Published in final edited form as:

Stat Med. 2008 June 30; 27(14): 2524–2535. doi:10.1002/sim.3055.

Tests for the homogeneity of two binomial proportions in extremely unbalanced 2×2 contingency tables

Seung-Ho Kang^{1,*†} and Chul W. Ahn²

¹Department of Statistics, Ewha Womans University, 11-1, DaeHyun-Dong, SeoDaeMun-Gu, Seoul 120-750, Korea

²Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX, U.S.A

SUMMARY

Asymptotic tests such as the Pearson chi-square test are unreliable for testing the homogeneity of two binomial probabilities in extremely unbalanced cases. Two exact tests (conditional and unconditional) are available as alternatives and can be implemented easily in StatXact 6.0. In equal sample cases it is well known that the unconditional exact test is more powerful than the conditional exact test. However, in this paper, we show that the opposite result holds in extremely unbalanced cases. The reason is that the peaks of the type I error occur at the extremes of the nuisance parameter when the imbalance among the sample sizes becomes severe. After we show that the conditional exact test is more powerful than the unconditional exact test in extremely unbalanced cases whose sample ratio is greater than 20, we compare the conditional exact test with the Berger and Boos approach (*J. Amer. Stat. Assoc.* 1994; **89**:1012–1016) in which the supremum is taken over a confidence interval for the nuisance parameter. The Berger and Boos approach turns out to be slightly more powerful than the conditional exact test in extremely unbalanced data. A real example is provided.

Keywords

size; type I error; epidemiology; exact power

1. INTRODUCTION

In this paper we focus on testing the homogeneity of two binomial proportions in extremely unbalanced 2×2 contingency tables obtained from independent binomial sampling. Here ‘extremely unbalanced’ means that the sample ratio n_2/n_1 is at least 20. We often encounter such extremely unbalanced 2×2 contingency tables in some epidemiological studies when we would like to compare the prevalence rates of a disease between a very large population and a small population.

For example, Arnett *et al.* [1] showed that the prevalence of scleroderma or systemic sclerosis (SSc) in full-blooded Choctaw Indians residing in eight southeastern Oklahoma counties was at least 8/1704 over the 4-year interval 1990–1994 and was significantly higher than that among non-full-blooded Choctaws (6/19 551). Choctaw Native American Indians in southeastern Oklahoma have the highest prevalence of scleroderma or SSc yet found

(469/100 000). Here, we apply the homogeneity test of two binomial proportions in an extremely unbalanced 2×2 contingency table to examine if the HLA DQB1*0201 allele frequency of a small population of Oklahoma Choctaw SSc cases is significantly different from that of a large population of non-Choctaw SSc cases. The frequencies of the HLA DQB1*0201 allele are presented in Table I.

In the data provided in Table I, the p -value of the asymptotic Pearson chi-square test is 0.073, while the p -values of the conditional and unconditional exact Pearson chi-square test are 0.093 and 0.117, respectively. Which p -value is the most reliable?

Simulation studies showed that, in unbalanced cases, asymptotic tests such as the Pearson chi-square test could produce anticonservative type I errors (for example, see [2–6]). Exact tests are often employed as alternatives of asymptotic tests. In testing the homogeneity of two binomial proportions, two exact tests have been developed: namely, the conditional exact test and the unconditional exact test. Since two exact tests guarantee that the type I errors are controlled under the nominal level, the next question is : ‘Which exact test produces greater power?’. Suissa and Shuster [7] showed that the unconditional exact test has greater power than the conditional exact test for comparing two binomial proportions in equal sample cases with a sample size as large as 150 per group. Does the same conclusion hold in extremely unbalanced cases? Mehta and Hilton [8] pointed out that the conditional exact test might be conservative due to the discreteness of the conditional distribution of the test statistic, and that the unconditional exact test might be conservative due to the need for eliminating the nuisance parameter by considering the worst possible case. They also concluded that the result of Suissa and Shuster [7] was due to the discreteness of the conditional distribution, and the difference in power between two exact tests decreases gradually as the sample size increases.

Recently we have observed some clues indicating that the opposite conclusion might hold true in extremely unbalanced 2×2 tables; in other words, the conditional exact test is more powerful than the unconditional exact test. Loh [9] and Loh and Yu [10] proved theoretically that the sizes of the Pearson chi-square test and the likelihood ratio test are greater than the nominal level in infinite samples. The key idea in their proof is to consider the cases in which the Poisson approximation to the binomial distribution is more accurate than normal approximation to the binomial distribution. Note that, since the size is defined as the supremum of the type I error over the nuisance parameter space, inflation of the sizes does not mean that the type I error is inflated at all values of the nuisance parameter. Recently, Kang *et al.* [11] showed that the inflation of the sizes increases as the sample ratio n_2/n_1 becomes large, and also obtained the asymptotic lower bound of the sizes of the Pearson chi-square test and the likelihood ratio test. Table II, which is a part of Table II in Kang *et al.* [11], presents the lower bounds of the sizes of the Pearson chi-square test and the likelihood ratio test in infinite samples.

