Colloquium Biometricum 47
2017, 16-23

COMPARISON OF BINOMIAL PROPORTIONS: NEW TEST

Stanisław Jaworski
Wojciech Zieliński

Department of Econometrics and Statistics
Warsaw University of Life Sciences
Nowoursynowska 159, PL-02-787 Warszawa
e-mail: stanislaw_jaworski@sggw.pl
e-mail: wojciech_zielinski@sggw.pl

Summary

In the problem of comparison of two probabilities of success the most widely used is approximate test based on de Moivre-Laplace theorem. In the paper a test based on likelihood ratio is proposed. Those tests are compared due to probability of an error of the first kind. A medical example is presented.

Keywords and phrases: binomial proportions, comparison of probabilities of success

Classification AMS 2010: 62F03, 62P10

1. Introduction

Let $\xi_1 \sim Bin(n_1, \theta_1)$ and $\xi_2 \sim Bin(n_2, \theta_2)$ be independent binomially distributed random variables. Let $\vartheta = \theta_1 - \theta_2$. Consider a problem of testing

$$H : \vartheta = 0 \text{ vs } K : \vartheta > 0. \tag{H}$$

Statistical model for (ξ_1, ξ_2) is

$$\left(\mathcal{X}, \{Bin(n_1, \theta_1) \times Bin(n_2, \theta_2), 0 < \theta_1, \theta_2 < 1\}\right),$$

where $\mathcal{X} = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\}$. Since difference $\vartheta = \theta_1 - \theta_2$ is a parameter of interest the model is reparametrized

$$(\mathcal{X}, \{Bin(n_1, \theta_1) \times Bin(n_2, \theta_1 - \vartheta), -1 < \vartheta < 1, a(\vartheta) < \theta_1 < b(\vartheta)\}),$$

where

$$a(\vartheta) = \max\{0, \vartheta\}, \quad b(\vartheta) = \min\{1, 1 + \vartheta\}.$$

Let $l(\vartheta) = b(\vartheta) - a(\vartheta)$.

In the problem (H) probability θ_1 is a nuisance parameter. It will be eliminated by appropriate averaging. Hence the statistical model under consideration has the form

$$(\mathcal{X}, \{P_\vartheta, -1 < \vartheta < 1\}),$$

where

$$P_\vartheta(k_1, k_2) = \frac{1}{l(\vartheta)} \int_{a(\vartheta)}^{b(\vartheta)} bin(n_1, k_1; \theta_1) bin(n_2, k_2; \theta_1 - \vartheta) d\theta_1,$$

$$bin(m, l; q) = \binom{m}{l} q^l (1 - q)^{m-l}, \text{ for } l = 0, 1, \ldots, m.$$

Note that, if verified hypothesis is true then

$$P_0(k_1, k_2) = \int_0^1 bin(n_1, k_1, \theta) bin(n_2, k_2, \theta) d\theta = \frac{1}{n_1 + n_2 + 1} \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n_1+n_2}{k_1+k_2}}.$$

2. Classical test for large sample sizes

The test is based on statistic (see for example https://onlinecourses.science.psu.edu/stat414/node/268)

$$W(\xi_1, \xi_2) = \frac{\xi_1/n_1 - \xi_2/n_2}{\sqrt{\frac{\xi_1+\xi_2}{n_1+n_2}\left(1 - \frac{\xi_1+\xi_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

This test is based on normal approximation of the distribution of $\hat{\vartheta} = \frac{\xi_1}{n_1} - \frac{\xi_2}{n_2}$. Let $w^* = W(k_1, k_2)$ be observed value of $W(\xi_1, \xi_2)$ and let

$$lev_W(\vartheta; k_1, k_2) = P_\vartheta \left\{ W(\xi_1, \xi_2) > w^* \right\} = \sum_{i,j:\ W(i,j) > w^*} P_\vartheta\left(i, j\right).$$

Hypothesis H is rejected if $lev_W(0, k_1, k_2) < \alpha$, where α is assumed significance level.

3. Test based on likelihood ratio

The test is based on likelihood ratio

$$\Lambda(\xi_1, \xi_2) = \frac{\sup_{\vartheta > 0} P_\vartheta(\xi_1, \xi_2)}{P_0(\xi_1, \xi_2)}.$$

Let $\Lambda^* = \Lambda(k_1, k_2)$ be observed value of $\Lambda(\xi_1, \xi_2)$ and let

$$lev_\Lambda(\vartheta; k_1, k_2) = P_\vartheta \left\{ \Lambda(\xi_1, \xi_2) > \Lambda^* \right\}.$$

Hypothesis H is rejected if $lev_\Lambda(0; k_1, k_2) < \alpha$.

