Journal of the Korean Statistical Society (2007), **36***: 3, pp 397–410*

ESTIMATING THE SIMULTANEOUS CONFIDENCE LEVELS FOR THE DIFFERENCE OF PROPORTIONS FROM MULTIVARIATE BINOMIAL DISTRIBUTIONS

HYEONG CHUL JEONG[1], MYOUNGSHIC JHUN[2] AND JAE WON LEE[3]

ABSTRACT

For the two groups data from multivariate binomial distribution, we consider a bootstrap approach to inferring the simultaneous confidence level and its standard error of a collection of the dependent confidence intervals for the difference of proportions with an *experimentwise error rate* at the α level are presented. The bootstrap method is used to estimate the simultaneous confidence probability for the difference of proportions.

AMS 2000 subject classifications. Primary 62G05; Secondary 62H12.
Keywords. Bootstrap, multivariate binomial distribution, simultaneous confidence levels.

1. INTRODUCTION

Data having multivariate Bernoulli structure are encountered in many areas such as genetic disease studies, analyses of system reliability, carcinogenesis bioassay studies and others. In the studies of the presence or absence of some cancer tissues with multiple groups of female mice, for example, the number of occurrences of a particular lesion at a particular tissue may be modeled as binomial, and the vector of such frequencies may be considered multivariate binomial with unknown dependence structure.

The statistical inferences of the multivariate binomial distribution are described by Hochberg (1988), Holland and Copenhaver (1987) and SAS multiple test procedure (MULTTEST). One of the more useful methods for statistically inferencing the multivariate Bernoulli data was proposed by Westfall (1985) and

Received October 2006; accepted April 2007.
[1]Corresponding author. Department of Applied Statistics, The University of Suwon, Kyunggi 445-743, Korea (e-mail: jhc@suwon.ac.kr)
[2]Department of Statistics, Korea University, Seoul 136-701, Korea (e-mail: jhun@korea.ac.kr)
[3]Department of Statistics, Korea University, Seoul 136-701, Korea (e-mail: jael@korea.ac.kr)

Westfall and Young (1989, 1993). To estimate the simultaneous confidence level, Westfall and Young (1989, 1993) proposed that experimentwise significance level be controlled by adjusting all p-values using vector based resampling methods including bootstrap and permutation methods.

In this paper, when the two-group observable data may be modeled as having come from multivariate binomial distributions, we consider two problems. First, a bootstrap method for assessing the simultaneous confidence level of a collection of confidence intervals for the difference of proportions are considered. This analysis is an extension of the Westfall (1985)'s bootstrap simultaneous confidence levels with one group multivariate Bernoulli data to the two groups proportions. In addition, standard errors of estimates of simultaneous coverage level are estimated by double bootstrap method. Second, which marginal confidence intervals have been employed to estimate the simultaneous confidence level are considered. The estimates of simultaneous coverage level are affected by the construction methods of marginal confidence interval for the difference of the binomial parameters from two groups. Many of the existing methods for constructing the marginal confidence interval for the difference of proportions, including the exact, asymptotically-based and bootstrap confidence interval are evaluated by estimating true simultaneous probability via Monte Carlo method.

The rest of this paper is organized as follows. Section 2 introduces the bootstrap simultaneous levels for the difference of proportions between two groups. Section 3 presents the simulation results of the estimating values to the simultaneous confidence levels given the various marginal confidence intervals with for the difference of proportions. We compared the bootstrap confidence interval with classical methods. Section 4 concludes the paper.

2. Simultaneous Confidence Levels

2.1. Estimating the simultaneous confidence levels

Assume that the observable data vectors $Y' = (Y_1, \ldots, Y_k)$ with binomial marginal distributions $Y_j \sim B(n, p_j)$ may be modeled as having come from multivariate binomial distributions. Define the notation

$$Y \sim MVB_k(P, n, D),$$

where $P = (p_1, \ldots, p_k)$ denotes population proportions vector for a k-component and dependence structure specified by D.

When $n = 1$, the distribution $MVB_k(P, 1, D)$ is multivariate Bernoulli distribution. If there are two treatment groups with n_1, n_2 experimental units respectively, the available data vectors are $X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}$, where $X'_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijk})$, $(i = 1, 2)$ has an independent and identical distribution (*i.i.d.*) as $MVB_k(P_i, 1, D_i)$ and X_{ijl} is the response falling in l^{th} variable of the j^{th} observation for the i^{th} group. Let the i^{th} group proportion vector $P_i = (p_{i1}, p_{i2}, \ldots, p_{ik})$ then $E(X_{ij}) = P_i$. Let the i^{th} group frequency vector $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ik})$ then $Y_i = \sum_j^{n_i} X_{ij}$ is distributed as $MVB_k(P_i, n_i, D_i)$. The maximum likelihood estimators of the probability p_{ij} is $\hat{p}_{ij} = Y_{ij}/n_i$. When n is large, $100(1-\alpha)\%$ large-sample marginal confidence intervals for $p_{1j} - p_{2j}$, $j = 1, \ldots, k$, can be obtained as

