A NEW METHOD FOR CALCULATING *P*-VALUE UNDER UNCONDITIONAL EXACT TEST

by

Tao JIANG, Ting CHEN, Xuexue MENG, Han MENG, and Min XIAO*

School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China

Original scientific paper
https://doi.org/10.2298/TSCI200220129J

An unconditional exact test is a classic method to test the significant difference between two independent binomial proportions or multinomial distributions. The p-value based on the unconditional exact test is computed by maximizing the probability of the tail region. The grid search method and polynomial method are able to find the maximum with sophisticated enough partition of the parameter space, while they require a rather long time to compute and those methods are computationally intensive for a study beyond two groups. In this paper, we propose a new method to obtain the solution of the global maximum which can diminish the computing time based on the fixed-point iterative algorithm. Additionally, both simulation and experiment indicate that this method is more competitive compared with the grid search and the polynomial method on the basis of guaranteed accuracy.

Key words*: p-value, unconditional exact test, monotonic condition, boundary, fixed-point iterative*

Introduction

In a hypothesis testing problem in clinical trials, such as cancer studies [1], AIDS research studies [2], and gastroesophageal research studies [3], the outcome is usually binary and the tail probability function is fundamentally important to guarantee that the actual Types I and II error rates of their associated parameter spaces. In a study of comparing two binary outcome proportions, the exact unconditional test computes the *p*-value by maximizing the tail probability over the nuisance parameter which is the common response rate from the two groups. Unlike the commonly used exact conditional test, *e. g.* Fisher's exact test [4], that assumes both marginal total sofa contingency table fixed, exact unconditional tests always assume only one marginal total or only the total sum fixed. Therefore, exact unconditional tests are increasingly used in practice for categorical data to increase the power of a study and to make the data analysis approach being consistent with the study designs. Unconditional exact tests become popular in categorical data analysis to compare two or more independent proportions [5-7], dependent proportions [8, 9], count data [10], and soon. For more references on the exact and the exact unconditional test, see [11-15].

* Corresponding author, e-mail: xiaomin90224@163.com

Specifically, in the test of whether there is a significant difference in the population parameter of the binomial distribution, the null hypothesis and the alternative hypothesis are:

$$H_0 : \delta = \pi_1 - \pi_2 \le \delta_0 \ \nu \ H_1 : \delta = \pi_1 - \pi_2 \le \delta_0 \tag{1}$$

where $0 < \delta_0 < 1$, π_1 is the response rate of the control group, and π_2 is the response rate of the experimental group.

Traditionally, the grid search method, numeric search methods and polynomial method are utilized in finding the maximum of the tail probability curve. When the curve is erratic, the grid search method has better performance than numerical search methods, since the latter depends on multiple factors in finding the accurate global maximum: initial value, search range, search algorithm, and convergence issue. In the accurate unconditional test method, there are two ways to search the maximum value of the tail probability in the parameter space. The first is grid search method, when the plain grid search method is used to search the maximum value of the tail probability in the parameter space, the approximate maximum, not the accurate global maximum is obtained. In addition, the grid search method becomes computationally intensive in a study with multiple groups or multiple stages [16-18]. For example, in the case of binomial distribution, Chan [19] discussed that when the number of grids in the grid search algorithm is greater than 1000. The results of Shan [16, 20] simulation and empirical analysis show that the accurate unconditional test method has better performance, but it is difficult to popularize because of its computational density. Then, the polynomial method is proposed, which converts the tail probability formula and divides the parameter space into some intervals of the same length in the form of segmentation, but requires the number of intervals *k* is unstable in different sample spaces. However, the result of the tail probability is also affected by *k*, when *k* is large enough, a more accurate tail probability can be obtained. Therefore, the calculation method of ten consumes a lot of calculation time, and the result is unsatisfactory. Similarly, in multiple group and multiple stage studies, the intensive nature of polynomial method can lead to computational difficulties also [16, 20]. Therefore, for sake of improving the defects of traditional methods, further optimizing the calculation and saving considerable time, this paper proposes a new algorithm based on the fixed-point iterative method.

Methods

In order to test the previous hypothesis (1), suppose the number of samples of the control group and the experimental group are n_1, n_2, respectively. At the same time, the response numbers are x_1 and x_2, which independently follow the 2×2 binomial distribution with parameters (n_1, π_1) and (n_2, π_2). The results can be summarized in the contingency table [4, 21] in the tab. 1, where $y_i = n_i - x_i$, $a_1 = x_1 + x_2$, $a_2 = y_1 + y_2$, and $n = n_1 + n_2$.

