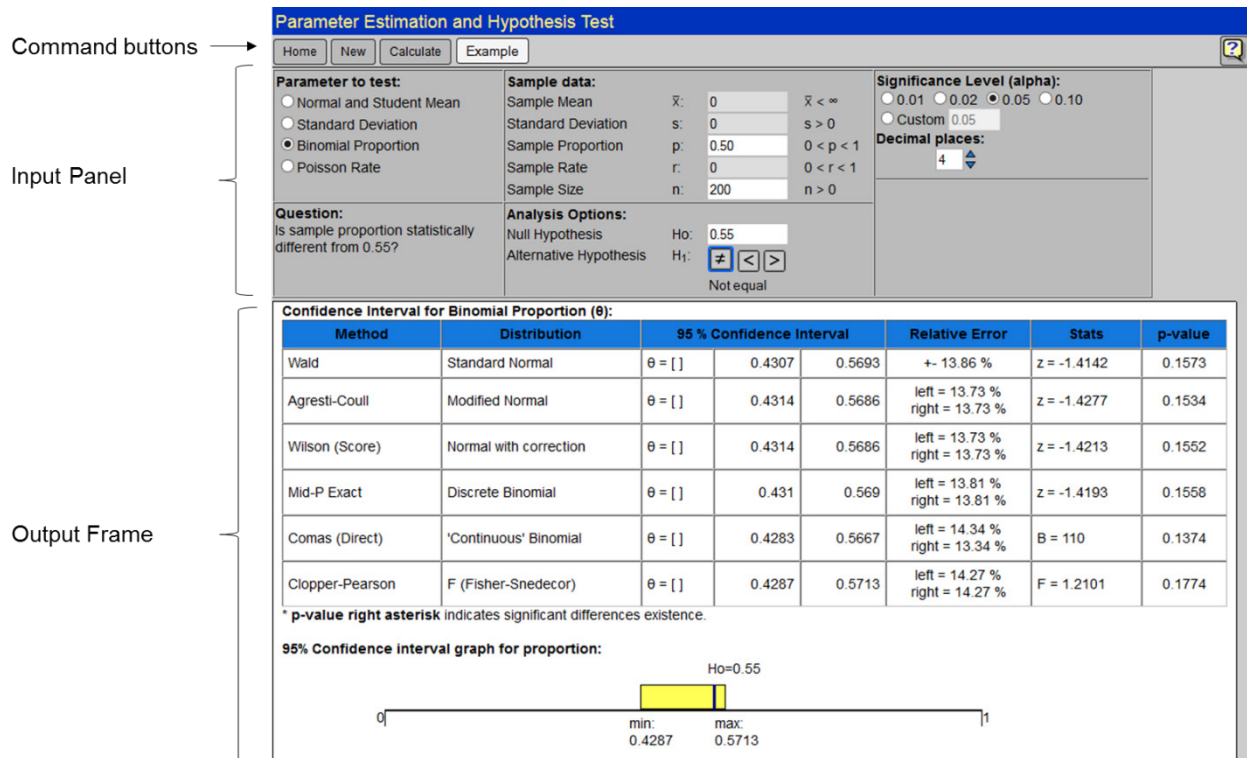
For the right tail test (H_1: $p_{Ho} > \ldots$)

  P-value = CB* (x_0, n, p)

Analysis and Discussion

The implementation of the «Direct» method was developed in open source by means of JavaScript programming and it can be found in the on-line statistical laboratory DynStats.com. It also incorporates a calculator of distribution functions with simulation of continuity for various discrete distributions. (DynStats, 2016).

Figure 4. DynStats output for the estimation of proportion confidence interval.

Parameter Estimation and Hypothesis Test

Command buttons → Home | New | Calculate | Example

Input Panel

Parameter to test:
- ○ Normal and Student Mean
- ○ Standard Deviation
- ● Binomial Proportion
- ○ Poisson Rate

Sample data:
Sample Mean	$\bar{x}$:	0	$\bar{x} < \infty$
Standard Deviation	s:	0	s > 0
Sample Proportion	p:	0.50	0 < p < 1
Sample Rate	r:	0	0 < r < 1
Sample Size	n:	200	n > 0

Significance Level (alpha):
○ 0.01 ○ 0.02 ● 0.05 ○ 0.10
○ Custom 0.05
Decimal places: 4

Question:
Is sample proportion statistically different from 0.55?