For example, when $(f_1, f_2) = (0.01, 0.99)$, the size of the Pearson chi-square test is greater than 0.104 at 5 per cent significance level in infinite samples. Since the unconditional exact test removes the nuisance parameter by taking the supremum over the nuisance parameter space, the results in Table II imply that the unconditional exact test could be very conservative in extremely unbalanced 2×2 tables and could result in the loss of power. This clue motivates this study to compare the exact powers of the conditional and the unconditional exact test in extremely unbalanced 2×2 tables. With the help of StatXact 6.0 [12], we can compute the exact powers of the two exact tests easily. It turns out that the conditional exact test is more powerful than the unconditional exact test in extremely unbalanced 2×2 contingency tables whose sample ratio is at least 20.

Berger and Boos [13] proposed a new test procedure based on confidence interval p -value. Since the method takes the supremum over a $100(1-\gamma)$ per cent confidence interval of p instead of the whole parameter space of p , it could be less conservative and more powerful than the standard unconditional exact test if the confidence interval does not include the extremes of p where the peak of the type I error occurs. Since the conditional exact test is more powerful than the unconditional exact test in extremely unbalanced 2×2 contingency tables and Berger [14] also showed that the Berger and Boos approach is more powerful than the unconditional exact test for testing the homogeneity of two binomial proportions, the next question is to conduct the power comparison between the conditional exact test and the Berger and Boos approach. From the numerical computation in Section 3, it turns out that the Berger and Boos approach is slightly more powerful than the conditional exact test in all investigated cases of extremely unbalanced 2×2 contingency tables. Freidlin and Gastwirth [15] also showed that the Berger and Boos approach is more powerful than the conditional exact test in both the Cochran–Mantel–Haenszel test for stratified 2×2 tables and the trend test for $2 \times K$ contingency tables.

This paper is organized as follows. In Section 2 we introduce notations and review the asymptotic tests, the conditional exact test, the unconditional exact test and the Berger and Boos approach. In Section 3 we compare the exact powers of the conditional exact test, the unconditional exact test and the Berger and Boos approach in extremely unbalanced 2×2 tables. A real example is provided in Section 4.

2. NOTATIONS AND REVIEW

Let X_1 and X_2 be two independent binomial random variables, $X_k \sim B(n_k, p_k)$, $k = 1, 2$. Then we have the 2×2 contingency table as in Table III. We wish to test $H_0 : p_1 = p_2 = p$ versus $H_1 : p_1 \neq p_2$, where $0 < p < 1$.

2.1. Asymptotic tests

When the expected cell frequencies are greater than some required number (for example, 5), then the Pearson chi-square test is often employed, which rejects H_0 if $X^2 > \chi_{1,\alpha}^2$, where $\chi_{1,\alpha}^2$ is the upper α -quantile of the chi-squared distribution with one degree of freedom and

$$X^2 = \sum_{i=1}^2 \left[\frac{(X_i - n_i M/n)^2}{n_i M/n} + \frac{(n_i - X_i - n_i(n-M)/n)^2}{n_i(n-M)/n} \right] \quad (1)$$

The Pearson chi-square test is equivalent to the Z-test, which rejects H_0 if $|Z| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $\alpha/2$ -quantile of the standard normal distribution and

$$Z = \frac{X_1/n_1 - X_2/n_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}, \quad \hat{p} = (X_1 + X_2)/(n_1 + n_2)$$

The likelihood ratio test is also often used, which rejects H_0 if $G^2 > \chi_{1,\alpha}^2$, where

$$G^2 = 2 \sum_{i=1}^2 \left[X_i \log \left(\frac{n X_i}{n_i M} \right) + (n_i - X_i) \log \left(\frac{n(n_i - X_i)}{n_i(n-M)} \right) \right]$$

When the 2×2 contingency tables are unbalanced, there are some cells in which expected frequencies are smaller than some required number (for example, 5). In such cases it is well known that the Pearson chi-square test is unreliable. Kang *et al.* [11] investigated the type I errors of three popular asymptotic tests: the Pearson chi-square test with and without continuity correction and the likelihood ratio test. They showed that the peaks of the type I errors are greater than the nominal level and the peaks increase as the sample ratio n_2/n_1 increases. They further showed that the type I errors are close to the nominal level at most values of the nuisance parameter and the peaks of the type I errors occur only at the extremes of the nuisance parameter.