It is known that the *Z*-test statistic usually be used see Chan [19], however, in the case of small or medium sample size, the deviation between the asymptotically distribution and the real distribution is far. Therefore, the exact unconditional test method is recommended to be used in the case of small sample size. In the case where the null hypothesis is true and the two variables are independent, the probability of the sample (x_1, x_2) is:

$$p(x_1, x_2; \pi_1, \pi_2) = \binom{n_1}{x_1} \pi_1^{x_1} (1-\pi_1)^{(n_1 - x_1)} \binom{n_2}{x_2} \pi_2^{x_2} (1-\pi_2)^{(n_2 - x_2)} \tag{2}$$

where $\pi_1, \pi_2 \in [0,1]$, given a test statistic T and T_0 is observed by the sample, the tail region is $CR = [(x_1, x_2) \mid T > T_0]$, then the tail probability is:

$$\sum_{(x_1,x_2\in CR)} p(x_1,x_2;\pi_1,\pi_2)$$

The unconditional exact test computes the *p*-value by maximizing the tail probability in the parameter space from 0 to 1.

Table 1. The result of an experiment for comparing two independent binomial proportions

A_0	A_0	A_1	Total
Control group	x_1	y_1	n_1
Experimental group	x_2	y_2	n_2
Total	a_1	a_2	n

It is known that the *p*-value is calculated by maximizing the tail probability function in the parameters space. Berger and Sidik [22] research shows that under certain conditions, it is not need to maximize the tail probability function about the whole parameters space, and the maximum value of the tail probability function is taken on the boundary of null hypothesis H_0. In our hypothesis, X_1 and X_2 are independent binomial random variables. The tail region constructed by *Z*-test statistic conforms to the condition of "C" convex set in Sidik theorem. The maximum value of *p*-value, P_M, of standard unconditional test method is obtained at the boundary $[(\pi_1,\pi_2:\pi_1-\pi_2=\delta_0)]$.

Two independent binomial distribution

For testing the equality of parameters of two independent binomial distributions, under the null hypothesis $H_0:\pi_1-\pi_2=0$, the parameter space $\{x_1,x_2\}$ is a triangle space, and the actual maximum is obtained under the condition of $\pi_1=\pi_2=\pi$ by the Sidik theorem [22], since the Barnard convexity condition is satisfied. Then, the unconditional *p*-value of the observed result is:

$$\alpha(\pi) >= \max \sum_{(x_1,x_2)\in CR} p(x_1,x_2;\pi)$$

The $\alpha(\pi)$ is usually obtained by the grid search method and the polynomial method [10]. The *p*-value based on the exact unconditional Barnard test is calculated by maximizing the tail probability over a redundant parameter ranging from 0 to 1.

As it is previously mentioned, the exact unconditional test computes the *p*-value by maximizing the tail probability, and the objective function for calculating the exact *p*-value is:

$$\alpha(\pi)=\max \sum_{(x_1,x_2)\in CR} p(x_1,x_2;\pi)=\max \sum_{(x_1,x_2)\in CR}\binom{n_1}{x_2}\binom{n_2}{x_2}\pi^{(x_1+x_2)}(1-\pi)^{(n_1+n_2-x_1-x_2)}$$

Owing to the complexity of the tail probability function, it is difficult to directly calculate the maximum value of the tail region probability function through numerical methods. Therefore, we can transform the tail probability function through mathematical processing:

$$F(x_1,x_2;\pi)=\sum_{(x_1,x_2)\in CR}\exp[\ln p(x_1,x_2;\pi)] \tag{3}$$

Then, we take the partial derivatives:

$$\frac{\partial F(x_1, x_2; \pi)}{\partial \pi} = \sum_{(x_1, x_2) \in CR} \exp[\ln p(x_1, x_2; \pi)] \frac{\partial \ln p(x_1, x_2; \pi)}{\partial \pi} =$$

$$= \sum_{(x_1, x_2) \in CR} \exp[\ln p(x_1, x_2; \pi)] \frac{x_1 + x_2 - (n_1 + n_2)\pi}{\pi(1-\pi)} \tag{4}$$

Next, we use the iterative algorithm to calculate the solution with the derivative equal to zero. In numerical analysis, fixed-point iteration is a classical root-finding algorithm, which can quickly calculate the solution of the equation. The derivative equation can be converted to algebraic form: $\pi = H(\pi)$:

$$\pi = \frac{\sum_{(x_1, x_2) \in CR} p(x_1, x_2; \pi_k)(x_1 + x_2)}{\sum_{(x_1, x_2) \in CR} p(x_1, x_2; \pi_k)(n_1 + n_2)} \tag{5}$$

It follows that, given an initial value $\pi_0^1 \in [0,1], \pi_1^1, \pi_2^1, \pi_3^1, \ldots$, is generated by an iterative function. By setting a small threshold ε, the iteration converges to the constant π^1 which is the approximation of the root when $|\pi_{k+1}^1 - \pi_k^1| \leq \varepsilon$. In general, the fixed-point algorithm requires multiple initial guesses of the root to be the initial values, $\pi_0^2, \pi_0^3, \ldots$, then each iteration of the algorithm produces a successively more accurate approximation to the root, $\pi_2, \pi_3, \ldots$. When the number of initial values is large enough and the threshold is small enough, we can obtain a solution containing all possible local maxima besides the boundary values 0 and 1. Therefore, the global maximum of the tail probability is $\max(\pi^1, \pi^2, \pi^3, \ldots)$.