Analysis Options:
Null Hypothesis Ho: 0.55
Alternative Hypothesis H₁: ≠ | < | >
Not equal

Output Frame

Confidence Interval for Binomial Proportion (θ):

Method	Distribution	95 % Confidence Interval		Relative Error	Stats	p-value
Wald	Standard Normal	θ = [ ] 0.4307	0.5693	+- 13.86 %	z = -1.4142	0.1573
Agresti-Coull	Modified Normal	θ = [ ] 0.4314	0.5686	left = 13.73 % right = 13.73 %	z = -1.4277	0.1534
Wilson (Score)	Normal with correction	θ = [ ] 0.4314	0.5686	left = 13.73 % right = 13.73 %	z = -1.4213	0.1552
Mid-P Exact	Discrete Binomial	θ = [ ] 0.431	0.569	left = 13.81 % right = 13.81 %	z = -1.4193	0.1558
Comas (Direct)	'Continuous' Binomial	θ = [ ] 0.4283	0.5667	left = 14.34 % right = 13.34 %	B = 110	0.1374
Clopper-Pearson	F (Fisher-Snedecor)	θ = [ ] 0.4287	0.5713	left = 14.27 % right = 14.27 %	F = 1.2101	0.1774

* **p-value right asterisk** indicates significant differences existence.

95% Confidence interval graph for proportion:

Ho=0.55

0 ———————————— 1
min: 0.4287   max: 0.5713

The software calculates the confidence bounds according to the required significance level. Six methods are evaluated and shown on the table, it includes the relative error and the respective stats used to calculate the P-value. The confidence interval graph is also shown.

Various confidence intervals have been obtained for different proportions and sample sizes, the results are shown in Table 1.

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

Table 1. Examples of confidence bounds and P-value obtained by means of the Direct method.

p (sample)	n	L bound	U bound	p(H_0)	P-value
0.10	20	0	0.2206	0.15	0.2659
0.10	50	0.0164	0.1796	0.15	0.1707
0.10	100	0.0405	0.1572	0.15	0.0798
0.10	200	0.0580	0.1408	0.15	0.0190 *
0.10	400	0.0704	0.1290	0.15	0.0013*
0.20	20	0.0199	0.3612	0.15	0.8229
0.20	50	0.0859	0.3057	0.15	0.4916
0.20	100	0.1198	0.2759	0.15	0.2570
0.20	200	0.1436	0.2542	0.15	0.0860
0.20	400	0.1603	0.2386	0.15	0.0121*
0.50	20	0.2585	0.6915	0.55	0.5034
0.50	50	0.3521	0.6279	0.55	0.3965
0.50	100	0.3972	0.5928	0.55	0.2713
0.50	200	0.4283	0.5667	0.55	0.1374
0.50	400	0.4498	0.5477	0.55	0.0402 *
0.80	20	0.5888	0.9322	0.75	0.7407
0.80	50	0.6743	0.8941	0.75	0.4689
0.80	100	0.7141	0.8702	0.75	0.2627
0.80	200	0.7408	0.8514	0.75	0.0987
0.80	400	0.7589	0.8372	0.75	0.0172 *

* P-value right asterisk indicates significant differences.

The table shows the confidence intervals for four proportion levels. Five sample sizes are considered. The confidence level is established to 95% ($\alpha = 0.05$).

Observe the confidence intervals decrease as the size of the sample is increased. Likewise the P-value diminishes until being smaller than 0.05 and this coincides with the fact that the null hypothesis proportion is outside of the confidence bounds.

In the first row the lower bound is truncated to zero. This had been anticipated when the problem of the lost probabilities was treated for $n \cdot p < 5$.

Comparison

Confidence intervals were determined by means of each described method evaluating four proportion levels and five sample sizes. The results were plotted in order to obtain the curves for both confidence bounds.

Figure 5. Comparison of lower and upper confidence bounds for proportion obtained by various methods (α=0.05).
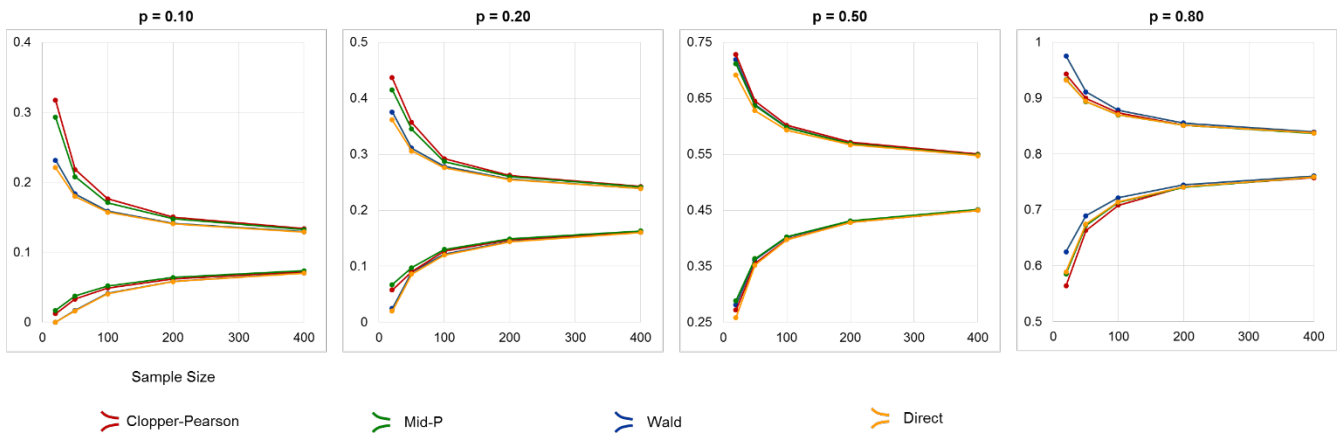
In the two first graphs (p <0.50) the Wald and Direct methods show a great similarity to each other (the curves of the lower bound are overlapped) and they return confidence bounds displaced toward zero regarding those obtained by means of Mid-P or Clopper-Pearson.

