DOI: 10.1515/bile-2015-0009

Biometrical Letters
Vol. 52 (2015), No. 2, 95 - 104

Power comparison of Rao′s score test, the Wald test and the likelihood ratio test in (2x*c*) contingency tables

Anita Dobek, Krzysztof Moliński, Ewa Skotarczak

Department of Mathematical and Statistical Methods, Poznań University of Life Sciences,
Wojska Polskiego 28, 60-637 Poznań, Poland, e-mail: andobek@up.poznan.pl

SUMMARY

There are several statistics for testing hypotheses concerning the independence of the distributions represented by two rows in contingency tables. The most famous are Rao′s score, the Wald and the likelihood ratio tests. A comparison of the power of these tests indicates the Wald test as the most powerful.

Key words: power of a test, simulation studies, test of independence.

1. Introduction

The well-known Rao′s score, Wald and likelihood ratio tests appeared in the literature in the last century. The likelihood ratio (LR) test was introduced by Neyman and Pearson (1928). Wald proposed his test (W) in a paper published in (1943). The score test (RS) was introduced by Rao in (1948). Since the RS test began to be used by econometricians in the late 1960s, the comparison of these tests under contiguous alternatives has received considerable attention.

Peers (1971) showed that none of these tests is uniformly superior to the other two when the power function is considered. Chandra and Joshi (1983) claimed that for a large sample size the RS test is more powerful. Madansky (1989) examined the case of a hypothesis that the variance of a normal distribution was equal to one. He stated that “none of the three tests dominates another, even locally". Based on a Monte Carlo study, Sutradhar and Bartlett (1993a) examined the behavior of these tests for a simple hypothesis, a one-dimensional hypothesis as well as a multi-dimensional composite hypothesis. Their conclusions are favorable to RS only in the last case. For the other hypotheses LR always takes second position, while RS and W exchange positions. The same authors, in another paper (1993b), claimed that for testing linear regression with autocorrelated errors, Rao's size-adjusted test is uniformly more powerful then the LR and W tests. Li (2001) compared RS, LR and W by testing their sensitivity to nuisance parameters. He concluded that they are equally sensitive. In a paper by Yi and Wang (2011) the problem is again examined for a special class of designs, namely response adaptive designs. Based on simulation studies, the authors claimed that the power of W is greater than that of RS and LR for small and medium sample sizes.

Concluding his considerations regarding power comparisons, Rao (2005) stated that "further investigations of power properties of LR, W and RS would be of interest". Following this suggestion, we performed some simulation studies for a very simple example, namely a (2xc) contingency table, to verify which test has the greatest power when the independence of distribution in contingency tables is analyzed. We focus on this example because many practical applications involve the problem of comparison of two populations with regard to a discrete variable (for instance: case-controls and genotypes).

2. Method

Let $\mathbf{N} = (n_{11}, n_{12}, \ldots, n_{1c}, n_{21}, n_{22}, \ldots, n_{2c})'$ be a simple sample from a contingency table. Let $P(\mathbf{N},\mathbf{p})$ denote a joint multivariate distribution probability function with a vector of parameters $\mathbf{p} = (p_{11}, p_{12}, \ldots, p_{1c}, p_{21}, p_{22}, \ldots, p_{2(c-1)})'$ and p_{2c} = 1 – $\mathbf{1}'\mathbf{p}$. As usual, $L(\mathbf{p}/\mathbf{N})$ will denote the log-likelihood function, namely $\ln(P(\mathbf{N},\mathbf{p}))$. We are interested in a verification of the hypothesis of independence of the distributions represented by two rows of our table, which can be written as H$_0$: $\mathbf{p} = \mathbf{p}^0$, where the elements of $\mathbf{p}^0$ are equal to $p_{ij}^0 = n_{i.}n_{.j}/n^2$ with i =1, 2, j = 1, 2, …, c and

$$n_{i.} = \sum_{j=1}^{c} n_{ij}, \quad n_{.j} = n_{1j} + n_{2j} \quad \text{and} \quad n = \sum_{i=1}^{2}\sum_{j=1}^{c} n_{ij}$$

In testing the hypothesis H$_0$ we can use one of the following statistics (Rao, 2005).