Two independent multinomial distribution populations

The comparison of the total ratio vectors under two independent multinomial distributions is also common in practical applications. For example, in the industrial quality control, the comparison of the defective rate of products from different suppliers. The comparison of the drug response rate in the drug dose experiment of clinical medicine, *etc*. In these studies, the results of interest are ordered and multi-classified.

Suppose X_1 and X_2 are two independent k-multinomial distributions, $X_i \sim M(n_i, \pi_i)$, i = 1, 2, where, $\pi_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ik})$, i = 1, 2, which is defined as the success ratio of each part of each group. Researchers are usually interested in whether these ratios are equal, the problem is equivalent to the following test:

$$H_0 : \pi_1 = \pi_2 \ \ vs. \ \ H_1 : \pi_1 \neq \pi_2 \tag{6}$$

Suppose $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ is the sample size vector observed of the i^{th} population experiment, namely, x_1 and x_2 are the real response sizes of the two groups, respectively. Define the observation sum of the j^{th} category of the two populations as $m_j = x_{1j} + x_{2j}, j = 1, 2, \ldots k$. Then, the data from the experiment can be expressed as a 2 × K contingency table, as follow the probability function of X_1 is expressed as:

$$M(n_1, x_1, \pi_1) = \frac{n_1!}{x_{11}! x_{12}! \ldots x_{1k}!} \pi_{11}^{x_{11}} \pi_{12}^{x_{12}} \ldots \pi_{1k}^{x_{1k}} \tag{7}$$

where $M(n_1, x_1, \pi_1)$ is expressed as a similar representation of the probability mass function of x_2. The sample space of the random vector (x_1, x_2) will be represented by a

$\Omega = \{0,1,...,n_1\}^k$. A classical test statistic is Wilcoxon's inter mediate rank sum test statistic for the ordered test of multinomial distribution population, which is:

$$W(x_1, x_2) = \sum_{j=1}^{k} \phi_j x_{1j} \tag{8}$$

where $\phi_1 = (m_1 + 1)/2, j = (m_1 + m_2 + ... + m_j + 1)/2, j = 2,3,...k$, and k is the middle rank of k-ordered category. Next, we will propose accurate and unconditional test methods for the problem of comparing the parameters of multiple distribution population, tab. 2. In the case of independence assumption, the joint distribution function of a specific sample observation vector (x_1, x_2) can be expressed as: $\Pr_k(x_1, x_2; \pi_1, \pi_2) = M(n_1, x_1, \pi_1)M(n_2, x_2, \pi_2)$.

Table 2. Two independent multiple distribution data

	ordered categorical outcome				
Response	1	2	...	k	Sample size
Population 1	x_{11}	x_{12}	...	x_{1k}	n_1
Population 2	x_{21}	x_{22}	...	x_{2k}	n_2
Total	m_1	m_2	...	m_k	$n_1 + n_2$

In the framework of our hypothesis problem, $W(x_1, x_2)$ is used as the hypothesis test statistic, and the original hypothesis is rejected for a large statistic value. Therefore, for a specific sample observation vector value (x_1^0, x_2^0), the tail region is expressed as: $\Omega(x_1^0, x_2^0) = [(x_1, x_2) | W(x_1, x_2) \geq W(x_1^0, x_2^0)]$. It is a function of unknown disturbance parameters π_1 and π_2. The standard unconditional test method is to eliminate the unknown disturbance parameters by maximizing the tail region probability function in the parameter definition domain. Therefore, the *p*-value of the standard unconditional test method can be defined:

$$P_M(x_1^0, x_2^0) = \sup_{(\pi_1, \pi_2) \in H_0} f(x_1^0, x_2^0) \tag{9}$$

In the case of the independence hypothesis of X_1 and X_2, for the test hypothesis problem of multinomial distribution population, the unconditional precision test P_M can take the maximum value on the boundary $[(\pi_1, \pi_2) : \pi_1 = \pi_2]$ of null hypothesis H_0. Given $\pi_1 = \pi_2 = \pi, \pi = (\pi_1, \pi_2, ..., \pi_k)$, the domain of the disturbance parameter can be expressed as $C_M = [(\pi_1, \pi_2, ..., \pi_k) | 0 \leq \pi_i \leq 1, i = 1,2,...k, \Sigma_{j=1}^k \pi_j = 1]$. Define:

$$f_\pi(x_1, x_2) = \frac{n_1}{x_{11}! x_{12}! ... x_{1k}!} \frac{n_2}{x_{21}! x_{22}! ... x_{2k}!} \pi_1^{x_{11}+x_{21}} \pi_2^{x_{12}+x_{22}} ... \pi_k^{x_{1k}+x_{2k}} \tag{10}$$

Then, the *p*-value P_M of the standard unconditional test method can be rewritten:

$$P_M(x_1^0, x_2^0) = \sup_{(\pi_1, \pi_2) \in C_M (x_1, x_2) \in (x_1^0, x_2^0)} f_\pi(x_1, x_2) \tag{11}$$

Record the P_M as the accurate *p*-value $P_M(x_1, x_2)$ of the standard unconditional test method. By the Sidik theorem, the *p*-value of the standard unconditional test method only needs to search the maximum value in the k-dimension parameter space, which greatly simplifies the calculation process. At the same time, P_M is an effective *p*-value due to $p[(\pi_1, \pi_2) \in H_0)(P_M \geq \alpha) \geq \alpha]$.

