TESTS OF INDEPENDENCE IN A SINGLE 2×2 CONTINGENCY TABLE WITH RANDOM MARGINS

By

Yuan Yu

A Thesis

Submitted to the Faculty

Of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

In

Applied Statistics

May 1, 2014

APPROVED:

_____

Professor Balgobin Nandram, Major Thesis Advisor

Abstract

In analysis of the contingency tables, the Fisher's exact test is a very important statistical significant test that is commonly used to test independence between the two variables. However, the Fisher' s exact test is based upon the assumption of the fixed margins. That is, the Fisher's exact test uses information beyond the table so that it is conservative. To solve this problem, we allow the margins to be random. This means that instead of fitting the count data to the hypergeometric distribution as in the Fisher's exact test, we model the margins and one cell using multinomial distribution, and then we use the likelihood ratio to test the hypothesis of independence. Furthermore, using Bayesian inference, we consider the Bayes factor as another test statistic. In order to judge the test performance, we compare the power of the likelihood ratio test, the Bayes factor test and the Fisher's exact test. In addition, we use our methodology to analyse data gathered from the Worcester Heart Attack Study to assess gender difference in the therapeutic management of patients with acute myocardial infarction (AMI) by selected demographic and clinical characteristics.

Acknowledgements

I would like to express my very great appreciation to my thesis advisor, Dr. Balgobin Nandram, who not only provided this meaningful topic but also put so much effort helping me understand essence of my thesis making it go the right way. I am also inspired and encouraged to explore and learn new knowledge during this period under his instruction. Thanks to Professor Sungsu Kim for taking so much time helping me examine the thesis carefully and patiently. I wish to acknowledge the great programming advices provided by Zhiqing Xu. Also, my special thanks are extended to all the professors who imparted to me with the valuable statistical knowledge and their assistance for the past three semesters.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Discussion about the Fisher's exact test

Usually, the Fisher's exact test is useful for categorical data that result from classifying objects in different ways; it is used to examine the significance of the association between the two kinds of classification, for example the famous problem of the Fisher's tea drinker. In the Fisher's tea drinker example, one criterion of classification could be whether milk or tea was put in the cup first; the other could be whether the lady thinks that the milk or tea was put in first. We want to know whether these two classifications are associated that is, whether the lady really can tell whether milk or tea was poured in first. To test, she was given 8 cups of tea, in four of which milk was added first. The null hypothesis is that there is no association between the true order of pouring and the lady's guess, and the alternative hypothesis is there is a positive association. Most uses of the Fisher test involve, like this example, a 2×2 contingency table. The p-value from the test is computed as if the margins of the table are fixed, i.e. as if, in the tea-tasting example, the lady knows the number of cups with each treatment (milk or tea first) and will therefore provide guesses with the correct number in each category. As pointed out by Fisher, this leads under a null hypothesis of independence to a hypergeometric distribution of the numbers in the cells of the table. Specifically, for the following table we calculated the Fisher's exact p-value of 0.2429, which means that in this situation we cannot reject the null hypothesis to make a conclusion that the

lady had such an ability.

	Guess-Milk	Guess-Tea	
Truth-Milk	3	1	4
Truth-Tea	1	3	4
	4	4	8

Although in practice it is employed when sample sizes are small, it is valid for all sample sizes. The Fisher's exact test is one of exact tests, since the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity. The approximation is inadequate when sample sizes are small, or the data are unequally distributed among the cells of the table. In fact, for small, sparse, unbalanced data, the exact and asymptotic p-values can be quite different and may lead to opposite conclusions. In contrast, the Fisher test is, as its name states, exact as long as the experimental procedure keeps the row and column totals fixed, and it can therefore be used regardless of the sample characteristics.

Despite the fact that Fisher's test gives exact p-values, some authors have argued that it is conservative, meaning that its actual rejection rate is below the nominal significance level. Fisherian basis of the exact test, that the marginal totals are "ancillary statistics" (Little 1989), and therefore provide no information respecting the configuration of the body of the table, is shown to be incorrect by Berkson (1978) through comparing the exact test for the one-sided case with the normal test for the nominal significance levels $\alpha=0.05$ and $\alpha=0.01$, and it is shown by direct computation that the effective level is closer to the nominal level with the normal test than with the exact test, and that the power of the normal test is considerably larger than that of the exact test, with the increasing power exceeding the change of effective level. Thus, a conclusion has been drawn that the exact test should not be used in preference to the normal test.

The traditional solution for testing the equality of two independent binomial distributions is to use (a) Pearson chi-squared test for large samples, (b) Chi-square test with Yates' continuity correction for intermediate size samples, and (c) the Fisher's exact test for small samples. However, several scientists, including Berkson (1978), Conover (1974), Grizzle (1967), Kempthorne (1979),

Upton (1982), have questioned the usefulness of this solution after a careful study of 22 alternative tests, and concluded that "the exact test of Fisher, and the corresponding Yates correction to Pearson's test, produce results that are both extremely conservative and inappropriate" (D'Agostino et. al 1988). He noted that the uncorrected chi-square test performs well and suggested that a preferred test is the scaled test $(N-1)X^2/N$, where N is the combined sample size and X^2 is the Pearson chi-square test statistic for a 2×2 contingency table.

While the Fisher's exact test perhaps creates an impression that exact implies best, Upton (1982) argued that we should be concerned more about the conservative nature or the poor power of Fisher's exact test and Yates' correction to the chi-square test rather than having an exact test. D'Agostino et. al (1988) confirmed Upton's results and further demonstrated that the uncorrected chi-squared test and the two-independent-sample t test with pooled variance are robust since their actual levels of significance are, in most situations, close to or smaller than the nominal levels and their maximum actual levels are close to the nominal levels.

Little (1989) also analysed two tables specifically between where the small one is $T_1 = \{a, b, c, d\} = \{3, 0, 0, 3\}$ (provided by Barnard 1945), and another table $T_2 = \{a, b, c, d\} = \{170, 2, 162, 9\}$ is of moderate size. With applying (a) (b) (c) methods (denoted as P, Y, F respectively) to one-sided tests for each of the two tables we observed that the p-values obtained by the last two methods are close and considerably larger than those for Pearson chi-square test even for a quite large table. Generally, the author suppose two independent random samples of size n_1 and n_2 are selected from a large population. The first sample is given treatment 1, yielding a successes and $b = n_1 - a$ failures. The second sample is given treatment 2, yielding c successes and $d = n_2 - c$ failures. Table 1 displays the data in a 2×2 contingency table, writing $m_1 = a + c$ as the total number of successes and $m_2 = b + d$ as the total number of failures.

Table 1. Notation for the 2 × 2 Table

Treatment	Outcome S	F	Total
1	a	b	n_1
2	c	d	n_2
Total	m_1	m_2	

The objective is to assess an evidence against the null hypothesis that the two treatments are equally effective;that is, $\pi_1 = \pi_2$, where π_j is the probability of success given treatment j. Table 2 shows the results of applying P, Y, and F to T_1, and T_2 ; p values are presented for both one-sided $(\pi_1 > \pi_2)$ and two-sided $(\pi_1 \neq \pi_2)$ alternatives.

Table 2. p Values (%) of Three Tests for Two Tables— T_1: {3, 0, 0, 3} and T_2: {170, 2, 162, 9}

Test	Table T_1 One-Sided	Two-Sided	Table T_2 One-Sided	Two-Sided
Pearson (P)	.7	1.4	1.6	3.2
Yates (Y)	5.1	10.2	3.2	6.4
Fisher (F)	5.0	10.0	3.0	3.5

For the one-sided tests, it is found that the p values for Y and F are close. In fact, Y is considered as an approximation to F; we can also find that p values for Y and F are considerably larger than those of P, even for T_2, where the sample size is quite large. It illustrated that even in quite respectable-sized samples, the choice of the three methods matters. Conventional wisdom favours the use of Y in moderate samples and F in small samples; however, D'Agostino et. al. (1988) challenged this position. Building on earlier work of Berkson (1978), Grizzle (1967), Upton (1982), they claimed that Y and F are much too conservative in small samples, and argued that the usual pooled t test for independent normal samples should replace Y and F. The t-test is essentially a studentized version of P, and will perform like P unless the combined sample sizes are less than, say, 30.

The basis for this conclusion is a comparison of empirical and nominal significance levels of the tests, using two independent samples from two binomial populations, where n_1 and n_2 are fixed. Table 3 shows the empirical rejection rates for P and F, for the one-sided tests at the nominal 5% level with null values of $\pi = .2, .5$, and $.8$, and samples size ranging from 5 to 200 in each group.

Table 3. Percent Rejection Rates Under H_0 of Pearson and Fisher Tests, for Nominal $\alpha = 5\%$
(H_0: $\pi_1 = \pi_2 = \pi$; H_a: $\pi_1 > \pi_2$)

| $n_1 = n_2$ | $\pi = .5$ | | $\pi = .2, .8$ | |
	Pearson	Fisher	Pearson	Fisher
5	5.5	1.1	2.2	.2
10	5.8	2.1	4.6	1.5
20	4.2	2.1	5.1	2.3
50	4.5	2.9	5.0	2.9
100	5.2	3.8	4.9	3.5
200	4.9	4.0	—	—
∞	5	5	5	5

Source: Berkson (1978), as summarized by Yates (1984).

Clearly F is conservative, that is, it has much smaller rejection rates than the nominal level; Y (not shown in the table) is similarly conservative. On the other hand, P has rejection rates similar to the nominal level, even in the small samples, except a table of the one-sided test when $n_1 = n_2 = 5$. D'Agostino et al. (1988) also showed that rejection rates for their studentized version of P are close to the nominal levels. Results here are based on exact calculation, not simulation.