2.1. Rao′s score test

The value of Rao′s score statistic is determined by the following formula:

$$\text{RS} = \mathbf{s}(\mathbf{p^0})'\mathbf{I}(\mathbf{p^0})^{-1}\mathbf{s}(\mathbf{p^0}) \tag{1}$$

where

$$\mathbf{s}(\mathbf{p^0}) = \frac{1}{P}\frac{\partial P}{\partial \mathbf{p}} = \left(s_{ij}(\mathbf{p}^0)\right).$$

In our case

$$s_{ij}(\mathbf{p}^0) = \frac{n_{ij}}{p_{ij}^0} - \frac{n_{2c}}{p_{2c}^0} \tag{2}$$

and the (2c-1) x (2c-1) Fisher’s information matrix $\mathbf{I}(\mathbf{p^0})$ is

$$\mathbf{I}(\mathbf{p^0}) = n\left[diag\left(\frac{1}{p_{ij}^0}\right) + \frac{1}{p_{2c}^0}\mathbf{11}'\right]$$

with the inverse

$$\mathbf{I}(\mathbf{p^0})^{-1} = \frac{1}{n}\left[diag\left(p_{ij}^0\right)(\mathbf{I} - \mathbf{11}')diag\left(p_{ij}^0\right)\right] \tag{3}$$

Substituting (2) and (3) into (1) gives

$$\text{RS} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{c}\frac{n_{ij}^2}{p_{ij}^0} - n\,. \tag{4}$$

It is easy to show that this formula is equivalent to the well-known Pearson's test for contingency tables.

2.2. The Wald test

The value of the Wald statistic is determined by the following formula:

$$\mathrm{W} = (\hat{\mathbf{p}} - \mathbf{p}^0)\mathbf{I}(\hat{\mathbf{p}})(\hat{\mathbf{p}} - \mathbf{p}^0) \tag{5}$$

Because

$$\hat{\mathbf{p}} = (\hat{p}_{ij}) = (\frac{n_{ij}}{n}) \tag{6}$$

the matrix $\mathbf{I}(\hat{\mathbf{p}})$ takes the form

$$\mathbf{I}(\hat{\mathbf{p}}) = n^2 \left[diag\left(\frac{1}{n_{ij}}\right) + \frac{1}{n_{2c}}\mathbf{11'} \right] \tag{7}$$

where *diag(*a_{ij}*)* means a diagonal matrix with elements a_{ij} on the main diagonal. Substituting (6) and (7) into (5) gives

$$\mathrm{W} = n^2 \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{(p_{ij}^0)^2}{n_{ij}} - n\,. \tag{8}$$

2.3. The likelihood ratio test

The value of the likelihood ratio statistic is determined by

$$\mathrm{LR} = 2\left[L(\hat{\mathbf{p}}/\mathbf{N}) - L(\mathbf{p}^0/\mathbf{N})\right],$$

which in our case takes the form

$$\mathrm{LR} = 2\sum_{i=1}^{2} \sum_{j=1}^{c} n_{ij} \ln\left(\frac{\hat{p}_{ij}}{p_{ij}^0}\right). \tag{9}$$

As noted by Rao (2005), all the statistics have an asymptotic χ^2 distribution with (c-1) degrees of freedom. It is clear that RS and LR are always applicable. W, by contrast, cannot be used in the case of empty classes.

To illustrate the differences between the three statistics we will consider a hypothesis concerning one parameter, i.e. H_0: $\theta = \theta^0$, and Figure 1 as given by Fox (1997).

