Published in final edited form as:

J Appl Stat. 2013 ; 40(3): 614–625. doi:10.1080/02664763.2012.750283.

Smooth bootstrap-based confidence intervals for one binomial proportion and difference of two proportions

Dongliang Wang^{a,*} and Alan D. Hutson^b

^aDepartment of Public Health and Preventive Medicine, State University of New York Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, USA

^bDepartment of Biostatistics, University at Buffalo, 3435 Main St., Buffalo, NY 14214-3000, USA

Abstract

Constructing confidence intervals (CIs) for a binomial proportion and the difference between two binomial proportions is a fundamental and well-studied problem with respect to the analysis of binary data. In this note, we propose a new bootstrap procedure to estimate the CIs by resampling from a newly developed smooth quantile function in [11] for discrete data. We perform a variety of simulation studies in order to illustrate the strong performance of our approach. The coverage probabilities of our CIs in the one-sample setting are superior than or comparable to other well-known approaches. The true utility of our new and novel approach is in the two-sample setting. For the difference of two proportions, our smooth bootstrap CIs provide better coverage probabilities almost uniformly over the interval $(-1, 1)$, particularly in the tail region as compared than other published methods included in our simulation. We illustrate our methodology via an application to several different binary data sets.

Keywords

bootstrap; quantile function; confidence interval; proportion; binary data

1. Introduction

Confidence interval (CI) estimation for a single binomial proportion and the difference between two binomial proportions are a fundamental problem in statistical practice. For instance, a large number of clinical trials are designed to investigate the response rate of a single treatment in the phase II trial setting. Similarly, in the phase III trial setting, the difference between two proportions, e.g. treatment versus control, is oftentimes of particular interest. The classic Wald CI, which is derived based on an asymptotic normal approximation for the sample proportions, is covered in most elementary statistical textbooks. Its relative popularity is due to the fact that it is easy to motivate, present and compute. However, the Wald CI has poor coverage probabilities under many scenarios, particularly when the sample size is small or the true proportion is close to 0 or 1; see details in [1].

*Corresponding author. wangd@upstate.edu.

A number of alternative CIs have been proposed for a single binomial proportion, including the exact interval [6], the Wilson score interval [15], the Agresti–Coull (A–C) interval [2] and the Jeffreys interval, to name a few. Newcombe [11] recommended CIs based on the exact tail areas and the Wilson score method in terms of minimizing the distance of either the minimum or the mean empirical coverage probability to the minimal $(1 - \alpha) \times 100\%$ level. Brown *et al.* [5] recommended the Wilson interval or equal-tails Jeffreys prior interval for small n and the A–C interval for large n . For the difference between two proportions, Newcombe [12] conducted a thorough comparison across 11 methods. Lin *et al.* [10] introduced a bootstrap procedure where the bootstrap distribution of the pseudo-observations are fully specified. The above-mentioned intervals are briefly described in Section 4.

The goal of this note is to propose a smooth bootstrap procedure to obtain CIs for a single binomial proportion or the difference between two proportions via resampling from the smooth discrete data-based sample quantile function estimator defined by Wang and Hutson [14]. As a background, we first describe the general definition of the smooth quantile function-based approach. Towards this end, consider a discrete random variable X with probability mass function $\Pr(X = x_{j:d}) = p_j$, where $x_{j:d}$ ($1 < d < \infty$) is the j th smallest distinct value. Wang and Hutson [14] introduced a smooth definition of population quantile function for X , given a fixed finite d as

$$Q_X(u) = \sum_{j=1}^d w_{j(u)} x_{j:d}, \quad (1)$$

where

$$w_{j(u)} = B_{d'u, d'(1-u)}(P_j) - B_{d'u, d'(1-u)}(P_{j-1}) \quad (2)$$

with $d' = d + 1$, $P_j = \sum_{i=1}^j p_i$, $P_0 = 0$ and $B_{p,q}(\cdot)$ is the beta cumulative distribution function with two shape parameters p and q . This novel approach allows us to distinguish the u th quantiles across a family of discrete distributions over the whole interval $u \in (0, 1)$. Furthermore, Wang and Hutson [14] demonstrated that the corresponding sample quantile estimators, obtained via plugging in the sample frequency of distinct values into Equation (1), possess proper asymptotic properties in terms of consistency and asymptotic normality properties.