$$I_\alpha(Y_{1j} - Y_{2j}) = (\hat{p}_{1j} - \hat{p}_{2j}) \pm Z_{(\alpha/2)}[n_1^{-1}\hat{p}_{1j}(1-\hat{p}_{1j}) + n_2^{-1}\hat{p}_{2j}(1-\hat{p}_{2j})]^{1/2}\,\forall j, \quad (2.1)$$

where $Z_{(\alpha/2)}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

Suppose that we are interested in simultaneous confidence level of a collection of confidence intervals for each of the parameter $p_{1j} - p_{2j}$, $j = 1, \ldots, k$. The exact simultaneous probability is determined by the underlying multivariate binomial distribution. In this case the goal is to estimate

$$\pi_\alpha = \Pr[p_{1j} - p_{2j} \in I_\alpha(Y_{1j} - Y_{2j}) \quad \text{for all} \quad j = 1, \ldots, k]. \quad (2.2)$$

This may be accomplished in analogous fashion to the one-sample case of Westfall (1985).

Let $X^*_{11}, X^*_{12}, \ldots, X^*_{1n_1}$ and $X^*_{21}, X^*_{22}, \ldots, X^*_{2n_2}$ be the *i.i.d.* bootstrap samples that put mass $1/n_i$ on each of the observed vectors $X_{11}, X_{12}, \ldots, X_{1n_1}$ and X_{21}, $X_{22}, \ldots, X_{2n_2}$ respectively. Thus the $Y^*_i = \sum_j^{n_i} X^*_{ij}$ are distributed as *i.i.d.* MVB_k $(\hat{P}_i, n_i, \hat{D}_i)$, where $\hat{P}_i = Y_i/n_i$ and $\hat{D}_i$ is the empirical probability measure of the observed k tuples in each group. In bootstrap samples, the complex dependence structure or other physical constraints are preserved. Therefore, the estimate of π_α is then computed as

$$\pi^*_\alpha = \Pr[\hat{p}_{1j} - \hat{p}_{2j} \in I_\alpha(Y^*_{1j} - Y^*_{2j}) \quad \text{for all } j = 1, \ldots, k].$$

The issue of consistency for π^*_α was addressed by Westfall (1985) and the use of bootstrap method in this instance is justifiable by Bickel and Freedman (1981) and Singh (1981).

2.2. The bias and standard error

In this section, the standard error and bias of the estimates of simultaneous confidence probability are considered. The simulation standard error is used to bound the true resample-based simultaneous confidence probability. In the course of nature, double bootstrap method is used for estimating standard error. Westfall (1985) discussed bootstrap-bootstrap method, but he did not attempt it because this approach was extremely expensive in terms of computer time. But the high-speed computer now makes it possible to estimate standard error of simultaneous confidence probability.

The standard error of π^*_α can be computed using the following algorithm.

Step 1. After selecting the bootstrap samples, generate two group double bootstrap sample $X^{**}_{11}, X^{**}_{12}, \ldots, X^{**}_{1n_1}$, and $X^{**}_{21}, X^{**}_{22}, \ldots, X^{**}_{2n_2}$ from bootstrap samples respectively, and compute the $100(1-\alpha)\%$ confidence intervals $I_\alpha(Y^{**}_{1j} - Y^{**}_{2j})$, $j = 1, \ldots, k$.

Step 2. Compute the $I(A^*)$ from $A^* = [\hat{p}^*_{1j} - \hat{p}^*_{2j} \in I_\alpha(Y^{**}_{1j} - Y^{**}_{2j})$ for all $j = 1, \ldots, k]$, where the maximum likelihood estimate from bootstrap sample $\hat{p}^*_{ij}$ is Y^*_{ij}/n_i and $I(A^*)$ is an indicator function.

Step 3. Step 1 and Step 2 are repeated M times to get a $\pi^{**}_\alpha = \sum I(A^*_i)/M$.

Step 1–3 are then repeated B times for each bootstrap sample to get the collection of $\pi^{**}_\alpha(b)$, $b = 1, \ldots, B$. Hence, the standard error and bias are given by

$$\hat{se}_{\pi^*} = \left\{ \sum_{b=1}^{B} (\pi^{**}_\alpha(b) - \bar{\pi}^{**}_\alpha)^2/(B-1) \right\}^{1/2}, \quad \text{bias} = \bar{\pi}^{**}_\alpha - \pi^*_\alpha,$$

where $\bar{\pi}^{**}_\alpha = \sum_{b=1}^{B} \pi^{**}_\alpha(b)/B$.

