Colloquium Biometricum 47
2017, 35 – 40

NEW TEST FOR COMPARISON OF BINOMIAL PROPORTIONS: POWER COMPARISON

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 test is approximate test based on de Moivre-Laplace theorem. In Jaworski and Zieliński (2017) a new test based on the likelihood ratio was proposed. In this paper those tests are compared due to their power.

Keywords: binomial proportions, comparison of probabilities of success.

Subject Classifications 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 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

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

where $\mathcal{X} = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\}$. Since the 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 $l(\vartheta) = b(\vartheta) - a(\vartheta) = 1 - |\vartheta|$.
In the problem (H) the 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

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

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

Let

$$\hat{\vartheta}_w = \frac{\xi_1}{n_1} - \frac{\xi_2}{n_2}$$

be the estimator of ϑ.
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 the statistic (see https:// online courses.science.psu.edu/stat414/node/268 for example)

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

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

$$lev_W(\vartheta; k_1, k_2) = P_\vartheta\{W(\xi_1, \xi_2) > w^*\}.$$

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 the 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 $lev_\Lambda(0;k_1,k_2)<\alpha$.

4. Power comparison

Jaworski and Zieliński (2017) showed that Λ test is more effective than classical one, i.e. its expected value of the probability of non rejecting true hypothesis is greater. In what follows power of Λ test will be compared with the power of the classical test. Let

$$pov_\Lambda(\vartheta;k_1,k_2)=\frac{lev_\Lambda(\vartheta;k_1,k_2)}{lev_\Lambda(0;k_1,k_2)}.$$

This is a measure of relative, with respect to the probability of rejecting true hypothesis H, power of the test Λ.
Similarly we define

$$pov_W(\vartheta;k_1,k_2)=\frac{lev_W(\vartheta;k_1,k_2)}{lev_W(0;k_1,k_2)}.$$

For comparison of Λ and W test the ratio

$$r(\vartheta)=E_\vartheta\left(\frac{pov_\Lambda(\vartheta;\xi_1,\xi_2)}{pov_W(\vartheta;\xi_1,\xi_2)}\right)$$

is applied. Values of $r(\vartheta)$ greater than one inform that the test Λ is more powerful that the W test.
The indicator $r(\vartheta)$ is calculated for $\vartheta\in\{0.1,0.2,\ldots,0.9\}$ and sample sizes $n_1,n_2\in\{5,10,15,25,30\}$ (Figure 1) or $n_1,n_2\in\{70,75,80,90,100\}$ (Figure 2). It may be concluded from the Figures that for samples $n_1=n_2\le 30$ the Λ test is more powerful then the W test for small ϑ. In other cases it is seen that $r(\vartheta)\ge 1$. So it follows that Λ test dominates W test except the case of small

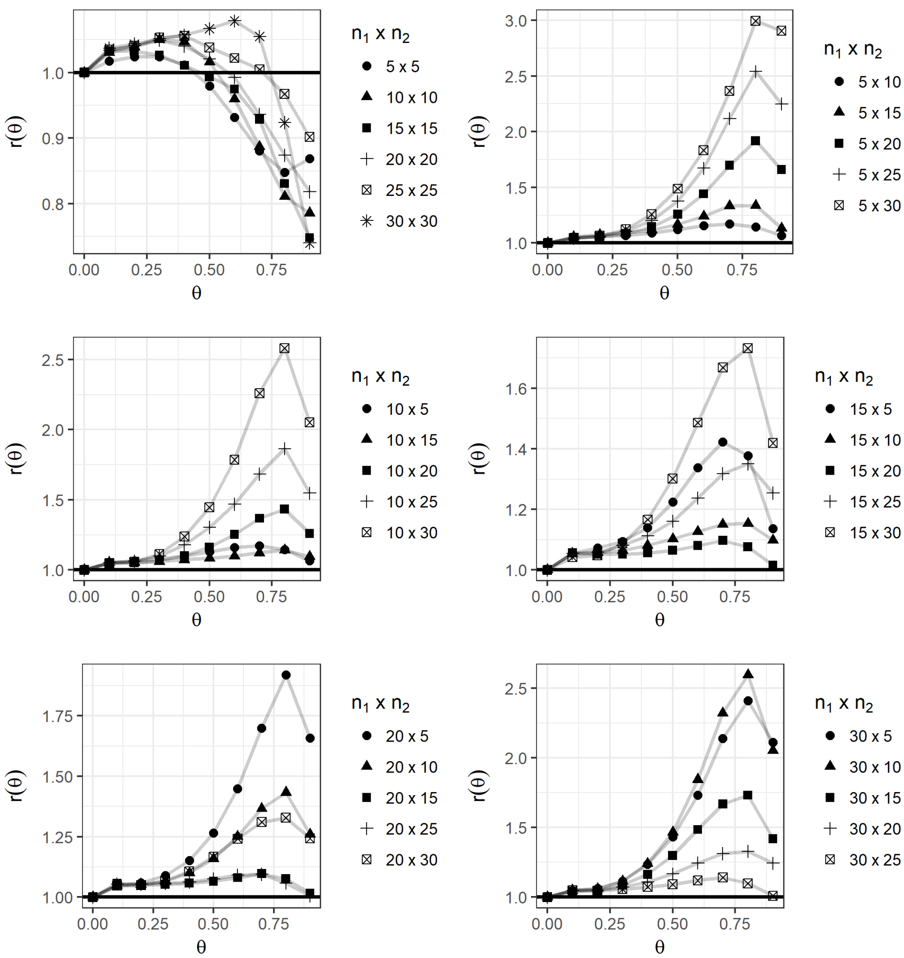

Figure 1. $r(\vartheta)$ for $\vartheta \in \langle 0, 0.9\rangle$ and $n_1, n_2 \in \{5, 10, 20, 15, 25, 30\}$

and equal sample sizes and large differences between binomial proportions. Note the dominance is increasing with n_1 and n_2 (For fixed ϑ we have greater $r(\vartheta)$ for larger sample sizes). For $n_1 = n_2 \geq 70$ (compare first panels of Figure 1 and Figure 2), for example, there is no advantage of W over Λ for every $\vartheta \in \langle 0, 0.9\rangle$.