The article is organized as follows. In Section 2, we illustrate how to utilize our smooth quantile function estimator in order to develop the corresponding bootstrap procedure for constructing a $100(1 - \alpha)\%$ CI for a single binomial proportion. We then extend this method to cases for constructing a CI for the difference of two proportions in Section 3. In Section 4, simulation results are presented for evaluating the coverage probabilities of our smooth bootstrap CIs in comparison with those of the CIs mentioned above. In Section 5, the proposed approach is applied to several different binary data sets.

2. A smooth bootstrap CI for one proportion

We start this section by restating the smooth quantile function defined by Wang and Hutson [14] for a Bernoulli distribution. Given a random variable X from a Bernoulli distribution $\text{Ber}(\pi)$, the smooth quantile function of X is defined as

$$Q_X(u) = 1 - B_{3u, 3(1-u)}(1 - \pi), \quad (3)$$

where $B_{3u, 3(1-u)}$ is the beta cumulative distribution function with two shape parameters $3u$ and $3(1-u)$. Equation (3) is a special case of the smooth quantile function in Equation (1) defined in [14] with $d = 2$.

Consider a random sample $\mathbf{X} = X_1, X_2, \dots, X_n$ from Bernoulli distribution $\text{Ber}(\pi)$. The sample counterpart of the quantile function in Equation (3) can be readily obtained by plugging $\hat{\pi}$, that is,

$$\hat{Q}_X(u) = 1 - B_{3u, 3(1-u)}(1 - \hat{\pi}), \quad (4)$$

where $\hat{\pi}$ can be any class of sample proportion estimators. The choice of the sample proportion estimator will be further discussed later on. The large sample properties of the sample quantiles defined above can be readily obtained by modifying Theorem 4.1 reported in [14]. Towards this end, we have the following:

Corollary 2.1

Let $K = \Gamma(d')/\Gamma(d'u)\Gamma(d'(1-u))$. Then, as $n \rightarrow \infty$,

$$\hat{Q}_X(u) \xrightarrow{p} Q_X(u) \quad (5)$$

$$n^{1/2}(\hat{Q}_X(u) - Q_X(u)) \sim \text{AN}(0, \sigma^2),$$

where $\sigma^2 = K^2 \pi^{6(1-u)-1} (1 - \pi)^{6u-1}$.

In reality, σ^2 can be estimated readily by substituting $\hat{\pi}$ for π .

In this note, we aim to construct a CI of the binomial proportion π via a bootstrap procedure using the smooth sample quantile function in Equation (4). Towards this end, we first derive the relationship between the proportion of interest π and the population mean $\mu_Q(x)$ as

$$\mu_Q(x) = EX|Q=1 - \int_0^{1-\pi} \frac{t^3 - (1-t)^3}{3t(1-t)\log(t/(1-t))} dt, \quad (6)$$

where $Q(\cdot)$ is the quantile function defined in Equation (3). Thus, given a pseudo-set of data $X_1^* = x_1^*, X_2^* = x_2^*, \dots, X_n^* = x_n^*$ generated from Equation (4), the bootstrapped proportion estimate π^* can be readily obtained from the sample mean $\mu^* = \bar{x}^*$ by numerically inverting the function (6). We further approximate the inversion of Equation (6) by a cubic B-spline regression, which is fitted by using the *smooth.spline* function in *R* with a generalized cross-

validation method; see more details for B-spline regression in [8]. The number of knots is manually selected until the cubic B-spline function provides a satisfactory approximation for the relationship between π and $\mu_Q(x)$ as shown in Figure 1. Thus, alternatively, given a single value of $\mu^* = x^*$ calculated from a bootstrap sample

$X^* = (X_1^* = x_1^*, X_2^* = x_2^*, \dots, X_n^* = x_n^*)$, we can readily estimate π^* by

$$\hat{\pi}^* = g(\bar{x}^*), \quad (7)$$

where $g(\cdot)$ is a standard cubic B-spline function with knots (0, 0, 0, 0, 0.293, 0.498, 0.699, 1, 1, 1, 1) and coefficients (0.005, -0.033, 0.159, 0.495, 0.835, 1.033, 0.996).