2.3. Other methods

Instead of using the bootstrap simultaneous confidence level from (2.2), we can use the Bonferroni simultaneous confidence level. Let $\pi^*_\alpha(j) = \Pr[\hat{p}_{1j} - \hat{p}_{2j} \in I_\alpha(Y^*_{1j} - Y^*_{2j})]$, and $\pi^*_\alpha(j)$ is the proportion of times the marginal bootstrap interval for $\hat{p}_{1j} - \hat{p}_{2j}$ is correct. The estimate of bootstrap Bonferroni simultaneous confidence level is $B^*_\alpha = \sum \pi^*_\alpha(j) - k + 1$, and the standard error of the bootstrap Bonferroni simultaneous confidence level is $\{\sum_{b=1}^{B}(B^{**}_\alpha(b) - \bar{B}^{**}_\alpha))^2/(B-1)\}^{1/2}$ by double bootstrap samples. However, in many applications this procedure is

TABLE 2.1 *Brown and Fears (1981) data showing numbers of tumor-bearing animals*

Site	0*ppm*	4*ppm*	8*ppm*	16*ppm*	50*ppm*
Liver	14	18	17	11	12
Lung	12	10	8	11	9
Lymph nodes	8	12	8	15	10
Cardiovascular	1	3	6	2	1
Pituitary	0	3	1	2	1
Ovary	3	1	2	2	5
Number in group	49	50	48	43	50

unnecessarily conservative due to the fact that all marginal coverage levels are conservative. The simultaneous confidence level assuming independence is $IA^*_\alpha = \prod_{j=1}^{k} \pi^*_\alpha(j)$, and the standard error of the bootstrap independent assumption lower bound is $\{\sum_{b=1}^{B}(IA^{**}_\alpha(b) - \bar{I}A^{**}_\alpha))^2/(B-1)\}^{1/2}$ by double bootstrap samples (Jhun *et al.*, 2007). These procedures do not consider the dependence structure and the discreteness of multivariate binomial distribution.

2.4. Example : an animal carcinogenicity study - Brown and Fears data

Brown and Fears (1981) reported data from a two-year bioassay involving female mice treated with five different levels of a compound. The group sizes, dose levels, and tumor incidence rate for six sites are given in Table 2.1. The raw data set contains 240 multivariate Bernoulli (six-component) vectors and follows five different multivariate binomial distribution with unknown dependent structure. This example data was also examined by Westfall and Young (1989) for multiple testing in multivariate binary application.

As noted by Brown and Fears (1981), when each dose is compared with the control (0 ppm), only one of the twenty-four tests is significant at the 5% level which is Lymph nodes between 0 ppm and 16 ppm. Therefore, we have examined the simultaneous confidence probability of collecting the experimentwise $100(1-\alpha)\%$ confidence intervals for the difference between 0 ppm and 16 ppm. The nominal coverages $1-\alpha = 0.8$, 0.9, 0.95 and 0.99 are considered. The marginal confidence intervals, bootstrap simultaneous confidence levels, standard errors and biases are given in Table 2.2.

The 80% marginal intervals show the Lymph nodes and Pituitary sites exhibit significant difference between two groups, however, these clearly may be spurious if one accepts the simultaneous confidence probability of 23.7% indicated by the

TABLE 2.2 *Estimate the marginal confidence intervals and simultaneous confidence interval for the difference of proportions between 0ppm and 16ppm. There were 1000 bootstrap and double bootstrap replications for the bootstrap simultaneous confidence probability and standard error respectively. (* : This mark means that the site is significant at the α level in marginal difference.)*

	Marginal prob.			*Marginal confidence level* $(1-\alpha)$			
Site	*0ppm*	*16ppm*	*difference*	80%	90%	95%	99%
Liver	.286	.256	.030	(−.089.149)	(−.123.182)	(−.152.212)	(−.209.269)
Lung	.245	.256	−.011	(−.127.105)	(−.160.138)	(−.188.167)	(−.244.222)
Lymph	.163	.349	−.186	(−.301 − .070)*	(−.333 − .038)*	(−.362 − .009)*	(−.417.046)
Cardio.	.020	.047	−.026	(−.075.023)	(−.089.036)	(−.100.048)	(−.124.072)
Pituitary	.000	.047	−.047	(−.088 − .005)*	(−.099.006)	(−.109.016)	(−.129.036)
Ovary	.061	.047	.015	(−.045.075)	(−.063.092)	(−.077.107)	(−.106.136)
Simultaneous conf. level estimate				0.237	0.451	0.582	0.791
Standard error				0.050	0.082	0.109	0.148
Bias				−0.005	−0.034	−0.038	−0.091
Bonferroni lower bound(S.E)				0.000(0.014)	0.261(0.138)	0.497(0.165)	0.781(0.182)
Independence assumption(S.E)				0.244(0.047)	0.453(0.080)	0.587(0.108)	0.795(0.148)
Independence lower bound $(1-\alpha)^k$				0.262	0.531	0.735	0.941

