Using Bootstrap for More Accurate Confidence Interval

Yuanjing Ma

Abstract: This article investigates the bootstrap methods for producing good approximate confidence intervals. The first section introduces the motivated example and compare different approximated confidence intervals with the exact intervals. Section two serves as an overview of different bootstrap methods. Inspired by Efron's paper (1994), this article further explores the possibility to apply the bootstrap method to common odds ratio in a series of 2 × 2 contingency tables. The third section explains the procedure and gives the results of Monte Carlo simulations.

1. Introduction

Confidence interval is a useful data-analytic tool in statistical research field. It combines point estimation and hypothesis testing, giving us important information on how confident we are to make inference about the parameters that we care about. Therefore, construction of more accurate confidence intervals are desperately required for decision making and risk management. This article discusses application of bootstrap methods in making refinement on coverage accuracy of confidence intervals. Inspired by Efron's remark H(1994), the article also explores the possibility of using bootstrap to make inference on the common odds ratio based on several contingency tables.

There are two commonly used approaches for confidence interval construction. In some cases, exact intervals can be derived. Agresti showed us few examples based on small samples from two-way contingency tables in his book *Categorical Data Analysis*. Even though exact intervals give the most accurate coverage probability, they require problem-based subtle thought and heavy computation, and are not suitable for many data sets. Therefore, exact intervals are not widely used in real applications. In fact, most confidence intervals are approximated ones, using the asymptotic properties of maximum likelihood estimators. These kind of intervals are called standard intervals.

$$\hat{\theta} \pm z^{(\alpha)} \hat{\sigma} \tag{1}$$

where $\hat{\theta}$ is the point estimate of the parameter and $\hat{\sigma}$ is an estimate of standard deviation. Compared to the exact intervals, standard intervals are much easier to calculate and do not need problem-by-problem based intervention. However, they can be quite inaccurate sometimes in real practice.

Here, We use an example from Efron (1996) to show the considerable difference between standard intervals and exact intervals, and give motivations to construct bootstrap confidence intervals.

Table 1 shows the cd4 data: 20 HIV-positive subjects received an experimental antiviral drug; cd4 counts in hundreds were recorded for each subject at baseline and after one year of treatment.

Table 1. the cd4 data

Subject	Baseline	Oneyear	Subject	Baseline	Oneyear
1	2.12	2.47	11	4.15	4.74
2	4.35	4.61	12	3.56	3.29
3	3.39	5.26	13	3.39	5.55
4	2.51	3.02	14	1.88	2.82
5	4.04	6.36	15	2.56	4.23
6	5.10	5.93	16	2.96	3.23
7	3.77	3.93	17	2.49	2.56
8	3.35	4.09	18	3.03	4.31
9	4.10	4.88	19	2.66	4.37
10	3.35	3.81	20	3.00	2.40

Each pair of data is recorded as $x_i = (B_i, A_i)$ for i = 1,2,...,20. The estimated correlation coefficient between two measurements $\hat{\theta}$ is 0.723. Assume (B_i, A_i) are i.i.d samples from bivariate normal distribution,

$$\begin{pmatrix} B_i \\ A_i \end{pmatrix} \sim_{i.i.d.} N(\lambda, \tau) \tag{2}$$

where λ and τ are the unknown mean vector and covariance matrix.

From the bivariate normal distribution model, we can calculate the exact interval, standard interval and bootstrap interval for $\hat{\theta}$. Figure 1 shows us the exact and approximated confidence intervals for correlation coefficient at different nominal coverage.

Figure 1 shows that the bootstrap-t intervals match perfectly to the exact intervals, while considerable differences exist between standard intervals and exact intervals. Standard intervals are always symmetric to the maximum likelihood