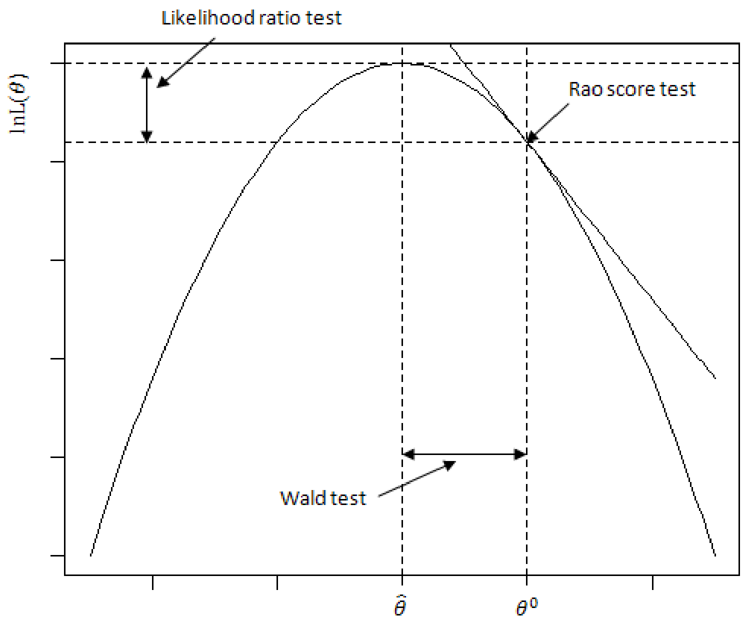


Figure 1. Interpretation of the RS, W and LR tests

Rao′s score test represents the slope at the hypothesized value θ^0. The Wald test compares the parameter estimate $\hat{\theta}$ and its hypothetical value θ^0, and the likelihood ratio test calculates the difference between the log-likelihood function estimated at those two points.

3. Simulation studies

Because it is impossible to compare the statistics described above in an analytical manner, to verify their properties we performed some simulation studies, taking as a criterion the power of the test. For several sets of probabilities in contingency tables (Table 1) we generated 5000 samples for $n = 50$, 80, 100, 200, 300 and 500, using the R platform with the rmultinom() function from the {stats} package (R Core Team (2013)). For each sample the values of the RS, W and LR tests were calculated using formulae (4), (8) and (9), and by taking 0.05 as the incidence level the power of these tests was established. Because the power function is the probability of rejection of a false hypothesis, the power is calculated as a percentage of rejections. Figure 2 contains the graphs of functions representing the size of the test (significance level). In cases **a1** and **b1** the hypothesis of independence is not rejected. Figure 3 contains the graphs of functions representing the power of the test. In all of these cases the hypothesis should be rejected.

A general conclusion which can be drawn from our results is that for sample sizes ranging from 50 to 300, Wald's test demonstrates superiority to the other tests in terms of power.

In cases **a1** and **b1**, as n increases, the size of the test (significance level) approaches the assumed value. For small n W significantly exceeds the value 0.05. The other two statistics behave similarly with values close to 0.05.

The next graphs illustrate a situation where in the contingency tables we are moving away from independence, with increasing Euclidean distance between **a1** (**b1**) and **a2** – **a5** (**b2** – **b5**). The power curves are increasing functions up to the maximal value (100%).

The superiority of Wald's test is also seen for larger contingency tables (**c1**, **c2**). Similar simulations were performed for 2 x 2 tables, and the results are the same as for larger tables. In the simulation studies we also changed the marginal probabilities, and again the results were the same.

Based on these simulation studies we can state that Wald's statistic has the greatest power as regards the problem of testing independence in (2 x c) contingency tables. It should be remembered, however, that its validity is subject to a significant limitation that does not apply to the other statistics, namely the requirement for a nonzero number of cases in each cell.

Table 1. Cell probabilities for different cases of (2 x *c*) contingency tables

a1	0.12	0.3	0.18	b1	0.12	0.24	0.06	0.18
	0.08	0.2	0.12		0.08	0.16	0.04	0.12
a2	0.1	0.3	0.2	b2	0.1	0.24	0.06	0.2
	0.1	0.2	0.1		0.1	0.16	0.04	0.1
a3	0.08	0.3	0.22	b3	0.08	0.24	0.06	0.22
	0.12	0.2	0.08		0.12	0.16	0.04	0.08
a4	0.06	0.3	0.24	b4	0.06	0.24	0.06	0.24
	0.14	0.2	0.06		0.14	0.16	0.04	0.06
a5	0.04	0.3	0.26	b5	0.04	0.24	0.06	0.26
	0.16	0.2	0.04		0.16	0.16	0.04	0.04
c1	0.18	0.2	0.06	0.1	0.06			
	0.08	0.12	0.04	0.06	0.1			
c2	0.18	0.12	0.08	0.06	0.1	0.06		
	0.08	0.06	0.06	0.04	0.06	0.1		