It is worth noting that the performance of our method is dependent on the estimator of π in that the sample quantile function (4) is solely determined by $\hat{\pi}$. For instance, our bootstrap method is meaningless when $\hat{\pi} = 0$ or 1, which is highly likely to occur for the maximum likelihood estimator (MLE) $\hat{\pi}_{MLE} = x$ when π is close to the boundary of the interval (0, 1). To address this deficiency, we consider alternative estimators of π , such as the median unbiased estimator (MUE). The MUE $\hat{\pi}_{MUE}$ of a binomial proportion was first defined in [9] as

$$\hat{\pi}_{MUE} = \frac{1}{2}(\hat{\pi}_{MUE}^L + \hat{\pi}_{MUE}^R), \quad (8)$$

where $\hat{\pi}_{MUE}^L$ and $\hat{\pi}_{MUE}^R$ are the cut-offs that satisfy $\Pr(\bar{X} \geq \bar{x} | \pi = \hat{\pi}_{MUE}^L) \geq 0.5$ and

$\Pr(\bar{X} \leq \bar{x} | \pi = \hat{\pi}_{MUE}^R) \leq 0.5$, respectively. Particularly, $\hat{\pi}_{MUE} = \frac{1}{2}(1 - 0.5^{1/n})$ when $X = 0$;

$\hat{\pi}_{MUE} = \frac{1}{2}(1 + 0.5^{1/n})$ when $X = 1$. The MUE has been also used in [10] for their exact bootstrap algorithm.

Another reasonable alternative deserved to be considered is the centre of the Wilson score interval as this method has been repeatedly recommended for interval estimation of a binomial proportion, including Newcombe [11] and Brown *et al.* [5]. Details about the derivation of Wilson score interval are provided in Section 4 in order to facilitate the comparison between different interval estimators. Our simulation results show that the utilization of this estimator yields wider CIs with more conservative coverage probability in comparison with the utilization of MUE (data not shown). Thus, for the rest of this note, we focus on the application of our bootstrap procedure in combination with the MUE.

In summary, our proposed smooth bootstrap $100(1 - \alpha)\%$ CI for one binomial proportion given a binary data set $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ can be calculated based on the following algorithm.

1. Calculate the sample proportion estimator $\hat{\pi}$ of π . A variety of estimators are available but we recommend the MUE in Equation (8) based on our unpublished simulation results.
2. For $b = 1, \dots, B$.

- Generate n uniformly distributed observations $U_1^*=u_1^*, U_2^*=u_2^*, \dots, U_n^*=u_n^*$ from $\text{unif}(0, 1)$.
 - Calculate the corresponding quantiles $X_i^*=x_i^*$ by using the estimated quantile function $x_i^*=Q_X(u_i^*|\hat{\pi})$ in Equation (4) and calculate the mean $\bar{x}^*$ of $X_1^*=x_1^*, X_2^*=x_2^*, \dots, X_n^*=x_n^*$
 - Calculate the proportion estimate by numerically inverting Equation (6) or by using the cubic B-spline function defined in Equation (7).
3. Calculate the $100(\alpha/2)$ percentile $\hat{\pi}_{\alpha/2}^*$ and the $100(1 - \alpha/2)$ percentile $\hat{\pi}_{1-\alpha/2}^*$. Then, $(\hat{\pi}_{\alpha/2}^*, \hat{\pi}_{1-\alpha/2}^*)$ is the $100(1 - \alpha)\%$ CI of π .

3. A smooth bootstrap CI for the difference between two proportions

Our bootstrap method for one proportion can be readily extended to the construction of $100(1 - \alpha)\%$ CI for the difference between two proportions. Let $X_1 = x_1, X_2 = x_2, \dots, X_{n_1} = x_{n_1}$ and $Y_1 = y_1, Y_2 = y_2, \dots, Y_{n_2} = y_{n_2}$ be two binary samples from Bernoulli distributions $\text{Ber}(\pi_1)$ and $\text{Ber}(\pi_2)$, respectively. The $100(1 - \alpha)\%$ CI for the difference $\delta = \pi_1 - \pi_2$ can be calculated based on the following algorithm.