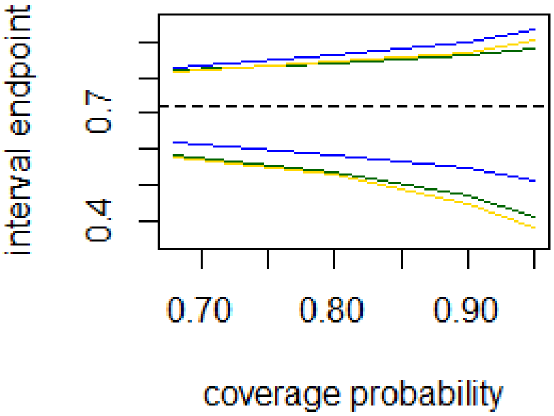


Figure 1. Exact and approximate confidence intervals for θ=corr(B,A) , assuming a bivariate normal model for the cd4 data. The black curves are exact intervals; blue curves are standard intervals; gold curves are Bootstrap-t intervals. The black dashed line indicates the maximum likelihood estimate of θ.The horizontal line represents the nominal coverage probability of 0.68, 0.80, 0.90, and 0.95. The vertical direction shows the upper and lower endpoints of the intervals from three different methods.

estimate of the parameter, which is not true in most situations. For example, in the bivariate normal cd4 data model, exact intervals are shifted leftwards. Standard intervals are too optimistic in ruling out the smaller endpoints and too pessimistic in getting rid of the upper endpoints. In section 2, we will briefly review the theoretical results of bootstrap intervals and give intuitive explanation on why bootstrap intervals are more accurate than standard intervals.

However, there is not much evidence supporting the bivariate normal model assumption. If (B_i,A_i) are i.i.d sample from some unknown distribution F, then will the bootstrap intervals still perform better than the standard intervals? Figure 2 shows us the standard intervals and bootstrap intervals for correlation coefficient at different nominal coverage based a nonparametric model. Figure 2 gives the similar results as we have seen in the parametric case. Section 2 will show theoretical results that the bootstrap intervals are second-order accurate.

Results from cd4 dataset has motivated us to think more on how to utilize the bootstrap method to make refinement on the standard intervals. Section 2 serves as an overview of different bootstrap confidence intervals.

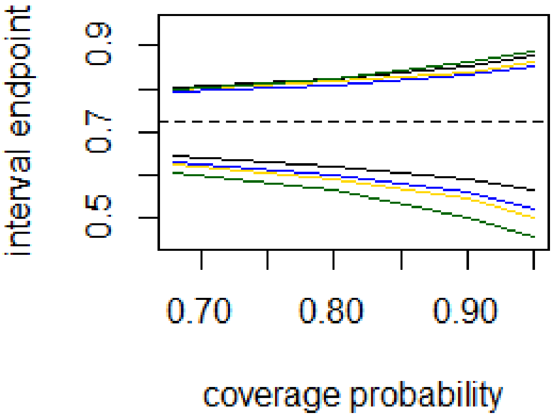


Figure 2. Standard and bootstrap confidence intervals for θ=corr(B,A) , assuming a nonparametric model. The black curves are standard intervals; blue curves are ABC intervals; gold curves are BCa intervals; darkgreen curves are Bootstrap-t intervals. The black dashed line indicates the maximum likelihood estimate of θ.The horizontal line represents the nominal coverage probability of 0.68, 0.80, 0.90, and 0.95. The vertical direction shows the upper and lower endpoints of the intervals from four different methods.

2. Methodology

We learned from bootstrap chapter (Horowitz, 2001) from the handbook of econometrics that bootstrap confidence intervals are second-order accurate. There are different ways to construct bootstrap intervals. In this section we briefly review three methods, BC_α intervals, the ABC method and the bootstrap-t method. For specifically detailed information, refer to Diciccio and Efron(1996).

2.1. The BC_a Intervals

Suppose θ is the parameter of interest. The BC_a method constructs confidence interval for θ from the percentiles of the bootstrap histogram. Suppose $\hat{\theta}(\mathbf{x})$ is an estimate of θ based on the observed data **x**; and $\theta^* = \theta^*(\mathbf{x}^*)$ is a bootstrap replication of $\hat{\theta}$ based on the resampled dataset $\mathbf{x}^*$ from an estimated distribution governing **x** (parametric case) or original sample **x** (nonparametric case). Let $\hat{G}(c)$ be the cumulative distribution function of B bootstrap replications $\hat{\theta}^*(b)$,

$$\hat{G}(c) = \sum I_{(\hat{\theta}^*(b)<c)}/B \tag{3}$$

The upper endpoint $\hat{\theta}_{BC_a}[\alpha]$ of a one-sided level-α BC_α is defined in terms of $\hat{G}$ and two numerical parameters z_0 and a.

$$\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1}\Phi(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}) \tag{4}$$

where z_0 serves as bias-correction and a is used to measure the speed that the standard error is changing on a normalized scale.