The above findings are impressive evidences against the use of Fisher's exact test or the chi-square test with the Yates' correction. There are two reasons why exact test is so conservative (Little 1989). One is that, the test statistic has a discrete null distribution, yielding discontinuities in the rejection rates. The other reason is, Fisher's exact test conditions on both fixed margins, whereas the computations in Table 3 only condition on one margin fixed by the sampling design.

The apparent contradiction stems from the combination of a discrete statistic with fixed significance levels. To be more precise, consider the following proposal for a significance test at the 5%-level which means rejecting the null hypothesis for each table to which the Fisher's test that assigns a p-value equal to or smaller than 5%. Because tables are discrete, there may not be a table

8

that would achieve the equality. While this problem occurs for any discrete statistic, it has been argued that the conservative effect is compounded by the fact that the Fisher's test assumes the fixed margins (Mehta and Senchaudhuri 2003). So in this thesis, by allowing random margins, we try to develop an exact method that is not as much conservative exact as the Fisher's exact test.

Chapter 2

A new probability distribution

To illustrate our idea, we start from a specific contingency table from Worcester Heart Attack Study, which contains the information about gender differences in the receipt of lidocaine therapy in acute myocardial infarction (AMI) patients with a history of hypertension and stroke whose age is greater than 75.

The most commonly used method is fitting the contingency table into two binomial distributions (Nandram B. et. al. 2013 and Hashemi L. et. al. 1997), in our case, we divide all our target patients into two groups of male and female, and denote the probability of therapy receiving of male as p_1, and the probability of therapy receiving of male as p_2. So the number of male receiver $X_1 \sim binomial(30, p_1)$, and the number of female receiver $X_2 \sim binomial(51, p_2)$. Thus our task of significance detecting based on constructing the null hypothesis as $p_1 = p_2$. Even though the model is more flexible with only one fixed margin, in comparison with Fisher's exact test which requires both margins fixed, it is still not very good by adding extra information. Hence we try to find a model free of the margin assumption.

	Receive	Not receive	
Males	8	22	30
Females	16	35	51
	24	57	81

2.1 Multinomial distribution

We denote the count of patient whose gender is male as random variable X, and denote the count of patients who received the therapy as random variable Y. We'll give our demonstration later on that actually both X and Y follow the binomial distribution with the correspondent parameter p and q. Here we view X and Y as the random variables in multinomial distribution represent the same meaning.

And then we try to explore the correlation between two random variables X and Y which follow the binomial distribution with parameter p and q. We consider about each patient, we can let random variable I indicates the gender difference (1 for male and 0 for female), and let random variable J represent if the patient received the therapy (1 for receive and 0 for not receive). Obviously the bivariate (I , J) variable decides totally four kinds of situation and the relationship between I, J and X, Y are

$$X = \sum_{k=1}^{N} I_k, \ Y = \sum_{k=1}^{N} J_k, \ Z = \sum_{k=1}^{N} I_k J_k$$

J

		1	0	
I	1	π_{11}	π_{10}	p
	0	π_{01}	π_{00}	$1-p$
		q	$1-q$	1

Then we get the multinomial distribution model of our table below

	Receive	Not receive	
Males	Z	X-Z	X
Females	Y-Z	N-X-Y+Z	N-X
	Y	N-Y	N

and the probability mass function is

$$P(Z, X-Z, Y-Z, N-X-Y+Z) = \frac{n!}{z!(x-z)!(y-z)!(n-x-y+z)!} \pi_{11}{}^{z} \pi_{10}{}^{x-z} \pi_{01}{}^{y-z} \pi_{00}{}^{n-x-y+z}$$

$$(2.1.1)$$

where $0 \leq x \leq n$, $0 \leq y \leq n$, $max(0, x + y - n) \leq z \leq min(x, y)$.

Let γ denote the correlation between I and J, and we can express γ using p q and π_{11} as below

$$\gamma = \frac{cov(I, J)}{\sqrt{Var(I)Var(J)}} \tag{2.1.2}$$

$$= \frac{E(IJ) - EIEJ}{\sqrt{Var(I)Var(J)}}$$

$$= \frac{\pi_{11} - pq}{\sqrt{p(1-p)q(1-q)}}.$$

We have already known three relationships among π_{ij} and parameter p and q, they are

$$\pi_{11} + \pi_{10} = p,$$

$$\pi_{11} + \pi_{01} = q,$$

$$\pi_{11} + \pi_{10} + \pi_{01} + \pi_{00} = 1. \tag{2.1.3}$$

So finally we get the expressions of π_{ij} with only one parameter γ introduced, the results are

$$\pi_{11} = pq + \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{10} = p(1-q) - \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{01} = q(1-p) - \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{00} = (1-p)(1-q) + \gamma\sqrt{p(1-p)q(1-q)}. \tag{2.1.4}$$

Further more, we can calculate the marginal distributions of X and Y

$$P(X = x) = \sum_{z=0}^{x} \sum_{y=z}^{n-x+z} P(X, Y, Z) \tag{2.1.5}$$

$$= \sum_{z=0}^{x} \sum_{y=z}^{n-x+z} \frac{n! \pi_{11}{}^{z} (p - \pi_{11})^{x-z} (q - \pi_{11})^{y-z} (1 - p - q + \pi_{11})^{n-x-y+z}}{z!(x - z)!(y - z)!(n - x - y + z)!}$$

$$= \sum_{z=0}^{x} \frac{n! \pi_{11}{}^{z} (p - \pi_{11})^{x-z}}{z!(x - z)!} \sum_{y=z}^{n-x+z} \frac{(p - \pi_{11})^{x-z} (q - \pi_{11})^{y-z} (1 - p - q + \pi_{11})^{n-x-y+z}}{(y - z)!(n - x - y + z)!}$$

$$= \frac{n! p^{x} (1 - p)^{n-x}}{x!(n - x)!} \sum_{z=0}^{x} \frac{x! (\frac{\pi_{11}}{p})^{z} (1 - \frac{\pi_{11}}{p})^{x-z}}{z!(x - z)!} \sum_{y-z=0}^{n-x} \frac{(n - x)! (\frac{q - \pi_{11}}{1-p})^{y-z} (1 - \frac{q - \pi_{11}}{1-p})^{n-x-y+z}}{(y - z)!(n - x - y + z)!}$$

$$= \binom{n}{x} p^{x} (1 - p)^{n-x} \ .$$

Similarly, we have

$$P(Y = y) = \sum_{z=0}^{y} \sum_{x=z}^{n-y+z} P(X, Y, Z) = \binom{n}{y} q^{x} (1 - q)^{n-y} \ . \tag{2.1.6}$$

The marginal distribution is binomial and X~Binomial(n, p), Y~Binomial(n, q).

Since X~Binomial(n, p), we can get EX and Var(X) , EY, Var(Y) easily from the feature of the binomial distribution,

$$EX = n(\pi_{11} + \pi_{10}) = np \tag{2.1.7}$$

$$VarX = n(\pi_{11} + \pi_{10})(\pi_{01} + \pi_{00}) = np(1 - p) \tag{2.1.8}$$

$$EY = n(\pi_{11} + \pi_{01}) = nq \tag{2.1.9}$$

$$VarX = n(\pi_{11} + \pi_{01})(\pi_{10} + \pi_{00}) = nq(1 - q) \ . \tag{2.1.10}$$

2.2 The MLE of the parameters

Our goal is to find the MLE of the three parameters p, q and γ . First of all, considering a particular 2×2 contingency table, which we can generate from our original contingency table

	Received	Not Received	
Male	n_{11}	n_{10}	$n_{1.}$
Female	n_{01}	n_{00}	$n_{0.}$
	$n_{.1}$	$n_{.0}$	n

In order to proceed a statistical optimization, we can set the initial value of the parameters as

$$\hat{p} = \frac{n_{11} + n_{10}}{n} \quad \hat{q} = \frac{n_{11} + n_{01}}{n} \quad \hat{\gamma} = \frac{\frac{n_{11}}{n} - \hat{p}\hat{q}}{\sqrt{\hat{p}(1 - \hat{p})\hat{q}(1 - \hat{q})}}. \tag{2.2.1}$$

Since they are a general straightforward estimation of p, q, and γ. We have already known that the relationships between p, q and γ are

$$pq + \gamma\sqrt{p(1 - p)q(1 - q)} \geq 0 \tag{2.2.2}$$

$$p - pq - \gamma\sqrt{p(1 - p)q(1 - q)} \geq 0 \tag{2.2.3}$$

$$q - pq - \gamma\sqrt{p(1 - p)q(1 - q)} \geq 0 \tag{2.2.4}$$

$$1 - p - q + pq + \gamma\sqrt{p(1 - p)q(1 - q)} \geq 0 \ . \tag{2.2.5}$$

So given the value of the either two of the parameters will give a restriction interval of the other one. Thus, we could obtain the value of the parameters generate from these intervals. If given the value of q, γ, in order to calculate the interval of p, noticing that the signal of γ matters, so we discuss in different situations with regard to $\gamma > 0$ and otherwise. If $\gamma > 0$, we only need to solve the inequalities (2.2.3) and (2.2.4). Otherwise, we only solve the inequalities (2.2.2) and (2.2.5). Finally, we obtain the constraints of p.