2.2. Conditional exact tests

In this subsection we review the conditional exact tests of the homogeneity of two binomial proportions. Under the null hypothesis, the likelihood of (X_1, X_2) is given by

$$P(X_1=x_1, X_2=x_2|H_0) = \prod_{i=1}^2 \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i} \quad (2)$$

In conditional exact tests we use the conditional null distribution of a test statistic, given a sufficient statistic of an unknown nuisance parameter, instead of the true distribution in (2), because the value of p is unknown. In testing the homogeneity of two binomial proportions, $M = X_1 + X_2$ is a sufficient statistic of the unknown nuisance parameter p under the null hypothesis. Hence, the following hypergeometric distribution is used to construct conditional exact tests:

$$P(X_1=x_1, X_2=x_2|M=m, H_0) = \binom{n_1}{x_1} \binom{n_2}{x_2} / \binom{n_1+n_2}{m}$$

because the above hypergeometric distribution does not depend on the unknown nuisance parameter p .

Let T be a test statistic on the sample space $S = \{0, 1, 2, \dots, n_1\} \times \{0, 1, 2, \dots, n_2\}$, whose large values indicate a deviation from the null hypothesis, and (x_1^0, x_2^0) be the observed vector of (X_1, X_2) and $m = x_1^0 + x_2^0$. When m , an observed value of M , is given, the conditional exact tests consider only the following set of 2×2 tables as a sample space:

$$\Gamma_m = \{(x_1, x_2) | x_1 + x_2 = m, 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2\}$$

In other words, the conditional exact tests deal with 2×2 contingency tables whose margins are the same as those of the observed table. Since the large value of T is significant, for the given m , the rejection region of the conditional exact tests is given by

$$\Gamma_m(t) = \{(x_1, x_2) \in \Gamma_m | T(x_1, x_2) \geq t\}$$

For computation of the exact power, we evaluate the exact null distribution of T :

$$P(T \geq t|m, H_0) = \sum_{(x_1, x_2) \in \Gamma_m(t)} \binom{n_1}{x_1} \binom{n_2}{x_2} / \binom{n_1+n_2}{m}$$

for each possible value of t . Let α be the given nominal level and $t_{\alpha(m)}$ be the smallest possible value such that

$$P(T \geq t_{\alpha(m)}|m, H_0) \leq \alpha$$

For a given m , the exact conditional test rejects H_0 if the observed value of T is greater than or equal to $t_{\alpha}(m)$. Then the exact power of the conditional exact test is

$$\begin{aligned} \beta^c(p_1, p_2) &= \sum_{m=0}^n P(T \geq t_{\alpha}(m)|m, p_1, p_2) P(M=m|p_1, p_2) \\ &= \sum_{(x_1, x_2) \in \Gamma_{\alpha}} \prod_{i=1}^2 \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i} \end{aligned} \quad (3)$$

where

$$\Gamma_{\alpha} = \bigcup_{m=0}^n \Gamma_m(t_{\alpha}(m))$$

2.3. Unconditional exact tests

In unconditional exact tests we consider the following reference set:

$$\Gamma = \bigcup_{m=0}^n \Gamma_m = \{(x_1, x_2) | 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2\}$$

For computing the power of the unconditional exact test, we evaluate the exact null distribution of T :

$$P(T \geq t|p, H_0) = \sum_{(x_1, x_2) \in \Gamma} \prod_{i=1}^2 \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i-x_i} I_{[T(x_1, x_2) \geq t]}$$

for each possible value of t . When $t = \chi_{1, \alpha}^2$, $P(T \geq t|p, H_0)$ is plotted in Figure 1 of Kang *et al.* [11]. Let α be the given nominal level and $t_{\alpha}(p)$ be the smallest possible value such that

$$P(T \geq t_{\alpha}(p)|p, H_0) \leq \alpha$$

The unconditional exact test rejects H_0 if the observed value of T is greater than or equal to an unconditional critical cutoff value, t_{α} , which is defined by

$$t_\alpha = \sup_{0 < p < 1} t_\alpha(p)$$

Then the power of the unconditional exact test can be computed by

$$\beta^u(p_1, p_2) = P(T \geq t_\alpha | p_1, p_2) = \sum_{(x_1, x_2) \in \Gamma} \prod_{i=1}^2 \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i} I_{[T(x_1, x_2) \geq t_\alpha]} \quad (4)$$