We can intuitively understand why the BC_a intervals are more accurate than standard intervals. The results from cd4 data show us that the true distribution of correlation coefficient is not normal, instead it is roughly left-skewed. The symmetric intervals given by standard method will be considerably biased. It is not hard to imagine that adding bias-correction parameter a into the model will give us more accurate intervals. Also the standard error may be different than the variance of normal distribution. a is used to correct the acceleration error and account for the nonconstant standard error. By adjusting for bias and nonconstant standard error, the BC_a method can give second-order accurate confidence intervals. When z_0 and a are equal to 0, then the BC_a intervals are the same as the standard intervals.

$$Prob\left\{\theta < \hat{\theta}_{BC_a}[\alpha]\right\} = \alpha + O(1/n) \tag{5}$$

$$Prob\left\{\theta < \hat{\theta}_{STAN}[\alpha]\right\} = \alpha + O(1/\sqrt{n}) \tag{6}$$

For computational details of z_0 and a, refer to Diciccio and Efron's paper (1996).

2.2. The ABC Method

ABC stands for approximated bootstrap confidence intervals. Deriving BC_a intervals require heavy computation and Monte-Carlo simulations. However, in some cases, it is possible to approximate the BC_a interval endpoints analytically, which largely reduces computational burden. The ABC intervals depend on five estimated parameters $(\hat{\theta}, \hat{\sigma}, \hat{a}, \hat{z_0}, \hat{c_q})$. $(\hat{a}, \hat{z_0}, \hat{c_q})$ corrects a deficiency of the standard method, making the ABC intervals second-order accurate. For computational details, refer to Diciccio and Efron's paper (1996).

2.3. Bootstrap-t Method

Bootstrap-t method is more bootstraplike and conceptually simpler than BC_a approach. The basic idea behind this method mimics the hypothesis testing and interval construction based on t-statistic.-The key for confidence intervals derived from this kind of method to be accurate is to find good approximate of the variance and accurate percentiles.

Suppose $\hat{\theta}$ is an estimate for a parameter of interest θ and $\hat{\sigma}$ is an estimate for the standard deviation of $\hat{\theta}$. Similar to t-statistic, we define

$$T = \frac{\hat{\theta} - \theta}{\hat{\sigma}} \tag{7}$$

Let $T^{(\alpha)}$ be the 100αth percentile of T. The lower endpoint of an α-level one-sided confidence interval for θ is

$$\hat{\theta} - \hat{\sigma} T^{(\alpha)} \tag{8}$$

However, unlike the t-statistic, we do not know the T-percentile here. The idea of bootstrap-t method is to estimate $T^{(\alpha)}$ by bootstrapping. We resample the data from the estimated distribution of original sample **x** or directly from **x** with replacement. And compute the T statistic for a large number of bootstrap replications B. Define $\hat{T}^{(\alpha)}$ = B $\cdot$ αth ordered value of $\{T^*(b), b = 1, 2, ..., B\}$.Then the lower endpoint can be approximated by

$$\hat{\theta}_T[\alpha] = \hat{\theta} - \hat{\sigma}\hat{T}^{(\alpha)} \tag{9}$$

Inspired by Efron's paper (1994), this article tries to apply bootstrap-t method to parameter of interest based on multiple data sets and explore whether refinement can be made. Common odds ratio (Mantel-Haenszel estimator) is an important parameter in both biostatistics and clinical trials. It gives us information on whether the treatment has pleasant effects or not. In applications, the common odds ratio is calculated based on two different situations: large sample size in each contingency table for fixed number of tables and sparse sample sizes in each table but with a large number of contingency tables. Hauck (1997) and Breslow (1981) gave the large sample approximations of estimators based on these two situations respectively. Breslow and Liang (1982) showed simulation results of comparing different variance computation approaches for Mantel-Haenszel estimator based on two different situations.

Inspired by Breslow and Liang's paper, this project will first re-conduct the simulation and then try to apply bootstrap-t method in interval construction. The next section will show the detailed simulation procedure and results.