That is,

$$if \ \gamma > 0, \quad \frac{\gamma^2 q}{1 - q + \gamma^2 q} \leq p \leq \frac{q}{\gamma^2(1 - q) + q} \tag{2.2.6}$$

and

$$if \ \gamma \leq 0, \quad \frac{\gamma^2(1 - q)}{q + \gamma^2(1 - q)} \leq p \leq \frac{1 - q}{\gamma^2 q + 1 - q} \ . \tag{2.2.7}$$

Similarly, we can calculate the range of q ,

$$if\ \gamma > 0, \qquad \frac{\gamma^2 p}{1 - p + \gamma^2 p} \leq q \leq \frac{p}{\gamma^2(1-p)+p} \qquad (2.2.8)$$

and

$$if\ \gamma \leq 0, \qquad \frac{\gamma^2(1-p)}{p + \gamma^2(1-p)} \leq q \leq \frac{1-p}{\gamma^2 p + 1 - p} \qquad . \qquad (2.2.9)$$

However, we need to solve all the inequalities to get the range of γ

$$max\{\frac{-pq}{\sqrt{p(1-p)q(1-q)}}, \frac{-(1-p)(1-q)}{\sqrt{p(1-p)q(1-q)}}\} \leq \gamma \leq min\{\frac{p(1-q)}{\sqrt{p(1-p)q(1-q)}}, \frac{q(1-p)}{\sqrt{p(1-p)q(1-q)}}\}$$
$$(2.2.10)$$

Furthermore, we can simplify it to

$$max\{-\sqrt{\frac{pq}{(1-p)(1-q)}}, -\sqrt{\frac{(1-p)(1-q)}{pq}}\} \leq \gamma \leq min\{\sqrt{\frac{p(1-q)}{(1-p)q}}, \sqrt{\frac{q(1-p)}{(1-q)p}}\} \qquad (2.2.11)$$

We can see from the above simplification that the two terms in max and min have a reciprocal relationship, so the maximum value is greater than -1 and the minimum value is less than 1. So with the initial value of p, q and γ, we can get a new set of parameters p, q and γ successively by being randomly picked up within their ranges. For example we generate 1100 sets of parameters and plug each set into the likelihood function $f(p, q, \gamma)$, which is

$$f(p,q,\gamma) = C \cdot (pq + \gamma\sqrt{p(1-p)q(1-q)})^{n_{11}} (q - pq - \gamma\sqrt{p(1-p)q(1-q)})^{n_{10}} \qquad (2.2.12)$$
$$(p - pq - \gamma\sqrt{p(1-p)q(1-q)})^{n_{01}} (1 - p - q + pq + \gamma\sqrt{p(1-p)q(1-q)})^{n_{00}}$$

where $C = \frac{n!}{n_{11}!n_{10}!n_{01}!n_{00}!}$ is the coefficient of the parameter distribution function of p, q, and γ.

With the above restrictions among the three parameters, we can apply a grid method to do the statistical optimization of the parameters in the likelihood function above by drawing 1100 sets of parameters p, q and γ. Considering the fact that the former drawn samples are not stable, we drop off the first 100 set of values and then plugging in the rest 1000 parameter sets to the likelihood function to find the ones making function reach the maximum value.

15

Applying to the example of gender differences in the receipt of lidocaine therapy in AMI patients (≥ 75) with a history of hypertension and stroke, we got MLEs of parameter p, q and γ, which are 0.3717696, 0.3012183, -0.05674575. The results are reasonable since they are close to the approximate estimations obtained directly from the table which are 0.3703704, 0.2962963, -0.04976726.

Chapter 3

Likelihood Ratio Test

In this chapter, we are going to use the likelihood ratio as test statistic to do the hypothesis testing, including developing methodology to compute the p-value and power function of the likelihood ratio test, and certainly for the purpose of comparison, we also need to compute the power function of the Fisher's exact test under our multinomial distribution model.

3.1 P-value of the Likelihood Ratio Test

Consider the two-sided hypothesis test,

$$H_0 : \gamma = 0 \qquad versus \qquad H_a : \gamma > 0$$

Now, we build our hypothesis test on our likelihood ratio test statistics defined as :

$$T = \frac{L(x,y,z|H_0)}{L(x,y,z|\Omega)} = \frac{L(x,y,z|\hat{p}_{mle0}, \hat{q}_{mle0}, \gamma = 0)}{L(x,y,z|\hat{p}_{mle}, \hat{q}_{mle}, \hat{\gamma}_{mle})} \tag{3.1.1}$$

where $L(x,y,z|H_0)$ represents the likelihood function under the null hypothesis, $L(x,y,z|\Omega)$ represents the likelihood function under the whole parameter space, and $\hat{p}_{mle0}, \hat{q}_{mle0}$ are the MLE under the null hypothesis where $\gamma = 0$, while $\hat{p}_{mle}, \hat{q}_{mle}, \hat{\gamma}_{mle}$ denote the MLE under the alternative hypothesis. MLE can be calculated by applying the above method in last section.

We try to calculate p-value of the hypothesis test follow the methodology below:

a) Draw 1000 samples under the null hypothesis, from which a multinomial distribution with parameter $\hat{p} = \frac{n_{11}+n_{12}}{n}$, $\hat{q} = \frac{n_{11}+n_{21}}{n}$, $\gamma = 0$.

b) For each sample drawn above, which we can denote as $(n_{11}^{(h)}, n_{12}^{(h)}, n_{21}^{(h)}, n_{22}^{(h)})$, calculate the the parameters with respect to H_0, which are $p_{mle0}^{(h)}, q_{mle0}^{(h)}, \gamma^{(h)} = 0$; and H_a, which are $p_{mle}^{(h)}, q_{mle}^{(h)}, \gamma_{mle}^{(h)}$, now we get all the parameters in Likelihood ratio statistics for each sample.

c) Plug in each sample, as well as the original table into their likelihood ratio test statistic and count the number of samples which are smaller than the value we got from the original table. So the p-value can be expressed as

$$p - value = P(T < T_{obs}|H_0) = \#(T < T_{obs})/1000 \tag{3.1.2}$$

where $T_{obs} = L(n_{11} + n_{10}, n_{11} + n_{01}, n_{11}|H_0)$.

Applying our specific data above, we get a p-value of 0.629, which fails to reject the null hypothesis compared to the p-value 0.8021 got from Fisher's Exact test and 0.6542 from chi-square test. In this case, the likelihood ratio test produces a consistant testing result with the other two standard test, even though a contradiction may occur on other tables and we will discuss more situations later.

3.2   Power comparison

So far, we have already got our newly defined method of likelihood ratio test, and got a brief comparison with Fisher's exact test and chi-square test. It is natural to think about comparing their power. Here we consider the power under three different test levels $\alpha = 0.01, 0.025, 0.05$, and the corresponding test statistics come respectively from the 1%, 2.5% and 5% lower quantile (T_α) of the 1000 samples drawn under the null hypothesis $\gamma = 0$ are 0.1350654, 0.07754259, 0.04561459.

In order to find the power of the test, in other words, we need to find the probability of rejection of null hypothesis under each value of $\gamma > 0$ within the range which is restrained by p and q (2.2.11). We divide the range into 85 pieces (each one is denoted as γ_i, grows in 0.01 from 0, where i=1, ..., 85) which can approximately express the whole alternative hypothesis. Since the value of γ is

18

constrained by p and q (see 2.2.11), so it cannot reach 1. Actually γ can only reach to 0.8460434. For each γ_i value greater than 1, the approach of finding the power is similar to the method of obtaining the p-value we discussed in the above section. Instead of drawing samples under the null hypothesis with $\gamma = 0$, we draw samples under each $\gamma_i > 0$, and then calculating the 1000 likelihood ratio statistics the same way.

Then p-value can be expressed as

$$p = \#(T_i < T_\alpha)/1000 \tag{3.2.1}$$

where T_α is the α quantile of the 1000 likelihood ratio test statistics generated from corresponding samples drawn above.

So we now get the power function of the likelihood ratio test, and now we intend to process a comparison to the power function of the Fisher's exact test.

Recall that the algorithm of the Fisher's exact test comes from hypergeometric distribution where we can obtain the p-value, which is the essence of the Fisher's exact test. In other words, through comparing the p-value we calculate with the test size (significance level), we can make a decision of rejecting the null hypothesis or not. Start from the definition of the power, which is the probability of correctly rejecting the null hypothesis, our interest is to obtain the power of the Fisher's exact test. However, the commonly referred power analysis of the Fisher's exact test focus on calculating the power with regard to different table size. Remember that in calculating the power of the likelihood ratio test, what we actually do is calculating the rejection probability point-by-point under the alternative hypothesis. So inspired by the same methodology, we need to find the point-by-point power with respect to each γ value under the alternative hypothesis, which in our case are the 85 values range equally from 0 to the maximum value of γ restrained by p and q. Thus we are thinking about applying our table to the non-central hypergeometric distribution. Hence, we start from introducing the probability mass function of the non-central hypergeometric distribution.

