One-tailed asymptotic inferences for a proportion

M. Álvarez* and A. Martín

Biostatistics, University of Granada, Spain; maria.alvarez@uvigo.es, amartina@ugr.es
**Corresponding author*

Abstract. *This paper evaluates several methods for obtaining a one-tailed confidence interval for a binomial proportion (most of them are new proposals) and comes to the conclusion that the optimal method is the Wilson's classic score method with continuity correction and with the original data (without any increases). A simpler option, though not so good as that just mentioned, is the new adjusted Wald method with the increase of data proposed by Borkowf.*

Keywords. *One-tailed asymptotic confidence interval; Continuity correction; Wald method and adjusted Wald methods; Wilson score method; Arcsine transformation.*

1 Introduction

The one-tailed confidence intervals (CI) are useful in quality control and in the non-inferiority test or superiority test of biomedical studies as mentioned by Pradhan *et al.* [6]. They must be analysed separately from the case of two-tailed because Cai [3] found that the performance of the same method may vary greatly from one case to another. Only two previous authors have specifically focused on the case of the one-tailed, but they analysed only a few methods with a 95% confidence (the first one) and with a 97.5% confidence (the second one).

For the inferences to be coherent the CI should be obtained by inverting the one-tailed test. This means two things: the first conclusion is that the definition for a method of inference can be made from the point of view of the test or from the point of view of the CI; the second conclusion is that evaluating a CI method is equivalent to evaluating its associated test method (if both are performed to the same nominal error α).

There is a simple and exact solution of the Clopper-Pearson (it is based on the distribution F-Snedecor), but it is too conservative and not a good choice in practice [3]. This is why many writers approach the problem from the approximate standpoint. They may be based on the real distribution of the variable x (the binomial distribution) or on its asymptotic distribution (usually the normal distribution). The seconds are usually simpler to apply and some of them are of great pedagogical interest.

From a Bayesian point of view, the CI depends on the distribution assigned to p. Various authors (such as [2]) select a prior beta distribution $B(0.5, 0.5)$, which results in Jeffreys' Bayesian method. This method is selected even in preference to the "quasi-exact" inferences, which is a good reason to include it in the present study.

The aim of this paper is to propose an asymptotic solution based on the normal distribution (that preserves as much as possible the objective error) and make it easier to justify. In addition, the Jeffreys' Bayesian method and the latest proposed method for the literature ([4],[7]) are included among the methods to be compared.

2 Methods for carrying out inferences

From both the exact and the approximate perspective, it is customary to obtain the one-tailed CI by inverting the one-tailed test. This one-tailed test could be a right tailed test $H : p \leq \pi$ *vs.* $K : p > \pi$ or a left tailed test $H : p \geq \pi$ *vs.* $K : p < \pi$ (with $0 < \pi < 1$). In this study, we have focused on the right tail test because the left one is equivalent. Thus, inverting the right-tailed test allows obtaining the CI by the left with the following lower limit: $p \geq p_L = \{\pi \,|\text{the previous test is not significant}\}$

2.1 Classic statistics and procedures to be used

Let $x \backsim B(n; p)$, a random binomial variable and let $y = n - x$, $\bar{p} = x/n$, $\bar{q} = 1 - \bar{p} = y/n$ and $q = 1 - p$, where p is the relevant parameter. From the point of view of the classic statistic, various statistics may be used to contrast $H : p \leq \pi$ *vs.* $K : p > \pi$ (with $0 < \pi < 1$). The more traditional ones are based on the typification of the sample proportion $z = \frac{\bar{p}-\pi}{\sqrt{p(1-p)/n}}$. The classic procedure in elementary textbooks consists of substituting $\bar{p}$ for p (an estimator which is not restricted by the null hypothesis), which yields the Wald procedure, where the test statistic is given by $z_W = \frac{\bar{p}-\pi}{\sqrt{\bar{p}(1-\bar{p})/n}}$.

Under the null hypothesis $p = \pi$, which is why Wilson proposed substituting in expression z the unknown parameter p by π, so obtaining the Wilson score procedure, where the test statistic is given by $z_S = \frac{\bar{p}-\pi}{\sqrt{\pi(1-\pi)/n}}$.

Other less traditional options are based on the arcsine transformation $z_A = 2\sqrt{n}\{sin^{-1}\sqrt{\bar{p}} - sin^{-1}\sqrt{\pi}\}$ and on the likelihood ratio test $z_R = \sqrt{2\{xln\frac{\bar{p}}{\pi} + (n-x)ln\frac{1-\bar{p}}{1-\pi}\}}$.

For the two-tailed inferences about a proportion, Guan [4] recommends the following CI (which is written for the one-tailed case): $p \geq p_L = \frac{n}{n+z_\alpha^2}\{\bar{p} + \frac{z_\alpha^2}{2n} - z_\alpha\sqrt{0.3(\frac{z_\alpha}{n})^2 + \frac{\bar{p}\bar{q}}{n}}\}$, where z_α is the $100 \cdot (1-\alpha)$ percentile of the standard normal distribution. In a more recent paper, Yu *et al.* [7] suggest the following CI (which is written for the one-tailed case): $p \geq p_L = 0.5 + \frac{n+z_\alpha^4/53}{n+z_\alpha^2}(\bar{p} - 0.5) - \frac{z_\alpha}{n+z_\alpha^2}\sqrt{n\bar{p}\bar{q} + \frac{z_\alpha^2}{4}}$.