1. Calculate the sample proportion estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ of π_1 and π_2 , respectively. Again, we recommend the MUE defined in Equation (8).
2. For $b = 1, \dots, B$.
 - Generate n_1 and n_2 uniformly distributed observations $U_i^*=u_i^*, 1 \leq i \leq n_1$ and $U_j^*=u_j^*, 1 \leq j \leq n_2$ from $\text{unif}(0, 1)$, respectively.
 - Calculate the corresponding quantiles $X_i^*=x_i^*, 1 \leq i \leq n_1$ and $Y_j^*=y_j^*, 1 \leq j \leq n_2$ by using the estimated quantile function in Equation (4) with respect to $\hat{\pi}_1$ and $\hat{\pi}_2$, respectively.
 - Calculate the mean $\bar{x}^*=(1/n_1)\sum_{i=1}^{n_1}x_i^*$ and $\bar{y}^*=(1/n_2)\sum_{j=1}^{n_2}y_j^*$. Calculate $\delta^* = g(\bar{x}^*) - g(\bar{y}^*)$, where $g(\cdot)$ is defined in Equation (7).
3. Calculate the $100(\alpha/2)$ percentile $\bar{\delta}_{\alpha/2}^*$ and the $100(1 - \alpha/2)$ percentile $\bar{\delta}_{1-\alpha/2}^*$. Then, $(\bar{\delta}_{\alpha/2}^*, \bar{\delta}_{1-\alpha/2}^*)$ is the $100(1 - \alpha)\%$ CI of $\delta = \pi_1 - \pi_2$.

Examples for the cases of one proportion and the difference between two proportions are provided in a later section.

4. Simulation study

Simulations studies are performed to evaluate performance of our smooth bootstrap-based CIs in comparison with several other well-known CIs.

4.1 One proportion

For one binomial proportion, the following methods are considered for comparison.

1. The Wald (Wald) CI is defined as $\hat{\pi} \pm z_{\alpha/2} \sqrt{\hat{\pi}(1 - \hat{\pi})/n}$, where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution. The interval is so called as the Wald interval as it is derived through Wald's large sample test.
2. The Wilson score (Wilson) interval is defined as

$$\frac{2n\hat{\pi} + z_{\alpha/2}^2 \pm z_{\alpha/2} \sqrt{4n\hat{\pi}(1 - \hat{\pi}) + z_{\alpha/2}^2}}{2n + z_{\alpha/2}^2}.$$

This interval is the inversion of the score test to the two-sided test $H_0 : \pi = \pi_0$ and can be derived by solving $|(\hat{\pi} - \pi) / \sqrt{\pi(1 - \pi)/n}| \leq z_{\alpha/2}$ for π ; see more details in [15].

3. The Wilson score (Wilson-CC) interval with continuity correction is derived by solving $|\pi| - 1/(2n) \leq z_{\alpha/2} \sqrt{\pi(1 - \pi)/n}$ for π ; see close-form expressions in [7].
4. The exact interval is derived by using 'exact' binomial tail areas; see more details in [6].
5. The A-C interval is derived by adding four additional pseudo-observations (two successes and two failures) and substituting the new data set into Wald interval; see more details in [2].
6. The Jeffreys (Jeffreys) prior interval is defined as the $1 - \alpha$ central probability interval of a beta distribution $B(n\hat{\pi} + 0.5, n - n\hat{\pi} + 0.5)$. This can be derived by imposing a prior beta distribution $B(0.5, 0.5)$ on π from a Bayesian point of view; see more discussions about Jeffreys interval in [4].

In our simulation study, we generate 1000 simulations for sample size of $n = 5, 10, 20, 40, 80$ and $\pi = 0.05-0.5$ by 0.05. The 95% CIs of π are calculated based on our bootstrap method and all other methods mentioned above. The empirical coverage probabilities for all combinations of n and π are presented in Figure 2. We further summarize the performance of each method upon different n , π and $n\pi$ in Table 1 (unaggregated tables are available upon request). According to the large sample theory, empirical coverage probability p follows an

asymptotic normal distribution such as $(\hat{p} - p) / \sqrt{p(1 - p)/n} \sim \text{AN}(0, 1)$, where p is the true coverage probability and n is the number of simulations. Thus, an observed empirical coverage probability lying between 0.940 and 0.960 can be considered as satisfactory at the 95% nominal confidence level. The first observation from Figure 2 and Table 1 is that the Wald method has absolutely worst performance among all methods, particularly when either n or π is small. Second, no method is uniformly better than the others. Under many parameter combinations with different n and π , our smooth bootstrap-based CIs provide superior performance in terms of the absolute distance from coverage probability to the nominal level in comparison with other methods.