For the test problem, the *p*-values of the standard unconditional test method is the probability functions of the maximum tail region in the disturbance parameter definition domain. However, the parameter domain span and the number of items in the multinomial distribution population increase exponentially. It is very difficult for the traditional grid search method to calculate the exact unconditional *p*-value of multi distribution population. Therefore, in order to ease the application difficulty of the accurate unconditional test method in multiple distributions, we will propose a fixed-point iteration method.

For a group of sample observations (x_1^0, x_2^0), the test problem $H_0 : \pi_1 \le \pi_2$ is considered, and the definition of $\Omega = [(x_1, x_2) \mid W(x_1, x_2) \ge W(x_1^0, x_2^0)]$ is the tail region set, where, $W(x_1, x_2)$ is the test statistic we use. Then, the *p*-value P_M of the exact unconditional test method is:

$$p = \sup_{(\pi_1, \pi_2) \in C} \sum_{(x_1, x_2) \in \Omega} f_\pi(x_1, x_2) \tag{12}$$

Next, similar to binomial distribution population, we construct a fixed-point iterative algorithm in the *p*-value calculation of the unconditional test method under multinomial distribution population. First, we rewrite the tail probability using the logarithmic property of the exponent:

$$p(\pi) = \sum_{(x_1, x_2) \in \Omega} f_\pi(x_1, x_2) = \sum_{(x_1, x_2) \in \Omega} \exp[\ln f_\pi(x_1, x_2)] \tag{13}$$

Then, the derivative of tail probability function (13) with respect to the disturbance parameter π is, where, $\pi = (\pi_1, \pi_2, ..., \pi_k)$ and $\Sigma_{i=1}^{k} \pi_i = 1$.

$$\begin{aligned}
\frac{\partial p(\pi)}{\partial \pi_1} &= \sum_{(x_1, x_2) \in \Omega} f_\pi(x_1, x_2) \left[\frac{x_{11} + x_{21}}{\pi_1} - \frac{n_1 + n_2 - \sum_{j=1}^{k-1} (x_{1i} + x_{2i})}{1 - \sum_{j=1}^{k-1} \pi_i} \right] \\
\frac{\partial p(\pi)}{\partial \pi_2} &= \sum_{(x_1, x_2) \in \Omega} f_\pi(x_1, x_2) \left[\frac{x_{12} + x_{22}}{\pi_2} - \frac{n_1 + n_2 - \sum_{j=1}^{k-1} (x_{1i} + x_{2i})}{1 - \sum_{j=1}^{k-1} \pi_i} \right] \\
&\cdots \\
\frac{\partial p(\pi)}{\partial \pi_{k-1}} &= \sum_{(x_1, x_2) \in \Omega} f_\pi(x_1, x_2) \left[\frac{x_{1k-1} + x_{2k-2}}{\pi_{k-1}} - \frac{n_1 + n_2 - \sum_{j=1}^{k-1} (x_{1i} + x_{2i})}{1 - \sum_{j=1}^{k-1} \pi_i} \right]
\end{aligned} \tag{14}$$

We can use the fixed-point iterative method to find the numerical solution of the linear equation, then the linear equations with derivative equal to zero are rewritten as fixed-point functions, $\pi = H(\pi)$. Therefore, fixed-point iterative functions can be written as $\pi^{m+1} = H(\pi^m)$, specifically:

$$\pi_1^{m+1} = \frac{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)(x_{11}+x_{21})\left(1-\sum_{j=1}^{k-1}\pi_i^m+\pi_1^m\right)}{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)\left[n_1+n_2-\sum_{j=1}^{k-1}(x_{1i}+x_{2i})+(x_{11}+x_{21})\right]}$$

$$\pi_2^{m+1} = \frac{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)(x_{12}+x_{22})\left(1-\sum_{j=1}^{k-1}\pi_i^m+\pi_2^m\right)}{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)\left[n_1+n_2-\sum_{j=1}^{k-1}(x_{1i}+x_{2i})+(x_{12}+x_{22})\right]} \quad (15)$$

$$\ldots$$

$$\pi_{k-1}^{m+1} = \frac{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)(x_{1k-1}+x_{2k-1})\left(1-\sum_{j=1}^{k-1}\pi_i^m+\pi_{k-1}^m\right)}{\sum_{(x_1,x_2)\in\Omega} f_\pi m(x_1,x_2)\left[n_1+n_2-\sum_{j=1}^{k-1}(x_{1i}+x_{2i})+(x_{1k-1}+x_{2k-1})\right]}$$

Given an initial point $\pi^0=(\pi^0,\pi_2^0,...,\pi_k^0)$ in the parameter definition domain, $\pi^1,\pi^2,...$ can be obtained through the previous iterative function, and the iterative parameter sequence will gradually shrink to a stable point. By defining a threshold that is small enough, we can get the numerical solution of a system of derivative equations. When the tail probability function surface has multiple maxima, multiple initial points in the domain are needed. Given several initial points and a small enough threshold in the domain, the numerical solution set of the iterative equation is defined as *S*, and then the *p*-value of the precision test without adjusting parts is:

$$p-\text{value}=\sup_{\pi\in s} p(x_1,x_2;\pi) \quad (16)$$

When a reasonable threshold and multiple initial iteration points are set, the fixed--point method can quickly find the disturbance parameter value and calculate the accurate *p*-value. Compared with the traditional grid search method, the fixed-point method can greatly reduce the calculation time and improve the efficiency of the test method.