a1 b1

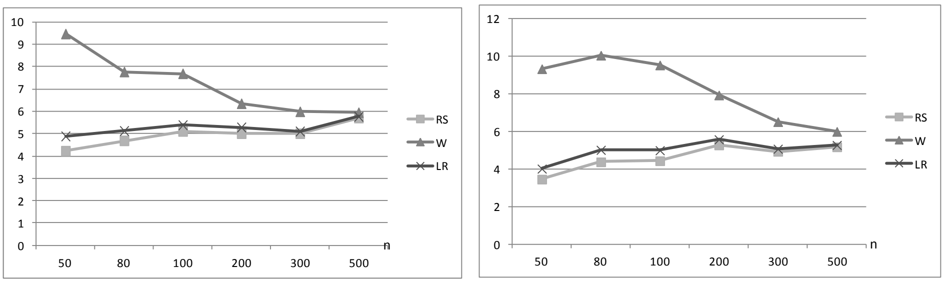
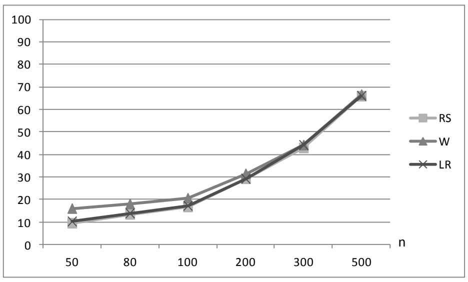
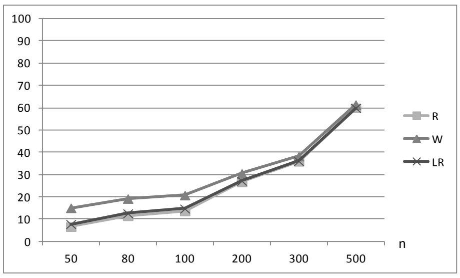