bootstrap estimate. At the 95% marginal levels, only the Lymph nodes site is different, with simultaneous confidence probability estimated at 58.2% and standard error estimated at 0.109. And at the marginal 99% level, for which the simultaneous probability approximately achieves the 80% level, none of the sites appears different between two groups.

We have presented the simultaneous confidence levels and standard errors for the marginal confidence levels 0.5 through 0.999 using Figure 2.1. The first graph shows the simultaneous confidence levels, the second graph shows their biases and the third graph shows their standard errors at given marginal confidence levels. Figure 2.1 indicates that the independence lower bound $(1-\alpha)^k$ were the highest of all methods. In contrast, the simultaneous level (B_α^*) by the bootstrapped Bonferroni methods were the lowest of all methods. The estimate of the second highest level was yielded by the independent bootstrap method (IA_α^*), while the estimate by the bootstrap method (π_α^*) was slightly lower than the estimate provided by the independent bootstrap method. If the marginal confidence intervals agree to the nominal level $(1-\alpha)$ then the independent bound $(1-\alpha)^k$ become the baseline meaning that it's a lower bound of all methods. But, in this context, Figure 2.1 shows opposite results because we use the Wald type interval (see the formula (2.1)) constructing the marginal confidence intervals. The coverage prob-

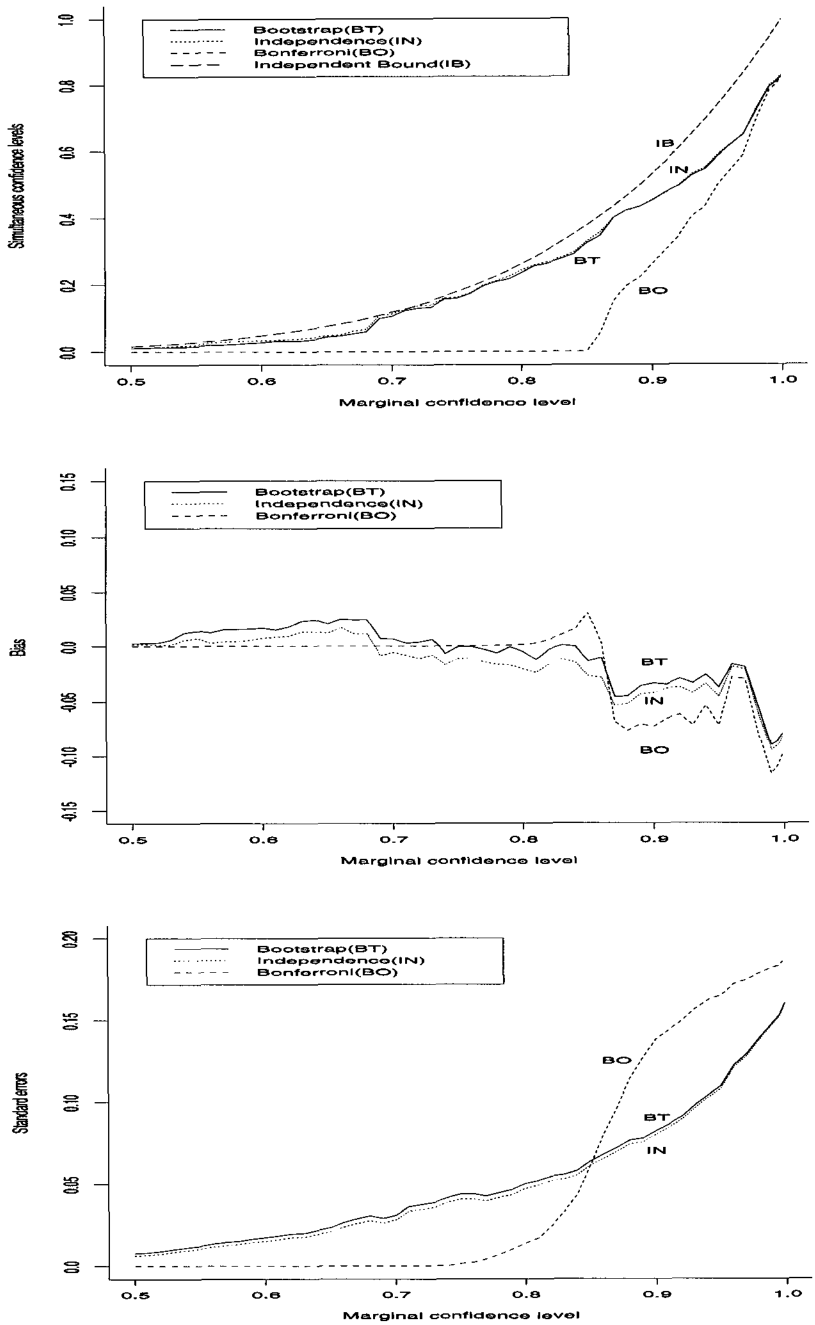

FIGURE 2.1 *Plots of simultaneous confidence levels, biases and standard errors at given marginal confidence levels.*

abilities for the Wald type interval tend to be severe underestimated the marginal nominal probabilities. If the marginal confidence intervals are constructed agreeing with the marginal nominal coverage $(1-\alpha)$, then a little more sites may be exhibit significant difference between two groups at the $(1-\alpha)$ level. Section 3 deals with the problem.