For a table like

	Receive	Not receive	
Males	x	$m_1 - x$	m_1
Females	$n - x$	$m_2 - n + x$	m_2
	n	$N - n$	N

The corresponding probability mass function is

$$fnchypg(x|n, m_1, N, w) = \frac{\binom{m_1}{x}\binom{m_2}{n-x}w^x}{\sum_{y=x_{min}}^{x_{max}} \binom{m_1}{y}\binom{m_2}{n-y}w^y} \tag{3.2.2}$$

where $m_1, m_2 \in \mathbb{N}, N = m_1 + m_2, n \in [0, N), w \in \mathbb{R}_+$, and $x \in [x_{min}, x_{max}]$, represent the number of the male patients who received the therapy, where $x_{min} = max(0, n - m_2), x_{max} = min(n, m_1)$, and w is the odds ratio of the male group over the female group, and the expression of the odds ratio is

$$w = \frac{\frac{\pi_{11}}{\pi_{11}+\pi_{10}}}{\frac{\pi_{01}}{\pi_{01}+\pi_{00}}} = \frac{\frac{pq+\gamma\sqrt{p(1-p)q(1-q)}}{p}}{\frac{q(1-p)-\gamma\sqrt{p(1-p)q(1-q)}}{1-p}} \tag{3.2.3}$$

which can be expressed as a function of γ, with $p = \frac{m_1}{N}$ and $q = \frac{n}{N}$ which are the straightforward estimation obtained directly from our observed table. Hence we can connect a specific γ value to some specific non-central hypergeometric distribution through the odds ratio w, and then we fit our table to that specific non-central hypergeometric distribution. As we try to run a upper-sided hypothesis test at a α significance level, first of all we need to find the upper (1- α)% critical value c_α of the table, and since the margins are fixed we only need to know the value of x. Because of the discrete nature of the distribution of X, we cannot strictly satisfy the condition $Pr(x \geq c_\alpha | H_0) = \alpha$. So we choose the c_α= $\{x : Pr(x \geq c_\alpha | H_0) - \alpha$ reaches the minimum$\}$ It is worth mentioning that the H_0 represent to the hypergeometric distribution that the odds ratio equals to 1, which can be straightly demonstrated by setting the value γ to 0.

$$power(\gamma) = Pr(x \geq c_\alpha | H_a) = Pr(x \geq c_\alpha | \gamma \neq 0) \tag{3.2.4}$$

20

$$= \sum_{x=c_\alpha}^{n} fnchypg(x|n, m_1, N, w(\gamma)) \tag{3.2.5}$$

Here we use R package 'BiasedUrn' for computing the non-central hypergeometric distribution. Henceforth, we can draw a comparative power plot through point description method, with regard to a test level of 0.01 (Figure 3.1) and 0.05 (Figure 3.2). We discover that the power plot of our likelihood ratio test (blue one) is almost above the power plot of the Fisher's exact test (red one), meaning that the likelihood ratio test is less likely to make a type II error compared to the Fisher's exact test. Additionally, from the fact that the intercept value which corresponds to $\gamma = 0$ of the power function is the test size, we can indirectly verify the rationality of our methodology of computing the power plot.

Figure 3.1: Power function of two tests at $\alpha = .01$
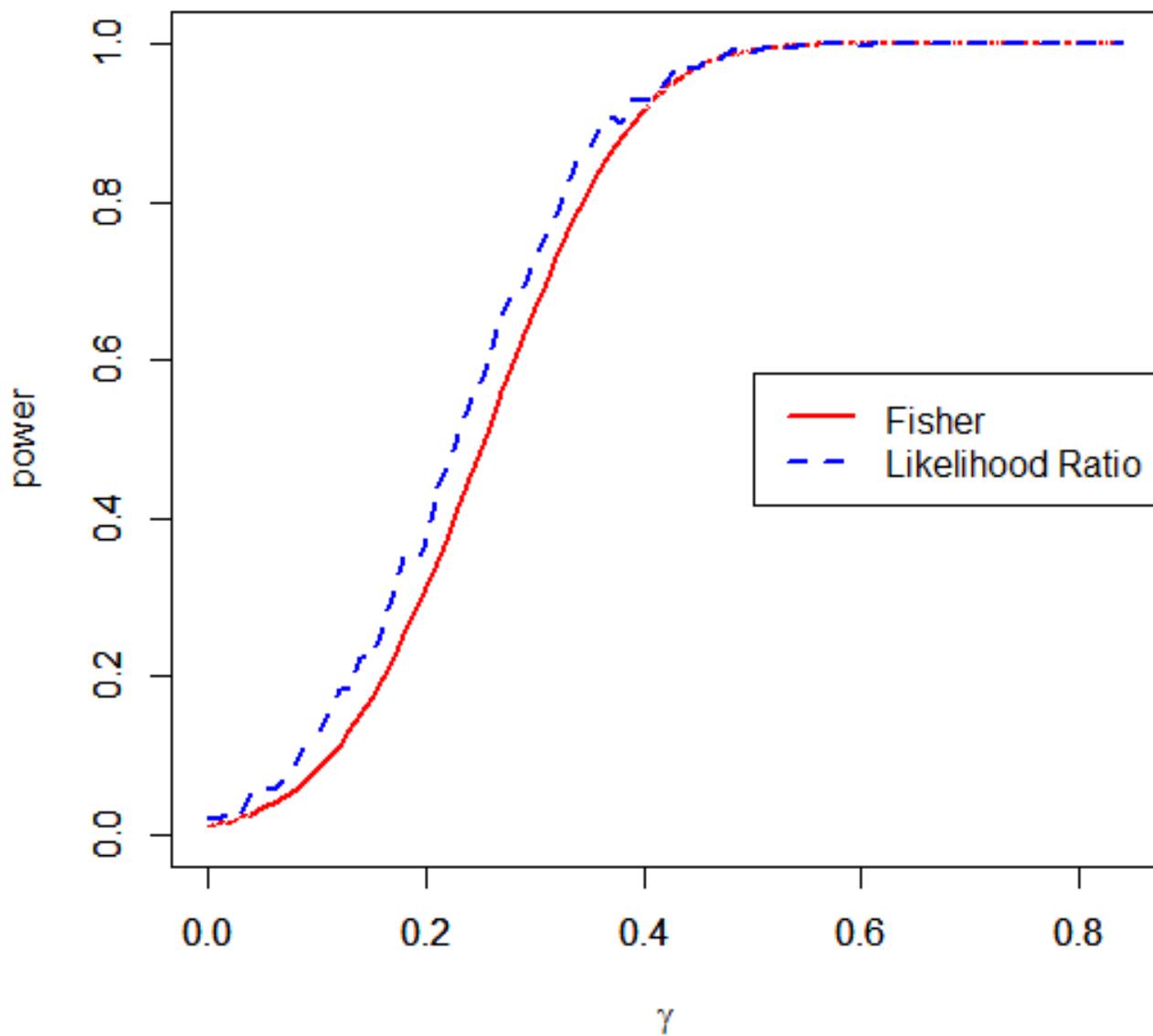
Figure 3.2: Power function of two tests at $\alpha = .05$
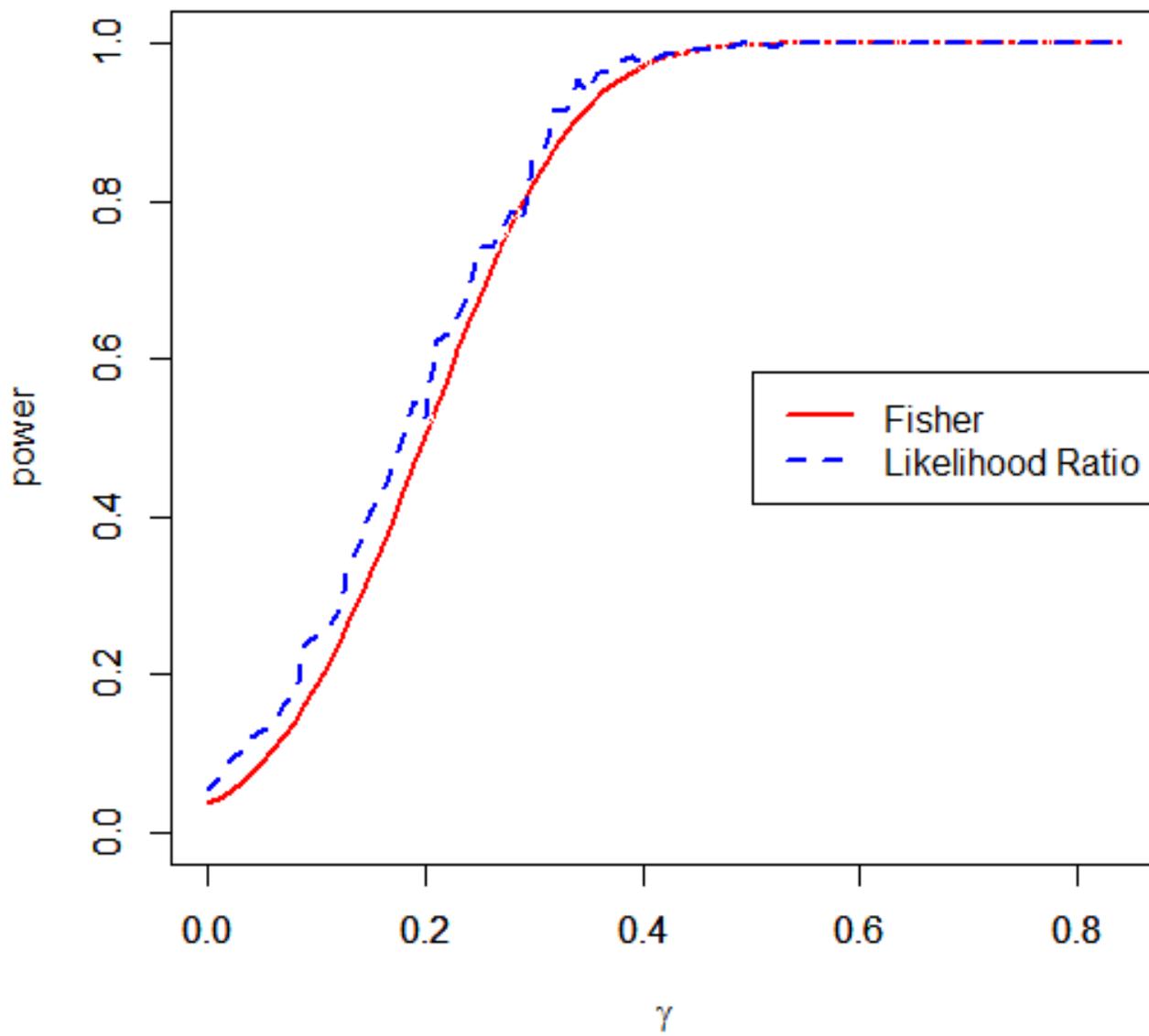
Chapter 4