2.4. The Berger and Boos approach

In the standard unconditional exact test, the p -value is defined by

$$p_T(x_1, x_2) = \sup_{p \in [0, 1]} P_p(T(X_1, X_2) \geq T(x_1, x_2))$$

for a test statistic T whose large value is significant. In the Berger and Boos [13] approach, which is also often called the confidence interval p -value, the supremum of $P_p(T(X_1, X_2) \geq T(x_1, x_2))$ is taken over a $100(1-\gamma)$ per cent confidence interval C_γ for p , so that the p -value is given by

$$p_C(x_1, x_2) = \sup_{p \in C_\gamma(x_1, x_2)} P_p(T(X_1, X_2) \geq T(x_1, x_2)) + \gamma$$

The rationale of this approach is to consider only plausible values of p that are supported by the data. The confidence interval employed in this paper is the Clopper and Pearson [16] interval as Berger [14] used. The interval is given by

$$\frac{a}{a + (b-a+1)F_{2(b-a+1), 2a, \gamma/2}} \leq p \leq \frac{(a+1)F_{2(a+1), 2(b-a), \gamma/2}}{b-a + (a+1)F_{2(a+1), 2(b-a), \gamma/2}}$$

where $a = x_1 + x_2$, $b = n_1 + n_2$ and $F_{\tau_1, \tau_2, \gamma/2}$ is the upper $100(\gamma/2)$ per cent percentile of an F distribution with (τ_1, τ_2) degrees of freedom. When $\gamma = 0$, $p_C(x_1, x_2)$ is the same as $p_T(x_1, x_2)$. The Berger and Boos [13] method rejects H_0 if $p_C(x_1, x_2) \leq \alpha$. This approach guarantees that the actual size does not exceed the nominal level [13]. In this paper we use $\gamma = 0.001$, which is suggested by Berger and Boos [13]. For computing the power of the Berger and Boos approach, we find the rejection region, which is given by

$$\Gamma_{B(\alpha)} = \{(x_1, x_2) : p_C(x_1, x_2) \leq \alpha, 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2\}$$

Then the power of the Berger and Boos approach can be computed by

$$\beta^B(p_1, p_2) = \sum_{(x_1, x_2) \in \Gamma_{B(\alpha)}} \prod_{i=1}^2 \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i} \quad (5)$$

3. POWER COMPARISONS

In this section we will compare the powers of the conditional exact test, the unconditional exact test and the Berger and Boos method. Note that, since the sizes of three tests are guaranteed not to exceed the nominal level, power comparisons are legitimate. For power comparisons between the conditional exact test and the unconditional exact test, we can rewrite equation (4) as

$$\beta^u(p_1, p_2) = \sum_{m=0}^n P(T \geq t_\alpha | m, p_1, p_2) P(M=m | p_1, p_2)$$

This equation is the same as equation (3), except that $t_\alpha(m)$ is replaced with t_α . Therefore, if $t_\alpha < \min_{0 \leq m \leq n} t_\alpha(m)$, we ensure that the unconditional exact test will be more powerful. On the other hand, if $t_\alpha > \max_{0 \leq m \leq n} t_\alpha(m)$, it guarantees that the conditional exact test is more powerful. In between these two extremes, either test could be more powerful.

Hereafter, for convenience we employ the Pearson chi-square test statistic in (1) to compare the exact powers of two exact tests. Computing $t_\alpha(m)$'s and t_α sheds some light on which exact test has greater power. For example, when $(n_1, n_2) = (50, 1000)$, the values of $t_\alpha(p)$ are computed at every value of p from 0.001 to 0.999 by increments of 0.001. The values of $t_\alpha(p)$ are 5.42 at $p = 0.002, 0.003$ and 0.004 . However, except in these cases, most values lie between 3.31 and 3.84 and their medium value is 3.82. Note that $t_\alpha = \sup_{0 < p < 1} t_\alpha(p) = 5.42$, although most values of $t_\alpha(p)$ range between 3.31 and 3.84. The values of $t_\alpha(m)$ are computed at each possible value of m and their percentiles are displayed in Table IV. If the extreme value 5.42 of $t_\alpha(p)$ does not exist, t_α would be 3.84, and would be smaller than 1 per cent percentile of the values of $t_\alpha(m)$, and the unconditional exact test would be more powerful than the conditional exact test. But, extreme imbalance among the sample sizes produces very large values of t_α ($t_\alpha = 5.42$). The value of 5.42 is greater than the 75 per cent percentile of the values of $t_\alpha(m)$, which results in loss of the power of the unconditional exact test. Therefore, the conditional exact test has greater power than the unconditional exact test when $(n_1, n_2) = (50, 1000)$.