As a measure of effectiveness of a test its expected value of probability of non rejecting true hypothesis is taken:

$$eff_W = 1 - E_0 lev_W(0; \xi_1, \xi_2) = 1 - \sum_{k_1, k_2} lev_W(0; k_1, k_2) P_0(k_1, k_2),$$
$$eff_\Lambda = 1 - E_0 lev_\Lambda(0; \xi_1, \xi_2) = 1 - \sum_{k_1, k_2} lev_\Lambda(0; k_1, k_2) P_0(k_1, k_2).$$

In Table 1 effectiveness of considered tests are presented for different sample size n_1 (rows) of the first sample and n_2 (columns) of the second sample.

Table 1. Effectiveness eff_Λ and eff_W

Test Λ

	5	10	15	20	25	50	100
5	0.6357	0.6336	0.6325	0.6244	0.6190	0.6164	0.6130
10	0.6336	0.6105	0.6275	0.6177	0.6193	0.6202	0.6157
15	0.6325	0.6275	0.6209	0.6248	0.6219	0.6204	0.6226
20	0.6244	0.6177	0.6248	0.6155	0.6232	0.6189	0.6280
25	0.6190	0.6193	0.6219	0.6232	0.6199	0.6228	0.6284
50	0.6164	0.6202	0.6204	0.6189	0.6228	0.6262	0.6326
100	0.6130	0.6157	0.6226	0.6280	0.6284	0.6326	0.6484

Test W

	5	10	15	20	25	50	100
5	0.5794	0.5373	0.5233	0.5165	0.5122	0.5052	0.5023
10	0.5373	0.5420	0.5134	0.5180	0.5080	0.5054	0.5021
15	0.5233	0.5134	0.5293	0.5074	0.5057	0.5028	0.5013
20	0.5165	0.5180	0.5074	0.5214	0.5046	0.5032	0.5023
25	0.5122	0.5080	0.5057	0.5046	0.5172	0.5067	0.5025
50	0.5052	0.5054	0.5028	0.5032	0.5067	0.5085	0.5032
100	0.5023	0.5021	0.5013	0.5023	0.5025	0.5032	0.5045

In Table 2 ratio of effectiveness is shown. It is seen that Λ test is more effective than W test up to almost 30%.

Table 2. Ratio of effectiveness eff_Λ/eff_W

	5	10	15	20	25	50	100
5	1.0970	1.1792	1.2087	1.2090	1.2086	1.2200	1.2203
10	1.1792	1.1264	1.2222	1.1924	1.2192	1.2272	1.2263
15	1.2087	1.2222	1.1730	1.2314	1.2297	1.2340	1.2421
20	1.2090	1.1924	1.2314	1.1805	1.2351	1.2299	1.2501
25	1.2086	1.2192	1.2297	1.2351	1.1986	1.2292	1.2506
50	1.2200	1.2272	1.2340	1.2299	1.2292	1.2315	1.2571
100	1.2203	1.2263	1.2421	1.2501	1.2506	1.2571	1.2852

4. Medical example

The aim of the investigation was comparing frequency of occurring the specific immunoglobulins G6 (*Phleum pratense* L.), D1 (*Dermatophagoides pteronyssinus*), E1 (*Felis capillum*) and M6 (*Alternaria tenuis*) in two sites: urban (represented by polish town Lublin) and rural area (represented by polish district Zamość). The investigation is a part of ECAP project (ecap.pl/eng_www/index_home.html) conducted by prof. Bolesław Samoliński (Warsaw Medical University). Presented data were obtained by his courtesy.

In Table 3 results of the experiment are presented. Those results were obtained in samples of sizes $n^{(m)} = 743$ and $n^{(w)} = 329$ from urban and rural area, respectively. The number of people with high concentration of immunoglobulin (at least 0.35 IU/ml) were counted.

Table 3. Observed

Immunoglobulin	$k^{(m)}$	$k^{(w)}$
IgE sp. D1 - *Dermatophagoides pteronyssinus*	107	50
IgE sp. E1 - *Felis capillum*	30	9
IgE sp. G6 - *Phleum pratense* L.	92	25
IgE sp. M6 - *Alternaria tenuis*	31	7

Let θ_m and θ_w denote percentages of people with high concentration of a immunoglobulin (at least 0.35 IU/ml) in town and in country, respectively. We are interested in testing

$$H : \theta_m = \theta_w \text{ vs } K : \theta_m > \theta_w,$$

i.e. it is of interest to check whether allergic indicators occur more frequently in town than in country. To the problem both above described tests were be applied. Results are presented in Table 4.