3. Bootstrap-t Intervals for M-H Estimator

Mantel-Haenszel is a well-established estimator for common odds ratio in a series of 2 × 2 contingency tables. Breslow and Liang proposed four variances for the logarithm of the Mantel-Haenszel estimator of the common odds ratio. Unconditional maximum likelihood method is useful to estimate the standard deviation for MH estimator when sample size for each table is ”infinite” with fixed number of contingency tables, however it performs poorly in the case where the table is sparse but with ”infinite” number of tables (Hauck,1979). In the sparse data case, instead, conditional maximum likelihood approached is used to give good approximation of the standard deviation (Breslow, 1981). However, in real biomedical research and clinical trial experiments, the situation is always between two cases mentioned above. Therefore, a weighted average of two variances is proposed to combine two situations.

Consider a series of K 2 × 2 contingency tables formed by pairs of independent binomial observations (X_i, Y_i) with denominators (n_i, m_i) and success probabilities (p_{1i}, p_{0i}) for $i = 1, 2, ..., K$. We assume that the odds ratio $\psi_i = p_{1i}(1 - p_{0i})/p_{0i}(1 - p_{1i})$ remains constant from table to table. Let $N_i = n_i + m_i$ and $N = \sum N_i$ denote the sample size in table i and the total sample sizes for K tables, respectively.

The Mantel-Haenszel estimator is defined by

$$\hat{\psi}_{MH} = \frac{\sum R_i}{\sum S_i} \tag{10}$$

where $R_i = X_i(m_i - Y_i)/N_i$ and $S_i = (n_i - X_i)Y_i/N_i$.

Since M-H estimator can only take positive values, the distribution of $\hat{\psi}$ is highly skewed to the right. Taking logarithm of the common odds ratio can alleviate the problem, making the distribution of $\hat{\psi}$ converging to normal distribution faster. $\hat{\theta}_{MH} = log(\hat{\psi}_{MH})$ is the transformed estimator. Three different variances of $\hat{\theta}_{MH}$ mentioned above are defined as

$$V_H = \frac{\sum S_i^2/W_i}{(S_i)^2} \tag{11}$$

$$V_B = \frac{\sum(R_i/\hat{\psi}_{MH} - S_i)^2}{(S_i)^2} \tag{12}$$

$$V_C = \frac{NV_H + K^2 V_B}{N + K^2} \tag{13}$$

The fourth variance estimator is based on the jackknife principle. The basic idea behind is that when the number of tables K is large, it is natural and reasonable to consider pseudo-values obtained by dropping each of the tables in turn from the calculations. Define $\hat{\psi}^{i}_{MH} = \sum_{j\neq i} R_i / \sum_{j\neq i} S_i$ be the Mantel-Haenszel estimator when the ith table is omitted. The pseudo-values are $\hat{\theta}_i = Klog\hat{\psi}_{MH} - (K-1)log\hat{\psi}^{i}_{MH}$. The jackknife estimator is defines by

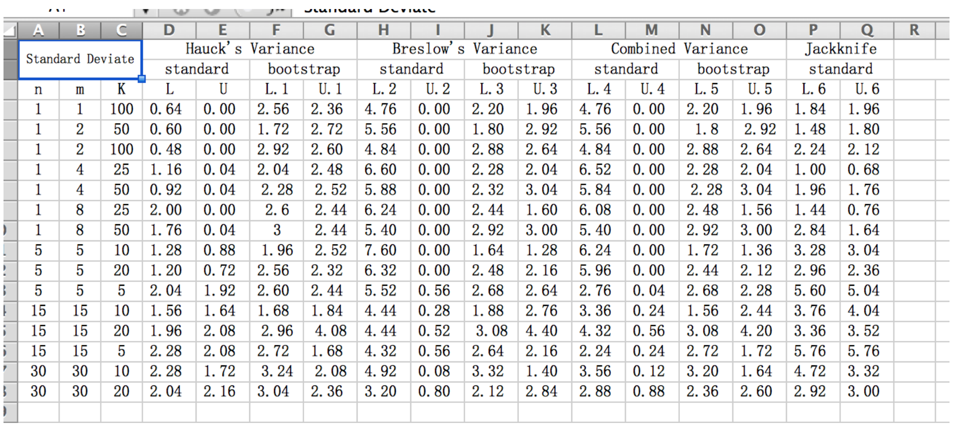
$$\hat{\theta}_J = K^{-1}\sum\hat{\theta}_i \tag{14}$$