Note that "Bonferroni lower bound" differs greatly from the "independence assumption" values at low α levels. Note also that the simultaneous confidence levels are close to the "independent assumption" values, which means that the correlation structure with 6 sites have nearly independent structure.

From bias plot of Figure 2.1, the estimated simultaneous confidence levels based on the double bootstrap samples are larger than the those based on the bootstrap samples, when marginal confidence level $(1-\alpha)$ is below the vicinity of 0.7. When marginal confidence level becomes large, such phenomenon is reversed. The biases of bootstrap method are close to the those of independent bootstrap method and we can also see that the biases of bootstrap methods yield lower than the those of independent bootstrap method when $(1-\alpha) > 0.70$. When $\sum \pi_\alpha^*(j) \leq k-1$, the estimate of bootstrap Bonferroni simultaneous confidence level becomes negative, therefore, we set the biases of Bonferroni method to zeros under the marginal confidence level 0.8 .

The simulation standard errors are used to bound the simultaneous confidence probabilities. From standard error plot of Figure 2.1, we can see that as marginal confidence level increase, the standard errors become large. At the 90% marginal level, the bootstrap simultaneous standard error is 0.082, then 95% confidence interval for the simultaneous confidence level is $0.451 \pm 1.96(0.082) = (0.290, 0.612)$. The Bonferroni standard errors are greater than the others above the marginal confidence level 0.85. The bootstrap simultaneous standard errors are similar to the independence standard errors.

Granting the estimates by the Bonferroni technique turned out to be rather conservative in various experiments, we could conclude the bootstrap method may be suggested as another statistically robust way of estimating the simultaneous confidence level.

3. Comparison of the Methods of Constructing the Marginal Confidence Intervals

The simultaneous confidence probability will be affected by the methods of constructing the marginal confidence intervals $p_{1j} - p_{2j}$, $j = 1, \ldots, k$ and the