Table 4. Results of testing

Immunoglobulin	$\hat{\theta}_m$	$\hat{\theta}_w$	lev_Λ	lev_W
IgE sp. D1	0.1440	0.1520	0.5215	0.6329
IgE sp. E1	0.0404	0.0274	0.1448	0.1468
IgE sp. G6	0.1238	0.0760	0.0086	0.0102
IgE sp. M6	0.0417	0.0213	0.0431	0.0472

Consider 0.01 as a significance level. For IgE sp. D1, IgE sp. E1 and IgE sp. M6 conclusions are obvious. In case of IgE sp. G6 we observe $lev_\Lambda < 0.01 < lev_W$. Because test Λ is more effective than W test hypothesis $H : \theta_m = \theta_w$ should be rejected.

5. Final remarks

The most commonly test used for hypothesis (H) is approximate W test based on de Moivre-Laplace theorem. In the paper a test Λ based on likelihood ratio is proposed (see Bartoszewicz (1989) or Lehmann (1959) for the general theory of testing statistical hypothesis). This test appears to be better than W test in the sense of greater probability of non rejecting true hypothesis. Unfortunately in considered statistical model likelihood ratio is not a monotone function of the difference of probabilities of success. Hence, its p-value may be calculated only numerically (an exemplary R code is included in Appendix).

Our calculations showed that proposed test Λ is more effective than classical W test. So it may be recommended to use this test in practise. Preliminary results concerning power comparison shows that Λ test is better than W test. Exhaustive results of power comparison will be published separately.

It should be noted that hypothesis testing, although it is a very useful approach in certain contexts, has some limitations. It gives evidence against the null hypothesis but does not indicate which of a family of alternatives is best supported by the data. For this reason the use of confidence intervals if possible is preferable. The reader interested in the relationship between hypotheses testing and confidence intervals is referred to Hirji (2006), where a unified and application-oriented framework, the distributional theory, statistical methods and computational methods for exact analysis of discrete data are presented. Newcombe (1998) investigated properties of confidence intervals for difference between probabilities of success in the classical statistical model, while Zieliński (2017a,b) constructed the confidence interval in the set up considered in the current paper.

References

Bartoszewicz J. (1989). *Wykłady ze statystyki matematycznej.* PWN, Warszawa.

Hirji K. F. (2006). *Exact analysis of discrete data.* Chapman & Hall.

Lehmann E. L. (1959). *Testing statistical hypotheses.* Wiley, New York.

Newcombe R. (1998). Interval Estimation for The Difference Between Independent Proportions: Comparison of Eleven Methods. *Statistics in Medicine* 17, 873-890.

Zieliński W. (2017a). Confidence Interval for the Weighted Sum of Two Binomial Proportions. submitted (for preprint see wojtek.zielinski.statystyka.info/Inne_informacje/spis_prace.html)

Zieliński W. (2017b). New Exact Confidence Interval for the Difference of Two Binomial Proportions. submitted (for preprint see wojtek.zielinski.statystyka.info/Inne_informacje/spis_prace.html)

Appendix

```
lev=function(k,n){
  intL=function(k,n,vartheta){
    f=function(theta){dbinom(k[1],n[1],theta)*
      dbinom(k[2],n[2],theta-vartheta)
    }
  a=max(0,vartheta)
  b=min(1,1+vartheta)
  integrate(f,a,b)$value/(b-a)
  }
  stat.Lambda=function(k,n){
    w0=dhyper(k[1],n[1],n[2],sum(k))/(sum(n)+1)
    w1=optimize(intL, interval=c(0,1), maximum=TRUE,
       k=k,n=n)$objective
    max(w1/w0,1)
  }
  g=function(x,y){stat.Lambda(c(x,y),n)}
  net=expand.grid(0:n[1],0:n[2])
  matrix(mapply(g,net$Var1,net$Var2),ncol=n[2]+1)->tabL
  condition=(tabL>tabL[k[1]+1,k[2]+1])
  g=function(x,y){dhyper(x,n[1],n[2],x+y)/(sum(n)+1)}
  Pzero=outer(0:n[1],0:n[2],g)
  # p-value
```

```
  sum(Pzero[,][condition])
}

#example of application
k=c(107,50) #number of successes
n=c(743,329) #sample sizes
k/n #proportions
lev(k,n) #p-value
```