Journal of Statistical Planning and Inference 148 (2014) 38–48

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Review

Better than you think: Interval estimators of the difference of binomial proportions

Luke A. Prendergast, Robert G. Staudte *

Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia

ARTICLE INFO

Article history:
Received 10 October 2013
Received in revised form
19 November 2013
Accepted 21 November 2013
Available online 11 December 2013

Keywords:
Confidence contours
Kullback–Leibler divergence
Risk difference

ABSTRACT

The paper studies explicitly defined interval estimation of the difference in proportions arising from independent binomial distributions for small to moderate sample sizes. In particular, the interval proposed by Agresti and Caffo is compared with the Newcombe interval, the KMS interval of Kulinskaya, Morgenthaler and Staudte, the Wald interval and the 'Jeffreys' interval proposed by Brown and Li. Our comparative contour plot summaries empirical studies help to identify where each of the methods performs best in terms of coverage and width. For example, for very unbalanced designs we recommend the Newcombe intervals. For obtaining the nominal coverage, the KMS intervals are recommended, providing coverages nearly always between 95% and 97%. Two new summary scores for interval coverage are introduced. In addition to comprehensive empirical findings, this paper also connects the mean value of the KMS variance stabilized statistic to the Kullback–Leibler symmetrized divergence, which helps to explain the good coverage properties of the interval based on it.

Contents

* Corresponding author.
E-mail address: r.staudte@latrobe.edu.au (R.G. Staudte).

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2013.11.012

1. Introduction

Given independent, binomially distributed X_1, X_2 with respective parameters (n_1, p_1), (n_2, p_2), we study five explicitly defined confidence intervals for $\Delta = p_1 - p_2$, called the *risk difference* in the medical literature. We include the Wald interval because of its widespread usage; the interval favored by Newcombe (1998) after careful examination of 11 methods; the Agresti and Caffo (2000) interval; and the 'Jeffreys' interval proposed by Brown and Li (2005). The latter authors considered many intervals, and recommended, if simplicity and conservatism of coverage are most valued, their Jeffreys interval and that of Agresti and Caffo (2000). We also include an interval based on variance stabilization by Kulinskaya et al. (2010). All 5 intervals are simple to compute, and all but the Wald interval have already proven themselves in earlier studies to have good coverage for moderate sample sizes. The surprising thing is how well four of them perform for sample sizes down to 6.

The articles mentioned above contain extensive references and review material. They also contain summary plots of simulation studies, but these plots are for specific cross-sections of the unit square, which are less informative than the contour plots of Section 3. We also introduce two score functions that may be used to quickly assess the performance, on average, of each of the methods over the parameter space. In addition, in Section 3 we compare the power of associated two-sided tests for rejecting the null $\Delta = 0$.

In Section 4 we give further theoretical justification for the KMS interval, by relating the mean of the KMS test statistic to the Kullback–Leibler symmetrized divergence between null and alternative distributions. Finally, a brief summary of findings is reported in Section 5.

2. The five intervals under consideration

2.1. Wald interval

Let $\hat{p}_i = X_i/n_i$ define the maximum likelihood estimator (MLE) of p_i, $i=1,2$. Further let $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$ be the MLE for Δ, and $\mathrm{SE}[\hat{\Delta}] = \{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2\}^{1/2}$ the MLE of its standard error. The asymptotic normality of the studentized $\hat{\Delta}$ leads to a level-α two-sided test for $\Delta = \Delta_0$, called the Wald test. By inversion of the family of such tests indexed by Δ_0, a $100(1-\alpha)\%$ confidence interval for Δ is obtained and abbreviated $\hat{\Delta} \pm z_{1-\alpha/2}\,\mathrm{SE}[\hat{\Delta}]$. Here $z_\alpha = \Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution. This interval is simple to motivate, but it is *deficient* in coverage for finite samples, see Newcombe (1998), Agresti and Caffo (2000), Brown and Li (2005) and Table 2.

2.2. Jeffreys interval

Brown et al. (2001) showed that in the one-sample problem, assuming Jeffreys' prior Beta(1/2,1/2) leads to the good performing Bayes estimator $\hat{p}_{i,0.5} = (X_i+0.5)/(n_i+1)$ of p_i, $i=1,2$. On this basis Brown and Li (2005) suggested substituting these estimates into the two-sample Wald interval, and called this pseudo-Bayesian interval the 'Jeffreys interval'. For a genuine Bayes estimator of $(p_1 - p_2)$, see Brown and Li (2005, p. 364).

The modification of $\hat{\Delta}$ to $\hat{\Delta}_{\mathrm{J}} = \hat{p}_{1,0.5} - \hat{p}_{1,0.5}$ and its standard error do not affect the asymptotic normality of the studentized $\hat{\Delta}_{\mathrm{J}}$, because of Slutsky's Theorem (DasGupta, 2008, p. 4); therefore the nominal $100(1-\alpha)\%$ Jeffreys' confidence interval is given by

$$[L,U]_{\mathrm{J}} = \hat{\Delta}_{\mathrm{J}} \pm z_{1-\alpha/2}\left\{\frac{\hat{p}_{1,0.5}(1-\hat{p}_{1,0.5})}{n_1} + \frac{\hat{p}_{2,0.5}(1-\hat{p}_{2,0.5})}{n_2}\right\}^{1/2}. \tag{1}$$

The Wald interval can produce degenerate intervals (e.g. if $X_1 = 0 = X_2$, but this modification of it cannot, because the $\hat{p}_{i,0.5}$ are bounded away from 0 and 1. Further, because they are biased toward 1/2, the products $\hat{p}_{i,0.5}(1-\hat{p}_{i,0.5})$ will be larger than the corresponding $\hat{p}_i(1-\hat{p}_i)$, ensuring wider intervals than the Wald procedure. Other authors have studied estimators of the form $p_{i,a} = (X_i + a)/(n_i + 2a)$, see Böhning and Viwatwongkasem (2005).

2.3. Agresti–Caffo interval

Agresti and Caffo (2000) used simulation studies for n_1, n_2 in the range 10–30 and p_1, p_2 selected at random from [0,1] to show that replacing each X_i by X_i+1 and each n_i by n_i+2 in the formula for the Wald interval led to considerable improvement in coverage probabilities. The point estimator is $\hat{\Delta}_{\mathrm{AC}} = \hat{p}_{1,1} - \hat{p}_{2,1}$, where $\hat{p}_{i,1} = (X_i+1)/(n_i+2)$; this $\hat{p}_{i,1}$ is the Bayes estimator in the one-sample problem with the uniform prior. The nominal $100(1-\alpha)\%$ Agresti–Caffo confidence interval is

$$[L,U]_{\mathrm{AC}} = \hat{\Delta}_{\mathrm{AC}} \pm z_{1-\alpha/2}\left\{\frac{\hat{p}_{1,1}(1-\hat{p}_{1,1})}{n_1+2} + \frac{\hat{p}_{2,1}(1-\hat{p}_{2,1})}{n_2+2}\right\}^{1/2}. \tag{2}$$

Agresti and Caffo (2000) also promote this interval for its simplicity, which makes it readily acceptable to beginning students of statistics, as well as its overall effectiveness in achieving nominal coverage.