and the jackknife variance is (Tukey,1958)

$$V_J = \frac{\sum(\hat{\theta}_i - \hat{\theta}_J)^2}{K(K-1)} \tag{15}$$

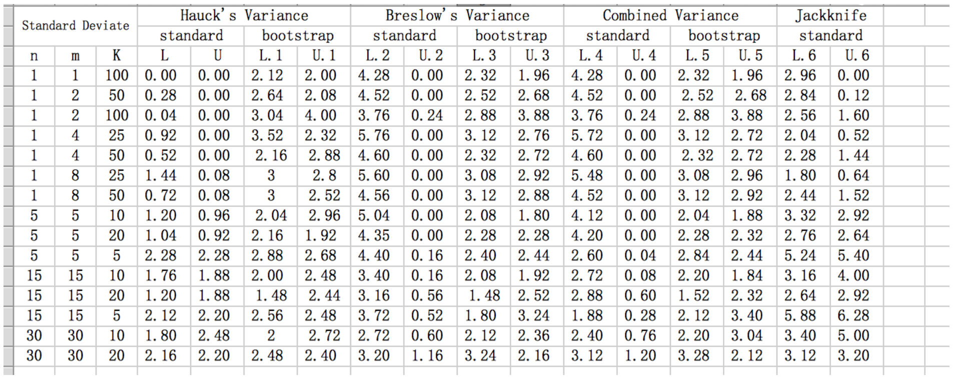
In this article, we use the same simulation logic as what is used in Breslow and Liang's paper. Here, K, n_i, m_i, ψ and p_{0i} are chosen to be representative of situations which are usually encountered in biomedical research. We conduct two series of Monte Carlo simulations. The first series of simulations investigate the performance of four different variances in matched sets, that is "sparse data with large numbers of tables" setting. The treatment group consists of 1 subject and control group consists of m subjects (ranging from 1 to 8). Depending on the value of m, the number of tables are 25, 50 or 100. The exposure probabilities range from 0.3 to 0.8, with different increments corresponding to different number of contingency tables. The second series of simulations investigate balanced experiments. The number of subjects in treatment and control groups are equal, ranging from 5 to 30. The number of tables are chosen to be 5, 10 or 20. The exposure probabilities still range from 0.3 to 0.8 with different increments corresponding to different number of contingency tables. The balanced settings are used to explore the performance of four variances in situations where sample size is large but with fixed number of contingency tables. Simulations in both series are conducted under the null hypothesis($\psi = 1$) and under the alternative hypothesis($\psi = 3.5$). The number of repeated samples is 2000 in each simulation. In addition, we also explore and compare the accuracy of confidence intervals for common odds ratio derived from standard method and bootstrap-t method. The number of bootstrap replication is 1500 in each round of simulation. The simulation results are shown in the following two tables.

The entries of two tables reflect the accuracy of two different large-sample confidence intervals for θ constructed from the M-H estimator and four different variances mentioned above. In detail, they show the percentage of simulated runs in which the true odds ratio θ fell beyond the standard intervals