The overall performance of our bootstrap method, as shown in column 1 of Table 1, is comparable to those of the Wilson score intervals and the Jeffreys prior intervals and much

better than the Wald method. Our bootstrap intervals are slightly more conservative than the Wilson score intervals and the Jeffreys prior intervals for small n , π and $n\pi$. We also observe that the CIs derived from the exact method and the Wilson method with continuity correction are more conservative than the others. The observed excessive conservativeness of the Clopper–Pearson exact CI is consistent with published literature. Agresti and Coull [2] even ‘believe it is inappropriate to treat this approach as optimal for statistical practice’.

In Table 1, we also provide the overall average length with standard deviation for individual CIs. If two CIs have similar coverage probabilities, the one with the smallest average length should be recommended. The Wald CI has the shortest length but it is not preferred because of its poor coverage probability. The Score-CC, CI and the Exact CI are the widest, which is the reason to their conservativeness. The length of our proposed bootstrap CI is overall comparable to the Score, A–C and Jeffreys’ CI and slightly wider for small sample size. It is worth noting that under several situations, the standard deviation of our bootstrap method is a bit larger than the other CIs. The extra variance could be due to the error arising from Monte Carlo resampling.

4.2 Difference of two proportions

For the difference of two proportions, we compare our smooth bootstrap method with the following methods:

1. The standard Wald (Wald) CI is defined as

$$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}},$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution.

2. The Wilson score CI with continuity correction (Score-CC) is defined as $(\hat{\pi}_1 - a, \hat{\pi}_1 + b)$ where

$$a = z_{\alpha/2} \sqrt{\frac{l_1(1 - l_1)}{n_1} + \frac{u_2(1 - u_2)}{n_2}},$$

$$b = z_{\alpha/2} \sqrt{\frac{u_1(1 - u_1)}{n_1} + \frac{l_2(1 - l_2)}{n_2}}.$$

Here l_i and u_i are the values delimiting the interval

$$\left\{ \pi_i : \left| \pi_i - \sum_{i=1}^{n_i} \frac{y_i}{n_i} \right| - \frac{1}{2n_i} \leq z_{\alpha/2} \sqrt{\frac{\pi_i(1 - \pi_i)}{n_i}} \right\}, i=1, 2.$$

The Score-CC interval is an extension of Wilson score method for one proportion [7].

3. The A–C interval is derived by adding four additional pseudo-observations (two successes and two failures) in each group and substituting the new data set into Wald interval; see more details in [2].
4. The exact bootstrap (ExactB–MUE) CI is derived where the bootstrap distribution of δ^* is fully specified; see more details in [10].

In our simulation study, we generate 1000 simulations for sample size of $n_1 = 5, 10, 40, n_2 = 5, 10, 40, \pi_1 = 0.05\text{--}0.95$ by 0.05 and $\pi_2 = 0, 1, 0.3, 0.5$. The 95% CIs of $\delta = \pi_1 - \pi_2$ are calculated based on our bootstrap method and all other methods mentioned above. The simulation results are presented in Figures 3–5 for $\pi_2 = 0.1, 0.3, 0.5$, respectively (unaggregated tables are available upon request). The Wald method has obviously worst performance. The Wilson score interval with continuity correction improves the performance a bit. The A–C method provides conservative intervals. The exact bootstrap method is comparable to our smooth bootstrap method when δ is close to 0. However, we realize that the exact bootstrap method could meet a problem when $|\delta|$ is close to 1. For instance, the coverage probability of the exact bootstrap method can be 0 as given in Table 2 for the case of $\pi_1 = 1$ and $\pi_2 = 0.1$ with $\delta = 0.9$. For illustration, we calculate the CIs derived

from different method given all combinations of $\sum_{i=1}^5 x_i = 5$ and $\sum_{j=1}^5 y_j = 0$ to 5 by 1, as given in Table 3. None of the exact bootstrap CIs covers the true $\delta = 0.9$. We further

calculate the bootstrap distribution of δ^* in Table 4 given $\sum_{i=1}^5 x_i = 5$ and $\sum_{i=1}^5 x_i = 0$. It is clear that under this setting, the maximum of δ^* is 0.871 so that the exact bootstrap CIs will never cover the true $\delta = 0.9$.