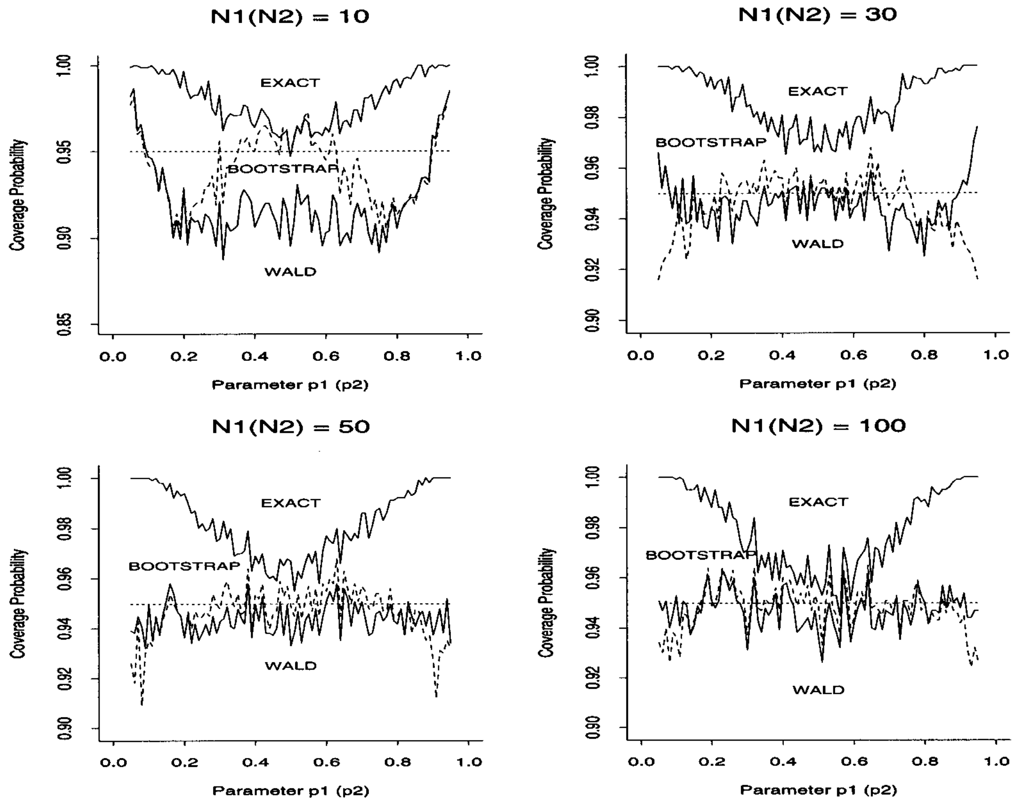

FIGURE 3.1 *Actual coverage probabilities (as functions of the true $p_{1j} = p_{2j}$) for nominal $(1-\alpha)$ confidence interval using Exact interval(EXACT), Wald interval(WALD) and bootstrapped Wald interval(BOOTSTRAP). The four different sample sizes are given, with $n_1 = n_2$ denoting Group 1 and Group 2 sample size. Although it is sometimes biased, the actual coverage probabilities of the bootstrap interval better approximates the nominal 95% level.*

sample sizes taken from two groups. It is meaningful to study which methods for obtaining marginal $100(1-\alpha)\%$ confidence intervals for the difference of proportions has been employed to estimate the joint level of confidence for the entire collection of intervals. A number of marginal confidence intervals are compared in this section.

Most introductory statistics textbooks present the confidence interval for the difference of proportions based on the asymptotic normality of the sample proportions and estimating the standard error. This $100(1-\alpha)\%$ confidence intervals for $p_{1j} - p_{2j}$, $j = 1, \ldots, k$ are given by (2.1). This is called the Wald confidence interval for the difference of proportions. The exact confidence interval for the difference of proportions, in contrast to asymptotic confidence intervals, are described by Santner and Snell (1980) and StatXact software. The main problem with using the exact confidence interval is the difficulty in computing the cumulative joint probability mass function of two independent binomial random variables. This procedure is too conservative, therefore the actual coverage probability can be much larger than the nominal confidence level unless the sample sizes are quite larger.

However, the actual coverage probabilities for the Wald interval tend to be much too small (see Figure 3.1) and a number of compromise asymptotic confidence intervals are described in Beal (1987) and Newcombe (1998). Those described include some intervals by Mee (1984), Miettinen and Nurminen (1985), Beal's Jeffrey-Perk and Haldane interval, profile method and Newcombe's methods. Another method of constructing confidence interval for the difference of proportions is bootstrap. Bootstrap confidence interval is similar to the Wald confidence interval except for using the bootstrap distribution. To study the performance of the methods introduced, a Monte Carlo investigation was done. There may be other methods, but this study is limited to the marginal confidence intervals for the difference of proportions explained by Wald, Exact, Mee, Miettinen and Nurminen(M-N), Jeffrey-Perk(J-P), Haldane, Profile method, Newcombe's method 10 and bootstrap.

We limit our attention to the following proportions and correlation matrix based on specific case. Two groups follow the same multivariate Bernoulli distribution with $P = \{0.3, 0.5, 0.75, 0.5, 0.3\}$ and equal correlation value 0.3. We consider sample sizes $n_1(n_2) = 10, 30, 50$ and 100, and nominal coverages $1-\alpha = 0.80, 0.90, 0.95$ and 0.99. For each sample size, a random sample is generated and summary measures of performance are calculated. For the measure of performance, the true simultaneous confidence probability π_α will be calculated. But it is a intractable since the intervals I_α are not easily expressed and have a complicated dependence structure. Therefore it will be estimated by simulation and the estimate value $\hat{\pi}_\alpha$ will be used the true simultaneous confidence probability. In order to estimate π_α, $n_1(n_2) = 100,000$ pseudo random samples are generated from the given multivariate bernoulli distribution and π_α is estimated by Monte Carlo simulation repeated 100,000 times. The algorithm of Park *et al.* (1996) was used to generate the pseudo random samples of the multivariate Bernoulli distribution with pre-defined proportions and dependence structure. The estimate values of π_α are 0.3598, 0.6223, 0.7950 and 0.9566 for $1-\alpha = 0.8$, 0.9, 0.95 and 0.99 respectively. To compare the marginal confidence intervals, we compute the following measures.