To compute the powers of the Berger and Boos approach, we note that

$$\sup_{p \in C_\gamma(x_1, x_2)} P_p(T(X_1, X_2) \geq T(x_1, x_2)) \leq \sup_{p \in [0, 1]} P_p(T(X_1, X_2) \geq T(x_1, x_2)) \quad (6)$$

If

$$\sup_{p \in [0, 1]} P_p(T(X_1, X_2) \geq T(x_1, x_2)) \leq \alpha - \gamma \quad (7)$$

from equation (6), we have

$$\sup_{p \in C_\gamma(x_1, x_2)} P_p(T(X_1, X_2) \geq T(x_1, x_2)) \leq \alpha - \gamma$$

In other words, any sample point (x_1, x_2) that satisfies condition (7) belongs to the rejection region of the Berger and Boos approach. Let $t_{\alpha-\gamma} = \sup_{0 < p < 1} t_{\alpha-\gamma}(p)$ be the unconditional critical cutoff value that is introduced in Section 2.3. We can compute $t_{\alpha-\gamma}$ easily, because we do not have to consider all possible values of p in $[0, 1]$. Since the peak of the type I error occurs at the extremes of p [11], it is enough to examine some small intervals of p that contain the extremes of p . Any sample point (x_1, x_2) with $T(x_1, x_2) \leq t_{\alpha-\gamma}$ satisfies equation (7) and is contained in the rejection region. We saved a lot of computing time by using the procedure described above. For the remaining sample points (x_1, x_2) with $T(x_1, x_2) > t_{\alpha-\gamma}$, we find (x_1^*, x_2^*) with $p_C(x_1^*, x_2^*) \simeq 2\alpha$ and compute $p_C(x_1, x_2)$ for each (x_1, x_2) in $\{(x_1, x_2) | T(x_1, x_2) \geq T(x_1^*, x_2^*)\}$. Then the power is computed based on equation (5).

By using commercial software StatXact 6.0 [12], we computed the exact powers of the conditional exact test and the unconditional exact test based on the Pearson chi-square test statistic in (1) at various values of (p_1, p_2) when $(n_1, n_2) = (100, 1000)$, $(50, 1000)$, $(20, 1000)$ and $(20, 2000)$. The powers of the Berger and Boos approach based on the Pearson chi-square test statistic in (1) are calculated with the SAS program by employing the procedure described above.

The numerical results are displayed in Tables V and VI. From Table V when $(n_1, n_2) = (100, 1000)$ the exact powers of the two exact tests are similar. But when $(n_1, n_2) = (50, 1000)$ it is clear that the conditional exact test is more powerful than the unconditional exact test at every value of (p_1, p_2) . For example, the difference of the exact power at $(p_1, p_2) = (0.1, 0.2)$ is 36.8–23.4 per cent = 13.4 per cent. From Tables V and VI, we see that the difference of the exact power between the conditional exact test and the unconditional exact test increases as the imbalance among the sample sizes increases. For instance, when $(n_1, n_2) = (50, 1000)$, $(20, 1000)$ and $(20, 2000)$, the difference of the exact power at $(p_1, p_2) = (0.2, 0.4)$ is 8.6 per cent (81.8 *versus* 73.2 per cent), 21.6 per cent (39.9 *versus* 18.3 per cent) and 33.5 per cent (40.9 *versus* 7.4 per cent), respectively. It coincides with the fact discovered by Kang *et al.* [11] that the peaks of the type I error of the Pearson chi-square test become larger as the imbalance among the sample sizes increases. The powers of the Berger and Boos approach are slightly larger than those of the conditional exact test in all investigated cases. For example, when $(n_1, n_2) = (50, 1000)$ and $(p_1, p_2) = (0.3, 0.5)$, the powers of the two tests are 80.6 and 78.6 per cent, respectively. The conditional exact test does not seem to be the most powerful due to the discreteness of the conditional distribution. The highest powers of the Berger and Boos approach imply that the approach reduces the conservatism of the unconditional exact test.

So far we have used the Pearson chi-square test statistic as a test statistic. Now we investigate whether the same conclusions hold for the likelihood ratio test. The powers of the three methods based on the likelihood ratio test are displayed in Table VII when $(n_1, n_2) = (50, 1000)$. The same conclusions as for the Pearson chi-square test hold for the likelihood ratio test. In other words, the powers are decreasing in the order: of the Berger and Boos approach, the conditional exact test and the unconditional exact test. For example, when $(p_1, p_2) = (0.4, 0.5)$, the powers of the Berger and Boos approach, the conditional exact test and the unconditional exact test are 27.8, 24.6 and 18.7 per cent, respectively.