In the third graph (p = 0.50) is observed the biggest coincidence among all methods. The vertical displacement of the bounds obtained by the Direct method remains.

For p> 0.50 the confidence intervals Mid-P and Direct show a clear tendency to the overlapping and they are near the values returned by the Clopper-Pearson method. While the Wald method returns intervals displaced upwards regarding the other.

The Direct method always offers the smallest upper bound regarding the other for all the proportion levels and sample size.

The confidence bounds tend to coincide for n > 400 with absolute differences among the similar bounds less than 1%.

Conclusion

The results obtained through the proposed method are coherent with other procedures, being observed a high likeness with the Wald method for p < 0.50 and with Mid-p method for p > 0.50.

The Direct method reaches bigger certainty for proportions greater than 0.50, approaching Mid-P and Clopper-Pearson.

The Direct method is the less conservative one for the upper confidence bound. While for the lower bound it is the opposite when the proportion of the sample is smaller than 0.50.

The biggest dispersion among the confidence intervals is observed for small samples and also for small proportions, being able to obtain lower bounds equal to zero mainly when n p <5. The analyzed methods show a great coincidence for samples greater than 400, this way they can be used almost indistinctly for p> 0.10.

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

Acknowledgment

References

Agresti, A. (2003). Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research*(12), 3-21.

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

Agresti, A., & Gottard, A. (2005). Comment: Randomized Confidence Intervals and the Mid-P Approach. *Statistical Science, 20*(4), 367-371. doi:10.1214/088342305000000403

Atkinson, K. E. (1989). *Introduction to Numerical Analysis.* New York: John Wiley and Sons.

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

Borovkov, A. A. (1972). *Course of Probability Theory.* Moscow: Nauka.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*(16), 101-117.

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

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

Demidovich, B. P., & Maron, I. A. (1980). *Computacional Mathematics.* Moscow: Mir.

DynStats. (2016). *Online Statistical Laboratory*, 0.3.2. Retrieved October 10, 2016, from http://dynstats.com

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). New York: John Wiley & Sons.

Habtzghi, D., Midha, C. K., & Das, A. (2014, December). Modified Clopper-Pearson Confidence Interval for Binomial Proportion. *Journal of Statistical Theory and Applications, 13* (4), 296-310.

Hazewinkel, M. (2001). The Stirling Formula. In *Encyclopaedia of Mathematics.* New York: Springer.

Koroliuk, V. S. (1981). *Handbook of probability Theory and Mathematical Statistics.* Moscow, USSR: Mir.

Larson, H. J. (1974). *Introduction to Probability Theory and Statistical Inference.* New York: Wiley International Edition.

Minitab, Inc. (2016). MINITAB 17. Philadelphia: Pennsylvania State University.

Newcombe, R. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*. Retrieved September 15, 2016

Newcombe, R. G. (2011). Measures of location for confidence intervals for proportions. *Communications in Statistics - Theory and Methods* (40), 1743-1767.

Newcombe, R. G., & Nurminen, M. M. (2011). In defence of score intervals for proportions and their differences. *Communications in Statistics - Theory and Methods*(40), 1271-1282.

Neyman, J. (1935). On the problem of confidence limits. *The Annals of Mathematical Statistics*, 111-116. Retrieved August 25, 2016

Direct Estimation of Confidence Intervals for Proportion by Means of Continuity Simulation of Binomial Distribution

NIST/SEMATECH. (2012). *e-Handbook of Statistical Methods*. Retrieved December 15, 2015, from http://www.itl.nist.gov/div898/handbook/

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods*. Ames: Iowa University Press.

StatPoint Technologies, Inc. (2015). STATGRAPHICS Centurion XVII. Warrenton, VA, USA: StatPoint Technologies, Inc.

Stevens, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika*, 117-129.

Stuart, D. G., & Kendall, M. G. (1973). *The Advanced Theory of Statistics. Vol 2: Inference and Relationship* (Vol. 2). London: Griffin.

Sullivan, K. M., & Soe, M. M. (2006, January 5). *Documentation for Confidence Intervals for a Proportion, OpenEpi Project*. Emory University. Retrieved September 20, 2016

Thulin, M. (2013). *The cost of using exact confidence intervals for a binomial proportion*. Uppsala: Uppsala University.

Wadsworth, G. P. (1960). *Introduction to Probability and Random Variables*. New York: McGraw-Hill.

Walck, C. (1996). *Handbook on statistical distributions for experimentalists*. Stockholm University, Stockholm.

Wallis, S. (2013). Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics, 20* (3), 178-208. doi: 10.1080/09296174.2013.799918

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