Bayes factor

We have already developed a natural way of judging the independence or the dependence situation of the two variables of a 2×2 contingency table by testing the null hypothesis $\gamma = 0$. Additionally, if we want to know more about the exact (positive or negative) relationship between the two variables, we can just change the alternative hypothesis from $\gamma \neq 0$ to $\gamma > 0$ (or $\gamma < 0$) respectively. And all of the work is done within the knowledge of classical statistics. Naturally, we are thinking about extended our idea to Bayesian statistics method, thinking of the fact that Bayes factor can be treated as Bayesian alternative to classical hypothesis testing. The Bayes factor is defined as the ratio of the marginal likelihoods of the two models M1 and M2, which respectively denote the probability distribution model under the null hypothesis and alternative hypothesis. Considering that the essence of our problem is still with regard to testing against two hypothesis, so Bayes factor could be automatically applied to our problem. The general idea remains the same, since Bayes factor is defined as the ratio of the marginal likelihoods, we can treat Bayes factor similarly like the likelihood ratio above as the test statistics, and run the test same way. In the following section, we will compute the Bayes factor by drawing samples from the given observed table.

4.1   Computing the Bayes factor

Under the null hypothesis when $\gamma = 0$, the probability distribution function of X, Y and Z have the simple form when we plug in $\gamma = 0$, which is

M_1:

$$p(x, y, z|p, q) = \frac{n!}{z!(x-z)!(y-z)!(n-x-y+z)!}p^x(1-p)^{n-x}q^y(1-q)^{n-y}$$

While under the null hypothesis when $\gamma \neq 0$, the probability distribution function of X, Y and Z have the form

M_2:

$$p(x, y, z|p, q, \gamma) = \frac{n!}{z!(x-z)!(y-z)!(n-x-y+z)!}\pi_{11}{}^z\pi_{10}{}^{x-z}\pi_{01}{}^{y-z}\pi_{00}{}^{n-x-y+z}$$

where

$$\pi_{11} = pq + \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{10} = p(1-q) - \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{01} = q(1-p) - \gamma\sqrt{p(1-p)q(1-q)},$$

$$\pi_{00} = (1-p)(1-q) + \gamma\sqrt{p(1-p)q(1-q)}.$$

A reasonable assumption for the prior distributions are $\pi(p) = 1, \pi(q) = 1$ and $\pi(\gamma)=1$, since p, q and γ ranges from 0 to 1. Thus, the marginal likelihood for M_1 can be expressed as

$$
\begin{aligned}
f(x, y, z|M_1) &= \int_0^1 \int_0^1 p(x, y, z|p, q)\pi(p)\pi(q) \; dp \; dq \qquad\qquad (4.1.1)\\
&= \int_0^1 \int_0^1 C \cdot p^x(1-p)^{n-x}q^y(1-q)^{n-y} \cdot 1 \cdot 1 \; dp \; dq \\
&= C \cdot B(x+1, n-x+1)B(y+1, n-y+1)
\end{aligned}
$$

where $B()$ represents the beta function, and $C = \frac{n!}{z!(x-z)!(y-z)!(n-x-y+z)!}$ is the constant after the integration. Similarly, the marginal likelihood for M_2

$$f(x, y, z|M_2) = \int_{a_0(p,q)}^{a_1(p,q)} \int_0^1 \int_0^1 p(x, y, z|p, q, \gamma)\pi(p)\pi(q)\pi(\gamma) \; dp \; dq \; d\gamma \qquad (4.1.2)$$

where in our case that H_a is $\gamma > 0$, $a_0(p, q) = 0$, $a_1(p, q) = min(\sqrt{\frac{(p(1-q)}{(1-p)q}}, 1/\sqrt{\frac{(p(1-q)}{(1-p)q}})$ is a function of p, q indicates the upper bound of γ given p and q. So the Bayes factor, the ratio of the marginal likelihood of M_2 to M_1 is

$$T_{BF} = \frac{f(x, y, z|M_2)}{f(x, y, z|M_1)} \quad (4.1.3)$$

In order to calculate Bayes factor, the pivotal procedure is calculating the integration of both nominator and the denominator. As we can see above derivation, the denominator integration can be easily calculated from beta distribution, obtaining the result of a multiplication of the two beta function. However, unlike the denominator, the nominator has a more complicated integrand so that the integration cannot be calculated through normal way. Therefore, we consider using a Monte Carlo method to get the an approximation to our integration. There are some methods work for computing the marginal likelihood, including applying the reduced Metropolis- Hastings samplers proposed by Chib and Jeliazkov (2001), which is also implemented by Nandram and Kim (2002) in the paper studying the marginal likelihood for a class of bayesian generalized models. Here we apply complete importance sampling into the procedure computing the integration. We rewrite the integration in a new form

$$f(x, y, z|M_2) = \int_{a_0(p,q)}^{a_1(p,q)} \int_0^1 \int_0^1 \left\{ \frac{p(x, y, z|p, q, \gamma)\pi(p)\pi(q)\pi(\gamma)}{\pi_a(p, q, \gamma|x, y, z)} \right\} \pi_a(p, q, \gamma|x, y, z) \, dp \, dq \, d\gamma \quad (4.1.4)$$

which means that instead of integrating over its prior density, we integrate over an approximation to the posterior density $\pi_a(p, q, \gamma|x, y, z)$. By drawing samples from the posterior density, we then are able to apply the Monte Carlo method to compute the integration. Thus our next task is obtaining the exact expression of our posterior density distribution $\pi_a(p, q, \gamma|x, y, z)$. Inspired by the above deduction that $X \sim B(n, p)$, $Y \sim B(n, q)$, we naturally consider the fact that the conjugate distribution of binomial distribution is beta distribution, thus we can make an assumption that the posterior distribution for p is beta distribution and so as q. So given the data (x, y, z), we can draw p from $\pi_a(p|x, y, z) = \text{beta}(x + 1, n - x + 1)$, and draw q from $\pi_a(q|x, y, z) = \text{beta}(y + 1, n - y + 1)$.

And the approximate posterior density can be decomposed as :

$$\pi_a(p, q, \gamma | x, y, z) = \pi_a(p | x, y, z)\ \pi_a(q | x, y, z)\ \pi_a(\gamma | p, q, x, y, z) \tag{4.1.5}$$

Acknowledging the fact that we can draw p and q respectively from beta distribution, we certainly expect that the posterior density for γ has the same distribution, otherwise is close to beta distribution to the most extent. Notice that beta distribution require the random variable ranges from 0 to 1, which actually can be satisfied by making a little adjustment according to the range of γ, we proceed with this thought to get the specific expression of posterior density $\pi_a(\gamma | p, q, x, y, z)$ where we can draw γ.

We will list the detailed procedure,

a) Given the count data (x, y, z), firstly we draw p from beta$(x + 1, n - x + 1)$; draw q from beta$(y + 1, n - y + 1)$, which can be denoted as $p^{(k)}$ and $q^{(k)}$, where $k = 1, 2, ..., M$. In our case, we choose a reasonable sample size M $= 1000$.

b) Apply the same $p^{(h)}$, $q^{(h)}\gamma^{(h)}$ generated from the process of computing the MLE in section 2.2 , and calculate the related range $a_0(p^{(h)}, q^{(h)})$ and $a_1(p^{(h)}, q^{(h)})$. Thus we have 1000 sets of $p^{(h)}$, $q^{(h)}$, $\gamma_{(h)}$, $a_0(p^{(h)}, q^{(h)})$ and $a_1(p^{(h)}, q^{(h)})$

c) Fit M $\gamma^{(h)}$ samples to the model :

$$\gamma_1^{(h)} = \frac{\gamma_{(h)} - a_0(p^{(h)}, q^{(h)})}{a_1(p^{(h)}, q^{(h)}) - a_0(p^{(h)}, q^{(h)})} \sim beta(\mu\tau, (1 - \mu)\tau) \tag{4.1.6}$$

where $0 < \mu < 1, \tau > 0$.