5. Example

We utilize the examples presented in [11] to illustrate our bootstrap method in comparison with other methods briefly described in Section 4 for the one proportion case. These data are originally introduced in [3,13]. Table 5 lists 95% CIs for four combination of n and r calculated by different methods. For $r/n = 81/263$ with $\pi \approx 0.3$, all methods yield acceptable CIs. For $r/n = 15/148$ with $\pi \approx 0.1$, all intervals are still comparable, since r is pretty large. It is crucial to choose the right intervals for the last two cases $r/n = 0/20$ and $r/n = 1/29$ with a very small π . The Wald method gives a clearly short interval. But as shown in Figure 2, it is associated with a significantly low empirical coverage probability. The CIs derived from our bootstrap method are closer to 0 than all other intervals. We utilize the example included in [10] to illustrate our smooth bootstrap method for the two proportions case. The 95% CIs for the difference between two proportions regarding either the primary tumour shrinkage endpoint or the toxicity endpoint are presented in Table 6. The A–C method tends to provide more conservative intervals. The Wald methods, either with or without continuity correction, yield non-reasonable intervals when no success is observed. The CIs derived from our smooth bootstrap method are a bit wider than those from exact bootstrap method.

6. Summary and discussion

In this note, we develop a novel method to construct CIs for one binomial proportion and the difference between two proportions. The CIs are derived through the classic bootstrap principle where the pseudo-observations are resampled from a smooth quantile function defined by Wang and Hutson [14]. The definition of smooth quantile function enables us to differentiate Bernoulli distributions from the whole interval (0, 1) and the sample counterpart possesses remarkable asymptotic properties.

The performance of our method is examined via simulation studies. In the case of one binomial proportion, our method is superior than the Wilson score method and the Jeffreys prior methods under a variety of configurations of n and π and overall it is comparable to them. In the case of the difference between two proportions, our method is almost dominantly the best among all methods included in our simulation, particularly when $|\delta|$ is close to 1 and the sample size is small. The execution of our method requires utilization of standard cubic B-spline function.

Acknowledgments

The authors thank the Executive Editor and the two anonymous reviewers for their helpful suggestions.

References

1. Agresti, A. Categorical Data Analysis. New York: Wiley; 2002.
2. Agresti A, Coull BA. Approximate is better than 'exact' for interval estimation of binomial proportions. *Amer. Statist.* 1998; 52:119–126.
3. Altman, DG. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
4. Berger, JO. Statistical Decision Theory and Bayesian Analysis. 2nd. New York: Springer; 1985.
5. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statist. Sci.* 2001; 16:101–133.
6. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934; 26:404–413.
7. Fleiss, JL. Statistical Methods for Rates and Proportions. 2nd. New York: Wiley; 1981.
8. Hastie, TJ.; Tibshirani, RJ. Generalized Additive Models. New York: Chapman and Hall; 1990.
9. Hirji KF, Tsiatis AA, Mehta CR. Median unbiased estimation for binary data. *Amer. Statist.* 1989; 43:7–11.
10. Lin Y, Newcombe RG, Lipsitz S, Carter RE. Fully specified bootstrap confidence intervals for the difference of two independent binomial proportions based on the median unbiased estimator. *Stat. Med.* 2009; 28:2876–2890. [PubMed: 19691015]
11. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* 1998; 17:857–872. [PubMed: 9595616]
12. Newcombe RG. Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat. Med.* 1998; 17:873–890. [PubMed: 9595617]
13. Turnbull PJ, Stimson GV, Dolan KA. Prevalence of HIV infection among ex-prisoners in England. *BMJ.* 1992; 204:90–91. [PubMed: 1737147]
14. Wang D, Hutson AD. A fractional order statistic towards defining a smooth quantile function for discrete data. *J. Statist. Plann. Inference.* 2011; 141:3142–3150.
15. Wilson EB. Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc.* 1927; 22:209–212.

Figure 1.
The relationship between $\mu_Q(X)$ and π via B-spline fit.

Figure 2.
Empirical coverage probabilities of a variety of CIs for one proportion.

Figure 3. Empirical coverage probabilities of a variety of CIs for $\pi_1 - \pi_2$ given $\pi_2 = 0.1$.

Figure 4.