From the Bayesian point of view, the CI depends on the distribution which is assigned to p. Brown *et al.* [2] compare various CI and selects the Jeffreys Bayesian CI for small values of n. The Jeffreys prior interval is defined as $p_L = B_\alpha(x+0.5; n-x+0.5)$, where $B_\alpha(a;b)$ is the αth quantile ($0 \leq \alpha \leq 1$) of the beta distribution with parameters a and b.

In this paper the statistics which are obtained from the logarithmic transformation like *ln* and like *logit* are excluded, because their results are not coherent in the case of two-sided [5].

In addition, when a discrete distribution is approximate by another continuous distribution, a correction for continuity (*cc* in the following) is frequently carried out; the theoretical justification for this can be seen in the works of Cox. This is what occurs in the present case in which a binomial distribution is approximate by a normal distribution. Haber proposes that a *cc* should consist of adding to and subtracting from the variable half its average jump. In the case of the statistics z_W (similarly for the statistic z_S, the *cc* will be the classic one of $c = 1/2n$. For the statistic z_A a similar procedure can be made, hence the *cc* will be $c = 3.1416/4n$.

2.2 Sample data to be used

Because the classic and simple Wald procedure does not perform as well as its traditional rivals, many writers have tried to improve using these statistics according to the original data with a given quantity h added on. The increases $h = 0.5$ (in a different context), $h = 2$ and $h = z_\alpha^2/2$ are traditional. Other possibilities for increases are those suggested in a more general context by Martín *et al.*: $h = z_\alpha^2$ if $\bar{p} = 1$ or $h = z_\alpha^2/2$ if $\bar{p} \neq 1$.

Moreover, Borkowf [1] suggested the following increase: if $\bar{p} > \pi$ in the test (or on determining p_L in the CI), then substitute $\bar{p}$ for $x/(n+1)$. And, in the case of the arcsine transformation, it is common to increase the data by $h = 3/8$ (Anscombe's transformation).

The increases above may be applied to any of the procedures defined in the previous sector yielding a large number of methods for possible inferences.

3 Evaluation of the methods for inference

In order to make a comparative evaluation of the above-mentioned methods, certain parameters must be obtained to synthesize the quality of each method. To achieve this objective, the following steps must be taken:

1. Select a triplet of values (α, π, n) from among the values: $\alpha = 1\%$, 5% and 10%; $\pi = 0.05$, 0.1 (0.1) 0.9, 0.95; $n = 20$ (20) 100 and 200.
2. Construct the critical region CR $= \{x \mid \text{the test is statistically significant}\}$ and calculate the real error of the test $\alpha^* = 100 \cdot \sum_{RC} \binom{n}{x} \pi^x (1-\pi)^{n-x}$ and the increase $\Delta\alpha = \alpha - \alpha^*$ of the nominal error with regard to the real error.
3. Calculate the value of the power $\theta = 100 \cdot$(number of points in the set CR)$/(n+1)\%$, where $(n+1)$ is the total number of points in the same space.

4. Determine whether the method "fails", that is if $\Delta\alpha \leq -1\%, -2\%$ or -4% for $\alpha = 1\%$, 5% or 10%, respectively, in order to check the number of times that the test is too liberal.
5. Calculate the total number of failures (F) and the average values of $\Delta\alpha$ ($\overline{\Delta\alpha}$) and θ ($\overline{\theta}$) in all the combinations (n, π) that are considered.

Once the results have been obtained, the optimal method will be the one that has few failures (preferably with none), a $\overline{\Delta\alpha}$ closest to 0 and a greater value of $\overline{\theta}$.

4 Conclusions

In the literature, to make inferences about one proportion, the interest is normally focused on the two-tailed confidence intervals. Surprisingly, the number of publications that have focused on one-tailed confidence intervals is comparatively limited, despite its usefulness.

In this paper several methods are evaluated of which some of them are new proposals based on classic procedures and including the most relevant published ones. Various conclusions have been obtained.

In the first place, the classic Wald method has a very poor performance, according to the results of Cai [3], but the behaviour of Jeffreys' Bayesian method is only slightly better than the classic Wilson score method and they are not recommended because of their many failures. In general terms, the optimal method is the score method with continuity correction, which is consistent with the conclusions in Pradhan *et al.* [6]. Finally, a simpler option, which behaves only slightly worse than those already mentioned, is the classic adjusted Wald method with the increase of data suggested by Borkowf [1].

Furthermore, we remark that conclusions about the optimal method for the one-tailed inferences are different from conclusions about the two-tailed inferences.

References

[1] Borkowf CB. (2006). Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failure or success. *Statistics in Medicine* **25**, 3679–3695.

[2] Brown LD, Cai TT and DasGupta A. (2001). Confidence intervals for a binomial proportion. *Statist Science* **16 (2)**, 101–133.

[3] Cai TT. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference* **131**, 63–88.

[4] Guan Y. (2012). A generalized score CI for a binomial proportion. *Journal of Statistical Planning and Inference* **142**, 785–793.

[5] Martín A. and Álvarez M. (2014). Two-tailed asymptotic inferences for a proportion. *Journal of Applied Statistics* **0 (0)** (published online), 1–14.

[6] Pradhan V., Evans J.C. and Banerjee T. (2013). Binomial confidence intervals for testing non-inferiority or superiority: a practitioner's dilemma. *Statistical Methods in Medical Research* **0 (0)** (published online), 1–11.

[7] Yu W., Gou X. and Xu W. (2014). An improved score interval with a modified midpoint for a binomial proportion. *Journal of Statistical Computation and Simulation* **84 (5)**, 1022–1038.