4. EXAMPLE

Scleroderma (SSc) is a connective tissue disease of unknown etiopathogenesis, characterized by prominent cutaneous and visceral fibrosis. Choctaw Native Americans residing in southeastern Oklahoma have a strikingly high prevalence rate of SSc [1]. Tan *et al.* [17] investigated whether the HLA allele frequency in Choctaw SSc cases is significantly different from that in Choctaw controls. Dr Tan kindly provided us the HLA allele frequencies of Choctaw SSc cases and non-Choctaw SSc cases. Here, we investigate whether the HLA DQB1*0201 allele frequency of Choctaw SSc cases is significantly different from that of non-Choctaw SSc cases. Table I shows the DQB1*0201 allele frequencies of Choctaw and non-Choctaw SSc cases. Suppose that we have decided to use the Pearson chi-square test. The p -values of the asymptotic Pearson chi-square test, the Berger and Boos approach, the conditional and unconditional exact chi-square test are 0.073, 0.082, 0.093 and 0.117, respectively. From the results of Section 3, the p -value of the Berger and Boos approach (0.082) is recommended.

5. CONCLUSION AND DISCUSSION

It is known that, in equal-sample cases, the unconditional exact test is more powerful than the conditional exact test [7]. However, in this paper we show that the opposite result holds in extremely unbalanced 2×2 contingency tables. Extreme imbalance among sample sizes makes the unconditional exact test very conservative and lose a lot of power. Kang *et al.* [11] showed that the peak of the type I error occurs only at the extremes of p in extremely unbalanced data. Most confidence intervals computed in the Berger and Boos approach do not include the extremes of p so that the Berger and Boos approach can be somewhat more powerful than the conditional exact test.

We conjecture that the same conclusion will hold in some problems if data are extremely unbalanced. For example, for testing the Hardy–Weinberg law, Kang and Shin [18] showed that the sizes of the chi-square test exceed the nominal level when rare alleles are present. Since both the conditional and the unconditional exact test have been developed [19, 20], it will be interesting to compare the exact powers of two exact tests and the Berger and Boos approach in a future study.

In comparing two Berger and Boos approaches based on the Pearson chi-square test and the likelihood ratio test, the results in Tables V–VII show that one does not uniformly outperform the other. Martín and Silva [21] presented several possible candidates of test statistics. It would require a future study to investigate which test statistic produces the greatest power in extremely unbalanced 2×2 contingency tables if the Berger and Boos approach is used.

Acknowledgments

Contract/grant sponsor: Korean Government (MOEHRD); contract/grant number: KRF-2005-070-C00021

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-070-C00021).

References

1. Arnett FC, Howard RF, Tan F, Moulds JM, Bias WB, Durban E, Cameron HD, Paxton G, Hodge TJ, Weathers PE, Reveille JD. Increased incidence of systemic sclerosis in a native American trib in Oklahoma: association with an Amerindian HLA haplotype. *Arthritis and Rheumatism*. 1996; 39(8):1362–1370. [PubMed: 8702445]