Empirical coverage probabilities of a variety of CIs for $\pi_1 - \pi_2$ given $\pi_2 = 0.3$.

Figure 5.
Empirical coverage probabilities of a variety of CIs for $\pi_1 - \pi_2$ given $\pi_2 = 0.5$.

Table 1
Empirical coverage probabilities and average lengths of CIs for a single binomial proportion π .

Method	Coverage probability											
	All	n	10	n	40	π	0.1	0.4	π	0.5	$n\pi$	40
Boot-MUE	0.957	0.965	0.965	0.951	0.977	0.947	0.960	0.945	0.960	0.945	0.945	0.945
Wald	0.846	0.742	0.742	0.930	0.682	0.920	0.814	0.944	0.814	0.944	0.944	0.944
Score	0.953	0.956	0.956	0.950	0.941	0.957	0.954	0.944	0.954	0.944	0.944	0.944
Score-CC	0.976	0.985	0.985	0.967	0.984	0.973	0.980	0.968	0.980	0.968	0.968	0.968
Exact	0.976	0.987	0.987	0.968	0.983	0.971	0.981	0.966	0.981	0.966	0.966	0.966
A-C	0.960	0.962	0.962	0.955	0.963	0.962	0.963	0.949	0.963	0.949	0.949	0.949
Jeffreys	0.954	0.956	0.956	0.953	0.975	0.948	0.957	0.944	0.957	0.944	0.944	0.944
Average length (SD)												
Method	All	n	10	n	40	π	0.1	0.4	π	0.5	$n\pi$	40
Boot-MUE	0.38 (0.19)	0.55 (0.16)	0.55 (0.16)	0.22 (0.064)	0.26 (0.16)	0.44 (0.19)	0.43 (0.20)	0.22 (0.0047)	0.43 (0.20)	0.22 (0.0047)	0.22 (0.0047)	0.22 (0.0047)
Wald	0.34 (0.22)	0.46 (0.28)	0.46 (0.28)	0.21 (0.068)	0.16 (0.17)	0.44 (0.20)	0.37 (0.24)	0.22 (0.0040)	0.37 (0.24)	0.22 (0.0040)	0.22 (0.0040)	0.22 (0.0040)
Score	0.35 (0.15)	0.51 (0.10)	0.51 (0.10)	0.21 (0.057)	0.27 (0.14)	0.40 (0.15)	0.40 (0.15)	0.21 (0.0038)	0.40 (0.15)	0.21 (0.0038)	0.21 (0.0038)	0.21 (0.0038)
Score-CC	0.40 (0.18)	0.59 (0.12)	0.59 (0.12)	0.23 (0.060)	0.32 (0.18)	0.45 (0.18)	0.46 (0.18)	0.22 (0.0042)	0.46 (0.18)	0.22 (0.0042)	0.22 (0.0042)	0.22 (0.0042)
Exact	0.41 (0.20)	0.60 (0.14)	0.60 (0.14)	0.23 (0.064)	0.31 (0.18)	0.47 (0.19)	0.47 (0.20)	0.22 (0.0039)	0.47 (0.20)	0.22 (0.0039)	0.22 (0.0039)	0.22 (0.0039)
A-C	0.37 (0.16)	0.53 (0.094)	0.53 (0.094)	0.22 (0.055)	0.30 (0.16)	0.41 (0.15)	0.42 (0.16)	0.21 (0.0038)	0.42 (0.16)	0.21 (0.0038)	0.21 (0.0038)	0.21 (0.0038)
Jeffreys	0.36 (0.17)	0.51 (0.14)	0.51 (0.14)	0.21 (0.061)	0.25 (0.15)	0.41 (0.16)	0.40 (0.17)	0.21 (0.0039)	0.40 (0.17)	0.21 (0.0039)	0.21 (0.0039)	0.21 (0.0039)

Table 2
Empirical coverage probabilities of CIs of $\delta = \pi_1 - \pi_2$ at a nominal level of $\alpha = 0.05$ given $\pi_1 = 1$ and $\pi_2 = 0.1$.

n_1	n_2	Boot-MUE	EB-MUE	Wald	Score-CC	A-C
5	5	0.989	0.000	0.412	0.999	0.579
5	10	0.951	0.000	0.636	0.999	0.736
5	40	0.954	0.402	0.912	1.000	0.790
10	5	0.994	0.000	0.405	1.000	0.589
10	10	0.987	0.927	0.632	0.634	0.744
10	40	0.955	0.796	0.905	0.984	0.796
40	5	0.99	0.99	0.406	1.000	0.914
40	10	0.99	0.928	0.687	0.687	0.928
40	40	0.968	0.953	0.912	0.915	0.910

Table 3

The 95% CIs for $\delta = \pi_1 - \pi_2$ given $n_1 = n_2 = 5$, $\pi_1 = 1$ and $\pi_2 = 0.1$.