Step 1. Generate random sample $X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}$ from that multivariate Bernoulli distribution and calculate the marginal confidence intervals of $p_{1j} - p_{2j}$, $j = 1, \ldots k$, using the described nine methods for marginal confidence intervals.

Step 2. Use the bootstrap procedure to obtain the simultaneous confidence probabilities denoted by $\hat{\pi}^*_\alpha$ (based on $B = 1000$) with the same nominal $1 - \alpha$ coverage.

Step 3. Step 1–2 are repeated $M = 100$ times to get a collection of summary measures,

$$\bar{\pi}^* = \sum_{m=1}^{M} \hat{\pi}^*_\alpha(m)/M, \quad \text{RMSE} = \sqrt{\sum_{m=1}^{M} (\hat{\pi}^*_\alpha(m) - \hat{\pi}_\alpha)^2/(M-1)},$$

where $\hat{\pi}^*_\alpha(m)$ denotes the bootstrap estimated simultaneous confidence probability of m^{th} generated data set.

The results of this simulation are shown in Table 3.1. When $n_1(n_2) \leq 50$, the mean of simultaneous confidence probabilities of the collections of Wald confidence intervals have the lower estimated simultaneous confidence probabilities. And the simultaneous confidence probabilities of the collections of exact confidence intervals have the greater estimated simultaneous confidence probabilities. Increasing sample size does not help the overestimation problem. When $n_1(n_2) \leq 30$, the means for the Mee, Miettinen and Nurminen, Jeffrey-Perk, Haldane, Profile, Newcombe and Bootstrap methods are not significantly different. Note that the estimate root mean squared error (RMSE) of Mee and bootstrap are generally small. When $n_1(n_2) \geq 30$, the means of simultaneous confidence probabilities of the collection of bootstrap confidence intervals are close to the estimated simultaneous confidence probabilities. The RMSE of bootstrap methods is smaller than that of the other methods and the bootstrap method has a slight edge over the other methods. As a result of these findings, if one wants to compute a simultaneous confidence probability, except Wald and exact methods, there are no significant difference among the methods used. But if one wants to compute a more exact simultaneous confidence probability, bootstrap intervals for the difference of two proportions are good choices.

TABLE 3.1 *The means and root mean squared errors of simultaneous confidence levels*

		$1-\alpha=0.8$		$1-\alpha=0.9$		$1-\alpha=0.95$		$1-\alpha=0.99$	
$n_1=n_2$	*method*	*Mean*	*RMSE*	*Mean*	*RMSE*	*Mean*	*RMSE*	*Mean*	*RMSE*
10	*WALD*	.30903	.06183	.52721	.10598	.67139	.13326	.84997	.12026
	EXACT	.74115	.38278	.89540	.27488	.91665	.12271	.99135	.03496
	MEE	.33299	.05302	.66056	.05385	.85226	.06226	.96964	.01502
	M-N	.34243	.05033	.66099	.05345	.85712	.06545	.97439	.01938
	J-P	.38122	.04864	.69278	.07841	.81712	.04252	.96342	.01562
	HALDANE	.36892	.04461	.63788	.05965	.78040	.06477	.91501	.08063
	PROFILE	.32903	.05508	.62281	.05287	.77628	.05257	.94833	.03072
	NEWCOMBE	.38006	.04819	.67848	.06910	.83500	.04913	.96314	.01733
	BOOTSTRAP	.41385	.07561	.63623	.06125	.77703	.06778	.91970	.07479
30	*WALD*	.32912	.03689	.60007	.03008	.76291	.03525	.93297	.02506
	EXACT	.58161	.22351	.74236	.12254	.91179	.11760	.98645	.03010
	MEE	.34704	.02743	.64992	.03845	.80511	.01794	.95689	.00696
	M-N	.35498	.02531	.65541	.04213	.81329	.02371	.96143	.00839
	J-P	.35218	.02488	.65927	.04446	.81118	.02218	.95882	.00709
	HALDANE	.34537	.02706	.64698	.03645	.80828	.02064	.95544	.00730
	PROFILE	.34252	.02827	.63973	.03153	.79720	.01545	.95131	.00916
	NEWCOMBE	.34723	.02665	.65217	.03913	.81094	.02183	.95882	.00677
	BOOTSTRAP	.37983	.03044	.63035	.02192	.79286	.01479	.95731	.00866
50	*WALD*	.36323	.02218	.60597	.02660	.77031	.02907	.94162	.01702
	EXACT	.44045	.08456	.71231	.09311	.87883	.08500	.97695	.02094
	MEE	.35431	.02607	.61195	.02296	.76783	.03237	.94485	.01433
	M-N	.35889	.02602	.61834	.02011	.77059	.02973	.94562	.01364
	J-P	.35725	.02583	.61644	.02123	.77025	.02994	.94568	.01372
	HALDANE	.35624	.02576	.61478	.02198	.76611	.03389	.94455	.01472
	PROFILE	.35343	.02619	.60861	.02453	.76329	.03636	.94209	.01679
	NEWCOMBE	.35607	.02616	.61393	.02202	.76811	.03216	.94541	.01392
	BOOTSTRAP	.35787	.02104	.62460	.02002	.78837	.01666	.95647	.00756
100	*WALD*	.35958	.01809	.61491	.01806	.77846	.02144	.94557	.01296
	EXACT	.47229	.11401	.74826	.12700	.86882	.07487	.97149	.01559
	MEE	.36133	.02051	.61052	.02147	.77654	.02322	.95109	.00858
	M-N	.36825	.02159	.61211	.02068	.77808	.02193	.95208	.00786
	J-P	.36285	.02020	.61271	.02057	.77773	.02215	.95172	.00801
	HALDANE	.35992	.02104	.61083	.02138	.77634	.02359	.95095	.00858
	PROFILE	.34407	.02863	.60917	.02192	.77532	.02432	.94914	.00997
	NEWCOMBE	.36115	.01916	.61175	.02101	.77730	.02257	.95041	.00887
	BOOTSTRAP	.35978	.01806	.62220	.01637	.78834	.01612	.95417	.00710
	$\hat{\pi}_\alpha$	.3598		.6223		.7950		.9566	