According to the property of the beta distribution, the mean of beta $(\mu\tau, (1-\mu)\tau)$ is $\frac{\mu\tau}{\mu\tau + (1-\mu)\tau} = \mu$, and the variance is $\frac{\mu(1-\mu)}{\tau+1}$. Thus we can estimate the beta distribution parameters by equalling the sample mean to the real mean and equal the sample variance to the real variance. By solving the equations:

$$E(\gamma_1^{(h)}) = \frac{\sum_{h=1}^{M} \gamma_1^{(h)}}{M} = \mu \tag{4.1.7}$$

27

$$Var(\gamma_1^{(h)}) = \frac{\sum_{h=1}^{M}(\gamma_1^{(h)} - E(\gamma_1^{(h)}))^2}{M-1} = \frac{\mu(1-\mu)}{\tau+1} \tag{4.1.8}$$

We get an estimation of μ and τ which are:

$$\hat{\mu} = E(\gamma_1^{(h)}) \tag{4.1.9}$$

$$\hat{\tau} = \frac{\mu(1-\mu)(M-1)}{\sum_{h=1}^{M}(\gamma_1^{(h)} - E(\gamma_1^{(h)}))^2} - 1 \tag{4.1.10}$$

d) Draw $\gamma_1^{(k)}$ from beta $(\hat{\mu}\hat{\tau}, (1-\hat{\mu})\hat{\tau})$, then $\gamma = h(\gamma_1^{(k)}) = [a_1(p^{(k)}, q^{(k)}) - a_0(p^{(k)}, q^{(k)})] \gamma_1 + a_0(p^{(k)}, q^{(k)})$, where $p^{(k)}$ and $q^{(k)}$ are drawn from beta distribution we mentioned above.

e) Get the expression of $\pi_a(p, q, \gamma|x, y, z)$:

The probability density function of γ_1 is

$$f(\gamma_1) = \frac{\gamma_1^{\hat{\mu}\hat{\tau}}(1-\gamma_1)^{(1-\hat{\mu})\hat{\tau}}}{B(\hat{\mu}\hat{\tau}, (1-\hat{\mu})\hat{\tau})} \tag{4.1.11}$$

Due to the rule of the density function of a transformed random variable, the probability density function of γ is

$$\pi_a(\gamma|p, q, x, y, z) = f(h^{-1}(\gamma)) \tag{4.1.12}$$

$$= \frac{[h^{-1}(\gamma)]^{\hat{\mu}\hat{\tau}}[1-h^{-1}(\gamma)]^{(1-\hat{\mu})\hat{\tau}}}{B(\hat{\mu}\hat{\tau}, (1-\hat{\mu})\hat{\tau})} \cdot \frac{d(h^{-1}(\gamma))}{d\gamma}$$

$$= \frac{\left[\frac{\gamma - a_0(p,q)}{a_1(p,q) - a_0(p,q)}\right]^{\hat{\mu}\hat{\tau}}\left\{1 - \left[\frac{\gamma - a_0(p,q)}{a_1(p,q) - a_0(p,q)}\right]\right\}^{(1-\hat{\mu})\hat{\tau}}}{B(\hat{\mu}\hat{\tau}, (1-\hat{\mu})\hat{\tau})} \cdot \frac{1}{a_1(p,q) - a_0(p,q)}$$

Together with

$$\pi_a(p|x, y, z) = \frac{p^x(1-p)^{n-x}}{B(x+1, n-x+1)} \tag{4.1.13}$$

$$\pi_a(q|x, y, z) = \frac{q^y(1-q)^{n-y}}{B(y+1, n-y+1)} \tag{4.1.14}$$

Therefore,

$$\pi_a(p, q, \gamma | x, y, z) = \pi_a(p | x, y, z) \pi_a(q | x, y, z) \pi_a(\gamma | p, q, x, y, z) \tag{4.1.15}$$

$$= \frac{p^x (1-p)^{n-x}}{B(x+1, n-x+1)} \frac{q^y (1-q)^{n-y}}{B(y+1, n-y+1)} \frac{[h^{-1}(\gamma)]^{\hat{\mu}\hat{\tau}} [1 - h^{-1}(\gamma)]^{(1-\hat{\mu})\hat{\tau}}}{B(\hat{\mu}\hat{\tau}, (1-\hat{\mu})\hat{\tau})} \cdot \frac{d(h^{-1}(\gamma))}{d\gamma}$$

Through all the work above, we can go back to our test statistics T_{BF}

$$T_{BF} = \frac{f(x, y, z | M_2)}{f(x, y, z | M_1)} \tag{4.1.16}$$

$$= \frac{\int_{a_0(p,q)}^{a_1(p,q)} \int_0^1 \int_0^1 \left\{ \frac{p(x,y,z|p,q,\gamma)\pi(p)\pi(q)\pi(\gamma)}{\pi_a(p,q,\gamma|x,y,z)} \right\} \pi_a(p, q, \gamma | x, y, z) \, dp \, dq \, d\gamma}{C \cdot B(x+1, n-x+1) B(y+1, n-y+1)}$$

$$= \frac{\frac{1}{M} \sum_{k=1}^M \frac{C \pi_{11}^z \pi_{10}^{x-z} \pi_{01}^{y-z} \pi_{00}^{n-x-y+z} \pi(p)\pi(q)\pi(\gamma)}{\pi_a(p,q,\gamma|x,y,z)}}{C \cdot B(x+1, n-x+1) B(y+1, n-y+1)}$$

f) Plug 1000 sets of $p^{(k)}$, $q^{(k)}$, $\gamma^{(k)}$ into (4.1.16) to obtain the Bayes factor.

Since it is a Bayes factor based on M_2 to M_1, a Bayes factor value of greater than 1 will indicate a rejection of null hypothesis. Similarly, we can get the power function of Bayes factor test just by treating the Bayes factor as the likelihood ratio test statistics in the former section. A widely sited judging table is provided by Kass and Raftery (1995):

$2 \ln(B_{10})$	(B_{10})	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

where B_{10} is the Bayes factor of M_2 to M_1.

Chapter 5

Conclusions

In this chapter we summarize our methods, give results on Worcester Heart Attack study, and we give an extension to a general $r \times c$ contingency table. We have been interested in a test of independence in a 2×2 contingency table when the margins are random. This has been a long-standing problem of the Fisher's exact test.

5.1 Summary

We have built a model based on the multinomial distribution to accommodate random margins in a 2×2 contingency table. We have constructed two competitive tests which take care of the random margins. The first test is based on the likelihood ratio and the second test is based on the Bayes factor. We have obtained the p-values for these tests; we use the Bayes factor for the corresponding test. These tests naturally remove the conservative nature of the Fisher's exact test. We have also obtained the power functions of the three tests for different sizes. We found that likelihood ratio test is always better than the Fisher's exact test and the based on the Bayes factor may also be better.

5.2 Results

By applying our tests to the data from three examples of the Worcester Heart Attack Study and calculating the p-values with respect to the different method, we obtain two kinds of p-values and the Bayes factor for each group listed in the table below. We can see that generally the p-value of likelihood ratio test and Fisher's exact test are close. In addition p-value of the Fisher's exact test tends to be greater (17 groups in total 18 groups in all the three examples), which can be understood as a result of the conservatism of Fisher's exact test.

For 14 groups (the last five groups in Example 1, the 1st, 3rd, 4th and 6th group in Example 2, and the 4th group in Example 3), we can get a consistent conclusion of no rejection which means male patients are not more likely to receive the therapy. This is true because both kinds of the p-values are greater than 0.05 and the Bayes factor is small than 1. For the first group in Example 1 and the second group in Example 2, both likelihood ratio test and Bayes factor reject while Fisher's exact test does not. For the 5th group in Example 3 all three tests tend to reject; for the 4th group in Example 3, only Bayes factor rejects.

To deal with this apparent contradictory conclusion, we should look at the power functions of these tests. A higher power plot indicates that we are less likely to making error of failing to reject the null hypothesis when the alternative hypothesis is true. Figure 5.1 and 5.2 shows that the power function of likelihood ratio test generally goes above the Fisher's exact test under test level of $\alpha = 0.01$ and $\alpha = 0.05$ and we also list some of the powers in Table 5.1 and 5.2, thus one may trust the result of likelihood ratio test. The gain in power of our likelihood ratio test may come from our more flexible model setting than the Fisher's exact test. We also look at the power function for the Bayes factor test to see if we have confidence to reject the null hypothesis of the 4th group in Example 3.

Example 1: Table I. Gender differences in the receipt of lidocaine therapy in AMI patients with a history of hypertension and stroke, Worcester Heart Attack Study

(*a*) Number of patients who did (+) and did not (−) receive lidocaine therapy (Rx)

Age(years)	Males Lidocaine Rx		Females Lidocaine Rx	
	(+)	(−)	(+)	(−)
< 55	4	4	0	5
55–59	2	3	1	2
60–64	6	5	2	4
65–69	4	3	6	6
70–74	8	11	3	12
⩾ 75	8	22	16	35

Age (years)	p-value (Likelihood ratio test)	p-value (Fisher's exact test)	Bayes factor
<55	0.01	0.0979	2.65415
55-59	0.456	0.7143	0.3760477
60-64	0.214	0.3733	0.5483426
65-69	0.355	0.57	0.3151118
70-74	0.104	0.1592	0.7962698
≥75	0.54	0.7562	0.09215492

Example 2: Table II. Gender differences in the receipt of antiplatelet therapy in AMI patients with a history of hypertension and stroke, Worcester Heart Attack Study

(*a*) Number of patients who did (+) and did not (−) receive antiplatelet therapy (Rx)

Age(years)	Males Antiplatelet Rx		Females Antiplatelet Rx	
	(+)	(−)	(+)	(−)
< 55	3	5	2	3
55–59	3	2	0	3
60–64	5	6	1	5
65–69	3	4	8	4
70–74	14	5	5	10
⩾ 75	16	14	22	29

Age (years)	p-value (Likelihood ratio test)	p-value (Fisher's exact test)	Bayes factor
<55	0.803	0.7506	0.265819
55-59	0.036	0.1786	1.894009
60-64	0.12	0.2613	0.9032894
65-69	0.979	0.9326	0.1273676
70-74	0.012	0.02191	5.554933
≥75	0.178	0.2554	0.3041774