Simulation and analysis

In this section, simulations analysis are operated in order to show the superiority of the new proposed method. We compare the new fixed-point iteration method with the network search method and polynomial method in the case of two independent binomial distributions and two independent multinomial distributions. We mainly consider the hypothesis test problem in the special classical case of $\delta_0=0$, and compare the accuracy and efficiency of grid search method, polynomial method and proposed fixed-point algorithm in different sample spaces. Due to the excellence of the unconditional test method, we will use the standard unconditional test method to examine the performance of the three calculation methods and the one-sided test is performed for the simulation.

Two independent binomial distributions

As an accurate calculation method, we first study the accuracy of the proposed method. Therefore, firstly, the fixed-point method is calculated based on the accuracy to illus-

trate that the fixed-point method is more accurate and effective than the traditional methods. We use grid search as a benchmark to compare the fixed-point method and the grid search method. In the grid search method, as the number of grids is set to 1000, the grid method can obtain a more accurate maximum tail probability, so we set the number of grids to 1000 in the simulation comparison. There are a variety of test statistics that can be adopted, without loss of generality, the *Z*-statistic is selected for the sake of convenience.

We enumerate the possible values in the sample space with the different sample sizes, and apply the fixed-point method and grid search method to calculate *p*-value. The fixed-point method is performed on by given the threshold ε and the number of initial values *k*. Given the threshold value ε = 0.001, 0.0001, respectively, we research the accuracy level of the fixed-point method with *k* value ranging from 1 to 10. What is more, the sample size of (n_1, n_2) = (30, 50), (50, 50), (50, 100), (100, 100) are used to generate the different sample space and significance level of the test was set as α = 0.05. Tables 3 and 4 show the percentage of tail probability samples of our proposed method in the grid search method in the same sample space. It is obvious that this percentage increases as the number of initial value *k* increases. When the threshold value was set to 0.0001 and *k* is no less than 10, the proposed fixed-point method can achieve the accurate *p*-value for all sample points in the sample space.

Table 3. The percentage of the sample point that the proposed fixed-point method and the grid search method have the same tail probability with the threshold value 0.001

(n_1 ,n_2)	k = 1	k = 2	k = 3	k = 4	k = 5
(30,30)	0.7323	0.7424	0.8377	0.8742	0.8905
(30,50)	0.6003	0.8789	0.8852	0.9042	0.9117
(50,30)	0.6003	0.8789	0.8852	0.9042	0.9117
(50,50)	0.7627	0.7702	0.8262	0.8760	0.8624
(50,100)	0.6370	0.8306	0.8329	0.8368	0.8453
(100,50)	0.6370	0.8306	0.8329	0.8368	0.8453
(100,100)	0.7883	0.7902	0.8438	0.8415	0.8651
(n_1, n_2)	k = 6	k = 7	k = 8	k = 9	k = 10
(30,30)	0.8945	0.9047	0.9108	0.9168	0.9128
(30,50)	0.9130	0.9130	0.9130	0.9231	0.9130
(50,30)	0.9130	0.9130	0.9130	0.9231	0.9130
(50,50)	0.8836	0.8957	0.8927	0.9033	0.9002
(50,100)	0.8460	0.8510	0.8514	0.8560	0.8610
(100,50)	0.8460	0.8510	0.8514	0.8560	0.8610
(100,100)	0.8632	0.8739	0.8687	0.8842	0.8788

Another key point is related to the computational time. The experiment is performed on a personal computer with an Intel CPU E5-2630 v3 processor with 32 cores at the processing speed of 2.40 GHz and 128 GB of main memory. The test environment is under the 64-bit R, version 3.41 of the Window 7 operating system. We use the sample sizes (n_1, n_2) = (30, 30), (30, 50), (50, 50) and (50,100) to generate the different sample space. For

Table 4. The percentage of the sample points that the proposed fixed-point method and the grid search method have the same tail probability with the threshold value 0.0001