	0	1	2	3	4	5
$\sum_{i=1}^{n_2} y_i$	0.59	0.328	0.0729	0.0081	0.00045	1e-05
$\Pr(\sum_{i=1}^{n_2} y_i \pi_2)$						
Boot-MUE	(0.401, 1.000)	(0.181, 0.985)	(0.000, 0.904)	(-0.149, 0.790)	(-0.266, 0.595)	(-0.345, 0.337)
ExactB-MUE	(0.384, 0.863)	(0.178, 0.851)	(0.000, 0.796)	(-0.157, 0.714)	(-0.185, 0.477)	(-0.342, 0.198)
Wald	(1.000, 1.000)	(0.449, 1.000)	(0.171, 1.000)	(-0.029, 0.829)	(-0.151, 0.551)	(0.000, 0.000)
Score-CC	(0.800, 1.000)	(0.249, 1.000)	(-0.029, 1.000)	(-0.229, 1.000)	(-0.351, 0.751)	(0.000, 0.000)
A-C	(0.181, 0.950)	(0.040, 0.865)	(-0.086, 0.765)	(-0.199, 0.652)	(-0.299, 0.525)	(-0.384, 0.384)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4
Bootstrap distribution of δ^* given a simulated data set $(\sum_{i=1}^{n_1=5} x_i=5, \sum_{i=1}^{n_2=5} y_i=0)$.

Order	$\sum_{i=1}^{n_1=5} x_i$	$\sum_{i=1}^{n_2=5} y_i$	δ^*	PMF	CDF
1	0	5	- 0.871	1.29e-12	1.29e-12
2	0	4	- 0.714	9.32e-11	9.45e-11
:	:	:	:	:	:
:	:	:	:	:	:
24	4	3	0.185	0.000587	0.000922
25	2	1	0.185	0.000587	0.00151
:	:	:	:	:	:
:	:	:	:	:	:
35	5	1	0.714	0.177	0.488
36	5	0	0.871	0.512	1

Table 5The 95% CIs for π .

<i>n</i>	263	148	20	29
<i>r</i>	81	15	0	1
Boot-MUE	(0.255, 0.365)	(0.058, 0.157)	(0.000, 0.111)	(0.000, 0.138)
Wald	(0.252, 0.364)	(0.053, 0.150)	(0.000, 0.000)	(0.000, 0.101)
Score	(0.255, 0.366)	(0.062, 0.160)	(0.000, 0.161)	(0.006, 0.172)
Score-CC	(0.254, 0.368)	(0.060, 0.164)	(0.000, 0.200)	(0.0012, 0.196)
Exact	(0.253, 0.368)	(0.058, 0.162)	(0.000, 0.168)	(0.001, 0.178)
A-C	(0.255, 0.366)	(0.061, 0.161)	(0.000, 0.190)	(0.000, 0.186)
Jeffreys	(0.255, 0.366)	(0.060, 0.158)	(0.000, 0.117)	(0.004, 0.150)

Table 6

The 95% CIs for $\delta = \pi_1 - \pi_2$.

	$\sum_{i=1}^{n_1=14} x_{i=0}$	$\sum_{i=1}^{n_1=14} x_{i=2}$
	$\sum_{i=1}^{n_1=11} y_{i=0}$	$\sum_{i=1}^{n_1=11} y_{i=1}$
Boot-MUE	(− 0.163, 0.143)	(− 0.226, 0.308)
ExactB-MUE	(− 0.168, 0.118)	(− 0.201, 0.260)
Wald	(0.000, 0.000)	(− 0.198, 0.302)
Score-CC	(0.000, 0.000)	(− 0.250, 0.354)
A–C	(− 0.247, 0.203)	(− 0.257, 0.301)