Example 3: Table III. Gender differences in the receipt of beta blocker therapy AMI patients with a history of congestive heart failure and hypertension, Worcester Heart Attack Study

(a) Number of patients who did (+) and did not (−) receive beta blocker therapy (Rx)

Age(years)	Males Beta blocker Rx		Females Beta Blocker Rx	
	(+)	(−)	(+)	(−)
< 55	0	1	3	2
55–59	0	6	2	1
60–64	2	4	1	5
65–69	4	3	2	7
70–74	7	12	6	6
⩾ 75	7	21	16	66

Age (years)	p-value (Likelihood ratio test)	p-value (Fisher's exact test)	Bayes factor
<55	0.910	1	0.2023028
55-59	0.798	1	0.1023002
60-64	0.283	0.5	0.4739417
65-69	0.094	0.1818	1.197311
70-74	0.902	0.8635	0.1320737
≥75	0.288	0.3559	0.2263033

Figure 5.1: Power functions of the three tests at $\alpha = .01$
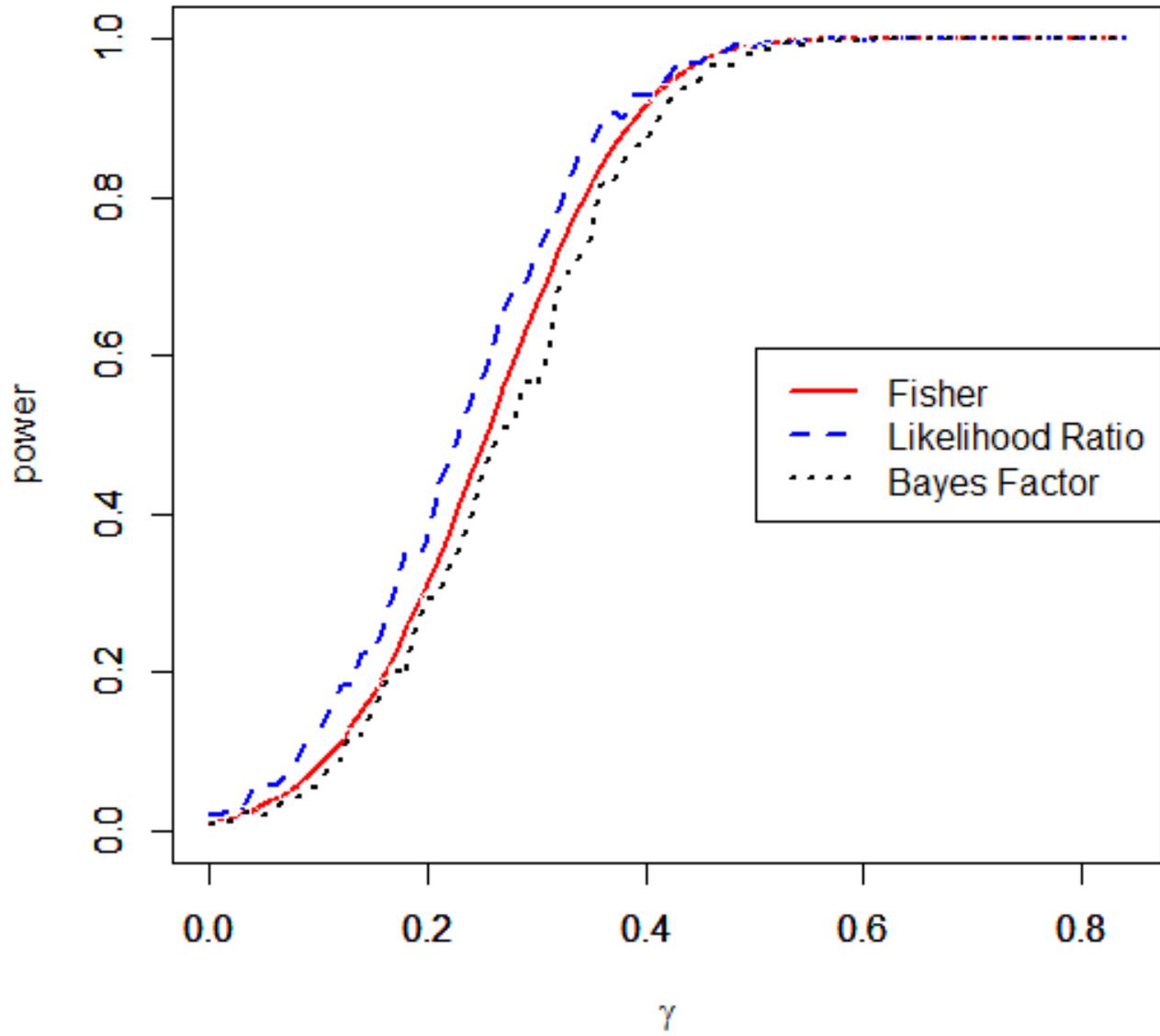
Table 5.1: Power values of the three tests of $\alpha = .01$ for γ ranges from 0 to 0.8460434

γ	Likelihood ratio test	Fisher's exact test	Bayes factor
0.00	0.022	0.01057581	0.008
0.05	0.060	0.03241211	0.018
0.10	0.128	0.08177870	0.057
0.15	0.224	0.17304464	0.152
0.20	0.368	0.31191125	0.292
0.25	0.573	0.48604699	0.453
0.30	0.731	0.63083935	0.561
0.35	0.868	0.81579139	0.749
0.40	0.927	0.91705728	0.869
0.45	0.973	0.97063244	0.951
0.50	0.990	0.99222457	0.986
0.55	0.997	0.99791878	0.997
0.60	0.998	0.99983293	0.996
0.65	1.000	0.99998964	1.000
0.70	1.000	0.99999975	1.000
0.75	1.000	1.00000000	1.000
0.80	1.000	1.00000000	1.000

Figure 5.2: Power functions of the three tests at $\alpha = .05$
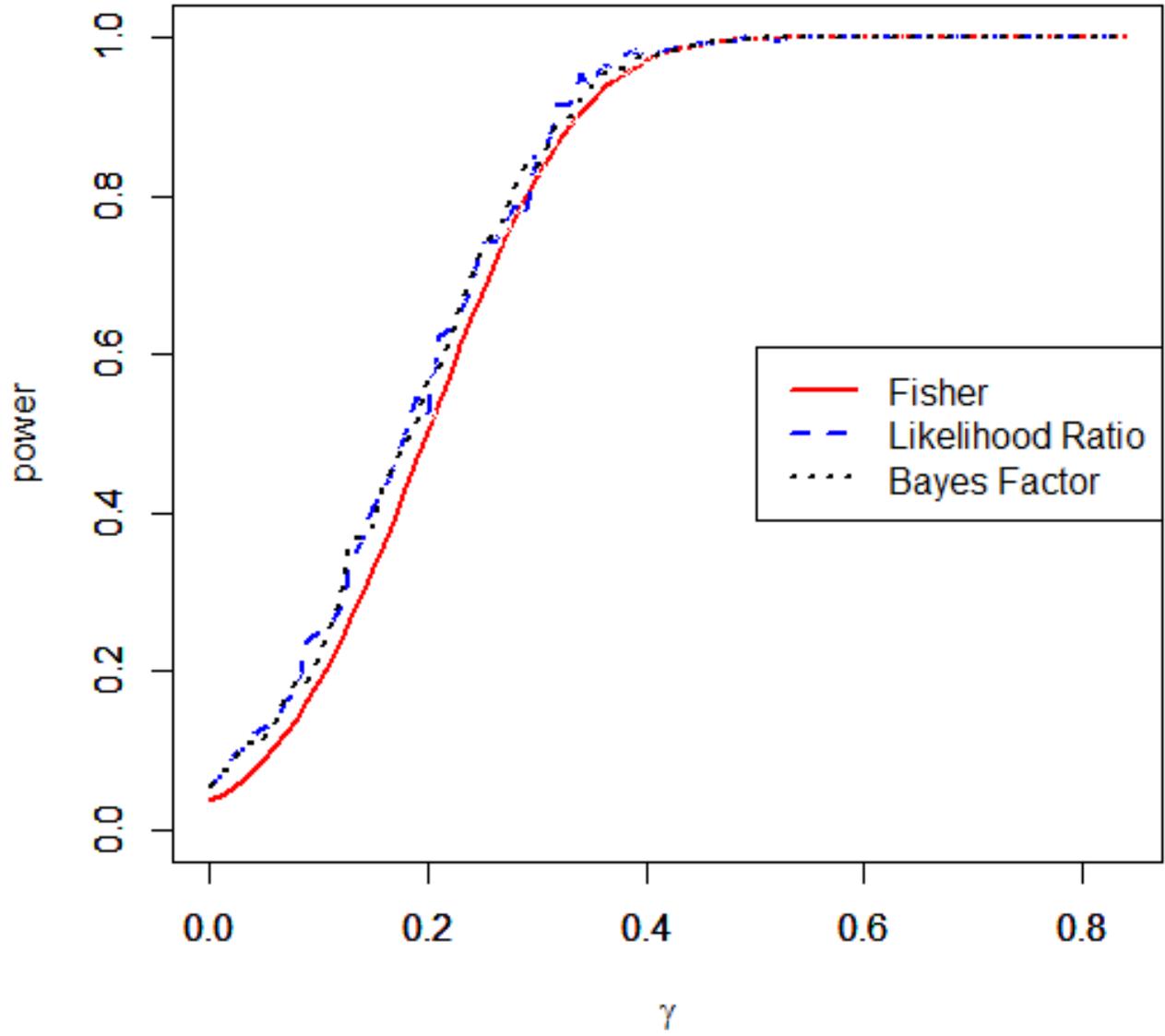
Table 5.2: Power values of the three tests at $\alpha = .05$ for γ ranges from 0 to 0.8460434