2. Berry KJ, Mielke PW. Monte Carlo comparisons of the asymptotic chi-square and likelihood-ratio tests with the nonasymptotic chi-square test for sparse $r \times c$ tables. *Psychological Bulletin*. 1988; 103(2):256–264.
3. Bradley DR, Bradley TD, McGrath SG, Cutcomb SD. Type I error rate of the chi-square test of independence in $R \times C$ tables that have small expected frequencies. *Psychological Bulletin*. 1979; 86(6):1290–1297.
4. Kraemer DF, Woolson RF. A comparison of tests of homogeneity for sparse contingency tables. *Communications in Statistics: Computation and Simulation*. 1987; 16(2):465–483.
5. Lawal HB, Upton GJG. Comparisons of some chi-squared tests for the test of independence in sparse two-way contingency table. *Biometrical Journal*. 1990; 32(1):59–72.
6. Tate MW, Hyer LA. Inaccuracy of the χ^2 test of goodness of fit when expected frequencies are small. *Journal of the American Statistical Association*. 1973; 68:836–841.
7. Suissa S, Shuster JJ. Exact unconditional sample sizes for the 2×2 binomial trial. *Journal of the Royal Statistical Society A*. 1985; 148(4):317–327.
8. Mehta CR, Hilton JF. Exact power of conditional and unconditional tests: going beyond the 2×2 contingency table. *The American Statistician*. 1993; 47(2):91–98.
9. Loh WY. Bounds on the size of the χ^2 test of independence in a contingency table. *Annals of Statistics*. 1989; 17:1709–1722.
10. Loh WY, Yu X. Bounds on the size of the likelihood ratio test of independence in a contingency table. *Journal of Multivariate Analysis*. 1993; 45:291–304.
11. Kang SH, Lee Y, Park E. The sizes of the three popular asymptotic tests for testing homogeneity of two binomial proportions. *Computational Statistics and Data Analysis*. 2006; 51:710–722.
12. Cytel. StatXact, version 6.0, Software for Exact Nonparametric Statistical Inference with Continuous or Categorical Data. Cytel Software; Cambridge, MA: 2006.
13. Berger RL, Boos DD. P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*. 1994; 89:1012–1016.
14. Berger RL. More powerful tests from confidence interval p values. *The American Statistician*. 1996; 50(4):314–318.
15. Freidlin B, Gastwirth JL. Unconditional versions of several tests commonly used in the analysis of contingency tables. *Biometrics*. 1999; 55:264–267. [PubMed: 11318165]
16. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26:404–413.
17. Tan FK, Stivers DN, Arnett FC, Charkabarty R, Howard R, Reveille JD. HLA haplotypes and microsatellite polymorphisms in and around the major histocompatibility complex region in a Native American population with a high prevalence of scleroderma (systemic sclerosis). *Tissue Antigens*. 1999; 53(1):74–80. [PubMed: 10082433]
18. Kang SH, Shin DW. The size of the chi-square test for the Hardy-Weinberg law. *Human Heredity*. 2004; 58:10–17. [PubMed: 15604560]
19. Haber M. An exact unconditional test for the Hardy-Weinberg equilibrium. *Biometrical Journal*. 1994; 36:741–749.
20. Aoki S. Network algorithm for the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrical Journal*. 2003; 45(4):471–490.
21. Martín Andrés A, Silva Mato A. Choosing the optimal unconditional test for comparing two independent proportions. *Computational Statistics and Data Analysis*. 1994; 17:555–574.

Table I

DQB1*0201 allele frequencies in Choctaw SSc cases *versus* non-Choctaw SSc cases.

DQB1*0201	Choctaws	Non-Choctaws
Present	1	251
Absent	18	866
	19	1117

Table II

Large-sample lower bounds of the sizes of the Pearson chi-square test (χ^2) and the likelihood ratio test (G^2).

(f_1, f_2)	χ^2	G^2	(f_1, f_2)	χ^2	G^2
(0.5, 0.5)	0.053	0.097	(0.1, 0.9)	0.050	0.089
(0.4, 0.6)	0.051	0.077	(0.05, 0.95)	0.071	0.104
(0.3, 0.7)	0.054	0.068	(0.01, 0.99)	0.104	0.130
(0.2, 0.8)	0.092	0.079			

Note: $\alpha = 0.05$, $f_i = n_i/(n_1 + n_2)$, $i = 1, 2$.

Table III

A 2×2 contingency table.

	Group 1	Group 2	Totals
Response	X_1	X_2	M
No response	$n_1 - X_1$	$n_2 - X_2$	$n - M$
Totals	n_1	n_2	n

Note: $M = X_1 + X_2$, $n = n_1 + n_2$.

Table IV

Percentiles of the critical values of the conditional exact test when $(n_1, n_2) = (50, 1000)$.

Percentile (per cent)	$t_{\alpha}(m)$
100	40.08
99	16.34
95	7.66
90	5.80
75	4.91
50	4.60
25	4.37
10	4.20
5	4.12
1	3.98

Table V

Exact powers based on the Pearson chi-square test statistic.