5. Final remarks

In the paper a test Λ based on likelihood ratio is compared with the known approximate test W with respect to their relative powers. The proposed Λ test is better than W test in the sense of greater probability of non rejecting true hypothesis (see

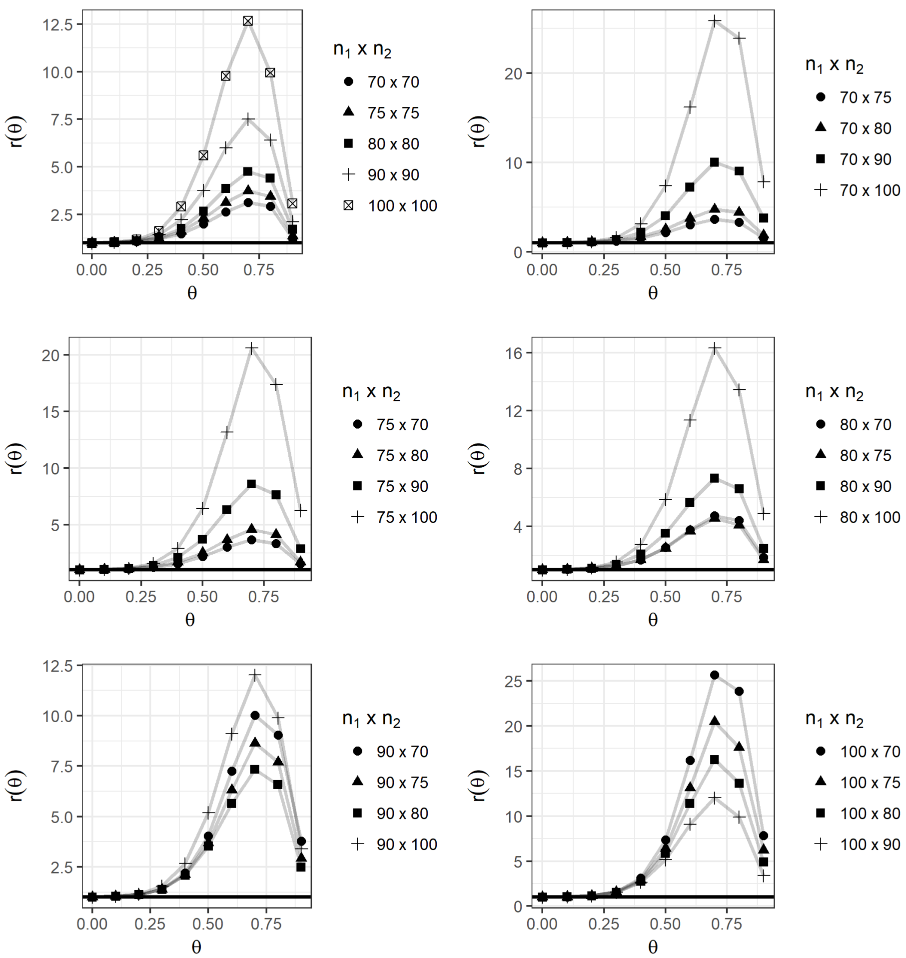

Figure 2. $r(\vartheta)$ for $\vartheta \in \langle 0, 0.9\rangle$ and $n_1, n_2 \in \{70, 75, 80, 90, 100\}$

Jaworski and Zieliński (2017)). Our calculations showed that except some cases of equal sample sizes the test Λ is more powerful too. So it can be recommended for use in practise, although it is not a uniformly most powerful test.

In statistical hypothesis testing a uniformly most powerful test is a hypothesis test which has the greatest power among all possible tests of a given size (see Bartoszewicz (1989) or Lehmann (1959) for the general theory of uniformly most powerful tests). In the paper the test Λ is actually compared only with one test, namely the classical W test. Since P_ϑ is a probability mass function but not a density function we compare the tests with respect to their relative power.

Assuming that, after observing $\{\xi_1 = k_1,\ \xi_2 = k_2\}$, $lev_\Lambda(\vartheta; k_1, k_2) < \alpha$ holds, $pov_\Lambda(\vartheta; k_1, k_2)$ measures how many times the probability of rejecting hypothesis by the Λ test is as likely when the hypothesis is false as when it is true. On the other side if $lev_\Lambda(\vartheta; k_1, k_2) \geq \alpha$ then $pov_\Lambda(\vartheta; k_1, k_2)$ measures how the test is close to rejecting false hypothesis. Hence $r(\vartheta)$, a relative power averaged with respect to probability measure P_ϑ, should properly reflect willingness of the Λ test to reject false hypothesis.
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 (2017) and Zieliński (2018) constructed the confidence interval in the setup considered in the current paper.

References

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

Jaworski, S., Zieliński, W. (2017), *Comparison of Binomial Proportions: New Test*, Colloquium Biometricum 47, 16-23.

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

Zieliński, W. (2018), *Confidence Interval for the Weighted Sum of Two Binomial Proportions*, Applicationes Mathematicae (to appear).