Figure 2. Size (in percentages) of the RS, W and LR tests

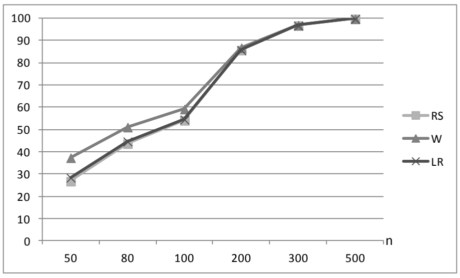
a2



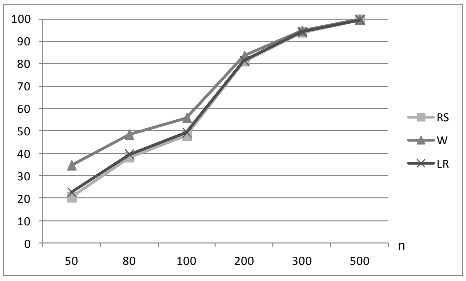
b2



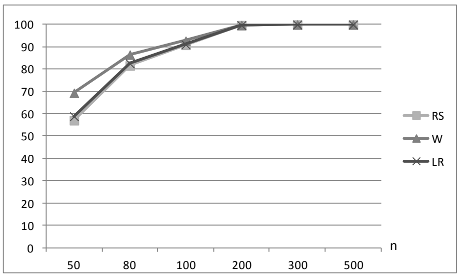
a3



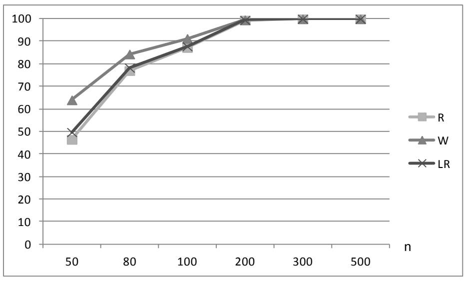
b3



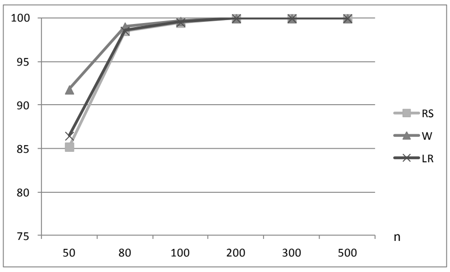
a4



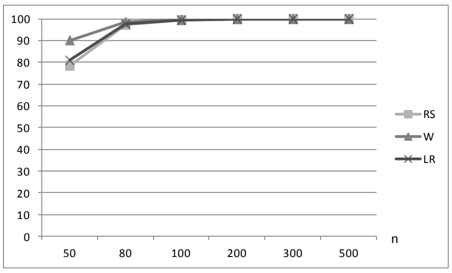
b4



a5



b5



c1 c2

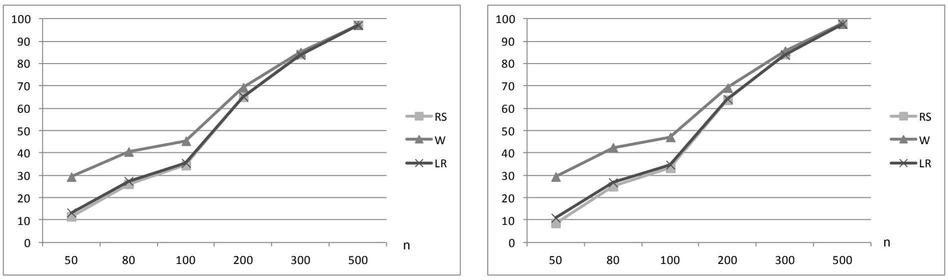


Figure 3. Power (in percentages) of the RS, W and LR tests

REFERENCES

Chandra T.K., Joshi S.N. (1983): Comparison of likelihood ratio, Rao's and Wald's tests and a conjecture of C.R. Rao. Sankhya A, 45: 226-246.

Fox J. (1997): Applied regression analysis, linear models, and related methods. Thousand Oaks, CA, US: Sage Publications, Inc.

Li B. (2001): Sensitivity of Rao's score test, the Wald test and the likelihood ratio test to nuisance parameters. J. Statistical Planning and Inference 97: 57-66.

Madansky A. (1989): A comparison of the Likelihood Ratio, Wald, and Rao tests. In: Contributions to Probability and Statistics, I.J. Gleser et al. (eds.), Springer, New York: 465-471.

Neyman J., Pearson E.S. (1928): On the use and interpretation of certain test criteria. Biometrika 20A: 175-240.

Peers H.W. (1971): Likelihood ratio and associated test criteria. Biometrika 58: 577-587.

R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rao C.R. (1948): Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. Proccedings of the Cambridge Philosophical Society 44: 50-57.

Rao C.R. (2005): Score test: historical review and recent developments. Advances in Ranking and Selection, Multiple Comparisons and Reliability. In: Statistics for Industry and Technology, Balakrishnan N., Kannan N., Nagaraja H.N. (eds.): 3-20.

Sutradhar B.C., Bartlett R.F. (1993a): Monte Carlo comparison of Wald's, likelihood Ratio and Rao's tests. J.Statist. Comput. Simul. 46: 23-33.

Sutradhar B.C., Bartlett R.F. (1993b): A small and large sample comparison of Wald's, Likelihood Ratio and Rao's tests for testing linear regression with autocorrelated errors. Sankhya: The Indian Journal of Statistics 55 B: 186-198.

Wald A. (1943): Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54: 426-482.

Yi Y., Wang X. (2011): Comparison of Wald, Score, and Likelihood Ratio Tests for Response Adaptive Designs. J. Statistical Theory and Applications 10(4): 553-569.