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Standard Deviate			Hauck’s Variance				Breslow’s Variance				Combined Variance				Jackknife	
			standard		bootstrap		standard		bootstrap		standard		bootstrap		standard	
n	m	K	L	U	L. 1	U. 1	L. 2	U. 2	L. 3	U. 3	L. 4	U. 4	L. 5	U. 5	L. 6	U. 6
1	1	100	0. 64	0. 00	2. 56	2. 36	4. 76	0. 00	2. 20	1. 96	4. 76	0. 00	2. 20	1. 96	1. 84	1. 96
1	2	50	0. 60	0. 00	1. 72	2. 72	5. 56	0. 00	1. 80	2. 92	5. 56	0. 00	1. 8	2. 92	1. 48	1. 80
1	2	100	0. 48	0. 00	2. 92	2. 60	4. 84	0. 00	2. 88	2. 64	4. 84	0. 00	2. 88	2. 64	2. 24	2. 12
1	4	25	1. 16	0. 04	2. 04	2. 48	6. 60	0. 00	2. 28	2. 04	6. 52	0. 00	2. 28	2. 04	1. 00	0. 68
1	4	50	0. 92	0. 04	2. 28	2. 52	5. 88	0. 00	2. 32	3. 04	5. 84	0. 00	2. 28	3. 04	1. 96	1. 76
1	8	25	2. 00	0. 00	2. 6	2. 44	6. 24	0. 00	2. 44	1. 60	6. 08	0. 00	2. 48	1. 56	1. 44	0. 76
1	8	50	1. 76	0. 04	3	2. 44	5. 40	0. 00	2. 92	3. 00	5. 40	0. 00	2. 92	3. 00	2. 84	1. 64
5	5	10	1. 28	0. 88	1. 96	2. 52	7. 60	0. 00	1. 64	1. 28	6. 24	0. 00	1. 72	1. 36	3. 28	3. 04
5	5	20	1. 20	0. 72	2. 56	2. 32	6. 32	0. 00	2. 48	2. 16	5. 96	0. 00	2. 44	2. 12	2. 96	2. 36
5	5	5	2. 04	1. 92	2. 60	2. 44	5. 52	0. 56	2. 68	2. 64	2. 76	0. 04	2. 68	2. 28	5. 60	5. 04
15	15	10	1. 56	1. 64	1. 68	1. 84	4. 44	0. 28	1. 88	2. 76	3. 36	0. 24	1. 56	2. 44	3. 76	4. 04
15	15	20	1. 96	2. 08	2. 96	4. 08	4. 44	0. 52	3. 08	4. 40	4. 32	0. 56	3. 08	4. 20	3. 36	3. 52
15	15	5	2. 28	2. 08	2. 72	1. 68	4. 32	0. 56	2. 64	2. 16	2. 24	0. 24	2. 72	1. 72	5. 76	5. 76
30	30	10	2. 28	1. 72	3. 24	2. 08	4. 92	0. 08	3. 32	1. 40	3. 56	0. 12	3. 20	1. 64	4. 72	3. 32
30	30	20	2. 04	2. 16	3. 04	2. 36	3. 20	0. 80	2. 12	2. 84	2. 88	0. 88	2. 36	2. 60	2. 92	3. 00

Figure 3. Percentage of samples for which the standardized deviate based on a log odds ratio estimation procedure fell beyond the upper or lower 2.5th percentile of the normal distribution and bootstrap-t distribution,$\theta = 0$ ($\psi = 1$).



Standard Deviate			Hauck’s Variance				Breslow’s Variance				Combined Variance				Jackknife	
			standard		bootstrap		standard		bootstrap		standard		bootstrap		standard	
n	m	K	L	U	L. 1	U. 1	L. 2	U. 2	L. 3	U. 3	L. 4	U. 4	L. 5	U. 5	L. 6	U. 6
1	1	100	0. 00	0. 00	2. 12	2. 00	4. 28	0. 00	2. 32	1. 96	4. 28	0. 00	2. 32	1. 96	2. 96	0. 00
1	2	50	0. 28	0. 00	2. 64	2. 08	4. 52	0. 00	2. 52	2. 68	4. 52	0. 00	2. 52	2. 68	2. 84	0. 12
1	2	100	0. 04	0. 00	3. 04	4. 00	3. 76	0. 24	2. 88	3. 88	3. 76	0. 24	2. 88	3. 88	2. 56	1. 60
1	4	25	0. 92	0. 00	3. 52	2. 32	5. 76	0. 00	3. 12	2. 76	5. 72	0. 00	3. 12	2. 72	2. 04	0. 52
1	4	50	0. 52	0. 00	2. 16	2. 88	4. 60	0. 00	2. 32	2. 72	4. 60	0. 00	2. 32	2. 72	2. 28	1. 44
1	8	25	1. 44	0. 08	3	2. 8	5. 60	0. 00	3. 08	2. 92	5. 48	0. 00	3. 08	2. 96	1. 80	0. 64
1	8	50	0. 72	0. 08	3	2. 52	4. 56	0. 00	3. 12	2. 88	4. 52	0. 00	3. 12	2. 92	2. 44	1. 52
5	5	10	1. 20	0. 96	2. 04	2. 96	5. 04	0. 00	2. 08	1. 80	4. 12	0. 00	2. 04	1. 88	3. 32	2. 92
5	5	20	1. 04	0. 92	2. 16	1. 92	4. 35	0. 00	2. 28	2. 28	4. 20	0. 00	2. 28	2. 32	2. 76	2. 64
5	5	5	2. 28	2. 28	2. 88	2. 68	4. 40	0. 16	2. 40	2. 44	2. 60	0. 04	2. 84	2. 44	5. 24	5. 40
15	15	10	1. 76	1. 88	2. 00	2. 48	3. 40	0. 16	2. 08	1. 92	2. 72	0. 08	2. 20	1. 84	3. 16	4. 00
15	15	20	1. 20	1. 88	1. 48	2. 44	3. 16	0. 56	1. 48	2. 52	2. 88	0. 60	1. 52	2. 32	2. 64	2. 92
15	15	5	2. 12	2. 20	2. 56	2. 48	3. 72	0. 52	1. 80	3. 24	1. 88	0. 28	2. 12	3. 40	5. 88	6. 28
30	30	10	1. 80	2. 48	2	2. 72	2. 72	0. 60	2. 12	2. 36	2. 40	0. 76	2. 20	3. 04	3. 40	5. 00
30	30	20	2. 16	2. 20	2. 48	2. 40	3. 20	1. 16	3. 24	2. 16	3. 12	1. 20	3. 28	2. 12	3. 12	3. 20