4. CONCLUSION

For the two groups data from multivariate Bernoulli distribution, the techniques for evaluating the simultaneous confidence level of the collections of the marginal confidence intervals for the difference of proportions and the comparisons for estimating the simultaneous confidence level given the specific marginal confidence intervals have been presented. The bootstrap methods have been employed to estimate the joint level of confidence for the entire collection of the marginal confidence intervals of $p_{1j} - p_{2j},\ j = 1, \ldots, k$. Especially, the marginal bootstrap confidence intervals are more accurate, at least in terms of estimating the true simultaneous confidence probability. In consideration of this preceding results, bootstrap methods are more reasonable and efficient way than the competing methods since they consider the dependence structure of the multivariate Bernoulli distribution.

ACKNOWLEDGEMENTS

Myoungshic Jhun and Jae Won Lee were supported by Korea Research Foundation Grant funded by Korea government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-070-C00020). And Hyeong Chul Jeong was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2006-521-C00031).

REFERENCES

BEAL, S. L. (1987). "Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples", *Biometrics*, **43**, 941–950.

BICKEL, P. J. AND FREEDMAN, D. A. (1981). "Some asymptotic theory for the bootstrap", *The Annals of Statistics*, **9**, 1196–1217.

BROWN, C. C. AND FEARS, T. R. (1981). "Exact significance levels for multiple binomial testing with application to carcinogenicity screens", *Biometrics*, **37**, 763–774.

HOCHBERG, Y. (1988). "A sharper Bonferroni procedure for multiple tests of significance", *Biometrika*, **75**, 800–802.

HOLLAND, B. S. AND COPENHAVER, M. D. (1987). "An improved sequentially rejective Bonferroni test procedure", *Biometrics*, **43**, 417–423.

JHUN, M., JEONG, H. C. AND BAHNG, J. S. (2007). "Simultaneous confidence intervals for the mean of multivariate poisson distribution: a comparison", *Communications in Statistics-Simulation and Computation*, **36**, 151–164.

MEE, R. W. (1984). "Confidence bounds for the difference between two probabilities", *Biometrics*, **40**, 1175–1176.

MIETTINEN, O. AND NURMINEN, M. (1985). "Comparative analysis of two rates", *Statistics in Medicine*, **4**, 213–226.

NEWCOMBE, R. G. (1998). "Interval estimation for the difference between independent proportions: comparison of eleven methods", *Statistics in Medicine*, **17**, 873–890.

PARK, C. G., PARK, T. S. AND SHIN, D. W. (1996). "A simple method for generating correlated binary variates", *The American Statistician*, **50**, 306–310.

SANTNER, T. J. AND SNELL, M. K. (1980). "Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 contingency tables", *Journal of the American Statistical Association*, **75**, 386–394.

SINGH, K. (1981). "On the asymptotic accuracy of Efron's bootstrap", *The Annals of Statistics*, **9**, 1187–1195.

WESTFALL, P. (1985). "Simultaneous small-sample multivariate Bernoulli confidence intervals", *Biometrics*, **41**, 1001–1013.

WESTFALL, P. H. AND YOUNG, S. S. (1989). "p-value adjustments for multiple tests in multivariate binomial models", *Journal of the American Statistical Association*, **84**, 780–786.

WESTFALL, P. H. AND YOUNG, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, John Wiley & Sons, New York.