(n_1, n_2)	k = 1	k = 2	k = 3	k = 4	k = 5
(30,30)	0.8641	0.8803	0.9614	0.9899	0.9979
(30,50)	0.6847	0.9559	0.9672	0.9874	0.9887
(50,30)	0.8662	0.8821	0.9410	0.9849	0.9803
(50,50)	0.8862	0.8821	0.9410	0.9849	0.9803
(50,100)	0.7506	0.9673	0.9723	0.9850	0.9896
(100,50)	0.7506	0.9673	0.9723	0.9850	0.9896
(100,100)	0.9182	0.9238	0.9621	0.9689	0.9780
(n_1, n_2)	k = 6	k = 7	k = 8	k = 9	k = 10
(30,30)	1.00	1.00	1.00	1.00	1.00
(30,50)	1.00	1.00	1.00	1.00	1.00
(50,30)	0.9940	0.9970	1.00	1.00	1.00
(50,50)	0.9940	0.9970	1.00	1.00	1.00
(50,100)	0.9931	0.9977	0.9981	0.9981	1.00
(100,50)	0.9931	0.9977	0.9981	0.9981	1.00
(100,100)	0.9914	0.9924	0.9947	0.9985	1.00

the fixed-point iterative algorithm, we use the results of precision study $k = 10$ and $\varepsilon = 0.0001$ as initial parameter settings to calculate *p*-value. For the polynomial method, Shan *et al.* [17] proposes to divide the parameter space into 150 intervals to study to ensure the accuracy of the results. Time efficiency performance is shown in tab. 5 where from we can see the average calculation time of the grid search method, the polynomial method and the fixed-point method under different *p*-values. The results manifest that the fixed-point algorithm is only 1/5 of the grid search method and 1/1000 of the polynomial method in calculating the exact *p*-value. Therefore, the fixed-point method is more effective than the two others from the simulation analysis results.

Two independent multivariate distributions

In this subsection, we will examine the performance of the fixed-point calculation method under the test of multiple parameter comparisons of two independent multivariate distributions.

First of all, similar to the comparison with two independent binomial distributions, under the assumption $\pi_1 = \pi_2$ to comparison of two independent multinomial distribution populations, we find that the fixed-point iteration method can calculate the maximum value of the tail region accurately when $\varepsilon = 1e^{-4}$, $k = 5$, therefore, the initial value of the fixed-point method is set to 5 in each dimension, and the threshold value is $\varepsilon = 1e^{-4}$. Besides, it is impossible to use transformation to get the expression by polynomial method at present, thus, in the simulation study, we do not consider polynomial method. In the simulation comparison, the grid search method takes 50 equal grids in each dimension of the parameter. Time efficiency performance of our proposed method and grid search method under multinomial distribution is shown in tab 6.

Table 5. The average computation time of the grid search method, polynomial method and the fixed-point method under different *p*-values

(n_1, n_2)	*p*-value	Fixed-point	Grid search	Polynomial
(30,30)	0.1	0.5413	2.5028	24.4820
	0.2	0.8254	2.7791	26.9314
	0.3	0.7583	2.9625	26.2669
	0.4	1.0887	3.2458	32.0612
	0.5	0.9908	3.5238	35.1463
(30,50)	0.1	0.8300	4.2302	101.7787
	0.2	0.7217	4.7404	114.1079
	0.3	0.7086	4.8725	119.0676
	0.4	0.4684	5.2659	127.2398
	0.5	0.3731	5.7098	137.5676
(50,50)	0.1	1.194	7.6016	212.8992
	0.2	1.3001	8.1104	229.8615
	0.3	1.5638	8.5436	223.8243
	0.4	2.0583	9.1090	254.5283
	0.5	2.4523	9.6721	261.9922
(50,100)	0.1	1.6352	15.6535	503.3533
	0.2	1.0836	17.1546	547.7028
	0.3	1.05716	17.9197	571.2815
	0.4	1.1375	18.9958	596.8898
	0.5	1.0610	20.1298	632.6630

We can see clearly that the grid search method needs more time to get the maximum value of a tail region than the fixed-point method in all samples. Especially, the grid search method needs more than 700 seconds when (n_1, n_2) = (20, 40). The grid search method consumes a lot of time is extremely infeasible in the high-dimensional parameter space and larger sample size, that is because the traditional grid search method is to search the maximum of the tail probability region by traversing all possible values of the parameters in the parameter space, therefore, with the complexity of the parameter domain and the more sample points in the tail region, the time consumption of the grid search method increases exponentially. While, the fixed-point method can get accurate results and the time consumed is only 1/30 of grid search method. Under the accurate unconditional test method, from the perspective of accuracy and efficiency, the new fixed-point method is a more efficient algorithm compared with the traditional grid search method of comparing the parameters of two independent multinomial populations.