Figure 4. Percentage of samples for which the standardized deviate based on a log odds ratio estimation procedure fell beyond the upper or lower 2.5th percentile of the normal distribution and bootstrap-t distribution,$\theta = 1.253$ ($\psi = 3.5$).

$\hat{\theta} \pm Z^{(1-\alpha)} V^{\frac{1}{2}}$ and the bootstrap-t intervals $\hat{\theta} \pm T^{(1-\alpha)} V^{\frac{1}{2}}$. Theoretically, the percentage should be 2.5% all the time.

From Figure 3, we can see that with common odds ratio equal to 1, Hauck’s Variance performs well when the sample size for each table is large. However, it tends to underestimate the percentage when the sample size in each table is small, especially for the values falling beyond the upper endpoint. Breslow and combined variances do not perform well either; the true odds ratio of $\psi = 1$ is less than the lower confidence bound in almost 5 to 7% of samples, whereas it should be close to 2.5% nominally. Jackknife performs well when the number of tables is sufficiently large, but tends to overestimate the percentage when K is small. From Figure 4, we can see that with common odds ratio equal to 3.5, Hauck's Variance performs well when the sample size for each table is large. However, it tends to underestimate the percentage when the sample size in each table is small, both for the upper endpoint and the lower endpoint. Breslow and combined variances do not perform well either. Jackknife performs well when the number of tables is large, but tends to overestimate the percentage when K is small.

To further compare the confidence intervals constructed by the standard method and the bootstrap-t method, we make two scatter plots to visualize those percentages using Hauck's, Breslow's and combined variances. It is obvious that if we use percentiles from the bootstrap distribution to normalize the common odds ratio, the percentages are closer to the theoretical results (2.5%, 2.5%).

Therefore, the best strategy to construct confidence intervals for the true common odds ratio is to select the reasonable variance form corresponding to different situations and the percentiles calculated from the bootstrap distribution.

4. Discussion

Before we apply the bootstrap method to the common odds ratio in a series of K 2×2, we take one table (Mendenhall et.all, 1984) as a pilot example. Exact, standard and bootstrap intervals are constructed. But for simplicity, we omit the results for this data set in this article. R codes for this part is attached in the appendix. In the future work, we may consider to compare the accuracy of confidence intervals for common odds ratio derived from different bootstrap approaches, such as BC_a and ABC methods etc. We may also consider to apply bootstrap methods to different data forms which consist of K-samples other than K 2×2 contingency tables.

References

Efron, B. (1994). Missing data, imputation, and the bootstrap. Journal of the American Statistical Association, 89(426), 463-475.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. Statistical science, 189-212.

DiCICCIO, T. H. O. M. A. S., & Efron, B. (1992). More accurate confidence intervals in exponential families. Biometrika, 231-245.