γ	Likelihood ratio test	Fisher's exact test	Bayes factor
0.00	0.054	0.03531537	0.053
0.05	0.128	0.08867190	0.113
0.10	0.249	0.18531512	0.217
0.15	0.409	0.32869365	0.384
0.20	0.526	0.50372789	0.565
0.25	0.740	0.67948445	0.736
0.30	0.860	0.82388455	0.837
0.35	0.938	0.91994950	0.940
0.40	0.974	0.97085519	0.977
0.45	0.992	0.99183238	0.994
0.50	1.000	0.99832841	0.999
0.55	1.000	0.99976793	1.000
0.60	0.999	0.99998043	1.000
0.65	1.000	0.99999916	1.000
0.70	1.000	0.99999999	1.000
0.75	1.000	1.00000000	1.000
0.80	1.000	1.00000000	1.000

5.3 Extension

It is natural to think about how to generalize our methodology to higher dimensional tables. In particular, in our situation we need to discuss how to test for independence in $r \times c$ contingency tables with larger values of r and c and specifically for random margins. Unlike the 2×2 contingency table where we have used the correlation coefficient between rows and columns as a test parameter, we cannot do so for the general $r \times c$ table. So we need to construct a new test procedure.

Consider the general $r \times c$ contingency table, one for the population and the other for the sample.

Table 5.3: Population table

	1	2	$\cdots$	c	
1	$\gamma_{11} p_1 q_1$	$\gamma_{12} p_1 q_2$	$\cdots$	$p_1 q_c$	p_1
2	$\gamma_{21} p_2 q_1$	$\gamma_{22} p_2 q_2$	$\cdots$	$p_2 q_c$	p_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
r-1	$\gamma_{r-1} p_{r-1} q_1$	$\gamma_{22} p_2 q_2$	$\cdots$	$p_{r-1} q_c$	p_{r-1}
r	$p_r q_1$	$p_r q_2$	$\cdots$	$p_r q_c$	p_r
	q_1	q_2	$\cdots$	q_c	1

Table 5.4: Population table

	1	2	$\cdots$	c	
1	n_{11}	n_{11}	$\cdots$	n_{1c}	x_1
2	n_{21}	n_{22}	$\cdots$	n_{2c}	x_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	x_r
	y_1	y_2	$\cdots$	y_c	n

Note that $0 < \gamma_{ij} p_i q_j < 1, i = 1, \ldots, r, j = 1, \ldots, c$ and $\sum_{i=1}^{r} p_i = \sum_{j=1}^{c} q_j = 1$

We wish to test and the null hypothesis is

$$H_0 : \gamma_{11} = \cdots = \gamma_{rc} = 1.$$

The alternative hypothesis is

$$H_a : \gamma_{ij} \neq 1 \text{ for some } 1 \leq i \leq r, 1 \leq j \leq c$$

.

There are a number of restrictions that must be satisfied among the $\gamma_{ij} p_i q_j$. For the first row in our model setting, there is a restriction between parameters, which is

$$\gamma_{11} p_1 q_1 + \cdots + \gamma_{r-1,1} p_{r-1} q_1 + p_r q_1 = q_1. \tag{5.3.1}$$

By dividing both sides by q_1 we get

$$\gamma_{11} p_1 + \cdots + \gamma_{r-1} p_{r-1} + p_r = 1.$$

Now replacing p_r with $1 - \sum_{i=1}^{r-1} p_i$ and we obtain

$$\gamma_{11}p_1 + \cdots + \gamma_{r-1}p_{r-1} + 1 - \sum_{i=1}^{r-1} p_i = 1 \tag{5.3.2}$$

$$\gamma_{11}p_1 + \cdots + \gamma_{r-1}p_{r-1} - \sum_{i=1}^{r-1} p_i = 0$$

$$\sum_{i=1}^{r-1} (\gamma_{i1} - 1)p_i = 0$$

Generally we can obtain $r + c$ restrictions exactly the same way, by making summation of rows and columns. Together with the global restriction $\sum_{i=1}^{r} p_i = \sum_{j=1}^{c} q_j = 1$, we totally have $r + c + 2$ restrictions and $r + c + (r-1)(c-1) - 2 = rc - 1$ parameters.

Hence we can rewrite our $r \times c$ table as So we can fit data from $r \times c$ into a multinomial

Table 5.5: A constrained $r \times c$ contingency table for the sample table

	1	2	$\cdots$	c-1	c	
1	n_{11}	n_{12}	$\cdots$	$n_{1,c-1}$	$x_1 - \sum_{j=1}^{c-1} n_{1j}$	x_1
2	n_{21}	n_{22}	$\cdots$	$n_{2,c-1}$	$x_2 - \sum_{j=1}^{c-1} n_{1j}$	x_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r-1	$n_{r-1,1}$	$n_{r-1,2}$	$\cdots$	$n_{r-1,c-1}$	$x_{r-1} - \sum_{j=1}^{c-1} n_{1j}$	x_{r-1}
r	$y_1 - \sum_{i=1}^{r-1} n_{i1}$	$y_2 - \sum_{i=1}^{r-1} n_{i1}$	$\cdots$	$y_{c-1} - \sum_{i=1}^{r-1} n_{i1}$	$n - \sum_{i=1}^{r-1} x_i - \sum_{j=1}^{c-1} y_j$	x_r
	y_1	y_2	$\cdots$	y_{c-1}	y_c	n

distribution,

$$(n_{11}, \cdots, n_{r-1,c-1}, x_1 - \sum_{j=1}^{c-1} n_{1j}, \cdots, y_1 - \sum_{i=1}^{r-1} n_{i1}, \cdots n - \sum_{i=1}^{r-1} x_i - \sum_{j=1}^{c-1} y_j) \sim \text{Multinomial}(n, vecp),$$

$$\tag{5.3.3}$$

where $vecp = (\gamma_{11}p_1q_1, \ldots, p_rq_c)$; see the population table.

Thus a more complicated multinomial distribution model is constructed and we can develop the likelihood ratio test and Bayes factor test to test for independence in $r \times c$ contingency tables. This is our future plan.

Bibliography

[1] Barnard G. A. (1945), *A New Test for 2×2 Tables.* Nature, Vol. 156, 177 &783, 86-105.

[2] Berkson J. (1978), *Do the marginal totals of the 2×2 table contain relevant information respecting the table proportions ?.* Journal of Statistical Planning and Inference Vol. 2, No. 3, 43–44.

[3] Chib, S., and Jeliazkov, I. (2001), *Marginal Likelihood From the MetropolisHastings Output.* Journal of the American Statistical Association, Vol. 96, 270-281.

[4] Conlon M. and Thomas R. G. (1993), *Algorithm AS 280: The Power Function for Fisher's Exact Test.* Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol. 42, No. 1, 258-260.

[5] Conover W. J. (1974), *Some Reasons for Not Using the Yates Continuity Correction on 22 Contingency Tables.* Journal of the American Statistical Association, Vol 69. No. 346, 374-376.

[6] D'Agostino R. B., Chase W. and Belanger A. (1988), *The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations.* The American Statistician, Vol. 42, No. 3, 198-202.

[7] Grizzle. J. E. (1967), *Continuity Correction in the chi- square test for 2×2 Tables.* The American Statistician Vol. 21, No. 4, 28-32.

[8] Hashemi L., Nandram B. and Goldberg R. (1997), *Bayesian Analysis for a Single 2×2 table.* Statistics in Medicine, Vol. 16, 1311-1328.

[9] Kass R. E. and Raftery A. E. (1995), *Bayes Factors*. Journal of the American Statistical Association, Vol. 90, No. 430, 773-795.

[10] Kempthorne O. (1979), *Sampling Inference, Experimental Inference, and Observation Inference*. Sankhya, Series B, Vol. 40, No. 3, 115145.

[11] Liddell D. (1976), *Practical Tests of 2×2 Contingency Tables*. Journal of the Royal Statistical Society. Series D (The Statistician), Vol. 25, No. 4, 295-304.

[12] Little R. J. A. (1989), *Testing the Equality of Two Independent Binomial Proportions*. The American Statistician, Vol. 43, No. 4 , 283-288.

[13] Mehta C. R. and Senchaudhuri P. (2003), *Conditional versus Unconditional Exact Tests for Comparing Two Binomials*. Cytel Software Corporation, 675 Massachusetts Avenue, Cambridge, MA 02139, www.cytel.com.

[14] Nandram B., Bhatta D. and Bhadra D. (2013), *Likelihood Ratio Test of Quasi-independence for Sparse Two-way Contingency Tables*. Journal of Statistical Computation and Simulation.

[15] Nandram B. and Kim H. (2002), *Marginal Likelihood for a Class of Bayesian Generalized Linear Models*. Journal of Statistical Computation and Simulation, Vol. 72, No. 4, 319-340.

[16] Upton G. (1982), *A comparison of alternative tests for the 2×2 comparative trial*. Journal of the Royal Statistical Society. Series A (General), 86-105.

[17] Yates F. (1984), *Tests of Significance for 2×2 Contingency Tables*. Journal of the Royal Statistical Society. Series A (General), Vol. 147, No. 3, 426-463.