Empirical study

To illustrate the accuracy and validity of the fixed-point method in real data, we apply this method to randomized drug clinical trials [23]. This was a Phase II, randomized, dou-

Table 6. Calculation time of (n_1, n_2) grid search method and fixed-point algorithm in different sample cases

	(x_1, x_2)			Grid search		Fixed-point
(n_1,n_2)	x_1	x_2	Grid search	Time	Fixed-point	Time
(10;10)	(0,6,4)	(1,8,1)	0.0578	10.6212	0.0591	0.3268
	(1,1,8)	(6,3,1)	0.0012	3.0250	0.0013	0.1182
	(0,2,8)	(2,5,3)	0.0113	6.8712	0.0106	0.2436
	(0,1,9)	(3,5,2)	0.0007	2.6480	0.0007	0.0932
	(0,0,10)	(0,3,7)	0.0437	9.6797	0.0435	0.3358
(20;20)	(2,2,16)	(4,8,8)	0.00960	147.5574	0.0090	4.9411
	(0,6,14)	(3,7,10)	0.0769	210.8356	0.0790	6.9865
	(0,19,1)	(13,0,7)	0.0403	181.9177	0.0413	6.0721
	(3,10,7)	(11,4,5)	0.0254	171.8585	0.0238	4.96216
	(3,2,15)	(0,14,6)	0.0192	163.3967	0.0196	5.2499
(20;30)	(1,14,5)	(5,23,2)	0.0243	375.3839	0.0237	11.408
	(3,7,10)	(12,11,7)	0.0151	349.8732	0.0124	10.1008
	(2,2,16)	(1,13,16)	0.0633	508.6135	0.0598	17.4253
	(1,18,1)	(5,25,0)	0.0645	508.4940	0.0602	15.9534
	(0,8,12)	(0,20,10)	0.0388	459.0562	0.0348	13.6328
(20;40)	(2,14,4)	(15,20,5)	0.0251	780.3429	0.0241	20.5737
	(2,8,10)	(10,21,9)	0.0225	724.9782	0.0219	18.9754
	(0,15,5)	(0,37,3)	0.0410	834.6929	0.0366	24.5432
	(3,0,17)	(11,8,21)	0.0225	745.9786	0.0225	87.8357
	(0,6,14)	(3,12,25)	0.0506	1991.131	0.0479	64.0939

ble-blind placebo controlled, subgroup trial designed to investigate the efficacy of sorafenib in the treatment of non-small cell lung cancer (NSCLC). The standard of treatment for non-small cell lung cancer is ergot isotonic added placebo. A total of 166 patients are randomized into two groups in a 2:1 rate, one receiving erlotinib added sorafenib and the other receiving erlotinib added placebo. At the end of the research, they conduct a blood test on 166 patients and measure their response rate and toxicity profile values. Among them, $n_1 = 111$ patients receive a combination of treatment in erlotinib added sorafenib, the number of patients whose condition is controlled after treatment is $x_1 = 60$, and the control rate is 54%. There are $n_2 = 55$ patients who receive erlotinib added placebo. The number of patients who are controlled after treatment is $x_2 = 21$, and the control rate is 38%. Therefore, the new method of erlotinib added sorafenib is superior to the gold standard (erlotinib added placebo).

Through using the *Z*-statistic, fig. 1 shows the tail region probability curve comparing the two treatments under the exact unconditional test. There exists a plurality of local maximum and minimum values in the tail region curve. By the fixed-point method, the *p*-value is

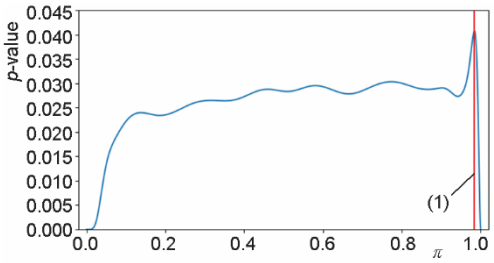

Figure 1. Tail probability curve from comparing the erlotinib added sorafenib group and erlotinib added placebo group of lung cancer patients. The red line (1) is the actual *p*-value

0.040743 when $\pi = 0.985618$. We also calculate the *p*-value using a grid search method with a grid width of 1/2000 and polynomial method of 100 equidistant. Table 7 shows the calculation results and runtime of similar results in different methods. The grid search method and the polynomial method can calculate an accurate *p*-value with an error of less than 0.001. In the calculation time, the fixed-point method time is 1.9 seconds, the grid search method is 54 seconds, and the polynomial method is 200 seconds, so the fixed-point method is more effective with less computing time.

Table 7. Calculation results and runtime of similar results in different methods

Method	*p*-value	Runtime [s]	Parameter
Fixed-point method	0.4074	2.5292	$k = 2$
Grid search method	0.4076	51.6947	$k = 1000$
Polynomial method	0.4076	524.9319	$k = 50$

Conclusion

In this paper, aiming at the comparison of two parameters of two independent binomial distributions and two independent multinomial distributions, the fixed-point algorithm is proposed to calculate the exact unconditional *p*-value. We have proved that the tail probability function conforms to the condition of the fixed-point method after a certain mathematical transformation. After solving it, the final solution of the tail probability function which reaches the maximum value is included in the solution set, and the global maximum of the solution set is calculated by the fixed-point iteration method. Compared with the grid search method and the polynomial method, the new method can greatly reduce the time required to calculate the exact *p*-value. Therefore, we conclude that the new method is more effective on the basis of ensuring accuracy from the results of simulation data and practical application data analysis, and it can be recommended for clinical trials because of the simplicity of the operation. In addition, we also make some preliminary attempts to calculate exact unconditional *p*-values in the case of multiple stage and multiple interference parameters (response rate, toxicity values), the results are also very optimistic.