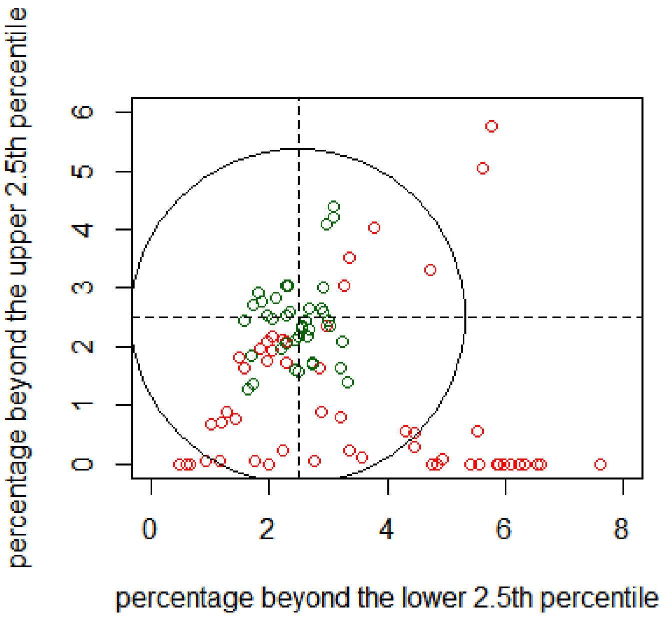


Figure 5. Scatter plot of percentage of samples for which the standardized deviate based on a log odds ratio estimation procedure fell beyond the upper or lower 2.5th percentile of the normal distribution and bootstrap-t distribution,$\theta = 0$ ($\psi = 1$). Red points and green points represent the percentages derived from the standard method and the bootstrap-t method, respectively.

Agresti, A., & Kateri, M. (2011). Categorical data analysis (pp. 206-208). Springer Berlin Heidelberg.

Mendenhall, C. L., Anderson, S., Weesner, R. E., Goldberg, S. J., & Crolic, K. A. (1984). Protein-calorie malnutrition associated with alcoholic hepatitis: Veterans Administration Cooperative Study Group on alcoholic hepatitis. The American journal of medicine, 76(2), 211-222.

Horowitz, J. L., Heckman, J. J., & Leamer, E. E. (2001). Handbook of econometrics.

Hauck, W. W. (1979). The large sample variance of the Mantel-Haenszel estimator of a common odds ratio. Biometrics, 817-819.

Breslow, N. (1981). Odds ratio estimators when the data are sparse. Biometrika, 73-84.

Breslow, N. E., & Liang, K. Y. (1982). The variance of the Mantel-Haenszel

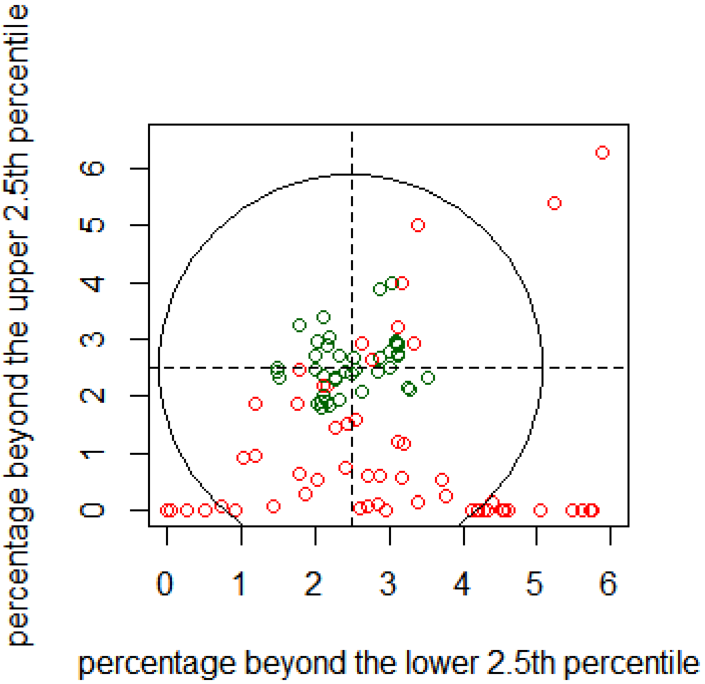


Figure 6. Scatter plot of percentage of samples for which the standardized deviate based on a log odds ratio estimation procedure fell beyond the upper or lower 2.5th percentile of the normal distribution and bootstrap-t distribution,$\theta = 1.253$ ($\psi = 3.5$). Red points and green points represent the percentages derived from the standard method and the bootstrap-t method, respectively.

estimator. Biometrics, 943-952.