$(n_1, n_2) = (100, 1000)$					$(n_1, n_2) = (50, 1000)$				
p_1	p_2	Condi	Uncondi	Berger	p_1	p_2	Condi	Uncondi	Berger
0.1	0.2	70.6	71.3	73.0	0.1	0.2	36.8	23.4	40.8
0.1	0.3	99.8	99.8	99.8	0.1	0.3	91.9	85.2	93.5
0.2	0.3	54.5	54.7	56.6	0.2	0.3	27.7	18.1	31.2
0.2	0.4	98.5	98.6	98.8	0.2	0.4	81.8	73.2	85.0
0.3	0.4	47.0	48.0	49.9	0.3	0.4	24.3	16.5	28.2
0.3	0.5	97.4	97.2	97.5	0.3	0.5	78.6	68.6	80.6
0.4	0.5	46.8	46.0	47.5	0.4	0.5	25.5	16.9	27.8
0.4	0.6	96.7	96.8	97.1	0.4	0.6	77.1	68.7	80.0
0.5	0.6	46.6	47.1	48.8	0.5	0.6	25.7	18.4	29.1
0.5	0.7	97.4	97.5	97.7	0.5	0.7	81.0	72.8	82.6
0.6	0.7	50.7	51.8	53.5	0.6	0.7	30.2	21.7	32.7
0.6	0.8	98.8	98.8	99.0	0.6	0.8	87.0	81.2	88.5
0.7	0.8	60.5	61.3	62.9	0.7	0.8	37.9	28.8	40.6
0.7	0.9	99.8	99.9	99.9	0.7	0.9	95.3	93.1	96.1
0.8	0.9	78.8	79.6	81.0	0.8	0.9	54.6	46.4	57.9
0.8	0.95	99.5	99.6	99.6	0.8	0.95	92.9	90.5	94.0
0.9	0.95	49.9	51.2	53.0	0.9	0.95	32.8	26.8	36.4
0.9	0.97	84.0	85.1	85.4	0.9	0.97	60.5	56.8	65.2

Note: Condi, Conditional exact test; Uncondi, Unconditional exact test; Berger, Berger and Boos approach.

Table VI

Exact powers based on the Pearson chi-square test statistic.

$(n_1, n_2) = (20,1000)$				$(n_1, n_2) = (20,2000)$					
p_1	p_2	Condi	Uncondi	Berger	p_1	p_2	Condi	Uncondi	Berger
0.1	0.3	49.1	13.4	49.2	0.1	0.3	43.5	8.5	46.0
0.1	0.4	85.2	62.8	87.3	0.1	0.4	86.4	40.2	86.9
0.2	0.4	39.9	18.3	42.9	0.2	0.4	40.9	7.4	41.6
0.2	0.5	79.1	55.1	80.4	0.2	0.5	80.1	41.3	80.4
0.3	0.5	40.5	19.1	42.1	0.3	0.5	41.4	10.9	41.7
0.3	0.6	76.0	55.8	77.9	0.3	0.6	77.0	43.0	77.6
0.4	0.6	40.6	21.9	43.0	0.4	0.6	41.4	13.5	42.2
0.4	0.7	77.5	60.6	79.6	0.4	0.7	78.0	51.8	79.1
0.5	0.7	44.8	26.6	47.4	0.5	0.7	44.9	20.0	46.6
0.5	0.8	83.6	71.1	85.5	0.5	0.8	84.7	62.1	86.3
0.6	0.8	53.7	36.8	56.6	0.6	0.8	55.2	27.8	57.6
0.6	0.9	92.0	86.3	93.6	0.6	0.9	92.6	82.9	94.3
0.7	0.9	69.3	56.9	73.1	0.7	0.9	70.7	51.4	74.7
0.7	0.95	89.8	88.0	93.1	0.7	0.95	89.5	85.6	93.1
0.8	0.9	30.2	19.0	34.0	0.8	0.9	31.6	15.6	35.6
0.8	0.95	60.4	56.9	69.9	0.8	0.95	59.5	52.6	69.7
0.9	0.95	14.8	12.7	23.5	0.9	0.95	13.9	10.8	23.3
0.9	0.97	31.8	30.4	34.5	0.9	0.97	32.3	27.9	38.0

Note: Condi, Conditional exact test; Uncondi, Unconditional exact test; Berger, Berger and Boos approach.

Table VII

Exact powers based on the likelihood ratio test statistic.

$(n_1, n_2) = (50, 1000)$									
p_1	p_2	Condi	Uncondi	Berger	p_1	p_2	Condi	Uncondi	Berger
0.1	0.2	41.5	35.1	47.2	0.5	0.7	78.0	72.2	80.9
0.1	0.3	92.8	90.0	94.7	0.6	0.7	26.9	21.1	30.3
0.2	0.3	29.8	24.0	34.5	0.6	0.8	83.5	79.1	86.1
0.2	0.4	82.2	77.5	85.9	0.7	0.8	32.2	26.1	36.0
0.3	0.4	24.9	19.9	29.5	0.7	0.9	92.2	90.9	93.4
0.3	0.5	77.6	71.3	80.7	0.8	0.9	43.3	40.2	47.1
0.4	0.5	24.6	18.7	27.8	0.8	0.95	85.2	86.2	86.5
0.4	0.6	75.1	69.4	79.2	0.9	0.95	18.6	19.6	20.2
0.5	0.6	23.8	18.8	28.1	0.9	0.97	57.7	45.3	45.6

Note: Condi, Conditional exact test; Uncondi, Unconditional exact test; Berger, Berger and Boos approach.