Acknowledgment

This paper is supported in part by Natural Science Foundation of China (No. 11971433) and First Class Discipline of Zhengjiang-A (Zhejiang Gongshang University-Statistics).

References

[1] Siefker-Radtke, A. O., *et al.*, A Phase 2 Clinical Trial of Sequential Neoadjuvant Chemotherapy with Ifosfamide, Doxorubicin, and Gemcitabine Followed by Cisplatin, Gemcitabine, and Ifosfamide in Locally Advanced Urothelial Cancer, *Cancer, 119* (2013), 3, pp. 540-547

[2] Zheng, L., The Design of Single-Arm Clinical Trials of Combination Antiretroviral Regimens for Treatment-Naive HIV-Infected Patients, *AIDS Res Hum Retrovir*, *29* (2012), 4, pp. 652-657
[3] Katz, P. O., *et al.* Guidelines for the Diagnosis and Management of Gastroesophageal Reflux Disease, *American Journal of Gastroenterology*, *108* (2013), 3, pp. 308-328
[4] Fisher, R. A., The Logic of Inductive Inference, *Journal of the Royal Statistical Society*, *98* (1935), 1, pp. 39-82
[5] Barnard, G. A., A New Test for 2 × 2 Tables, *Nature*, *156* (1945), 3954, 177
[6] Lloyd, C. J., Exact *P*-Values for Discrete Models Obtained by Estimation and Maximization, *Australian & New Zealand Journal of Statistics*, *50* (2008), 4, pp. 329-345
[7] Shan, G., *et al.*, An Efficient and Exact Approach for Detecting Trends with Binary Endpoints, *Statistics in Medicine*, *31* (2012), 2, pp. 155-164
[8] Lloyd, C. J., Moldovan, M. V., A More Powerful Exact Test of Noninferiority from Binary Matched-Pairs Data, *Statistics in Medicine*, *27* (2008), 18, pp. 3540-3549
[9] Shan, G., Wang, W., ExactCIdiff: An R Package for Computing Exact Confidence Intervals for the Difference of Two Proportions, *The R. Journal*, *5* (2013), 2, pp. 62-71
[10] Shan, G., Exact Unconditional Testing Procedures for Comparing Two Independent Poisson Rates, *Journal of Statistical Computation and Simulation*, *85* (2015), 5, pp. 947-955
[11] Tang, M. L., Tang, N. S., Exact Tests for Comparing Two Paired Proportions with Incomplete Data, *Biometrical Journal*, *46* (2004), 1, pp. 72-82
[12] Fisher, R. A., The Logic of Inductive Inference, *Journal of the Royal Statistical Society*, *98* (1935), 1, pp. 39-82
[13] Basu, D., On the Elimination of Nuisance Parameters, *Publications of the American, Statistical Association, Series B (Methodological)*, *72* (1977), 358, pp. 355-366
[14] Berger, R. L., Boos, D. D., *P*-Values Maximized Over a Confidence Set for the Nuisance Parameter, *Publications of the American Statistical Association*, *89* (1994), 427, pp. 1012-1016
[15] Lloyd, C. J., A New Exact and More Powerful Unconditional Test of No Treatment Effect from Binary Matched pAirs, *Biometrics*, *64* (2008), 3, pp. 716-723
[16] Shan, G., *et al.*, Boundary Problem in Simon's Two-Stage Clinical Trial Designs, *Journal of Biopharmaceutical Statistics*, *27* (2017), 1, pp. 25-33
[17] Shan, G., *et al.*, Optimal Adaptive Two-Stage Designs for Early Phase II Clinical Trials, *Statistics in Medicine*, *35* (2016), 8, pp. 1257-1266
[18] Shan, G., *et al.*, Accurate Unconditional *P*-Values for a Two-Arm Study with Binary Endpoints, *Journal of Statistical Computation and Simulation*, *88* (2018), 6, pp. 1200-1210
[19] Chan, I. S., Exact Tests of Equivalence and Efficacy with a Non-Zero Lower Bound for Comparative Studies, *Statistics in Medicine*, *17* (1998), 12, pp. 1403-1413
[20] Simon, R., Optimal 2-Stage Designs for Phase II Clinical Trials, *Controlled Clinical Trials*, *10* (1989), 1, pp. 1-10
[21] Barnard, G. A., Significance Tests for 2 × 2 Tables, *Biometrika*, *34* (1947), 1-2, pp. 123-138
[22] Berger, R. L., Sidik, K., Exact Unconditional Tests for a 2 × 2 Matched-Pairs Design, Statistical Methods in Medical Research, *12* (2003), 2, pp. 91-108
[23] Spigel, D. R., *et al.*, Randomized, Double-Blind, Placebo Controlled, Phase II Trial of Sorafenib and Erlotinib or Erlotinib Alone in Previously Treated Advanced Non-Small-Cell Lung Cancer, *Journal of Clinical Oncology*, *29* (2011), 18, pp. 2582-2589

Paper submitted: February 20, 2020
Paper revised: July 7, 2020
Paper accepted: July 7, 2020